LECTURE 16

LECTURE OUTLINE

- Enhanced Fritz John Conditions
- Pseudonormality
- Constraint qualifications

- Problem

$$
\begin{aligned}
&\text{minimize} \quad f(x) \\
&\text{subject to} \quad x \in X, \quad h_1(x) = 0, \ldots, h_m(x) = 0 \\
&\qquad\qquad\qquad\quad g_1(x) \le 0, \ldots, g_r(x) \le 0
\end{aligned}
$$

where f, h_i, $g_j : \Re^n \mapsto \Re$ are smooth functions, and X is a nonempty closed set.

- To simplify notation, we will often assume no equality constraints.

DEFINITION OF LAGRANGE MULTIPLIER

- Consider the Lagrangian function

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{j=1}^{r} \mu_j g_j(x).$$

Let x^* be a local minimum. Then λ^* and μ^* are Lagrange multipliers if for all j,

$$\mu_j^* \geq 0, \qquad \mu_j^* = 0 \text{ if } g_j(x^*) < 0,$$

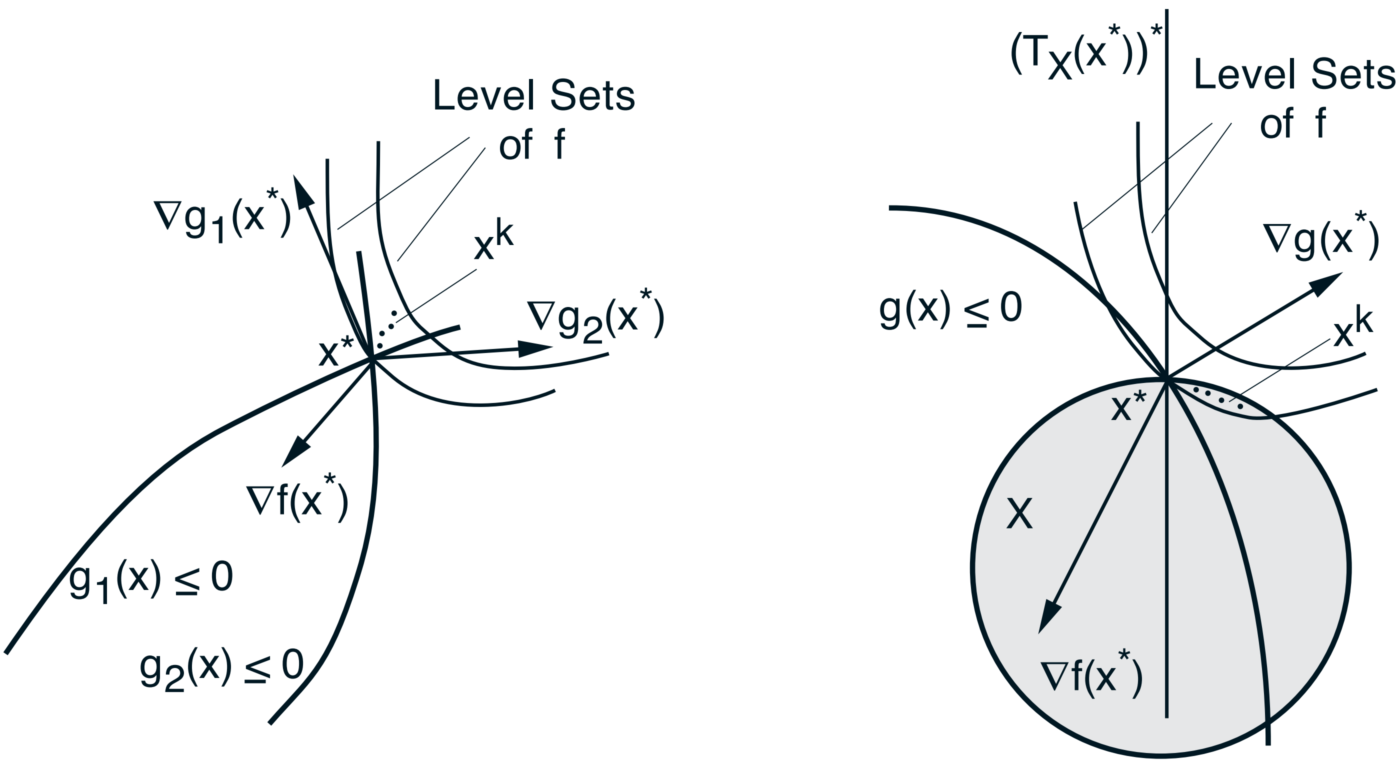
and the Lagrangian is stationary at x^*, i.e., has ≥ 0 slope along the tangent directions of X at x^* (feasible directions in case where X is convex):

