Lecture 20

Algorithms for VIs

Merit Functions for VIs

November 16, 2008

Outline

- Merit functions for VIs
- D-gap merit functions for VIs
- Note on constructing algorithms

Introduction

- Earlier we considered the merit function

$$\theta_{\text{gap}}(x) \equiv \sup_{y \in K} F(x)^T (x - y).$$

- $\theta_{\text{gap}}(x) \geq 0 \quad \forall x \in K$.

- Furthermore, a zero of $\theta_{\text{gap}}(x)$ that lies in K must solve VI(K,F)

- Problem: non-differentiable and extended real-valued in general(Note latter will not hold if K is bounded)

Loss of differentiability

Theorem 1 Let K be a nonempty closed subset and let Ω be a nonempty open set. Assume that $f : \Omega \times K \to \mathbb{R}$ is continuous on $\Omega \times K$ and that $\nabla_x f(x, y)$ exists and is continuous on $\Omega \times K$. Define the function $g : \Omega \to \mathbb{R} \cup \{\infty\}$ by

$$g(x) \equiv \sup_{y \in K} f(x, y), \quad x \in \Omega$$

and set

$$M(x) \equiv \{y \in K : g(x) = f(x, y)\}.$$

Let $x \in \Omega$ be a given vector. Suppose that a nbhd $\mathcal{N} \subset \Omega$ of x exists such that $M(x')$ is nonempty for all $x' \in \mathcal{N}$ and the set

$$\cup_{x' \in \mathcal{N}} M(x')$$

is bounded. The following two statements (a) and (b) are valid.

(a) The function g is directionally differentiable at x and

$$g'(x; d) = \sup_{y \in M(x)} \nabla_x f(x, y)^T d.$$

(b) If $M(x)$ reduces to a singleton, say $M(x) = \{y(x)\}$, then g is differentiable at x and

$$\nabla g(x) = \nabla_x f(x, y(x)).$$

Proof of (a):

- Let $d \in \mathbb{R}^n$ be given and suppose that $\bar{t} > 0$ is small enough that $x + td \in \mathcal{N}$ for $t \in (0, \bar{t}]$. By definition, we may write for every $y \in M(x)$,

$$g(x+td) - g(x) \geq f(x+td, y) - f(x,y)$$

- Dividing by t and passing to the limit, we have

$$\liminf_{t \downarrow 0} \frac{g(x+td) - g(x)}{t} \geq \sup_{y \in M(x)} \nabla_x f(x,y)^T d.$$

- Similarly for $y_t \in M(x+td)$ we have

$$g(x+td) - g(x) \leq f(x+td, y_t) - f(x, y_t) = t \nabla_x f(\bar{x}_t, y_t)^T d,$$

for some reasonable $\bar{x}_t \in (x, x+td)$. This implies that

$$\frac{g(x+td) - g(x)}{t} \leq \nabla_x f(\bar{x}_t, y_t)^T d, \qquad \forall y_t \in M(x+td).$$

- Since the union in the definition is bounded, then for every sequence $\{t_k\}$ of positive numbers converging to zero, if $\{y^k\}$ is any sequence of vectors such that $y^k \in M(x + t^k d)$ for every k, then $\{y^k\}$ is bounded and every limit point of this sequence belongs to $M(x)$. It follows that

$$\limsup_{t \downarrow 0} \frac{g(x+td) - g(x)}{t} \leq \sup_{y \in M(x)} \nabla_x f(\bar{x}_t, y_t)^T d, \qquad \forall y_t \in M(x+td).$$

This implies that (a) holds.

Regularized Gap function

Clearly the multiplicity of solutions to maximization problem of the merit function results in a loss of differentiability,

One way to rectify this is by adding a **strongly concave function** to $\theta_{\text{gap}}(x)$. This would imply that the resulting maximization problem would have a unique maximizer.

Definition 1 Let VI(K,F) be given with F defined an an open set Ω containing K. Let $c > 0$ and G be a symmetric positive definite matrix. Then the **regularized gap function** of VI(K,F) is defined as

$$\theta_c(x) \equiv \sup_{y \in K} \{F(x)^T(x-y) - \tfrac{1}{2}c(x-y)^T G(x-y)\}$$

for all $x \in \Omega$.

We note that

- $\theta_b(x) \leq \theta_a(x), \quad \forall x \in \Omega$ for $b > a > 0$.

- $\theta_c(x) \geq 0, \quad \forall x \in K$. (why?)

