Lecture 13

Newton-type Methods

A Newton Method for VIs

October 20, 2008

Outline

- Quick recap of Newton methods for composite functions
- Josephy-Newton methods for VIs
- A special case: mixed complementarity problems

Composite maps

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

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

- Example: The normal map of the VI(K,F), denoted by $\mathbf{F}_k^{\mathrm{nor}}$:

$$\mathbf{F}_K^{\mathrm{nor}}(v) \equiv F(\Pi_K(v)) + v - \Pi_K(v).$$

- A single-valued approximation for G:

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

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

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

Note that the inequality on the third line is by def. 7.2.2. (definition of Newton approximation)

- Property (b') can be proved if JS is Lipschitz around $N(\bar{x})$ (a strong

Newton approximation)

- Finally - establish nonsingularity of approximation may be established.

A Newton Method for VIs

- Newton method on composite equation
- Introduce a Newton method for VIs
  - K is a closed convex set
  - F is a continuously differentiable function
- Vehicle: the normal map for VIs: $\mathbf{F}_K^{\rm nor}(z)$, defined as

$$\mathbf{F}_K^{\rm nor}(z) \equiv F(\Pi_K(z)) + z - \Pi_K(z), \quad z \in \mathbb{R}^n$$

and $z^* = x^* - F(x^*)$ is a zero of $\mathbf{F}_K^{\rm nor}(z)$ and $x^* = \Pi_K(z^*)$.

A nonsingular Newton approximation

- To apply the Newton method to the

$$\mathbf{F}_K^{\rm nor}(z) = 0,$$

we need a nonsingular Newton approximation

- The normal map can be viewed as a composite function $\mathbf{F}_K^{\rm nor}(z) = S \circ N$, where

$$S : \mathbb{R}^{2n} \to \mathbb{R}^n, \quad S(a,b) = F(a) + b$$
$$N : \mathbb{R}^n \to \mathbb{R}^{2n}, \quad N(z) \equiv \begin{pmatrix} \Pi_K(z) \\ z - \Pi_K(z) \end{pmatrix}.$$

- Note that S is differentiable (since F is) while N is not (since it contains the projector Π_K)

- Recall that we employed the following approximation scheme for composite functions:

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

- Therefore if $z^{k+1} = z^k + d^k$ and $x^k = \Pi_K(z^k)$, we have

$$\begin{aligned} A(z,d) &= \begin{pmatrix} J(F(\Pi_K(z)) & I \end{pmatrix} \begin{pmatrix} \Pi_K(z+d) - \Pi_K(z) \\ d - \Pi_K(z+d) + \Pi_K(z) \end{pmatrix} \\ &= JF(\Pi_K(z))(\Pi_K(z+d) - \Pi_K(z)) + d - \Pi_K(z+d) + \Pi_K(z) \\ &= (JF(\Pi_K(z)) - I)(\Pi_K(z+d) - \Pi_K(z)) + d. \end{aligned}$$

- Given x^k, we observe that the Newton system to be solved reduces to

$$\begin{aligned}
0 &= \mathbf{F}_K^{\rm nor}(z^k) + A(z^k, d^k) \\
&= F(\Pi_K(z^k)) + z^k - \Pi_K(z^k) + JF(x^k)(\Pi_K(z^{k+1}) - x^k) \\
&+ \Pi_K(z^k) - \Pi_K(z^{k+1}) + z^{k+1} - z^k \\
&= F(x^k) + JF(x^k)(\Pi_K(z^{k+1}) - x^k) - \Pi_K(z^{k+1}) + z^{k+1}.
\end{aligned}$$

- One can observe that the last system is the normal map of $VI(K, F^k)$, where

$$F^k(x) \equiv F(x^k) + JF(x^k)(x - x^k),$$

is the linearization of $F(x)$ at x^k. In effect, the vector $x^{k+1} = \Pi_K(z^{k+1})$ is the solution to VI(K,F^k).

- By the nonexpansiveness of the Euclidean projector, we have

$$\|x^{k+1} - x^k\| \leq \|z^{k+1} - z^k\|$$

implying that if $\|z^{k+1} - x^k\| \leq \epsilon$, $\|x^{k+1} - x^k\| \leq \epsilon$.