$$\nabla_x L(x^*, \lambda^*, \mu^*)' y \geq 0, \quad \forall\, y \in T_X(x^*).$$

- **Note 1:** If $X = \Re^n$, Lagrangian stationarity means $\nabla_x L(x^*, \lambda^*, \mu^*) = 0$.

- **Note 2:** If X is convex and the Lagrangian is convex in x for $\mu \geq 0$, Lagrangian stationarity means that $L(\cdot, \lambda^*, \mu^*)$ is minimized over $x \in X$ at x^*.

ILLUSTRATION OF LAGRANGE MULTIPLIERS



(a) (b)

- (a) Case where $X = \Re^n$: $-\nabla f(x^*)$ is in the cone generated by the gradients $\nabla g_j(x^*)$ of the active constraints.
- (b) Case where $X \neq \Re^n$: $-\nabla f(x^*)$ is in the cone generated by the gradients $\nabla g_j(x^*)$ of the active constraints and the polar cone $T_X(x^*)^*$.

ENHANCED FRITZ JOHN NECESSARY CONDITIONS

If x^* is a local minimum, there exist $\mu_0^*, \mu_1^*, \ldots, \mu_r^*$, satisfying the following:

(i) $$-\left(\mu_0^* \nabla f(x^*) + \sum_{j=1}^{r} \mu_j^* \nabla g_j(x^*)\right) \in N_X(x^*)$$

(ii) $\mu_0^*, \mu_1^*, \ldots, \mu_r^* \geq 0$ and not all 0

(iii) If

$$J = \{j \neq 0 \mid \mu_j^* > 0\}$$

is nonempty, there exists a sequence $\{x^k\} \subset X$ converging to x^* and such that for all k,

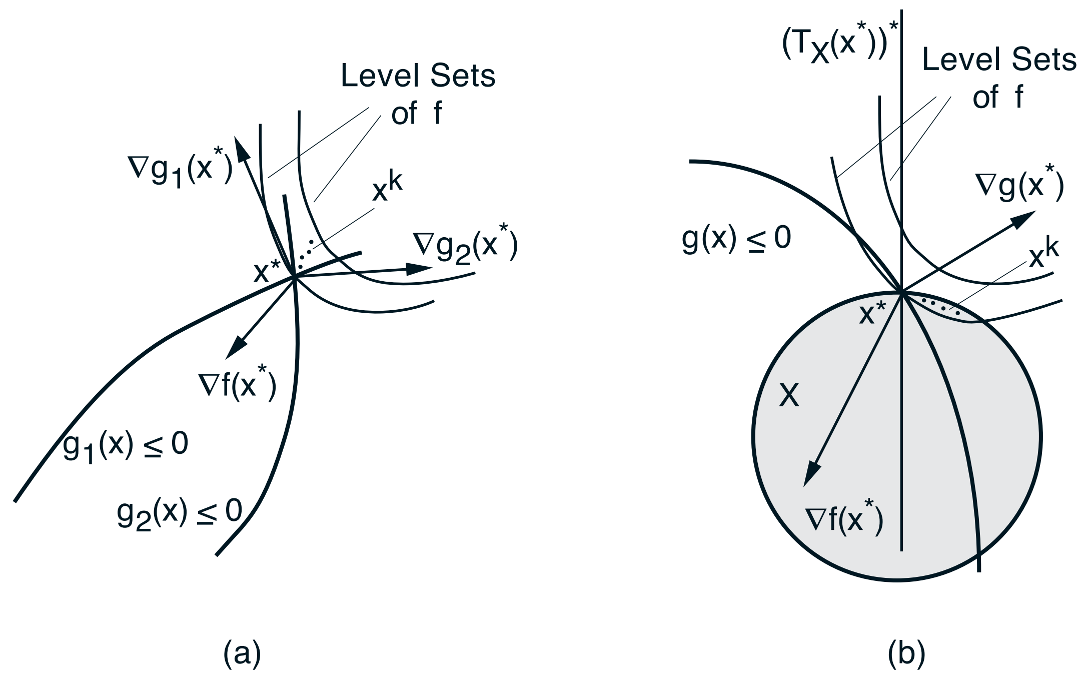
$$f(x^k) < f(x^*), \qquad g_j(x^k) > 0, \quad \forall\, j \in J,$$

$$g_j^+(x^k) = o\left(\min_{j \in J} g_j(x^k)\right), \ \ \forall\, j \notin J.$$