- Finally, by defining $\psi_c(x,y) \equiv F(x)^T(x-y) - \frac{1}{2}c(x-y)^T G(x-y), \quad (x,y) \in \Omega \times \mathbb{R}$, implying that $\theta_c(x) = \sup_{y \in K} \psi(x,y)$.

- With K being nonempty and closed for each $x \in \Omega$, there exists a unique $y_c(x) \in K$ that maximizes $\psi_c(x,.)$ on K

- Therefore the supremum is achieved in K and we can replace it by the maximum

- By continuous differentiability of F on Ω, the function $\theta_c(x)$ is differentiable on Ω (follows from the continuity of $y_c(x)$) - see next result

Differentiabilut of the regularized gap function

Theorem 2 (10.2.3) Let K be closed convex and $F : \Omega \to \mathbb{R}^n$ be continuous on the open set Ω. Let c be a positive scalar and let G be a symmetric positive definite matrix. Then the following four statements are valid:

(a) For every $x \in \Omega$, $y_c(x) = \Pi_{K,G}(x - c^{-1}G^{-1}F(x))$, where $\Pi_{K,A}(x)$ is the unique solution to the strictly convex program

$$\begin{cases} \text{min} & \frac{1}{2}(y-x)^T A(y-x) \\ \text{subject to} & y \in K. \end{cases}$$

(b) $\theta_c(x)$ is continuous on Ω and nonnegative on K.

(c) $\theta_c(x) = 0, x \in K$ if and only if $x \in SOL(K,F)$.

(d) If F is continuously differentiable on Ω, then so is θ_c; moreover

$$\nabla\theta_c(x) = F(x) + (JF(x) - cG)^T(x - y_c(x)).$$

Proof:

- Since

  $$\psi_c(x,y) = (2c)^{-1}F(x)^T G^{-1}F(x) - \tfrac{1}{2}c(x - c^{-1}G^{-1}F(x) - y)G(x - c^{-1}G^{-1}F(x) - y),$$

  the expression for $y_c(x)$ follows from the definition of the skewed Euclidean projector, namely:

  $$y_c(x) = \Pi_{K,G}(x - c^{-1}G^{-1}F(x))$$

  establishing (a).

  ‘

- Recall that $\Pi_{F,G}$ is a globally Lipschitz continuous function implying that $y_c(x)$ is a continuous function on Ω; hence so is θ_c implying part (b).

- In terms of $y_c(x)$, we have

  $$\theta_c(x) = F(x)^T(x - y_c(x)) - \tfrac{1}{2}c(x - y_c(x))^T G(x - y_c(x))$$

  . Suppose $\theta_c(x) = 0$ and $x \in K$ implying that

  $$F(x)^T(x - y_c(x)) - \tfrac{1}{2}c(x - y_c(x))^T G(x - y_c(x)) = 0$$
  $$F(x)^T(x - y_c(x)) = \tfrac{1}{2}c(x - y_c(x))^T G(x - y_c(x)) \geq 0. \quad (**)$$

On the other hand, by the variational principle of the maximization problem defining $\theta_c(x)$, we have

$$(z - y_c(x))^T (F(x) + cG(y_c(x) - x) \geq 0, \qquad \forall z \in K.$$

Substituting $z = x$ we have

$$(x - y_c(x))^T (F(x) + cG(y_c(x) - x) \geq 0,$$

which together with (**) implies that

$$\frac{1}{2}c(x - y_c(x))^T G(x - y_c(x)) - \frac{1}{2}c(x - y_c(x))^T G(x - y_c(x)) \geq 0$$
$$\implies y_c(x) = x.$$

By (a), we have

$$x = \Pi_{K,F}(x - c^{-1}G^{-1}F(x))$$

which shows that $x \in SOL(K, F)$.

- Conversely if $x \in SOL(K, F)$ we have $\theta_c(x) \leq 0$ implying that $\theta_c(x) = 0$. This establishes (c).

- Finally, the gradient formula follows from theorem 10.2.1. By part (b),

$$\nabla\theta_c(x) = \nabla_x \psi_x(x, y_c(x)).$$

Note that the nonnegativity of the regularized gap function is only on K; for a zero of θ_c to be a solution of VI(K,F), the zero needs to belong to K. Essentially θ_c is a valid merit function only on K.