- Algorithm of nonsmooth Newton method may be described in two ways:
  - Construct a sequence $\{z^k\}$ by iteratively solving

$$F(x^k) + JF(x^k)(\Pi_K(z^{k+1}) - x^k) - \Pi_K(z^{k+1}) + z^{k+1} = 0,$$

  where $x^k = \Pi_K(z^k)$.
  - Solve a sequence of VIs given by VI(K, F^k) to give a sequence $\{x^k\}$ – we use the latter in the Josephy Newton method.

Josephy-Newton Method for the VI

1. Given $x^0 \in K$ and $\epsilon > 0$

2. Set $k = 0$

3. If $x^k \in SOL(K, F)$, stop.

4. Let x^{k+1} be a solution of the VI(K, F^k) such that $x^{k+1} \in B(x^k, \epsilon)$.

5. $k := k + 1$ and return to 3.

Nonsingularity of Newton appproximation

- This requires the error function given by

$$\mathbf{e}_F(x) := F(x) - F(x^*) - JF(x^*)(x - x^*),$$

the residual of the first-order Taylor approximation.

- If F is strongly F-differentiable at x^*, the error function satisfies

$$\lim_{x^1 \neq x^2, (x^1, x^2) \to (x^*, x^*)} \frac{\mathbf{e}_F(x^1) - \mathbf{e}_F(x^2)}{\|x^1 - x^2\|} = 0;$$

and if JF is Lipschitz continuous in a neighborhood of x^*, then

$$\limsup_{x^* \neq x \to x^*} \frac{\|\mathbf{e}_F(x)\|}{\|x - x^*\|^2} < \infty.$$

Quick aside: Solution stability

Definition 1 (Def. 5.3.1)

- A solution x^* of the VI(K, F) is said to be **semistable** if for every open neighborhood $\mathcal{N}$ of x^* satisfying

$$cl(\mathcal{N}) \subset \mathcal{D} \text{ and } SOL(K,F) \cap cl(\mathcal{N}) = \{x^*\},$$

there exist two scalars c and ϵ such that for every G in $B(F;\epsilon,K_{\mathcal{N}})^*$ and every $x \in SOL(K,G) \cap \mathcal{N}$,

$$\|x - x^*\| \leq c\|e(x)\|,$$

*The set $B(H;\epsilon,S)$ is an $\epsilon-$nbhd of the function H restricted to the set S, comprising of all functions G such that $\|G-H\|_S \equiv \sup_{y\in S} \|G(y)-H(y)\| < \epsilon$.

where $e(v) \equiv F(v) - G(v)$ is the difference function.

- The solution x^* is said to be **stable** if in addition $SOL(K, G) \cap \mathcal{N} \neq \emptyset$.
- This solution is said to be **strongly-stable** if x^* is stable and for every nbhd $\mathcal{N}$ with scalars c and ϵ as above, and for any two continuous functions G and $\tilde{G}$ belonging to $B(F; \epsilon, K_{\mathcal{N}})$,

$$\|x - x'\| \leq c\|e(x) - \tilde{e}(x')\|,$$

for every $x \in SOL(K, G) \cap \mathcal{N}$ and $x' \in SOl(K, G') \cap \mathcal{N}$, where $\tilde{e}$ is the difference function between F and $\tilde{G}$; $\tilde{e}(v) = F(v) - \tilde{G}(v)$.

