Pseudonormality and Informative Lagrange Multipliers

Asuman Ozdaglar and Dimitri Bertsekas

Electrical Engineering and Computer Science Dept.

Massachusetts Institute of Technology

November 20, 2002

Optimality Conditions

- Let x^* be a local minimum of the problem:

$$\begin{aligned} &\text{minimize} \quad && f(x) \\ &\text{subject to} \quad && g_j(x) \leq 0, \ j = 1, \ldots, r. \end{aligned}$$

- Karush-Kuhn-Tucker Opt Conds: Under some conditions, $\exists$ scalars $\mu_j^* \geq 0$, called Lagrange multipliers, such that

$$\nabla f(x^*) + \sum_{j=1}^{r} \mu_j^* \nabla g_j(x^*) = 0,$$

where $\mu_j^* = 0$ for all j with $g_j(x^*) < 0$. (Complementary slackness)

- Main Issue: What is the structure of the constraint set that guarantees the existence of Lagrange multipliers?

Constraint Qualifications

- It is well-known that there exist Lagrange multipliers when the tangent cone has the form

$$\{y \mid \nabla g_j(x^*)'y \leq 0,\ \forall\ j \text{ with } g_j(x^*) = 0\}$$

$\Rightarrow$ Quasiregularity Condition (Proof via Farkas' Lemma)

- Classical work focused on constraint qualifications that guarantee quasiregularity
  - Requires complicated proofs (based on implicit function theorem)
  - This approach fails when there is an additional set constraint.

- Fritz John Optimality Conditions: There exist $\mu_0^* \geq 0$ and $\mu_j^* \geq 0$ such that

$$\mu_0^* \nabla f(x^*) + \sum_{j=1}^{r} \mu_j^* \nabla g_j(x^*) = 0$$

where $\mu_j^* = 0$ for all j with $g_j(x^*) < 0$.

  - *Question now becomes*: When is $\mu_0^* > 0$?

Our Development of Lagrange Multiplier Theory

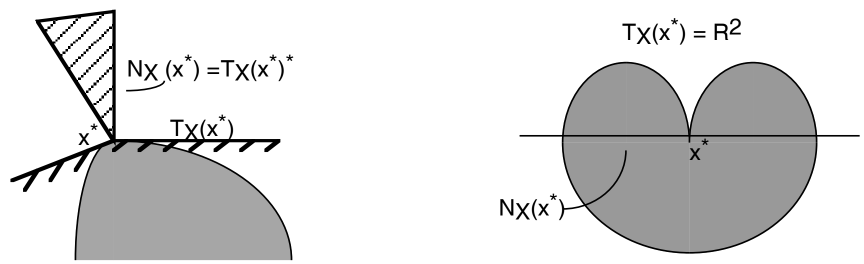
- Consider optimization problems of the form

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in X,\ g_j(x) \leq 0,\ j = 1, \ldots, r. \end{array}$$

- Simple and powerful line of analysis
  - Based on an enhanced set of Fritz John conditions
  - Special Lagrange multipliers that carry significant sensitivity information
  - New conditions that extend and unify constraint qualifications
  - Connections to exact penalty functions
- In the first part of the talk, assume that f and the g_j are smooth functions, and X is a closed set.
- When X is not convex, need machinery from nonsmooth analysis.

Conical Approximations of Constraint Sets

- Polar cone of a set T : $T^* = \{y \mid y'x \le 0,\ \forall\, x \in T\}$
- Two cones to go into the heart of issues about Lagrange multipliers
  - Tangent cone, $T_X(x^*)$: characterizes directions along which some feasible sequence converges to x^*
  - Normal cone, $N_X(x^*)$: cone of limiting normal vectors [$z \in N_X(x)$ if $\exists$ sequences $\{x_k\} \subset X$ and $\{z_k\}$ such that $x_k \to x$, $z_k \to z$, and $z_k \in T_X(x_k)^*$]



- Crucial property of constraint sets: We say that X is regular at x^* when

$$T_X(x^*)^* = N_X(x^*)$$

Definition of Lagrange Multipliers

Let x^* be a local minimum. Then $\mu^* = (\mu_1^*, \ldots, \mu_r^*)$ is a Lagrange multiplier if $\mu_j^* \geq 0$ for all j, $\mu_j^* = 0$ for all j with $g_j(x^*) < 0$, and

- For $X = \Re^n$,

$$\nabla f(x^*) + \sum_{j=1}^{r} \mu_j^* \nabla g_j(x^*) = 0.$$

- For X any convex closed set,

$$\left(\nabla f(x^*) + \sum_{j=1}^{r} \mu_j^* \nabla g_j(x^*)\right)'(x - x^*) \geq 0, \qquad \forall\, x \in X.$$

- For X any closed set,

$$\left(\nabla f(x^*) + \sum_{j=1}^{r} \mu_j^* \nabla g_j(x^*)\right)' y \geq 0, \qquad \forall\, y \in T_X(x^*).$$

Enhanced Fritz John Optimality Conditions

Proposition: Let x^* be a local minimum. Then there exist multipliers μ_0^*, μ_j^* s.t.

