A NEW LOOK AT CONVEX ANALYSIS AND OPTIMIZATION

Dimitri Bertsekas
M.I.T.

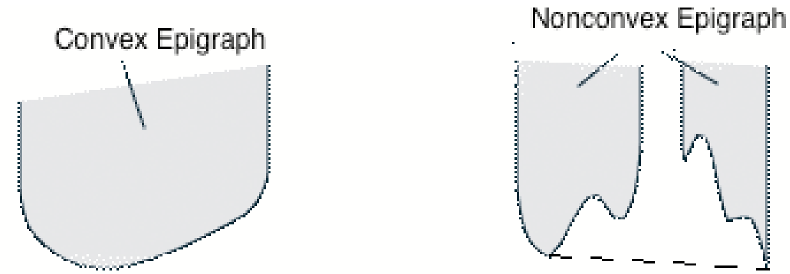
May 2007

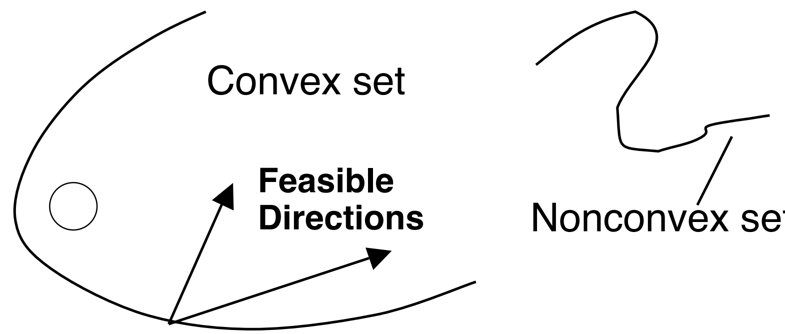
OUTLINE

- Convexity issues in optimization
- Historical remarks
- Our treatment of the subject
  - Math rigor enhanced by visualization
  - Unification and intuition enhanced by geometry
- Three unifying lines of analysis
  - Common geometrical framework for duality and minimax
  - Unifying framework for existence of solutions and duality gap analysis
  - Unification of Lagrange multiplier theory using an enhanced Fritz John theory and the notion of pseudonormality

WHY IS CONVEXITY IMPORTANT IN OPTIMIZATION I

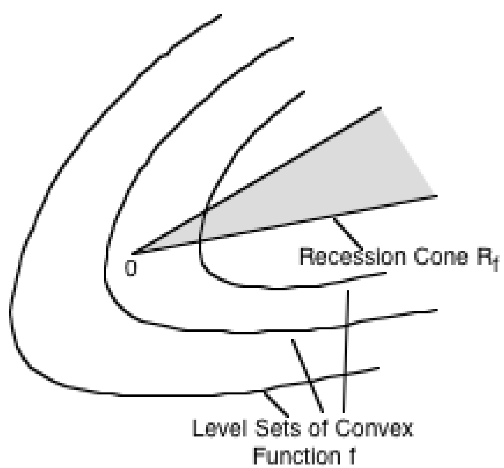
- A convex function has **no local minima** that are not global
- A nonconvex function can be "**convexified**" while maintaining the optimality of its minima
- A convex set has **nonempty relative interior**
- A convex set has **feasible directions** at any point

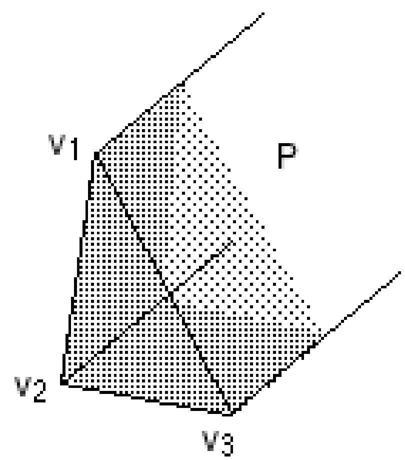




WHY IS CONVEXITY IMPORTANT IN OPTIMIZATION II

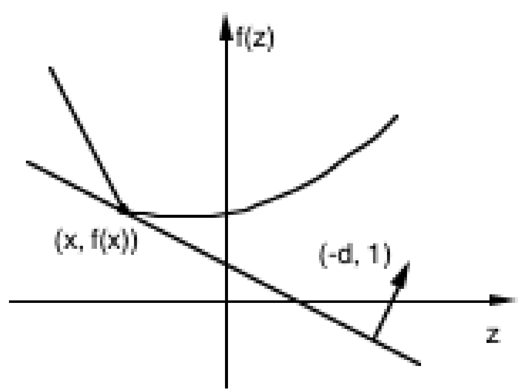
- **The existence of minima of convex functions is conveniently characterized using directions of recession**
- **A polyhedral convex set is characterized by its extreme points and extreme directions**