• **Note:** In the classical Fritz John theorem, condition (iii) is replaced by the weaker condition that

$$\mu_j^* = 0, \qquad \forall\, j \text{ with } g_j(x^*) < 0.$$

GEOM. INTERPRETATION OF LAST CONDITION



- **Note:** Multipliers satisfying the classical Fritz John conditions may not satisfy condition (iii).

- **Example:** Start with any problem $\min_{h(x)=0} f(x)$ that has a local min-Lagrange multiplier pair (x^*, λ^*) with $\nabla f(x^*) \neq 0$ and $\nabla h(x^*) \neq 0$. Convert it to the problem $\min_{h(x)\leq 0,\, -h(x)\leq 0} f(x)$. The (μ_0^*, μ^*) satisfying the classical FJ conditions:

$$\mu_0^* = 0,\ \ \mu_1^* = \mu_2^* \neq 0 \text{ or } \mu_0^* > 0,\ (\mu_0^*)^{-1}(\mu_1^* - \mu_2^*) = \lambda^*$$

The enhanced FJ conditions are satisfied only for

$$\mu_0^* > 0,\ \mu_1^* = \lambda^*,\ \mu_2^* = 0 \text{ or } \mu_0^* > 0,\ \mu_1^* = 0,\ \mu_2^* = -\lambda^*$$

PROOF OF ENHANCED FJ THEOREM

- We use a quadratic penalty function approach. Let $g_j^+(x) = \max\{0, g_j(x)\}$, and for each k, consider

$$\min_{X\cap S} F^k(x) \equiv f(x) + \frac{k}{2}\sum_{j=1}^{r}\big(g_j^+(x)\big)^2 + \frac{1}{2}||x - x^*||^2$$

where $S = \{x \mid ||x - x^*|| \leq \epsilon\}$, and $\epsilon > 0$ is such that $f(x^*) \leq f(x)$ for all feasible x with $x \in S$. Using Weierstrass' theorem, we select an optimal solution x^k. For all k, $F^k(x^k) \leq F^k(x^*)$, or

$$f(x^k) + \frac{k}{2}\sum_{j=1}^{r}\big(g_j^+(x^k)\big)^2 + \frac{1}{2}||x^k - x^*||^2 \leq f(x^*).$$

Since $f(x^k)$ is bounded over $X \cap S$, $g_j^+(x^k) \to 0$, and every limit point $\overline{x}$ of $\{x^k\}$ is feasible. Also, $f(x^k) + (1/2)||x^k - x^*||^2 \leq f(x^*)$ for all k, so

$$f(\overline{x}) + \frac{1}{2}||\overline{x} - x^*||^2 \leq f(x^*).$$

- Since $\overline{x} \in S$ and $\overline{x}$ is feasible, we have $f(x^*) \leq f(\overline{x})$, so $\overline{x} = x^*$. Thus $x^k \to x^*$, and x^k is an interior point of the closed sphere S for all large k.

PROOF (CONTINUED)

- For k large, we have the necessary condition $-\nabla F^k(x^k) \in T_X(x^k)^*$, which is written as

$$-\left(\nabla f(x^k) + \sum_{j=1}^{r} \zeta_j^k \nabla g_j(x^k) + (x^k - x^*)\right) \in T_X(x^k)^*,$$

where $\zeta_j^k = k g_j^+(x^k)$. Denote

$$\delta^k = \sqrt{1 + \sum_{j=1}^{r} (\zeta_j^k)^2}, \quad \mu_0^k = \frac{1}{\delta^k}, \quad \mu_j^k = \frac{\zeta_j^k}{\delta^k}, \quad j > 0.$$

Dividing with δ^k,

$$-\left(\mu_0^k \nabla f(x^k) + \sum_{j=1}^{r} \mu_j^k \nabla g_j(x^k) + \frac{1}{\delta^k}(x^k - x^*)\right) \in T_X(x^k)^*$$

Since by construction $(\mu_0^k)^2 + \sum_{j=1}^{r} (\mu_j^k)^2 = 1$, the sequence $\{\mu_0^k, \mu_1^k, \ldots, \mu_r^k\}$ is bounded and must contain a subsequence that converges to some limit $\{\mu_0^*, \mu_1^*, \ldots, \mu_r^*\}$. This limit has the required properties ...