Regularized gap program

- We may now define a regularized gap program as follows:

$$\begin{aligned} \min \quad & \theta_c(x) \\ \text{subject to} \quad & x \in K. \end{aligned}$$

- What are the necessary and sufficient conditions for a stationary point of this problem to be a solution to VI(K,F).

- Recall that $(-F(x))^*$ represents the dual cone of the singleton $\{-F(x)\}$, i.e.

$$(-F(x))^* \equiv \{d \in \mathbb{R}^n : d^T F(x) \leq 0\}.$$

For a vector $x \in K$.

$$T_c(x; K) \equiv \mathcal{T}(x : K) \cap (-\mathcal{T}(y_c(x); K))$$

and

$$T_c(x; K, F) \equiv T_c(x : K) \cap (-(F(x))^*.$$

- Recall that $\mathcal{T}(x; K)$ represents the tangent cone of K at a vector $x \in K$ where

$$\mathcal{T}(x : K) = \{d \in \mathbb{R}^n : \exists \{y^\nu\} \subset K, \tau^\nu > 0, \lim_{\nu \to \infty} y^\nu = x, \lim_{\nu \to \infty} \tau_\nu = 0 \text{ and } \lim_{\nu \to \infty} \frac{y^\nu - x}{\tau^\nu} = d.\}$$

- Since K is closed and convex, $T_c(x;K)$ is a closed convex cone containing the vector $y_c(x)-x$; $T_c(x;K,F)$ is also a closed convex cone.

- When $x \in SOL(K,F)$, the space given by $T_c(x;K,F)$ reduces to a familiar space. When $x \in SOL(K,F)$, we have $\theta_c(x) = 0$ and thus $x = y_c(x)$ since $y_c(x)$ is the unique vector in K satisfying

$$\theta_c(x) = F(x)^T(x - y_c(x)) + \tfrac{1}{2}c(y_c(x)-x)^T G(y_c(x)-x).$$

- Therefore $T_c(x;K) = \mathcal{T}(x;K) \cap (-\mathcal{T}(x;K))$ which is equal to the lineality space of the tangent cone of K at x.*

- Moreover,

$$\mathcal{T}(x;K) \cap (-F(x))^* = \mathcal{T}(x;K) \cap (F(x))^{\perp}.$$

- The **critical cone** $\mathcal{C}(x;K,F)$ of the pair (K,F) is defined as

$$\mathcal{C}(x;K,F) \equiv \mathcal{T}(x;K) \cap F(x)^{\perp}.$$

. The elements of the critical cone are called the critical vectors of the pair (K,F) at x. In specifying the local uniquness requirements of a solution to a VI, the critical cone is an essential object.

*Recall that the intersection $C \cap (-C)$ is called the lineality subspace of C.

- With a little more effort, we may show that

$$T_c(x; K, F) = \mathcal{C}(x; K), F \cap (-\mathcal{C}(x; K, F)).$$

- Essentially, if $x \in SOL(K, F)$ we have that the $T_c(x; K, F)$ is the lineality space of the critical cone $\mathcal{C}(x; K, F)$ of VI(K,F).

- In the next result, which we state without proof $=$ the cone $T_c(x; K, F)$ needs to be contained in $F(x)^\perp$ for the a vector $x \in K$ to be a solution to $VI(K, F)$.

Theorem 3 Let K be closed convex and F is continuously differentiable on the open set Ω. Let c be a positive scalar and G be symmetric positive definite. Suppose that x is a stationary point of

$$\begin{aligned} \text{min} \quad & \theta_c(x) \\ \text{subject to} \quad & x \in K. \end{aligned}$$

Then the following are equivalent:

(a) x solves VI(K,F)

(b) $T_c(x;K,F)$ is contained in $F(x)^{\perp}$.

(c) The implication below holds:

$$d \in T_c(x;K,F), JF(x)^T d \in T_c(x;K,F)^* \implies d^T F(x) = 0.$$

A related corollary provides a necessary and sufficient condition for x to be a solution to the VI.

Corollary 1 Let K be a closed convex and F be continuously differentiable on the open set Ω. Let c be a positive scalar and let G be a symmetric positive definite matrix. A vector $x \in K$ solves VI(K,F) if and only if x is a stationary point of

$$\begin{aligned} \min \quad & \theta_c(x) \\ \text{subject to} \quad & x \in K. \end{aligned}$$

and

$$d \in T_c(x;K,F), JF(x)^T d \in -T_c(x;K,F)^* \implies d^T F(x) = 0$$

holds.

