Lecture 19

Algorithms for VIs

KKT Conditions-based Ideas

November 16, 2008
Outline for solution of VIs

- Algorithms for general VIs

- Two basic approaches:
  - First approach reformulates (and solves) the KKT conditions directly (as a complementarity problem) - useful when $K$ has a complex algebraic characterization
  - Second approach is inherently a primal framework that uses an appropriately defined merit function - a good idea when $K$ is simple (next and last lecture)
KKT Conditions based Methods

- Consider the VI(K,F) which requires an \( x^* \) such that
  \[
  (x - x^*)^T F(x^*) \geq 0, \quad \forall x \in K.
  \]

- Suppose \( K \) is finitely representable as
  \[
  K = \{ x \in \mathbb{R}^n : h(x) = 0, g(x) \leq 0 \},
  \]
  where \( h : \mathbb{R}^n \to \mathbb{R}^\ell \) and \( g : \mathbb{R}^n \to \mathbb{R}^m \) are continuously differentiable.

- Then the KKT conditions for VI(K,F) are
  \[
  (\text{KKT}) \quad F(x) + JH(x)^T \mu + Jg(x)^T \lambda = 0
  \]
  \[
  h(x) = 0
  \]
  \[
  0 \geq g(x) \perp \lambda \geq 0.
  \]

- For the remainder of this section we assume that
  - \( F \in C^1 \) on \( \mathbb{R}^n \)
  - \( h, g \in C^2 \) on \( \mathbb{R}^n \)
Using the FB function

• A reformulation of (KKT) that uses the FB function is given by

\[
0 = \Phi_{FB}(x, \mu, \lambda) \equiv \begin{pmatrix}
L(x, \mu, \lambda) \\
-h(x) \\
\psi_{FB}(-g_1(x), \lambda) \\
\vdots \\
\psi_{FB}(-g_m(x), \lambda)
\end{pmatrix},
\]

where \( L(x, \mu, \lambda) \equiv F(x) + JH(x)^T \mu + Jg(x)^T \lambda \) is the VI-Lagrangian function

• The natural merit function is given by

\[
\theta_{FB}(x, \mu, \lambda) \equiv \frac{1}{2} \Phi_{FB}(x, \mu, \lambda)^T \Phi_{FB}(x, \mu, \lambda).
\]

• By the results of the last 2 lectures:
  • \( \Phi_{FB} \) and \( \theta_{FB} \) are semismooth and continuously differentiable, respectively on \( \mathbb{R}^{n+\ell+m} \).
  • Moreover if \( JF, \nabla^2 g_i \) and \( \nabla^2 h_i \) are locally Lipschitz then \( \Phi_{FB} \) is strongly semismooth.
Nonsingularity of elements of Generalized Clarke Jacobian

- Next, we discuss the nonsingularity of all elements in $\partial \Phi_{FB}(x, \mu, \lambda)$
- We also relate these to the strong regularity of the KKT system given by (KKT)

**Proposition 1 (10.1.1)** The generalized Clarke Jacobian $\partial \Phi_{FB}(x, \mu, \lambda)$ is contained in the following family of matrices:

$$A(x, \mu, \lambda) \equiv \left\{ \begin{pmatrix} J_x L(x, \mu, \lambda) & J_h(x)^T & J_g(x)^T \\ -J_h(x) & 0 & 0 \\ -D_g(x, \lambda) J_g(x) & 0 & D_\lambda(x, \lambda) \end{pmatrix} \right\},$$

where

$$D_\lambda(x, \lambda) = \text{diag}(a_1(x, \lambda), \ldots, a_m(x, \lambda))$$

and

$$D_g(x, \lambda) = \text{diag}(b_1(x, \lambda), \ldots, b_m(x, \lambda))$$

are $m \times m$ diagonal matrices whose diagonal elements are given by

$$(a_i(x, \lambda), b_i(x, \lambda)) \begin{cases} \equiv \frac{\lambda_i - g_i(x)}{\sqrt{g_i^2(x) + \lambda_i^2}} - (1, 1), & (g_i(x), \lambda_i) \neq (0, 0) \\ \in clB(0, 1) - (1, 1), & (g_i(x), \lambda_i) = 0. \end{cases}$$
Moreover, $\mathcal{A}(x, \mu, \lambda)$ is a linear Newton approximation of $\Phi_{FB}$ at $(x, \mu, \lambda)$ which is strong if $JF$ and each $\nabla g_i^2$ and $\nabla^2 h_j$ are locally Lipschitz continuous at $x$.