CONSTRAINT QUALIFICATIONS

Suppose there do NOT exist $\mu_1, \ldots, \mu_r$, satisfying:

(i) $-\sum_{j=1}^{r} \mu_j \nabla g_j(x^*) \in N_X(x^*)$.

(ii) $\mu_1, \ldots, \mu_r \geq 0$ and not all 0.

• Then we must have $\mu_0^* > 0$ in FJ, and can take $\mu_0^* = 1$. So there exist $\mu_1^*, \ldots, \mu_r^*$, satisfying all the Lagrange multiplier conditions except that:

$$-\left(\nabla f(x^*) + \sum_{j=1}^{r} \mu_j^* \nabla g_j(x^*)\right) \in N_X(x^*)$$

rather than $-(\cdot) \in T_X(x^*)^*$ (such multipliers are called *R-multipliers*).

• If X is regular at x^*, R-multipliers are Lagrange multipliers.

• **LICQ (Lin. Independence Constr. Qual.):** There exists a unique Lagrange multiplier vector if $X = \Re^n$ and x^* is a *regular point*, i.e.,

$$\{\nabla g_j(x^*) \mid j \text{ with } g_j(x^*) = 0\}$$

are linearly independent.

PSEUDONORMALITY

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

(i) $-\left(\sum_{j=1}^r \mu_j \nabla g_j(x^*)\right) \in N_X(x^*)$.

(ii) $\mu_j \geq 0$, for all $j = 1, \ldots, r$, and $\mu_j = 0$ for all $j \notin A(x^*)$.

