Lecture 12

Newton-type Methods

October 15, 2008

Outline

- Inexact nonsmooth Newton methods
- Piecewise-smooth Newton methods
- Josephy-Newton methods for VIs

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

$$G(x^k) + A(x^k, d^k) = r^k$$ **inexact Newton equation**

where r^k is such that

$$\|r^k\| \le \eta^k \; \|G(x^k)\|$$

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 FP II)

Lemma 1 *Let $G : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^n$ with Ω open, be a locally Lipschitz function in a nbhd of $x^* \in \Omega$, satisfying $G(x^*) = 0$. Assume that G admits a nonsingular Newton approximation $\mathcal{A}$ at x^*. Then:*
For every $\epsilon \in (0, \epsilon_{\mathcal{A}}]$ and for every $\bar{\eta} > 0$, a neighborhood $B(x^*, \delta)$ of x^* exists such that for every $x^k \in B(x^*, \delta)$, every scalar $\eta_k \in (0, \bar{\eta}]$ and for every vector r^k satisfying $\|r^k\| \le \eta_k \|G(x^k)\|$, the equation

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

has a unique solution d^k.

Proof: omitted

- Lemma quarantees existence of direction d^k at each iteration k

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**

Proof:

- Since $A(x,.) \in \mathcal{A}(x)$, there exists a function Δ with $\lim_{t\to 0} \Delta(t) = 0$ suchthat for x sufficiently near x^* and $A(x,.) \in \mathcal{A}(x)$, we have

$$\frac{\|G(x) + \mathcal{A}(x, x^* - x) - G(x^*)\|}{\|x - x^*\|} \leq \Delta(\|x - x^*\|).$$

$$\begin{aligned}\|G(x) + \mathcal{A}(x, x^* - x) - G(x^*)\| &= \|G(x) + \mathcal{A}(x, x^* - x)\| \\ &\leq \Delta(\|x - x^*\|)\|(x - x^*)\|\end{aligned}$$

Let $\epsilon \in (0, \epsilon_{\mathcal{A}}]$ be given and suppose that $\eta_k \leq \bar{\eta}$ for every k. We can pick a $\delta > 0$ such that for every $x^k \in B(x^*, \delta)$, the following holds:

$$\begin{aligned} \| - G(x^k) + r_k)\| &\leq \|G(x_k)\| + \|r_k\| \\ &\leq \|G(x^k)\| + \eta_k \|G(x^k)\| \\ &\leq (1 + \bar{\eta})\|G(x^k)\| \\ &\leq (1 + \bar{\eta})\|x^k - x^*\| \\ &\leq (1 + \bar{\eta})\delta \\ &\leq \epsilon, \end{aligned}$$

where L is the Lipschitz constant and the final result follows by choice of δ.

- By the uniform Lipschitz homeomorphism property of $A^{-1}(x^k)$ with constant $L_{\mathcal{A}}$, we have

$$\begin{aligned}
\|x^{k+1} - x^*\| &= \|x^k - x^* + A^{-1}(x^k, -G(x^k) + r^k)\| \\
&= \|A^{-1}(x^k, -G(x^k) + r^k) - A^{-1}(x^k, A(x^k, x^* - x^k)))\| \\
&\le L_{\mathcal{A}}\| - G(x^k) + r^k + A(x^k, x^* - x^k) \\
&\le L_{\mathcal{A}}\| - G(x^k) + A(x^k, x^* - x^k)\| + L_{\mathcal{A}}\|r^k\| \\
&\le L_{\mathcal{A}}\|x^k - x^*\|\Delta(\|x^k - x^*\|) + L_{\mathcal{A}}\|r^k\| \\
&\le L_{\mathcal{A}}\|x^k - x^*\|\Delta(\|x^k - x^*\|) + L_{\mathcal{A}}\eta_k\|G(x^k)\| \\
&\le L_{\mathcal{A}}\|x^k - x^*\|\Delta(\|x^k - x^*\|) + L_{\mathcal{A}}\eta_k L\|x^k - x^*\| \\
&\le L_{\mathcal{A}}\|x^k - x^*\|\Delta(\|x^k - x^*\|) + L_{\mathcal{A}}\bar{\eta} L\|x^k - x^*\|
\end{aligned}$$

- If $\bar{\eta}$ and δ are chosen to be sufficiently small, we have $\|x^{k+1} - x^*\| \le \frac{1}{2}\|x^k - x^*\|$, implying that $\{x^k\}$ converges at least Q-linearly to x^*.