**Proof:** Omitted (follows easily from results in Ch. 7).
Index sets of relevance

- Next we introduce several index sets that will aid in the discussion:

\[ I_0 \equiv \{ i \in I : g_i(x) = 0 \leq \lambda_i \} \]
\[ I_< \equiv \{ i \in I : g_i(x) < 0 = \lambda_i \} \]
\[ I_R \equiv I \setminus (I_0 \cup I_<) \]

- Note that \( i \in I_R \) if and only if one of the following hold:

\[
\begin{cases}
  g_i(x) > 0 & \text{or infeasibility of g} \\
  \lambda_i < 0 & \text{or infeasibility of multipliers} \\
  (-g_i(x), \lambda_i) > 0 & \text{no complementarity}
\end{cases}
\]

- If \( 0 \geq g_i(x) \perp \lambda_i \geq 0 \), then \( I_R \) is empty. More generally

\[ \psi_{FB}(-g_i(x), \lambda) = 0 \iff i \in I_0 \cup I_<. \]
• We may also define $I_{00}, I_+ \subseteq I_0$ given by

\[
I_{00} = \{i \in I_0 : \lambda_i = 0\} \quad \text{no strict complementarity}
\]
\[
I_+ = \{i \in I_0 : \lambda_i > 0\}. \quad \text{strict complementarity}
\]

• We further refine $I_{00}$ with the particular elements $a_i(x, \lambda), b_i(x, \lambda)$ of the generalized Clarke Jacobian $\partial \Phi_{FB}(x, \mu, \lambda)^*$:

\[
I_{01} = \{i \in I_{00} : a_i(x, \lambda) = 0\}
\]
\[
I_{02} = \{i \in I_{00} : \max(a_i(x, \lambda), b_i(x, \lambda)) < 0\}
\]
\[
I_{03} = \{i \in I_{00} : b_i(x, \lambda) = 0\}
\]

• If we let $I_{R2} \equiv I_R \cup I_{02}$, then the following relationships hold:

\begin{itemize}
  \item $I = I_0 \cup I_\prec \cup I_R$ (the whole set of indices)
  \item $I_0 = I_{00} \cup I_+$
  \item $I_{00} = I_{01} \cup I_{02} \cup I_{03}$
  \item implying that $I = I_+ \cup I_{01} \cup I_{R2} \cup I_{03} \cup I_\prec$
\end{itemize}

*Recall that $(a_i, b_i) \in \text{cl}(B(0, 1) \setminus (1, 1))$
Moreover we have
\[
(D_g)_{ii} = 0 \quad \text{and} \quad (D_\lambda)_{ii} = -1, \quad \forall i \in I_\prec \cup I_{03} \quad \text{(from definition)}
\]
\[
i \in I_\prec \implies a_i = -1 \implies (D_\lambda)_{ii} = -1
\]
\[
i \in I_{03} \implies b_i = 0 \implies (D_g)_{ii} = 0
\]

Similarly, we have
\[
(D_g)_{ii} = -1 \quad \text{and} \quad (D_\lambda)_{ii} = 0, \quad \forall i \in I_+ \cup I_{01}
\]
\[
\max((D_g)_{ii}, (D_\lambda)_{ii}) < 0, \quad \forall i \in I_{R2} = I_R \cup I_{02}.
\]

- Useful to define a family of matrices closely related to the strong stability conditions.

