Enhanced Fritz John Optimality Conditions and Sensitivity Analysis

Dimitri P. Bertsekas

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology

March 2016

Constrained Optimization Problem

We focus on the problem

$$\text{minimize} \;\; f(x) \qquad \text{subject to} \qquad x \in X, \qquad g_j(x) \leq 0, \;\; j = 1, \ldots, r,$$

where $f : \Re^n \mapsto \Re$, $g_j : \Re^n \mapsto \Re$, and $X \subset \Re^n$. The Lagrangian function

$$L(x, \mu) = f(x) + \sum_{j=1}^{r} \mu_j g_j(x)$$

involves the multiplier vector $\mu = (\mu_1, \ldots, \mu_r)$.

We aim to find μ (Langange multipliers) that facilitate analysis or algorithms.

Salient Properties of Lagrange Multipliers

- They enter in optimality conditions: They convert the problem to an unconstrained or less constrained “optimization" of the Lagrangian
- They are central in sensitivity analysis: They quantify the rate of cost improvement as the constraint level is perturbed. Lagrange multipliers are some sort of “derivative" of the primal function

$$p(u) = \inf_{x \in X,\ g(x) \leq u} f(x)$$

This talk will revolve around these two properties.

Fritz John Conditions: A Classical Line of Development of L-Multiplier Theory

- There are several different lines of development of L-multiplier theory (forms of the implicit function theorem, forms of Farkas lemma, penalty functions, etc)
- FJ conditions is a classical line but not the most popular
- This talk will be in the direction of strengthening this line

Our starting point

- A more powerful version of the classical FJ conditions
- They include extra conditions that narrow down the candidate multipliers
- It is based on the combination of two related but complementary works: Hestenes (1975) and Rockafellar (1993)
- The line of proof is based on penalty functions; goes back to a 4-page paper by McShane (1973)
- Allows:
  - An “easy" development of unifying constraint qualifications
  - A direct connection to sensitivity

References for this Overview Talk

Joint and individual works with Asuman Ozdaglar and Paul Tseng

- Bertsekas and Ozdaglar, 2002. “Pseudonormality and a Lagrange Multiplier Theory for Constrained Optimization," J. Opt. Th. and Appl.
- Ozdaglar and Bertsekas, 2004. “The Relation Between Pseudonormality and Quasiregularity in Constrained Optimization," Optimization Methods and Software.
- Bertsekas, 2005. “Lagrange Multipliers with Optimal Sensitivity Properties in Constrained Optimization," in Proc. of the 2004 Erice Workshop on Large Scale Nonlinear Optimization, Erice, Italy, Kluwer.
- Bertsekas, Ozdaglar, and Tseng, 2006. “Enhanced Fritz John Optimality Conditions for Convex Programming," SIAM J. on Optimization.

An umbrella reference is the book

- Bertsekas, D. P., Nedić, A., and Ozdaglar, A. E., 2003. Convex Analysis and Optimization, Athena Scientific, Belmont, MA.

But it does not contain some refinements in the last three papers.

Outline

1. Enhanced FJ Conditions for Nonconvex/Differentiable Problems
2. Pseudonormality: A Unifying Constraint Qualification
3. Sensitivity for Nonconvex/Differentiable Problems
4. Convex Problems
5. Sensitivity Analysis for Convex Problems

$$\text{minimize } f(x) \quad \text{subject to} \quad x \in X, \quad g_j(x) \leq 0, \ \ j = 1, \ldots, r,$$

where $f : \Re^n \mapsto \Re$, $g_j : \Re^n \mapsto \Re$, are cont. differentiable, and $X \subset \Re^n$ is closed. The Lagrangian function is

$$L(x, \mu) = f(x) + \sum_{j=1}^{r} \mu_j g_j(x)$$

We focus at a local minimum x^*

A L-multiplier at x^* is a $\mu^* = (\mu_1^*, \ldots, \mu_r^*) \geq 0$ such that $L(\cdot, \mu^*)$ is stationary at x^* and

$$g_j(x^*) = 0 \qquad \forall\, j \text{ with } \mu_j^* > 0, \qquad \text{Complementary Slackness (CS)}$$

Meaning of stationarity

- Case $X = \Re^n$: $\nabla_x L(x^*, \mu^*) = 0$
- Case $X \neq \Re^n$: $\nabla_x L(x^*, \mu^*)' y \geq 0$ for all $y \in T_X(x^*)$, the tangent cone of X at x^* (we will define this later).