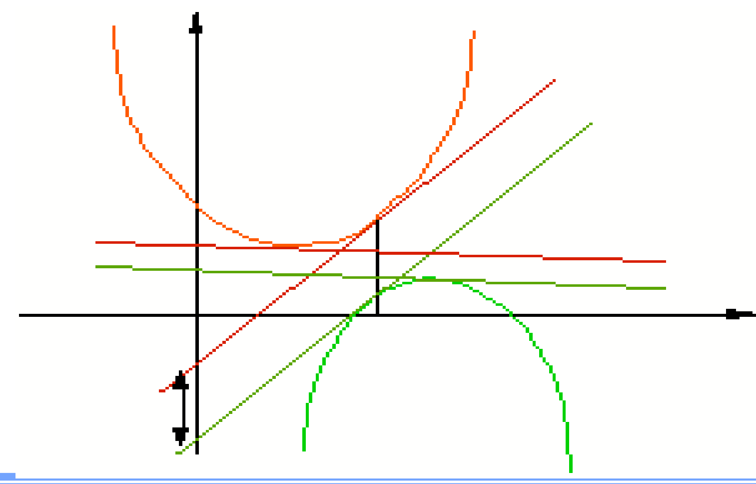




WHY IS CONVEXITY IMPORTANT IN OPTIMIZATION III

- A convex function is continuous and has **nice differentiability properties**
- Convex functions arise prominently in **duality**
- Convex, lower semicontinuous functions are self-dual with respect to **conjugacy**





SOME HISTORY

- **Late 19th-Early 20th Century:**
  - Caratheodory, Minkowski, Steinitz, Farkas
- **40s-50s: The big turning point**
  - Game Theory: von Neumann
  - Optimization-related convexity: Fenchel
  - Duality: Fenchel, Princeton group (Gale, Kuhn, Tucker)
- **60s-70s: Consolidation**
  - Rockafellar
- **80s-90s: Extensions to nonconvex optimization and nonsmooth analysis**
  - Clarke, Mordukovich, Rockafellar-Wets

ABOUT THE BOOKS

- **Convex Analysis and Optimization, by D. P. Bertsekas, with A. Nedic and A. Ozdaglar (March 2003 - extends to nonconvex analysis)**
- **Convex Optimization Theory, by D. P. Bertsekas (more narrowly/deeply focused on convexity - to appear in 2007-08)**
- **Aims to make the subject accessible through unification and geometric visualization**
- **Unification is achieved through several new lines of analysis**

NEW LINES OF ANALYSIS

I A unified geometrical approach to convex programming duality and minimax theory

- **Basis: Duality between two elementary geometrical problems**

II A unified view of theory of existence of solutions and absence of duality gap

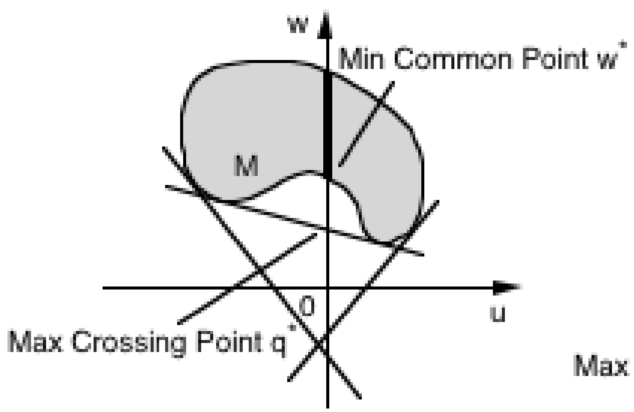
- **Basis: Reduction to basic questions on intersections of closed sets**

III A unified view of theory of existence of Lagrange multipliers/constraint qualifications

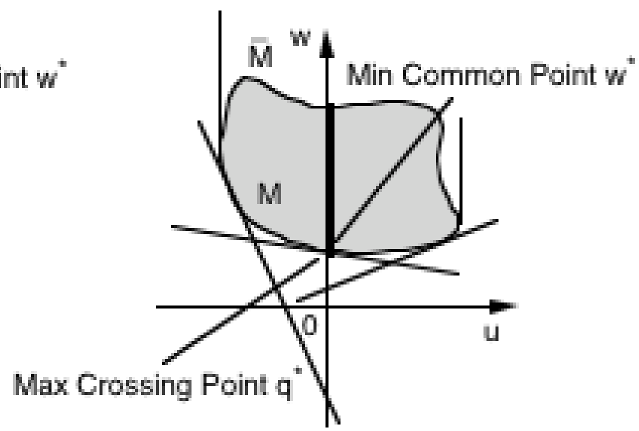
- **Basis: The notion of constraint pseudonormality, motivated by a new set of enhanced Fritz John conditions**