For any subset \( \mathcal{J} \subseteq I_{00} \cup I_R \),
\[
M_{FB}(\mathcal{J}) \equiv \begin{pmatrix}
J_x L & J_h^T & J_{g_+}^T & J_g^T_{\mathcal{J}} \\
-J_h & 0 & 0 & 0 \\
-J_{g_+} & 0 & 0 & 0 \\
-J_g & 0 & 0 & 0
\end{pmatrix}.
\]
Equivalent statement of strong stability

**Lemma 1 (10.1.2)** A KKT triple \((x^*, \mu^*, \lambda^*)\) is a strongly-stable KKT point of \(VI(K,F)\) if and only if for all subsets \(J \subseteq I_{00}\), the determinants of the matrices \(M_{FB}(J)\) have the same nonzero sign.\(^\dagger\)

**Proof:**

\((\Rightarrow)\) Suppose that \((x^*, \mu^*, \lambda^*)\) is a strongly stable KKT triple. By corollary 5.3.2., it follows that

\[
B \equiv \begin{pmatrix}
JL & Jh^T & Jg_+^T \\
-Jh & 0 & 0 \\
-Jg_+ & 0 & 0
\end{pmatrix}
\]

is nonsingular and the Schur complement

\[
C \equiv \begin{pmatrix}
Jg_{00} & 0 & 0
\end{pmatrix} B^{-1} \begin{pmatrix}
Jg_{00}^T \\
0 \\
0
\end{pmatrix}
\]

\(^\dagger\) Why do we only need to restrict ourselves to \(I_{00}\)?
is a P matrix. Let $\mathcal{J}$ be an arbitrary subset of $I_{00}$. By the Schur determinantal formula, we have

$$\det M_{FB}(\mathcal{J}) = \det B \det (M_{FB}(\mathcal{J})/B).$$

The Schur complement $(M_{FB}(\mathcal{J})/B)$ is a principal submatrix of $C$. But $C$ is a P-matrix implying that every principal submatrix has a positive determinant or $\det (M_{FB}(\mathcal{J})/B) > 0$. But this implies that $\text{sgn} \det M_{FB}(\mathcal{J}) = \text{sgn} \det B$ and is independent of $\mathcal{J}$.

$(\Leftarrow)$ Omitted.

- This suggests a possible definition for an arbitrary triple $(x, \mu, \lambda)$. If this triple is a KKT triple, then it is equivalent to the earlier definition. The ESSC defined below extends this to the case of strong-stability of a non-KKT triple.
**Extended Strong-Stability Condition (ESSC)**

- **Definition 1 (10.1.3)** The Extended Strong-Stability Condition (ESSC) is said to hold at a triple \((x, \mu, \lambda) \in \mathbb{R}^{n+\ell+m}\) if
  
  (a) \(\text{sgn det } M_{FB}(\mathcal{J})\) is a nonzero constant for all index sets \(\mathcal{J} \subseteq I_{00}\);
  
  (b) \(\det M_{FB}(\mathcal{J}) \det a M_{FB}(\mathcal{J}') \geq 0\) for any two index set \(\mathcal{J}\) and \(\mathcal{J}'\) that are subsets of \(I_{00} \cup I_{R}\).

- Equivalently, the ESSC holds at \((x, \mu, \lambda)\) if and only if
  
  (i) The matrix
  
  \[
  \mathbf{B} = \begin{pmatrix}
  J_x \mathbf{L} & J_h^T & J_g^T \\
  -J_h & 0 & 0 \\
  -J_g^+ & 0 & 0
  \end{pmatrix} = M_{FB}(\emptyset)
  \]

  is nonsingular

  (ii) the \(M_{FB}(\mathcal{J})\) is nonsingular for all nonempty subsets \(\mathcal{J} \subseteq I_{00}\).