1. $-\left(\mu_0^* \nabla f(x^*) + \sum_{j=1}^r \mu_j^* \nabla g_j(x^*)\right) \in N_X(x^*)$.
2. $\mu_j^* \geq 0$ for all j, and $\mu_0^*, \ldots, \mu_j^*$ are not all equal to 0.
3. If the set $J = \{j \neq 0 \mid \mu_j^* > 0\}$ is nonempty, arbitrarily close to x^*, $\exists$ some $x \in X$ s.t.:

$$f(x) < f(x^*), \quad g_j(x) > 0, \ \forall \ j \in J,$$

while violation of constraints with 0 multipliers is arbitrarily small.

- Proof using a quadratic penalty function approach.
- We call condition (3) the complementary violation (CV) condition $\Rightarrow$ Stronger than complementary slackness condition:

$$g_j(x^*) = 0, \qquad \forall \ j \in J.$$

New Condition - Pseudonormality

A feasible vector x^* is pseudonormal if there are no scalars $\mu_1, \ldots, \mu_r$, and no sequence $\{x^k\} \subset X$ such that

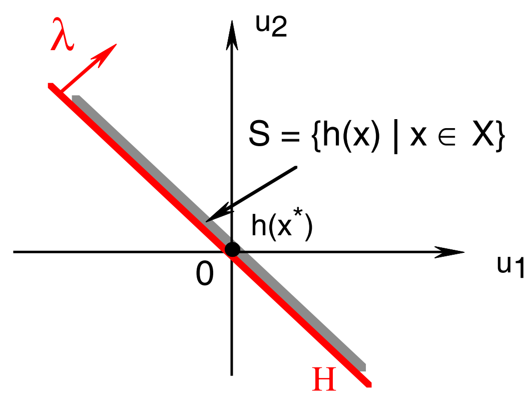
1. $-\left(\sum_{j=1}^r \mu_j \nabla g_j(x^*)\right) \in N_X(x^*)$.
2. $\mu_j \geq 0$, for all j, and $\mu_j = 0$ for all j with $g_j(x^*) < 0$.
3. $\{x^k\}$ converges to x^* and
$$\sum_{j=1}^r \mu_j g_j(x^k) > 0, \qquad \forall\ k.$$

- If x^* is a pseudonormal local minimum and X is regular at x^*, then there exist Lagrange multipliers (of special type).

- Classical CQs for $X = \Re^n$ (such as linear indep. of const. gradients, Slater cond.) can easily be shown to imply pseudonormality.
- Yields new CQs for $X \neq \Re^n$ (extended Mangasarian-Fromovitz cond., conds for problems with linear equality constraints).

Insight into Pseudonormality

- Consider the problem: minimize $f(x)$
  subject to $h_i(x) = 0,\ i = 1, \ldots, m.$
- x^* is pseudonormal if there are no scalars $\lambda_1, \ldots, \lambda_m$ and no sequence $\{x_k\}$ such that
  1. $\sum_{i=1}^m \lambda_i \nabla h_i(x^*) = 0.$
  2. $x_k \to x^*$ and $\sum_{i=1}^m \lambda_i h_i(x_k) > 0.$
- Linear independence of $\nabla h_i(x^*) \Rightarrow$ pseudonormality.
- h_i's affine $\Rightarrow$ pseudonormality.



Informative Lagrange Multipliers

- We say that a Lagrange multiplier μ^* is informative if, in addition to Lagrangian stationarity condition, it satisfies the CV condition, i.e., $\exists$ some $x \in X$ s.t.:

$$f(x) < f(x^*), \quad g_j(x) > 0, \ \forall \ j \text{ with } \mu_j^* > 0,$$

$$g_j(x) = o\left(\min_{j \in J} g_j(x)\right), \ \forall \ j \text{ with } \mu_j^* = 0.$$

  - Provide significant amount of sensitivity interpretation by indicating which constraints to violate to effect a cost reduction.