The D-Gap Merit Function

- We conssider a closed convex K without assuming that it is finitely representable

- The domain of definition of θ_c coincides with that of F - however, the regularized gap program is a constrained minimization problem

- Is there an equivalent unconstrained minimization problem - yes!

  Definition 2 Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be a given mapping and K a closed convex set. Let a, b be positive scalars such that $b > a > 0$. Then the D-gap function is defined as

$$\theta_{ab} \equiv \theta_a(x) - \theta_b(x), \quad \forall x \in \mathbb{R}^n,$$

  whhere D stands for difference.

- **Lemma 1** For every $x \in \mathbb{R}^n$ it holds that

$$\frac{b-a}{2}\|x - y_b\|_G^2 \leq \theta_{ab}(x).$$

Proof: By the definition of the D-gap function, theorem 10.2.3 and simple majorizations we have

$$\begin{aligned}\theta_{ab}(x) &= \sup_{x\in K}\{F(x)^T(x-y) - \tfrac{1}{2}a(x-y)^T G(x-y)\}\\ &- \sup_{x\in K}\{F(x)^T(x-y) - \tfrac{1}{2}b(x-y)^T G(x-y)\}\\ &\geq \{F(x)^T(x-y_b(x)) - \tfrac{1}{2}a(x-y_b(x))^T G(x-y_b(x))\}\\ &- \{F(x)^T(x-y_b(x)) - \tfrac{1}{2}b(x-y_b(x))^T G(x-y_b(x))\}\\ &= \tfrac{1}{2}(b-a)\|x-y_b\|_G^2.\end{aligned}$$

- The next result shows that the D-gap function is truly an unconstrained merit function of the VI(K,F).

Theorem 4 Let F be continuous and K be closed convex subset of $\mathbb{R}^n$. For $b > a > 0$, the D-gap function θ_{ab} is continuous on $\mathbb{R}^n$ and

(a) $\theta_{ab}(x) \geq 0$ for all $x \in \mathbb{R}^n$

(b) $\theta_{ab}(x) = 0$ if and only if $x \in SOL(K, F)$

(c) If F is continuously differentiable on $\mathbb{R}^n$, then so is the D-gap function θ_{ab} and

$$\nabla\theta_{ab}(x) = JF(x)^T(y_b(x) - y_a(x)) + aG(y_a(x) - x) - bG(y_b(x) - x).$$

Proof: Parts (a) and (c) need no proof (see earlier results). Similarly, if $x \in SOL(K, F)$ then $\theta_{ab}(x) = 0$. The necessity condition in (b) can be shown as follows. Assume that $\theta_{ab}(x) = 0$. By the earlier lemma, $x = y_b(x)$. But this implies $x \in K$ and $\theta_b(x) = 0$. But this implies that $x \in SOL(K, F)$ by theorem 10.2.3 (c).

- Finally, since $\theta_{ab})(x)$ is an unconstrained merit function for VI(K,F), we finally proceed to discuss when an unconstrained stationary point of θ_{ab} is a solution of VI. Note that a stationary point need not necessarily be an element of K since θ_{ab} is defined over the entire space (as opposed to constrained stationary points).

- The following result makes this clear.

Theorem 5 (10.3.4) Let F be continuous and K closed convex. Let a and b be scalars with $b > a > 0$ and G be a symmetric positive definite matrix. Suppose that x is a stationary point of $\theta_{ab}(x)$. Then the following are equivalent:

- $x \in SOL(K, F)$
- $T_{ab}(x; K, F)$ is contained in $F(x)^{\perp}$
- The following implication holds:

$$d \in T_{ab}(x; K, F), JF(x)^T d \in -T_{ab}(x; K, F)^* \implies d^T F(x) = 0$$

Merit function-based Algorithms

- A natural question is whether we can use the ideas discussed earlier to construct iterative descent type methods.
- However, we face some problems in that regard:
  - The evaluation of $\theta_a(x)$ and its gradient requires the computation of a Euclidean projection on the set K
  - The evaluation of $\theta_{ab}(x)$ and its gradient requires the computation of two Euclidean projections on the set K
  - Neither is naturally associated with a set of nonsmooth equations (specifically neither of these is associated with a systems of nonsmooth equations as the two-norm squared).
- If K is polyhedral, some of the computational questions are easier to deal with (projection is a strictly convex QP)
- Third point has implications in terms of constructing locally fast methods - thats the effort - find a family of Newton approximation that allow for obtaining directions