Lemma 1 *Let K be a closed convex set and F be continuously differentiable. Suppose that x^* is a strongly stable solution of VI(K, F). Then the approximation scheme of $\mathbf{F}_K^{nor}$ is nonsingular at z^*.*

Proof:

- x^* is a strongly stable solution of VI(K,F) if and only if z^* is a strongly stable zero of $\mathbf{F}_K^{\rm nor}$. (Th. 5.3.6)
- The approximation A(x^*, .) is globally Lipschitz (by Prop. 5.2.15 and 5.2.8), it suffices to show that $A(z^*, z - z^*)$ is a strong first-order approximation (FOA) of $\mathbf{F}_K^{\rm nor}$ at z^*. That is

$$\lim_{x^1 \neq x^2 (x^1, x^2) \to (x^*, x^*)} \frac{(\mathbf{e}_F(x^1) - \mathbf{e}_F(x^2))}{\|x^1 - x^2\|} = 0, \quad (1)$$

with

$$\mathbf{e}(z) = A(z^*, z - z^*) - \mathbf{F}_K^{\rm nor}(z).$$

- By substituting, we have

$$
\begin{aligned}
\mathbf{e}(z) &= A(z^*, z - z^*) - \mathbf{F}_K^{\mathrm{nor}}(z) \\
&= (JF(\Pi_K(z^*) - I)(\Pi_K(z) - \Pi_K(z^*)) + z - z^* - \mathbf{F}_K^{\mathrm{nor}}(z) \\
&= JF(x^*)(\Pi_K(z) - x^*) - \Pi_K(z) + \Pi_K(z^*) + z - z^* - \mathbf{F}_K^{\mathrm{nor}}(z) \\
&= JF(x^*)(\Pi_K(z) - x^*) - \Pi_K(z) + x^* + z - z^* \\
&- F(\Pi_K(z)) - z + \Pi_K(z) \\
&= JF(x^*)(\Pi_K(z) - x^*) - z^* + x^* - F(\Pi_K(z)) \\
&= -F(x^*) - z^* + x^* - \mathbf{e}_F(\Pi_K(z)).
\end{aligned}
$$

Since for $z^1 \neq z^2$, we have

$$\begin{aligned}
\frac{(\mathbf{e}(z^1) - \mathbf{e}(z^2))}{\|z^1 - z^2\|} &= \frac{(\mathbf{e}_F(\Pi_K(z^1)) - \mathbf{e}_F(\Pi_K(z^2))}{\|z^1 - z^2\|} \\
&= \frac{(\mathbf{e}_F(\Pi_K(z^1)) - \mathbf{e}_F(\Pi_K(z^2))}{\|\Pi_K(z^1) - \Pi_K(z^2)\|} \frac{\|\Pi_K(z^1) - \Pi_K(z^2)\|}{\|z^1 - z^2\|} \\
&\leq \frac{(\mathbf{e}_F(\Pi_K(z^1)) - \mathbf{e}_F(\Pi_K(z^2))}{\|\Pi_K(z^1) - \Pi_K(z^2)\|} (\text{ by nonexpansivity }) \\
&= \frac{(\mathbf{e}_F(x^1) - \mathbf{e}_F(x^2))}{\|\Pi_K(x^1) - \Pi_K(x^2)\|} = 0. \qquad (by \quad (1))
\end{aligned}$$

Convergence rate result

Theorem 1 Suppose that x^* is a strongly-stable solution of VI(K,F) with F continuously differentiable around x^*. There exists a scalar $\bar{\epsilon}$ such that for every $\epsilon \in (0, \bar{\epsilon}]$, there exists a scalar $\delta > 0$ such that for every $x^0 \in K \cap B(x^*, \delta)$, the Josephy-Newton method (defined earlier) generates a unique sequence $\{x^k\}$ in $B(x^*, \delta)$ that converges Q-superlinearly to x^*. If JF is Lipschitz continuous near x^*, the convergence rate is Q-quadratic.

Proof:

- By Lemma 1 (Lemma 7.3.2) and Th. 7.2.20, there exists $\bar{\epsilon}$ such that for every $\epsilon \in (0, \bar{\epsilon}]$, a scalar δ_z exists such that if z^1 is chosen from B(z^*, δ_z), a sequence $\{z^k\}$ is uniquely defined and converges to z^*, where z^{k+1} is given by

$$F(x^k) + JF(x^k)(\Pi_K(z^{k+1}) - x^k) - \Pi_K(z^{k+1}) + z^{k+1} = 0,$$

Moreover, the convergence rate is Q-quadratic if JF is Lipshitz near x^* - essentially, this result emerges from applying the convergence theory of Nonsmooth Newton methods on composite functions, specifically the normal map. We need to extend this result to the sequential VI approach.