There exists $(\mu_0^*, \mu^*) \geq (0, 0)$ such that $(\mu_0^*, \mu^*) \neq (0, 0)$ and

1. $\mu_0^* \nabla f(x^*) + \sum_{j=1}^{r} \mu_j^* \nabla g_j(x^*) = 0$
2. In every neighborhood of x^* there exists x such that

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

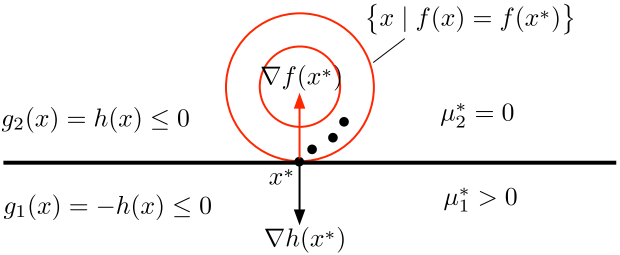
Condition (2) (called CV, for complementary violation) $\Longrightarrow$ CS

Two Cases

- $\mu_0^* = 0$. Then $\mu^* \neq 0$ and satisfies $\sum_{j=1}^{r} \mu_j^* \nabla g_j(x^*) = 0$ and the stronger CV condition. This greatly facilitates proofs that $\mu_0^* \neq 0$ under various constraint qualifications.
- $\mu_0^* = 1$. Then μ^* is a L-multiplier and the positive μ_j^* indicate the constraints j that need to be violated to effect cost improvement - a sensitivity property.

An Example

A problem with equality constraint $h(x) = 0$ split as $-h(x) \leq 0$ and $h(x) \leq 0$



- CS requires that $\mu_1^* \geq 0$ and $\mu_2^* \geq 0$ (there is an infinite number of these)
- CV requires that $\mu_1^* > 0$ and $\mu_2^* = 0$ (we cannot violate simultaneously both constraints)
- The multiplier satisfying CV is unique. Through its sign pattern indicates that the constraint $g_1(x) \leq 0$ should be violated for cost reduction.

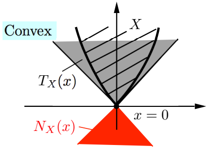
Extension of FJ conditions with the Lagrangian stationarity at the local min x^* expressed in terms of $N_X(x^*)$, the normal cone of X at x^*.

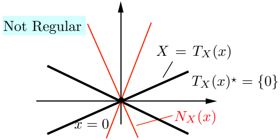
Definitions of Tangent Cone $T_X(x^*)$ and Normal Cone $N_X(x^*)$

The tangent cone $T_X(x)$ at some $x \in X$ is the set of all y such that $y = 0$ or there exists a sequence $\{x^k\} \subset X$ such that $x^k \neq x$ for all k and

$$x^k \to x, \qquad \frac{x^k - x}{\|x^k - x\|} \to \frac{y}{\|y\|}$$

The normal cone $N_X(x)$ at some $x \in X$ is the set of all z such that there exist sequences $\{x^k\} \subset X$ and $\{z^k\}$ such that $x^k \to x$, $z^k \to z$, and $z^k \in T_X(x^k)^\star$ [the polar of $T_X(x^k)$]. Note that $T_X(x^*)^\star \subset N_X(x^*)$. If equality holds, X is called regular at x^* (if X is convex, it is regular at all points).





Constrained optimization problem

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

where $f : \Re^n \mapsto \Re$, $g : \Re^n \mapsto \Re$, are cont. differentiable, and $X \subset \Re^n$ is closed.

FJ Conditions

There exists $(\mu_0^*, \mu^*) \geq (0, 0)$ such that $(\mu_0^*, \mu^*) \neq (0, 0)$ and

1. $-\left(\mu_0^* \nabla f(x^*) + \sum_{j=1}^r \mu_j^* \nabla g_j(x^*)\right) \in N_X(x^*)$
2. $g_j(x^*) = 0 \quad \forall\, j$ with $\mu_j^* > 0$, i.e., CS holds.

If $\mu_0^* = 1$ and X is regular at x^*, then $N_X(x^*) = T_X(x^*)^\star$, and condition (1) becomes

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

so μ^* is a L-multiplier satisfying CS.

Enhanced FJ Necessary Conditions for Case $X \neq \Re^n$ (B+O 2002)

Constrained optimization problem

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

where $f : \Re^n \mapsto \Re$, $g : \Re^n \mapsto \Re$, are cont. differentiable, and $X \subset \Re^n$ is closed.

There exists $(\mu_0^*, \mu^*) \geq (0, 0)$ such that $(\mu_0^*, \mu^*) \neq (0, 0)$ and

1. $-\left(\mu_0^* \nabla f(x^*) + \sum_{j=1}^r \mu_j^* \nabla g_j(x^*)\right) \in N_X(x^*)$
2. In every neighborhood of x^* there exists x such that

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

i.e., CV holds.

So if $\mu_0^* = 1$ and X is regular at x^*, then μ^* is a L-multiplier satisfying CV. We call such a multiplier informative.

