Enhanced Fritz John Optimality Conditions and Sensitivity Analysis

Dimitri P. Bertsekas

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology

March 2016

We focus on the constrained optimization problem

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

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 that facilitates the analytical or algorithmic solution.

Salient properties of L-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. L-multipliers are some sort of "derivative" of the primal function

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

This talk will revolve around both of these 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 version of FJ conditions
- They include extra conditions that narrow down the candidate multipliers
- The starting point is a largely overlooked 1975 work by Hestenes (which does not allow for a set constraint $x \in X$)
- Hestenes' line of proof goes back to a 4-page paper by McShane (1973) and is based on penalty functions
- 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.
- 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, with D. P., Nedić, A., and Ozdaglar, A. E., 2003. Convex Analysis and Optimization, Athena Scientific, Belmont, MA.

But it does not contain important parts of the last two papers.

Outline

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

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 $\mu^* = (\mu_1^*, \ldots, \mu_r^*) \ge 0$ such that

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

and $L(\cdot, \mu^*)$ is stationary at x^*.

Meaning of stationarity

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

FJ Necessary Conditions for Case $X = \Re^n$ (Hestenes 1975)

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

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

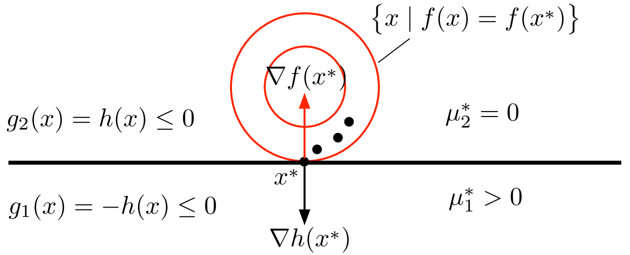
Condition (2) (called CV, for complementary violation) $\Longrightarrow$ CS

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 $-h(x) \le 0$ and $h(x) \le 0$



- CS requires that $\mu_1^* \ge 0$ and $\mu_2^* \ge 0$ (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 indicates through its sign the constraint $g_1(x) \le 0$ that should be violated for cost reduction.

Case $X \neq \Re^n$: The “Standard" Extension of FJ Conditions

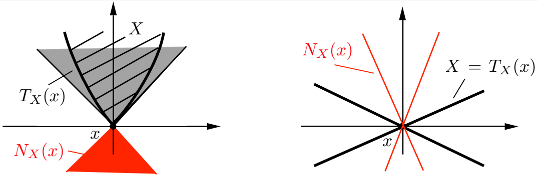
The standard FJ conditions with CS (not CV) express Lagrangian stationarity at the local min x^* in terms of $N_X(x^*)$, the normal cone of X at x^*.

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 $z^k \in T_X(x^k)^\perp$. Note that $T_X(x^*)^\perp \subset N_X(x^*)$. If equality holds X is called regular at x^* (if X is convex, it is regular at all points).



Extension of FJ Framework for Case $X \neq \Re^n$

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

Let x^* be a local minimum. Then 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.

So if $\mu_0^* = 1$ and X is regular at x^*, so that condition (1) becomes

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

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

Enhanced FJ Necessary Conditions for Case $X \neq \Re^n$ (B+O 2003)

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.

Let x^* be a local minimum. Then 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

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

   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.

How do we Prove that $\mu_0^* \neq 0$?

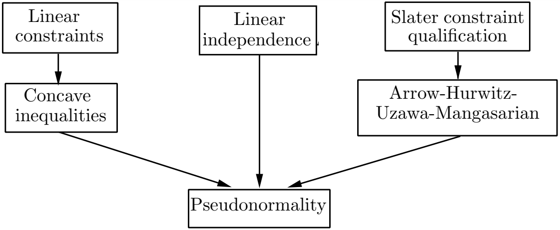
Pseudonormality: An umbrella constraint qualification (B+O 2003)

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

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

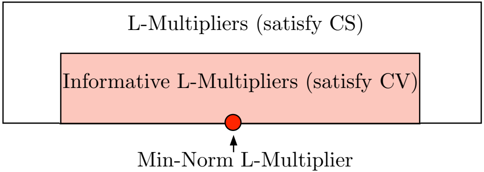
Note: If x^* is pseudonormal and X is regular at x^* there is an informative L-multiplier.

All the principal constraint qualifications for existence of L-multipliers and some new ones can be shown 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.



Sensitivity Result

The L-multiplier of min norm, call it μ^*, is informative. Moreover:

- For every sequence of infeasible vectors $\{x^k\}$ with $x_k \to x^*$ we have

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

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

- If $\mu^* \neq 0$, there exists infeasible sequence $\{x^k\}$ with $x_k \to x^*$ and such that

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

Thank you!