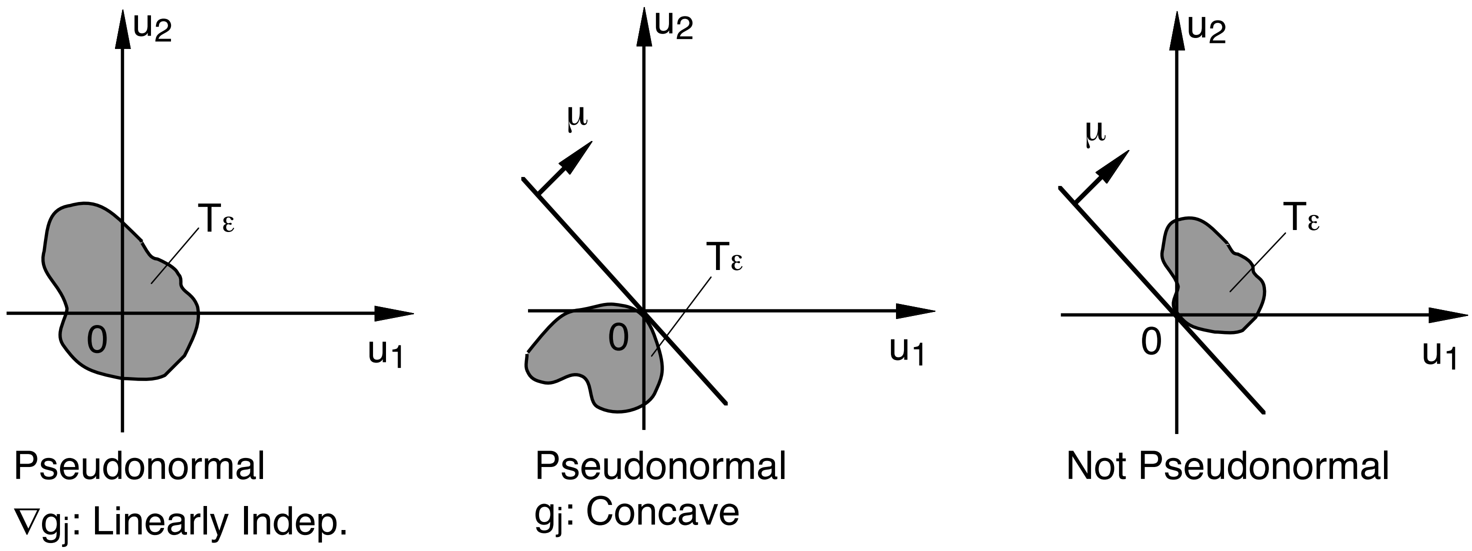
(iii) $\{x^k\}$ converges to x^* and

$$\sum_{j=1}^r \mu_j g_j(x^k) > 0, \qquad \forall\, k.$$

- From Enhanced FJ conditions:
  - If x^* is pseudonormal there exists an R-multiplier vector.
  - If in addition X is regular at x^*, there exists a Lagrange multiplier vector.

GEOM. INTERPRETATION OF PSEUDONORMALITY I

- Assume that $X = \Re^n$



- Consider, for a small positive scalar ϵ, the set

$$T_\epsilon = \big\{ g(x) \mid \|x - x^*\| < \epsilon \big\}$$

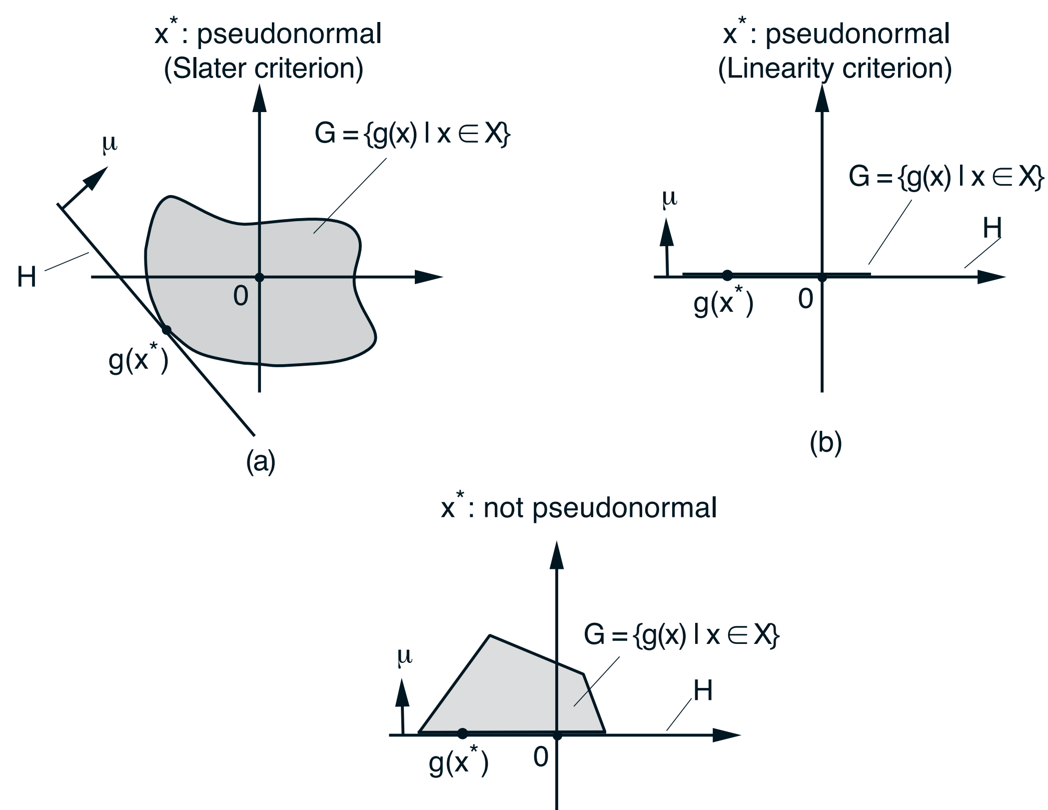
- x^* is pseudonormal if and only if either
  - (1) the gradients $\nabla g_i(x^*)$, $j = 1, \ldots, r$, are linearly independent, or
  - (2) for every $\mu \geq 0$ with $\mu \neq 0$ and such that $\sum_{j=1}^r \mu_j \nabla g_j(x^*) = 0$, there is a small enough ϵ, such that the set T_ϵ does not cross into the positive open halfspace of the hyperplane through 0 whose normal is μ. This is true if the g_j are concave [then $\mu' g(x)$ is maximized at x^* so $\mu' g(x) \leq 0$ for all $x \in \Re^n$].