- Proposition: Assume that $T_X(x^*)$ is convex. If the set of Lagrange multipliers is nonempty, then the Lagrange multiplier with minimum norm is an informative Lagrange multiplier.

Exact Penalty Functions

- Consider the penalized problem for some finite value of c

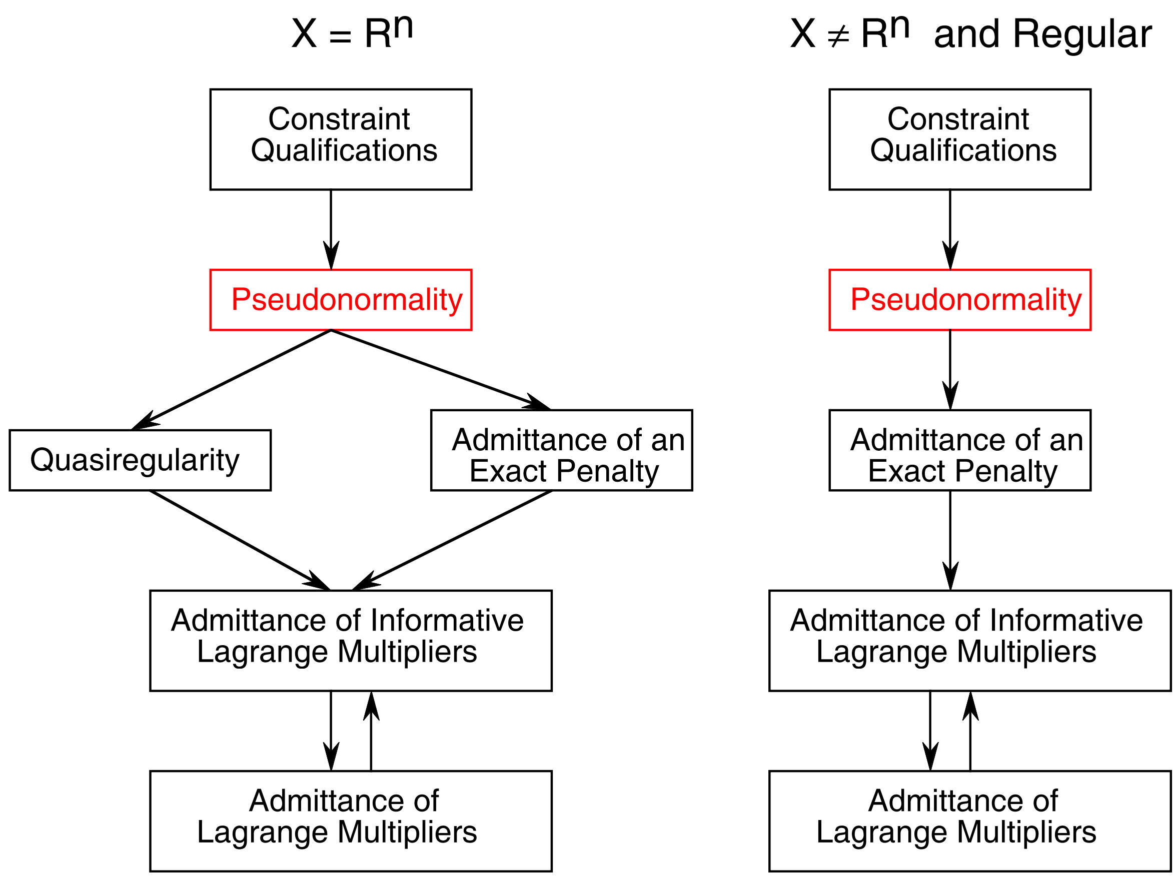
$$\begin{array}{ll} \text{minimize} & F_c(x) = f(x) + c\sum_{j=1}^{r} \max\{0,\ g_j(x)\} \\ \text{subject to} & x \in X. \end{array}$$

- There exists an exact penalty at x^*, if for every smooth f for which x^* is a strict local minimum of original problem, there is a $c > 0$, such that x^* is also a local minimum of F_c over X.

- Proposition: If x^* is pseudonormal, then there exists an exact penalty at x^*.

- Proves in a unified way the existence of an exact penalty for a much larger variety of constraint qualifications.

Unifying Role of Pseudonormality



- Pseudonormality also provides unification in the case when X is not regular.