- If $\{\eta_k\} \to 0$, superlinear convergence is implied by

$$\|x^{k+1} - x^*\| \le L_{\mathcal{A}}\|x^k - x^*\|\Delta(\|x^k - x^*\|) + L_{\mathcal{A}}L\eta_k\|x^k - x^*\|$$

$$\lim_{k\to\infty} \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} = \lim_{k\to\infty} \Big(L_{\mathcal{A}}\Delta(\|x^k - x^*\|) + L_{\mathcal{A}}L\eta_k\Big) = 0.$$

- Finally suppose that the approximation $\mathcal{A}$ us strong and η_k satisfies the condition $\eta_k \leq \tilde{\eta}\|G(x^k)\|$ for some positive $\tilde{\eta}$. Then as above we have

$$\begin{aligned}
\|x^{k+1} - x^*\| &= \|x^k - x^* + A^{-1}(x^k, -G(x^k) + r^k)\| \\
&= \|A^{-1}(x^k, -G(x^k) + r^k) - A^{-1}(x^k, A(x^k, x^* - x^k)))\| \\
&\leq L_{\mathcal{A}}\| - G(x^k) + r^k + A(x^k, x^* - x^k) \\
&\leq L_{\mathcal{A}}\| - G(x^k) + A(x^k, x^* - x^k)\| + L_{\mathcal{A}}\|r^k\| \\
&\leq L_{\mathcal{A}}L'\|x^k - x^*\|^2 + L_{\mathcal{A}}\|r^k\| \\
&\leq L_{\mathcal{A}}L'\|x^k - x^*\|^2 + L_{\mathcal{A}}\eta_k\|G(x^k)\| \\
&\leq L_{\mathcal{A}}L'\|x^k - x^*\|^2 + L_{\mathcal{A}}\eta_k L\|x^k - x^*\| \\
&\leq L_{\mathcal{A}}L'\|x^k - x^*\|^2 + L_{\mathcal{A}}\tilde{\eta}L^2\|x^k - x^*\|^2
\end{aligned}$$

Therefore, we have that

$$\lim_{k\to\infty} \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|^2} \le \left(L_{\mathcal{A}} L' + L_{\mathcal{A}} L^2 \tilde{\eta}\right) < \infty,$$

implying quadratic convergence.

Newton methods for piecewise smooth functions

Definition 1 *A continuous mapping* $G : \mathcal{D} \subseteq \mathbb{R}^n \to \mathbb{R}^m$ *is said to be* PC^1 *near the vector* $x \in \mathcal{D}$ *if there exists an open neighborhood* $\mathcal{N} \subseteq \mathcal{D}$ *of* x *and a finite family* $\{G^1, \ldots, G^p\}$ *of* C^1 *functions defined on* $\mathcal{N}$ *such that* $G(y)$ *is an element of* $\{G^1(y), \ldots, G^p(y)\}$ *for all* $y \in \mathcal{N}$*.*
Each function G^i *is called a* C^1 *piece of* G *at* x*.*
Let $\mathcal{P}(y)$ *denote the set of indices* $i \in \{1, \ldots, p\}$ *such that* $G(y) = G^i(y)$*.*

- Let $G : \mathbb{R}^n \to \mathbb{R}^n$ be a PC^1 mapping with x^* being a zero of the map. Moreover let $\{G^1, \ldots, G^p\}$ be the set of C^1 pieces of G at x^*