- Let $\epsilon \in (0, \bar{\epsilon}]$ be given. Need to show a δ exists such that if $x^0 \in K \cap B(x^*, \delta)$, then the Josephy-Newton method generates a unique vector x^1.

- By the C^1 nature of F near x^*, the affine function

$$F^*(x) \equiv F(x^*) + JF(x^*)(x - x^*)$$

is a strong first-order approximation (FOA) at x^*.

- Since x^* is a strongly stable solution of VI(K,F), it is also a strongly stable solution of VI(K,F^*). Thus there exists η, δ_x, c, $\delta_x < \epsilon/2$ such that for every $G \in B(F^*; \eta, K_{\mathcal{N}})$, $\mathcal{N} = B(x^*, \delta_x)$, it holds that $SOL(K,G) \cap \mathcal{N}$ is a singleton (this follows from Def. 5.2.7). Moreover, if x_G is this unique element, then

$$\|x_G - x^*\| \leq c\|e(x_G)\|,$$

where $e(x) \equiv F^*(x) - G(x)$.

- Furthermore,

$$\begin{aligned} F^*(x) - F^0(x) &= F(x^*) - F(x^0) + JF(x^*)(x - x^*) - JF(x^0)(x - x^0) \\ &= -\mathbf{e}_F(x^0) + (JF(x^*) - JF(x^0))(x - x^0). \end{aligned}$$

- Since JF is continuous near x^*, there exists a $\delta \in (0, \epsilon/2)$ such that with $x^0 \in B(x^*, \delta)$, we have $F^0 \in B(F^*; \eta, K_{\mathcal{N}})$. Thus $SOL(K, F^0) \cap B(x^*; \delta_x)$ is a singleton; let x^1 be its unique element and we have

$$\|x^1 - x^0\| \leq \|x^1 - x^*\| + \|x^* - x^0\| \leq \delta_x + \delta < \epsilon.$$

- Furthermore, we have

$$
\begin{aligned}
\|x^1 - x^*\| &\leq c\|e(x^1)\| \\
&= c\|F^*(x^1) - F^0(x^1)\| \\
&= c\|F(x^*) + JF(x^*)(x^1 - x^*) - F(x^0) - JF(x^0)(x^1 - x^0)\| \\
&= c\|F(x^*) + JF(x^*)(x^0 - x^*) + JF(x^*)(x^1 - x^0) - F(x^0) \\
&- JF(x^0)(x^1 - x^0)\| \\
&= c\| - e_F(x^0) + JF(x^*)(x^1 - x^0) - JF(x^0)(x^1 - x^0)\| \\
&\leq c\| - e_F(x^0)\| + c\|(JF(x^*) - JF(x^*))(x^1 - x^0)\| \\
&\leq c\|e_F(x^0)\| + c\|(JF(x^*) - JF(x^0))\|\|(x^1 - x^0)\| \\
&\leq c[\|e_F(x^0)\| + \epsilon\|(JF(x^*) - JF(x^0))\|].
\end{aligned}
$$

- Letting $z^1 = x^1 - F(x^1), x^1 = \Pi_K(z^1)$. Therefore, we have

$$\begin{aligned} \|z^1 - z^*\| &\leq \|x^1 - x^*\| + \|F(x^1) - F(x^*)\| \\ &\leq (1 + L)\|x^1 - x^*\|, \end{aligned}$$

  where L is the local Lipschitz modulus of F near x^*. Therefore by making δ arbitrarily small, we can ensure that $\|JF(x^*) - JF(x^0)\|$ and $\|\mathbf{e}_F(x^0)\|$ can be made arbitrarily small and therefore ensure that z^1 belongs to $B(z^*, \delta_z)$.

- What remains to be shown is that the convergence properties of $\{z^k\}$ to z^* translate into analogous properties relating convergence of $\{x^k\}$ to x^*. Since $x^{k+1} = \Pi_K(z^{k+1})$, we have

$$\|x^{k+1} - x^*\| \leq \|z^{k+1} - z^*\|.$$

which implies that $\{x^k\}$ converges to x^*.