Multipliers and Convex Programming

Consider the problem

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad x \in X, \; g_j(x) \le 0, \; j = 1, \ldots, r, \end{aligned}$$

where X is a convex set, and f and g_j are convex over X. Assume that optimal value of this problem f^* is finite.

- A vector μ^* is said to be a geometric multiplier if $\mu^* \ge 0$ and

$$f^* = \inf_{x \in X} \left\{ f(x) + {\mu^*}' g(x) \right\}.$$

- Fritz John type optimality conditions can be derived using convex set support/separation arguments.
  - These do not include conditions analogous to CV condition.

Enhanced Fritz John Conditions

Proposition: Assume that X is closed and convex, and the f and the g_j are convex and lower semicontinuous over X. Then there exist multipliers $\mu_0^*,\ \mu_j^*$ s.t.

1. $\mu_0^* f^* = \inf_{x \in X} \left\{ \mu_0^* f(x) + {\mu^*}' g(x) \right\}.$
2. $\mu_j^* \geq 0$ for all j, and $\mu_0^*, \ldots, \mu_j^*$ are not all equal to 0.
3. If $J = \{j \neq 0 \mid \mu_j^* > 0\}$ is nonempty, $\exists$ some sequence $\{x^k\} \subset X$ s.t.:

$$\lim_{k \to \infty} f(x^k) = f^*, \qquad \limsup_{k \to \infty} g(x^k) \leq 0$$

and for all k

$$g_j(x) > 0,\ \forall\ j \in J, \qquad g_j(x) = o\left(\min_{j \in J} g_j(x)\right),\ \forall\ j \notin J.$$

- Proof via saddle point theory and compactification arguments.

Pseudonormality of Constraint Set

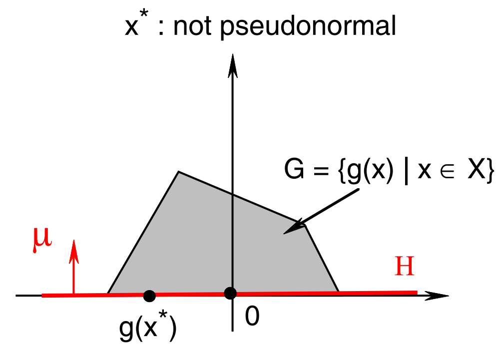
Consider the convex program under the closedness assumptions of the preceding proposition. The constraint set is pseudonormal if there are no nonnegative scalars $\mu_1, \ldots, \mu_r$, and no sequence $\{x^k\} \subset X$ such that

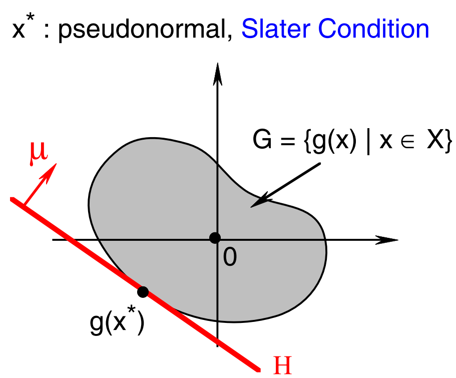
1. $0 = \inf_{x \in X} \mu' g(x)$.
2. $\limsup_{k \to \infty} g(x^k) \leq 0$ and $\mu' g(x^k) > 0$ for all k.

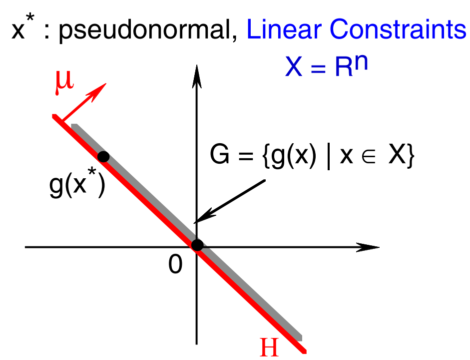
- If the problem has a convex lower semicontinuous cost function and a pseudonormal constraint set, then there exist geometric multipliers (of special type).

Geometric Interpretation of Pseudonormality

- Consider the set $G = \{g(x) \mid x \in X\}$ and hyperplanes that support this set at $g(x^*)$.







Summary

- A new approach to Lagrange multiplier theory based on an enhanced set of Fritz John conditions.
- Motivates the notion of "constraint pseudonormality" as the linchpin of a theory of constraint qualifications and the connection with exact penalty functions.
- Existence of informative Lagrange multipliers.
- Extension of pseudonormality to convex programming and geometric multipliers.
- Existence of informative geometric multipliers.