- The set $\mathcal{P}(x)$ of active "pieces" at x is given by

$$\mathcal{P}(x) = \{i : G(x) = G^i(x)\}.$$

- A Newton approximation scheme $\mathcal{A}$ at x^* consists of family $\mathcal{A}(x)$ (with x in some neighborhood of x^*) and $\mathcal{A}(x)$ given by

$$\mathcal{A}(x) = \{\nabla G^i(x) : i \in \mathcal{P}(x)\}.$$

Thus:

  - The family $\mathcal{A}(x)$ is finite and each $A(x, \cdot) \in \mathcal{A}(x)$ is linear.
  - The Newton system of equations reduces to finding a vector $d^k \in B(0, \epsilon)$ such that

$$G(x^k) + \nabla G^j(x^k) d^k = 0 \qquad \text{where } j \in \mathcal{P}(x^k).$$

  - Note that the equations are linear so that the Newton systems are easy to solve.

Piecewise smooth Newton method (PCNM)

1. Given x^0, $\texttt{tol}$, and $\epsilon > 0$

2. Set $k = 0$

3. If $\|G(x^k)\| \leq \texttt{tol}$, stop.

4. Select an i_k in $\mathcal{P}(x^k)$ and find a direction d^k such that

$$G(x^k) + \nabla G^{i_k}(x^k)d^k = 0.$$

5. Set $x^{k+1} = x^k + d^k$ and $k := k + 1$; go to step 2.

Convergence result

Theorem 1 *(Th. 7.2.15 in FP II)*
Let $G : \Omega \to \mathbb{R}^n$ *with be a* PC^1 *mapping with* C^1 *pieces given by* $\{G^i, i = 1, \ldots, p\}$ *over an open set* $\Omega \subseteq \mathbb{R}^n$*. Let* $x^* \in \Omega$ *be a zero of* G*. Suppose that* $\nabla G^i(x^*)$ *are nonsingular for all* $i \in \mathcal{P}(x^*)$*. Then:*

- *There exists a neighborhood* $B(x^*, \delta)$ *of* x^* *such that if* x^0 *belongs to this neighborhood, the PCNM method produces a sequence* $\{x^k\}$ *that converges Q-superlinearly to* x^**.*

- *If the Jacobians* $\nabla G^i(x)$ *of the active pieces near* x^* *are locally Lipschitz, then the convergence rate is Q-quadratic.*

Proof:

- For x sufficiently close to x^*, we have $\mathcal{P}(x) \subseteq \mathcal{P}(x^*)$.
- We now show that the family $\{\nabla G^i(x) : i \in \mathcal{P}(x)\}$ is a Newton approximation family, i.e.,
  - $A(x, 0) = 0$ **for every** $A(x, .) \in \mathcal{A}(x)$: This follows from linearity of approximation.
  - **For any** $x' \in \Omega$**,** $x \neq x'$ **and for any** $A(x, .) \in \mathcal{A}(x)$**,**

$$\frac{\|G(x) + A(x, x' - x) - G(x')\|}{\|x - x'\|} \leq \Delta(\|x - x'\|).$$

We show that for every sequence $\{y^k\}$ converging to x^* (with $y^k \neq x^*$ for all k) and for every $i_k \in \mathcal{P}(y^k)$,

$$\lim_{k \to \infty} \frac{\|G(y^k) + \nabla G^{i_k}(y^k)(x^* - y^k) - G(x^*)\|}{\|y^k - x^*\|} = 0.$$

Since $\mathcal{P}(y_k) \subseteq \mathcal{P}(x^*)$, we have $G(x^*) = G^{i_k}(x^*)$ for all k sufficiently large. Moreover, $G(y^k) = G^{i_k}(y^k)$. Therefore,

$$\lim_{k\to\infty} \frac{\|G(y^k) + \nabla G^{i_k}(y^k)(x^* - y^k) - G(x^*)\|}{\|y^k - x^*\|} =$$
$$\lim_{k\to\infty} \frac{\|G^{i_k}(y^k) + \nabla G^{i_k}(y^k)(x^* - y^k) - G^{i_k}(x^*)\|}{\|y^k - x^*\|}.$$

But there are finitely many pieces and each G^{i_k} is a C^1 function, implying that the above limit is zero.