Constraint Qualifications: Conditions Under Which $\mu_0^* \neq 0$

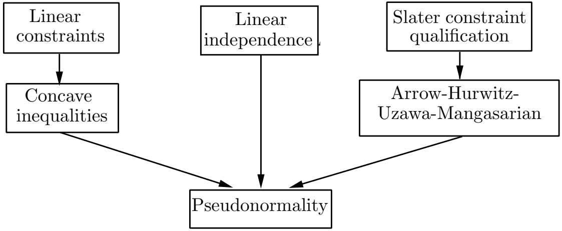
Pseudonormality: An umbrella constraint qualification (B+O 2002)

x^* is a pseudonormal local min if there is no $\mu \geq 0$ and sequence $\{x^k\} \subset X$ with $x^k \to x^*$ such that

$$-\left(\sum_{j=1}^{r} \mu_j \nabla g_j(x^*)\right) \in N_X(x^*), \qquad \sum_{j=1}^{r} \mu_j g_j(x^k) > 0, \quad \forall\, k.$$

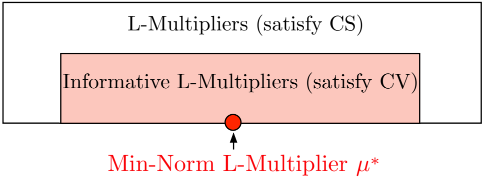
Note: If x^* is pseudonormal and X is regular at x^* there is an informative L-multiplier.

All the principal constraint qualifications for existence of L-multipliers and some new ones can be shown (very easily) to imply pseudonormality.



Sensitivity: A Hierarchy of L-Multipliers (B+O 2002, 2005)

Assume $T_X(x^*)$ is convex and the set of L-multipliers is nonempty. Then:



Sensitivity Result

- Cost improvement bound: For every sequence of infeasible vectors $\{x^k\} \subset X$ with $x_k \to x^*$ we have

$$f(x^*) - f(x^k) \leq \|\mu^*\| \, \big\|g^+(x^k)\big\| + o\big(\|x^k - x^*\|\big),$$

where $g^+(x) = \big(g_1^+(x), \ldots, g_r^+(x)\big)$: constraint violation vector.

- Optimal constraint violation direction: If $\mu^* \neq 0$, there exists infeasible sequence $\{x^k\} \subset X$ with $x^k \to x^*$ and such that

$$\lim_{k\to\infty} \frac{f(x^*) - f(x^k)}{\big\|g^+(x^k)\big\|} = \|\mu^*\|, \qquad \lim_{k\to\infty} \frac{g_j^+(x^k)}{\big\|g^+(x^k)\big\|} = \frac{\mu_j^*}{\|\mu^*\|}, \quad j = 1, \ldots, r.$$

The Convex Case (B+O+T 2006)

Consider the problem

$$\text{minimize } f(x) \quad \text{subject to} \quad x \in X, \quad g_j(x) \leq 0, \ \ j = 1, \ldots, r,$$

where $f : \Re^n \mapsto \Re$, $g : \Re^n \mapsto \Re$, are convex, and $X \subset \Re^n$ is convex and closed.

Primal function: $p(u) = \inf_{x \in X,\ g(x) \leq u} f(x)$, Dual function: $q(\mu) = \inf_{x \in X} L(x, \mu)$

Optimal primal value: $f^* = p(0)$, Optimal dual value: $q^* = \sup_{\mu \geq 0} q(\mu)$

Duality gap: $f^* - q^*$

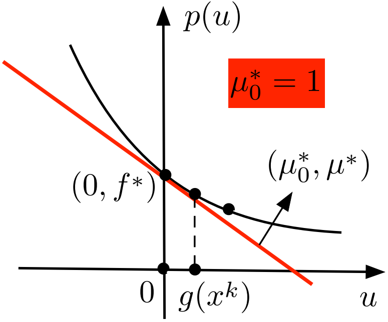
μ is Lagrange multiplier if $f^* - q^* = 0$ and μ is a dual optimal solution

Primal and dual FJ conditions adapted to convex programming

There exists $(\mu_0^*, \mu^*) \geq (0, 0)$ such that $(\mu_0^*, \mu^*) \neq (0, 0)$, and

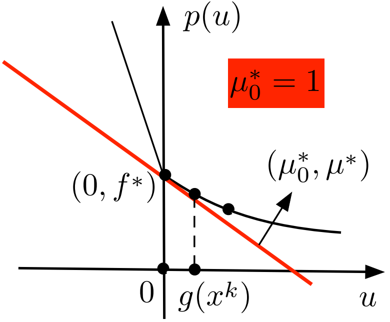
- $\mu_0^* f^* = \inf_{x \in X} \left\{ \mu_0^* f(x) + \sum_{j=1}^r \mu_j^* g_j(x) \right\}$ (Primal FJ Conditions)
- $\mu_0^* q^* = \inf_{x \in X} \left\{ \mu_0^* f(x) + \sum_{j=1}^r \mu_j^* g_j(x) \right\}$ (Dual FJ Conditions)

and CS holds (there are versions with CV also).