  (iii) for all subsets \(\mathcal{J} \subseteq I_{00} \cup I_{R}\) for which which \(M_{FB}(\mathcal{J})\) is nonsingular, we have

  \[
  \text{sgn } \det M_{FB}(\mathcal{J}) = \text{sgn } \det \mathbf{B}.
  \]
• By considering the nonsingularity of the matrix $M_{FB}(I_{00})$, it follows that if ESSC holds at $(x, \mu, \lambda)$ then the gradients

$$\{\nabla h_j, j = 1, \ldots, \ell\} \cup \{\nabla g_i : i \in I_0\}$$

are linearly independent.

• Essentially, the ESSC is a sufficient condition for the matrices in the generalized Jacobian to be nonsingular.

**Theorem 1** Suppose the ESSC holds at a triple $(x, \mu, \lambda)$. All the elements in $A(x, \mu, \lambda)$ (and therefore all matrices in the generalized Jacobian $\partial \Phi_{FB}(x, \mu, \lambda)$) are nonsingular.

**Proof sketch:** Consider an arbitrary matrix $H$ in $A(x, \mu, \lambda)$. By definition, such a
matrix has the following structure (by breaking into index sets):

\[
H = \begin{pmatrix}
J_x L & J_h^T & J g_+^T & J g_{01}^T & J g_{R2}^T & J g_{03}^T & J g_{<}^T \\
-J_h & 0 & 0 & 0 & 0 & 0 & 0 \\
-J g_+ & 0 & 0 & 0 & 0 & 0 & 0 \\
-J g_{01} & 0 & 0 & 0 & 0 & 0 & 0 \\
-(D_g)_{R2} J g_{R2} & 0 & 0 & 0 & (D_\lambda)_{R2} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -I & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -I \\
\end{pmatrix},
\]

where \((D_g)_{R2}\) and \((D_\lambda)_{R2}\) are negative definite diagonal matrices. Moreover \(H\) is
nonsingular if the following matrix is nonsingular:

\[
\bar{H} = \begin{pmatrix}
J_x \mathbf{L} & J_h^T & J_g_+^T & J_{g01}^T & J_{g_R2}^T \\
-Jh & 0 & 0 & 0 & 0 \\
-Jg_+ & 0 & 0 & 0 & 0 \\
-Jg_{01} & 0 & 0 & 0 & 0 \\
-Jg_{R2} & 0 & 0 & 0 & (D_g)^{-1}_{R2}(D_{\lambda})_{R2}
\end{pmatrix}
\]

\[
= \begin{pmatrix}
J_x \mathbf{L} & J_h^T & J_g_+^T & J_{g01}^T & J_{g_R2}^T \\
-Jh & 0 & 0 & 0 & 0 \\
-Jg_+ & 0 & 0 & 0 & 0 \\
-Jg_{01} & 0 & 0 & 0 & 0 \\
-Jg_{R2} & 0 & 0 & 0 & 0
\end{pmatrix} + \begin{pmatrix}
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & (D_g)^{-1}_{R2}(D_{\lambda})_{R2}
\end{pmatrix}
\]

\[
= \bar{M}_{FB}(\{I_{01} \cup I_{R2}\}) + D.
\]

- Next, we employ the following determinantal formula: For any square matrix \(A\) of order \(r\) and any diagonal matrix \(D\) of the same order,

\[
\det(D + A) = \sum_{\alpha} \det D_{\alpha\alpha} \det A_{\bar{\alpha}\bar{\alpha}},
\]

where the summation ranges for all subsets \(\alpha\) of \(\{1, \ldots, r\}\) with complements \(\bar{\alpha}\).
• It follows that

\[
\det \tilde{H} = \sum_{\alpha \subseteq I_{R^2}} \det(( (D_g)^{-1}(D_\lambda)_{R^2})_{\alpha\alpha} \det M_{FB}(\mathcal{J})_{\alpha'\alpha'},
\]

where the summation is over all subsets \( \alpha \) of \( I_{R^2} \) and \( \alpha' \equiv \mathcal{J}\setminus\alpha \).