- We next show that $\mathcal{A}$ is a family of uniformly Lipschitz homeomorphisms
  - Each $\nabla G^i(x^*)$ for $i \in \mathcal{P}(x^*)$ is nonsingular
  - Every $\nabla G^i(x^*)$ is a Lipschitz homemorphism with some constants L_i, L_i' (on a neigborhood V_X of x^*).

- For x near x^*, we have $\mathcal{P}(x) \subset \mathcal{P}(x^*)$, implying that every $A(x, \cdot) = \nabla G^i(x)$ for some $i \in \mathcal{P}(x^*)$
- Hence, every $A(x, \cdot)$ is nonsingular, and it is a Lipschitz homemorphism with constants L_i and L_i' when $A(x, \cdot) = \nabla G^i(x)$.
- Therefore, $\mathcal{A}(x)$ is a family of uniformly Lipschitz homeomorphisms (with constants $\max_{L_i, i \in \mathcal{P}(x^*)}$, $L_i' = \max L_i', i \in \mathcal{P}(x^*)$ on a neigborhood V_X of x^*).

Therefore we may employ **Theorem 7.2.5** which provides a convergence result for locally Lipschitz mappings which admit nonisingular Newton approximations – we have just shown that our mapping does indeed admit one.

- If the Jacobians $\nabla G^(x)$ of the active pieces for x near x^* are locally Lipschitz, then similar to the preceding we can show that there is a scalar $\bar{L}$ such that

$$\lim_{k\to\infty} \frac{\|G(y^k) + \nabla A(y^k, x^* - y^k) - G(x^*)\|}{\|y^k - x^*\|^2} =$$
$$\lim_{k\to\infty} \frac{\|G^{i_k}(y^k) + \nabla G^{i_k}(y^k)(x^* - y^k) - G^{i_k}(x^*)\|}{\|y^k - x^*\|^2} \le \bar{L}$$

fr any $\{y^k\}$ of points distinct from x^* and converging to x^*.

  - Hence the approximation $\mathcal{A}(x)$ is strong, and by **Theorem 7.2.5** the quadratic convergence follows.

Example

- Consider the nonsmooth equation $H(x) = 0$, where

$$H(x) = \min(F(x), G(x)) \qquad \forall x \in \mathbb{R}^n,$$

with F and G being continuously differentiable maps from $\mathbb{R}^n$ to $\mathbb{R}^n$.

- For an arbitrary vector x, the matrix $A(x)$ is defined row-wise as

$$A_{i,\cdot}(x) = \begin{cases} \nabla F_i(x)^T & F_i(x) < G_i(x) \\ \text{either } \nabla F_i(x)^T \text{ or } \nabla G_i(x)^T & F_i(x) = G_i(x) \\ \nabla G_i(x)^T & G_i(x) < F_i(x). \end{cases}$$

- If $\beta(x) = \{i : F_i(x) = G_i(x)\}$, then there are $2^{|\beta(x)|}$ matrices of this form for every x.

- Note that if $\beta_x = \emptyset$, then $A(x)$ is a singleton (why???).

- These matrices are the Jacobian matrices of the C^1 parts of the mapping H at x

- The PCNM algorithm generates a sequence $\{x^k\}$ in the following fashion:

$$H(x^k) + \widehat{A}(x^k)d^k = 0$$

where $\widehat{A}(x^k)$ is one of the $2^{|\beta(x^k)|}$ matrices defined above.

- Convergence of such a sequence is guaranteed based on earlier result.