The most favorable case: $\nabla p(0)$ exists, and $\mu^* = \nabla p(0)$ is the unique L-multiplier

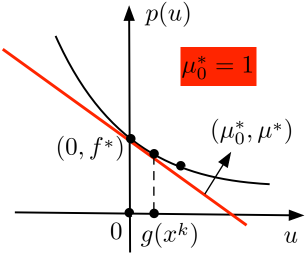
Rate of cost improvement: The slope of the support hyperplane at $(0, f^*)$

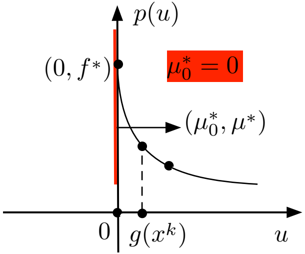


p is subdifferentiable at 0. Set of L-multipliers = $\partial p(0)$.

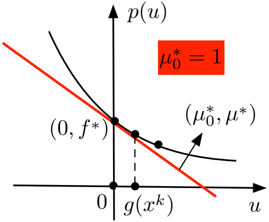
Optimal rate of cost improvement: The slope of the min norm support hyperplane at $(0, f^*)$

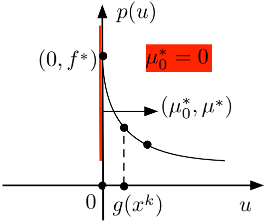
Visualization with the Primal Function $p(u) = \inf_{x \in X,\, g(x) \le u} f(x)$

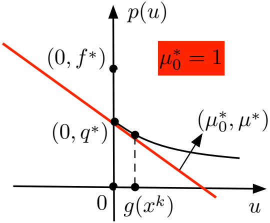


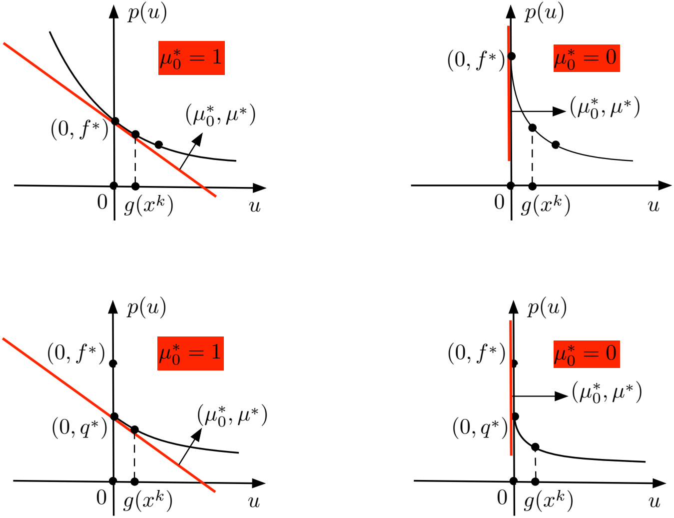


Visualization with the Primal Function $p(u) = \inf_{x \in X,\, g(x) \le u} f(x)$








$p(u)$
$\mu_0^* = 1$
(μ_0^*, μ^*)
$(0, f^*)$
0
$g(x^k)$
u
$p(u)$
$(0, f^*)$
$\mu_0^* = 0$
(μ_0^*, μ^*)
0
$g(x^k)$
u
$p(u)$
$(0, f^*)$
$\mu_0^* = 1$
(μ_0^*, μ^*)
$(0, q^*)$
0
$g(x^k)$
u
$p(u)$
$(0, f^*)$
$\mu_0^* = 0$
(μ_0^*, μ^*)
$(0, q^*)$
0
$g(x^k)$
u

Dual Sensitivity Result

Assume that $-\infty < q^* \leq f^* < \infty$, and that the set of dual optimal solutions is nonempty. Let μ^* be the min norm dual optimal solution (may not be a L-multiplier):

- Cost improvement bound: For all $x \in X$

$$q^* - f(x) \leq \|\mu^*\| \, \|g^+(x)\|$$

- If $\mu^* \neq 0$, there exists infeasible sequence $\{x^k\} \subset X$ such that

$$f(x^k) \to q^*, \qquad g^+(x^k) \to 0,$$

  the bound is asymptotically sharp,

$$\frac{q^* - f(x^k)}{\|g^+(x^k)\|} \to \|\mu^*\|.$$

  Also μ^* is the optimal constraint violation direction,

$$\frac{g^+(x^k)}{\|g^+(x^k)\|} \to \frac{\mu^*}{\|\mu^*\|}$$

  [Note: $\{g^+(x^k)\}$ may need to lie on a curve].

Thank you!