GEOM. INTERPRETATION OF PSEUDONORMALITY II

- Assume that X and the g_j are convex, so that

$$-\left(\sum_{j=1}^{r} \mu_j \nabla g_j(x^*)\right) \in N_X(x^*)$$

if and only if $x^* \in \arg\min_{x\in X} \sum_{j=1}^{r} \mu_j g_j(x)$. Pseudonormality holds if and only if for every hyperplane with normal $\mu \geq 0$ that passes through the origin and supports the set $G = \{g(x) \mid x \in X\}$, contains G in its negative halfspace.



SOME MAJOR CONSTRAINT QUALIFICATIONS

CQ1: $X = \Re^n$, and the functions g_j are concave.

CQ2: There exists a $y \in N_X(x^*)^*$ such that

$$\nabla g_j(x^*)'y < 0, \quad \forall\, j \in A(x^*)$$

- Special case of CQ2: The Slater condition (X is convex, g_j are convex, and there exists $\overline{x} \in X$ s.t. $g_j(\overline{x}) < 0$ for all j).

- CQ2 is known as the (generalized) Mangasarian-Fromowitz CQ. The version with equality constraints:

  (a) There does not exist a nonzero vector $\lambda = (\lambda_1, \ldots, \lambda_m)$ such that

  $$\sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) \in N_X(x^*).$$

  (b) There exists a $y \in N_X(x^*)^*$ such that

  $$\nabla h_i(x^*)'y = 0,\ \forall\, i,\ \ \nabla g_j(x^*)'y < 0,\ \forall\, j \in A(x^*)$$

CONSTRAINT QUALIFICATION THEOREM

- If CQ1 or CQ2 holds, then x^* is pseudonormal.

Proof: Assume that there are scalars μ_j, $j = 1, \ldots, r$, satisfying conditions (i)-(iii) of the definition of pseudonormality. Then assume that each of the constraint qualifications is in turn also satisfied, and in each case arrive at a contradiction.

Case of CQ1: By the concavity of g_j, the condition $\sum_{j=1}^r \mu_j \nabla g_j(x^*) = 0$, implies that x^* maximizes $\mu' g(x)$ over $x \in \Re^n$, so

$$\mu' g(x) \leq \mu' g(x^*) = 0, \qquad \forall\, x \in \Re^n$$

This contradicts condition (iii) [arbitrarily close to x^*, there is an x satisfying $\sum_{j=1}^r \mu_j g_j(x) > 0$].

Case of CQ2: We must have $\mu_j > 0$ for at least one j, and since $\mu_j \geq 0$ for all j with $\mu_j = 0$ for $j \notin A(x^*)$, we obtain

$$\sum_{j=1}^r \mu_j \nabla g_j(x^*)' y < 0,$$

for the vector y of $N_X(x^*)^*$ that appears in CQ2.

PROOF (CONTINUED)

Thus,

$$-\sum_{j=1}^{r} \mu_j \nabla g_j(x^*) \notin \left(N_X(x^*)^*\right)^*.$$

Since $N_X(x^*) \subset \left(N_X(x^*)^*\right)^*$,

$$-\sum_{j=1}^{r} \mu_j \nabla g_j(x^*) \notin N_X(x^*)$$

a contradiction of conditions (i) and (ii). **Q.E.D.**

- If $X = \Re^n$, CQ2 is equivalent to the cone $\{y \mid \nabla g_j(x^*)'y \leq 0,\ j \in A(x^*)\}$ having nonempty interior, which (by Gordan's theorem) is equivalent to conditions (i) and (ii) of pseudonormality.

- Note that CQ2 can also be shown to be equivalent to conditions (i) and (ii) of pseudonormality, even when $X \neq \Re^n$, as long as X is regular at x^*. These conditions can in turn can be shown in turn to be equivalent to nonemptiness and compactness of the set of Lagrange multipliers (which is always closed and convex as the intersection of a collection of halfspaces).