- Since $z^k = x^k - F(x^k)$ for $k \geq 1$, we have

$$\|z^k - z^*\| \leq (1 + L)\|x^k - x^*\|,$$

implying that

$$\frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} \leq (1 + L)^{-1} \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|},$$

which proves the Q-superlinear convergence of $\{x^k\}$ (since $\{z^k\}$ converges to z^* Q-superlinearly).

- Quadratic convergence may be shown with a similar inequality. ∎

Related results

- Similar results can be shown if we only require stability of the limit point (as opposed to stability) - give up uniqueness of the sequence in such instances

- Inexact results can also be proved.

Specialization to mixed complementarity problems

- In such instances, an MiCP is solved by a sequence of sub-MLCPs
- Consider the MiCP arising from KKT system of VI, where K is

$$K \equiv \{x \in \mathbb{R}^n : h(x) = 0, g(x) \leq 0\},$$

and g, h are C^1.

- Specifically consider the system

$$\begin{aligned} \mathbf{L}(x, \mu, \lambda) &= 0 \\ h(x) &= 0 \\ 0 \leq \lambda \perp g(x) &\leq 0, \end{aligned}$$

where $\mathbf{L}$ is the vector Lagrangian of the VI and is defined as

$$\mathbf{L}(x,\mu,\lambda) \equiv F(x) + \sum_{j=1}^{\ell} \mu_j \nabla h_j(x) + \sum_{j=1}^{m} \lambda_i \nabla g_j(x).$$

- The above KKT system is equivalent to a VI in which

$$F = \begin{pmatrix} \mathbf{L}(x,\mu,\lambda) \\ -h(x) \\ -g(x) \end{pmatrix}.$$

- Writing $z \equiv (x, \mu, \lambda)$, we have

$$
\begin{aligned}
F^k(z) &= F(z^k) + JF(z^k)(z - z^k) \\
&= \begin{pmatrix} \mathbf{L}(z^k) \\ -h(x^k) \\ -g(x^k) \end{pmatrix} + \begin{pmatrix} \nabla_x \mathbf{L}(z^k) & Jh(x^k)^T & Jg(x^k)^T \\ -JH(x^k) & 0 & 0 \\ -Jg(x^k) & 0 & 0 \end{pmatrix} \begin{pmatrix} x - x^k \\ \mu - \mu^k \\ \lambda - \lambda^k \end{pmatrix} \\
&= \begin{pmatrix} F(z^k) \\ -h(x^k) \\ -g(x^k) \end{pmatrix} + \begin{pmatrix} \nabla_x \mathbf{L}(z^k) & Jh(x^k)^T & Jg(x^k)^T \\ -JH(x^k) & 0 & 0 \\ -Jg(x^k) & 0 & 0 \end{pmatrix} \begin{pmatrix} x - x^k \\ \mu \\ \lambda \end{pmatrix}.
\end{aligned}
$$

- The sub-VI given by VI$(\mathbb{R}^{n+\ell} \times \mathbb{R}^m_+, F^k)$ may be considered as an MLCP which is the KKT system of the $AVI(K(x^k), q^k, \nabla_x \mathbf{L}(z^k))$, where

$$\begin{aligned} K(x^k) \equiv \{x \in \mathbb{R}^n : h(x^k) + Jh(x^k)(x - x^k) = 0 \\ g(x^k) + Jg(x^k)(x - x^k) \leq 0.\} \\ q(x^k) \equiv F(x^k) - \nabla \mathbf{L}(z^k)x^k. \end{aligned}$$

- Constraint set $K(x^k)$ is formed as a polyhedron obtained by linearization of the original VI. Instead of just using JF, linearization includes Hessians of the constraints:

$$F(x) \approx F(x^k) + \left(JF(x^k) + \sum_{j=1}^{\ell} \mu_j \nabla^2 h_j(x^k) + \sum_{j=1}^{m} \lambda_j \nabla^2 g_j(x^k) \right) (x - x^k).$$