Finite termination

Theorem 2 *Let* $F, G : \Omega \to \mathbb{R}^n$ *be continuously differentiable mappings defined on an open set* $\Omega \subseteq \mathbb{R}^n$*. Let* x^* *be a solution of the* $CP(F, G)$ *such that every row-representative matrix* M *of the pair* $(\nabla F(x^*), \nabla G(x^*))$ *satisfying*

$$M_i = \begin{cases} \nabla F_i(x^*) & \forall i, F_i(x^*) = 0,\ G_i(x^*) > 0 \\ \nabla G_i(x^*) & \forall i, F_i(x^*) > 0,\ G_i(x^*) = 0, \end{cases}$$

is nonsingular. Then, if F *and* G *are affine functions, the PC Newton method terminates in a finite number of iterations, i.e., there exists a* $\bar{k}$ *such that* $x^{\bar{k}} = x^*$*.*

Proof:

- Let $\bar{k}$ be such that for all $k \geq \bar{k}$,

$$0 = F_i(x^*) < G_i(x^*) \implies F_i(x^k) < G_i(x^k)$$
$$0 = G_i(x^*) < F_i(x^*) \implies G_i(x^k) < F_i(x^k).$$

- For such an index k, and for all i such that $0 = F_i(x^*) < G_i(x^*)$, we have

$$0 = F_i(x^k) + \nabla F_i(x^k)^T d^k = F_i(x^k + d^k) = F_i(x^{k+1})$$

- For such an index k, and for all i such that $0 = G_i(x^*) < F_i(x^*)$, we similarly have

$$0 = G_i(x^k) + \nabla G_i(x^k)^T d^k = G_i(x^k + d^k) = G_i(x^{k+1}).$$

Moreover for an index i such that $F_i(x^*) = G_i(x^*) = 0$, we have either $F_i(x^{k+1}) = 0$ or $G_i(x^{k+1}) = 0$. Thus $H(x^{k+1}) = 0$.

- By the nonsingularity of the row-representative matrix, it follows that x^{k+1} is unique.

- Furthermore, by the fact that x^* satisfies all these equations, it follows that $x^{k+1} = x^*$.

- Therefore, in a finite number of iterations, one achieves convergence.

Composite maps

- Consider a mapping G that is a composition of a smooth and Lipschitz continuous map

- Specifically $G(x) \equiv S \circ N(x)$ where $N : \mathbb{R}^n \to \Omega$ and $S : \Omega \to \mathbb{R}^n$, with $\Omega \subseteq \mathbb{R}^m$.

- Example: The normal map $\mathbf{F}_K^{\mathrm{nor}}$ of the $VI(K, F)$, given by
$$\mathbf{F}_K^{\mathrm{nor}}(x) = F(\Pi_K(x)) + v - \Pi_K(x)$$

- A single-valued approximation for G:

$$A(x,d) = JS(N(x))[N(x+d) - N(x)] \qquad \forall d \in \mathbb{R}^n.$$

  - Property (a): A(x,0) $=$ 0

- Property (b):

$$\begin{aligned}
&\limsup_{x\to\bar{x}} \frac{\|G(x) + A(x, \bar{x} - x) - G(\bar{x})\|}{\|x - \bar{x}\|} \\
&= \limsup_{x\to\bar{x}} \frac{\|S(N(x)) + JS(N(x))[N(\bar{x}) - N(x)] - S(N(\bar{x}))\|}{\|x - \bar{x}\|} \\
&\leq \limsup_{x\to\bar{x}} \frac{\|S(N(\bar{x})) + o(\|N(x) - N(\bar{x})\|) - S(N(\bar{x}))\|}{\|x - \bar{x}\|} \\
&\leq \limsup_{x\to\bar{x}} L' \frac{\|o(\|x - \bar{x}\|)\|}{\|x - \bar{x}\|} = 0.
\end{aligned}$$

- Finally, when $JS(N(\bar{x}))[N(\bar{x}+d) - N(\bar{x})]$ is a locally Lipschitz homeomorphism near $d = 0$, nonsingularity of the approximation $A(x, \cdot)$ may be established (see Ch. 7.2.2 of FP II).