• The final step requires invoking the ESSC to show that this determinantal form is indeed positive. (omitted)
Necessity of ESSC

• We see that ESSC is a sufficient condition for nonsigularity of the elements in the generalized Jacobian. If the triple satisfies the KKT conditions, then the ESSC is also necessary:

**Theorem 2** Let \((x^*, \mu^*, \lambda^*)\) be a KKT triple of \(K, F\). Then the following are equivalent:

(a) The ESSC holds at \((x^*, \mu^*, \lambda^*)\).

(b) All matrices in \(A(x^*, \mu^*, \lambda^*)\) are nonsingular.

(b) All matrices in \(\partial \Phi_{FB}(x^*, \mu^*, \lambda^*)\) are nonsingular.

**Proof:**

• From the earlier result, we have \((a) \implies (b) \implies (c)\). It suffices to prove that \((c) \implies (a)\).

• We assume that all the matrices in \(\partial \Phi_{FB}(x^*, \mu^*, \lambda^*)\) are nonsingular. We first show that the set of gradients

\[
\{\nabla g_i(x^*) : i \in I_{00}\}
\]
are linearly independent. Define a sequence \( \{\lambda^k\} \) as

\[
\lambda^k_i \equiv \begin{cases} 
\lambda_i^*, & \text{if } i \in I \setminus I_{00} \\
1/k, & \text{if } i \in I_{00}
\end{cases}
\]

The sequence \( \{\lambda^k\} \) converges to \( \lambda^* \); in addition for each \( k \), \( \Phi_{FB} \) is continuously differentiable at \( (x^*, \mu^*, \lambda^k) \). The sequence of matrices \( \{J\Phi_{FB}(x^*, \mu^*, \lambda^k)\} \) converges to a matrix \( H \in \partial PB(x^*, \mu^*, \lambda^*) \) with \( I_{01} = I_{00} \) and \( I_{02} = I_{03} = \emptyset \). The matrix \( H \) is nonsingular by definition and the linear independence of the gradients follows (see earlier proof.)

- Next, we show that for any subset \( J \subseteq I_{00} \), there exists an element in \( \partial \Phi_{FB}(x^*, \mu^*, \lambda^*) \) such that \( I_{01} = J \), \( I_{02} = \emptyset \) and \( I_{03} = I_{00} \setminus J \). Since the gradients are linearly independent, the classical implicit function theorem implies that there is a sequence \( \{x^k\} \) converging to \( x^* \) such that \( g_i(x^k) > 0 \) for all \( i \in I_{00}, \forall k \). For each \( k \), we define

\[
\lambda^k_i \equiv \begin{cases} 
\lambda_i^*, & \text{if } i \in I \setminus I_{00} \\
\sqrt{g_i(x^k)}, & \text{if } i \in J \\
g_i(x^k)^2, & \text{if } i \in I_{00} \setminus J.
\end{cases}
\]
The sequence \( \{ (x^k, \mu^*, \lambda^k) \} \) converges to \((x^*, \mu^*, \lambda^*)\) and \( \Phi_{FB} \) is continuously differentiable at \((x^k, \mu^*, \lambda^k)\) since there is no index \( i \) such that \( g_i(x^k) = \lambda^k_i = 0 \). By using prop. 10.1.1. and the continuity of the involved functions, it follows that \( \{ J \Phi_{FB}(x^k, \mu^k, \lambda^k) \} \) converges to an \( H \) with the desired properties.

- **Proposition 2 (7.1.18)** Let \( G : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^n \) be Lipschtiz in nbhd of \((\bar{x}, \bar{y})\) for which \( G(\bar{x}, \bar{y}) = 0 \). Assume that all matrices in \( \pi_x \partial G(\bar{x}, \bar{y}) \) are nonsingular. Then there exist open nbhds \( U \) and \( V \) of \( \bar{x} \) and \( \bar{y} \), respectively such that for every \( y \in V \), the equation \( G(x, y) = 0 \) has a unique solution \( x = F(y) \in U \), \( F(\bar{y}) = \bar{x} \) and the map \( F : V \to U \) is Lipschtiz.

- Since all the matrices in \( \partial \Phi_{FB}(x^*, \mu^*, \lambda^*) \) (a convex set) are nonsingular, it follows that such matrices have the same nonzero determinantal sign. Suppose not. Then if \( H_1 \) and \( H_2 \) are two matrices in \( \partial \Phi^*_{FB} \) such that \( \det H_1 < 0 \) and \( \det H_2 > 0 \). Then there exists an \( \bar{H} \) such that \( \det \bar{H} = 0 \) where \( \bar{H} \) is a convex combination of \( H_1 \) and \( H_2 \). This contradicts the nonsingularity of all elements in \( \partial \Phi_{FB}(x^*, \mu^*, \lambda^*) \).

- To complete the proof, we need so to show that the matrices \( M_{FB}(\mathcal{J}) \) have the same nonzero determinantal sign for all \( \mathcal{J} \subseteq I_{00} \). For each index set, let \( H(\mathcal{J}) \in \partial \Phi_{FB}(x^*, \mu^*, \lambda^*) \) such that \( \mathcal{J} = I_{01} \) and \( I_{03} = I_{00} \setminus I_{01} \). From the
structure of $H(J)$, we have

\[
sgn \det H(J) = (-1)^m sgn \det \begin{pmatrix}
JL & Jh^T & Jg_+^T & Jg_+^T \\
-Jh & 0 & 0 & 0 \\
-Jg_+ & 0 & 0 \\
-Jg_J & 0 & 0
\end{pmatrix}
\]

\[
= (-1)^m sgn \det M_{FB}(J)
\]

Since all the matrices in $\partial PB(x^*, \mu^*, \lambda^*)$ have determinants of the same nonzero sign, the ESSC holds.

**Corollary 1** A KKT triple $(x^*, \mu^*, \lambda^*)$ is strongly stable if and only all the matrices in $\partial \Phi_{FB}(x^*, \mu^*, \lambda^*)$ are nonsingular.
Example

• Let $n = 1$, $\ell = 0$, $m = 1$ and $g_1(x) \equiv -x$. Thus the VI is a 1-dimensional LCP(q,M).

• Case 1: $q = -1$, $M = 1 \implies F(x) \equiv x - 1$. Then

$$\Phi_{FB}(x, \lambda) = \begin{pmatrix} \frac{x - 1 - \lambda}{\sqrt{x^2 + \lambda^2}} - x - \lambda \end{pmatrix}.$$ 

Consider the KKT pair (0,0) implying that $I_{00} = \{1\}$, $I_+ = I_R = \emptyset$. Moreover, $M_{FB}(\emptyset) = 1$ and

$$M_{FB}(\{1\}) = \begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix}$$ 

implying that ESSC is satisfied. From the expression of $\Phi_{FB}$, the elements of $\partial \Phi_{FB}(0,0)$ are given by

$$\begin{pmatrix} 1 & -1 \\ a - 1 & b - 1 \end{pmatrix}, a^2 + b^2 \leq 1.$$ 

These matrices are nonsingular (can be seen by showing that the determinant is nonzero).
• Case 2: $q = 0, M = -1, F(x) = -x$ Then

$$\Phi_{FB}(x, \lambda) = \begin{pmatrix} x - \lambda \\ \sqrt{x^2 + \lambda^2} - x - \lambda \end{pmatrix}.$$  

Consider the KKT pair $(0,0)$ implying that $I_{00} = \{1\}, I_+ = I_R = \emptyset$. Moreover, $M_{FB}(\emptyset) = -1$ and

$$M_{FB}(\{1\}) = \begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix}$$

implying that ESSC is not satisfied (since the signs of the determinants are not the same). From the expression of $\Phi_{FB}$, the elements of $\partial \Phi_{FB}(0, 0)$ are given by

$$\begin{pmatrix} -1 & -1 \\ a - 1 & b - 1 \end{pmatrix}, a^2 + b^2 \leq 1.$$  

Singularity of this matrix follows when $a = b = 1/\sqrt{2}$. 