I
MIN COMMON/MAX CROSSING DUALITY

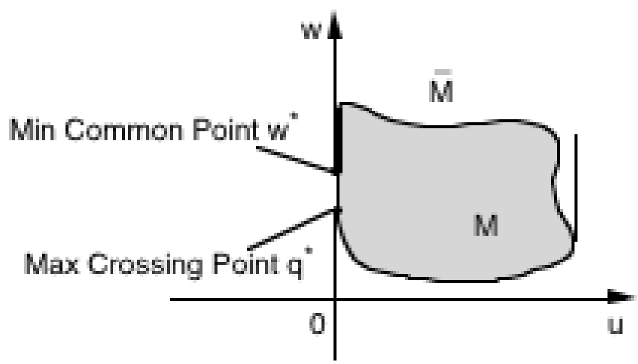
GEOMETRICAL VIEW OF DUALITY



(a)



(b)



(c)

ANALYTICAL APPROACH

- Prove theorems about the geometry of M
  - Conditions on M that guarantee $w^* = q^*$
  - Conditions on M that guarantee existence of a max crossing hyperplane
- Specialize the min common/max crossing theorems to duality and minimax theorems
- Special choices of M apply to:
  - Constrained optimization problem
  - Minimax (zero-sum game) problem
  - Others (e.g., Fenchel duality framework)

CONVEX PROGRAMMING DUALITY

- Primal problem:

  min $f(x)$ subject to $x \in X$ and $g_j(x) \le 0,\ j=1,\dots,r$

- Dual problem:

  max $q(\mu)$ subject to $\mu \ge 0$

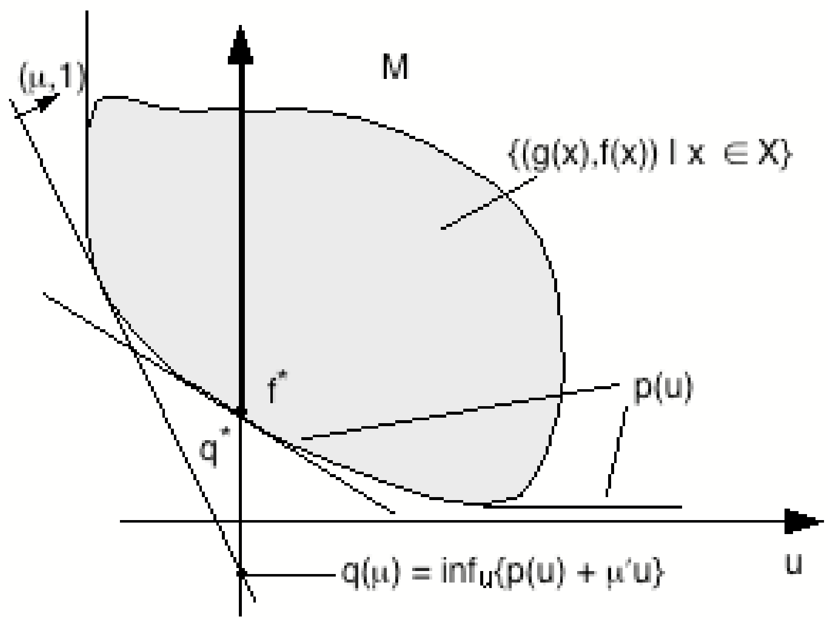
  where the dual function is

$$q(\mu) = \inf_{x \in X} L(x,\mu) = \inf_{x \in X} \{f(x) + \mu' g(x)\}$$

- Optimal primal value = $\inf_{x \in X} \sup_{\mu \ge 0} L(x,\mu)$
- Optimal dual value = $\sup_{\mu \ge 0} \inf_{x \in X} L(x,\mu)$
- Min common/max crossing framework:

$$p(u) = \inf_{x \in X,\ gj(x) \le uj} f(x), \qquad M = \text{epi}(p)$$

VISUALIZATION



MINIMAX / ZERO SUM GAME THEORY ISSUES

- Given a function $\Phi(x,z)$, where $x \in X$ and $z \in Z$, under what conditions do we have

$$\inf_x \sup_z \Phi(x,z) = \sup_z \inf_x \Phi(x,z)$$

- Assume convexity/concavity, semicontinuity of Φ
- Min common/max crossing framework:

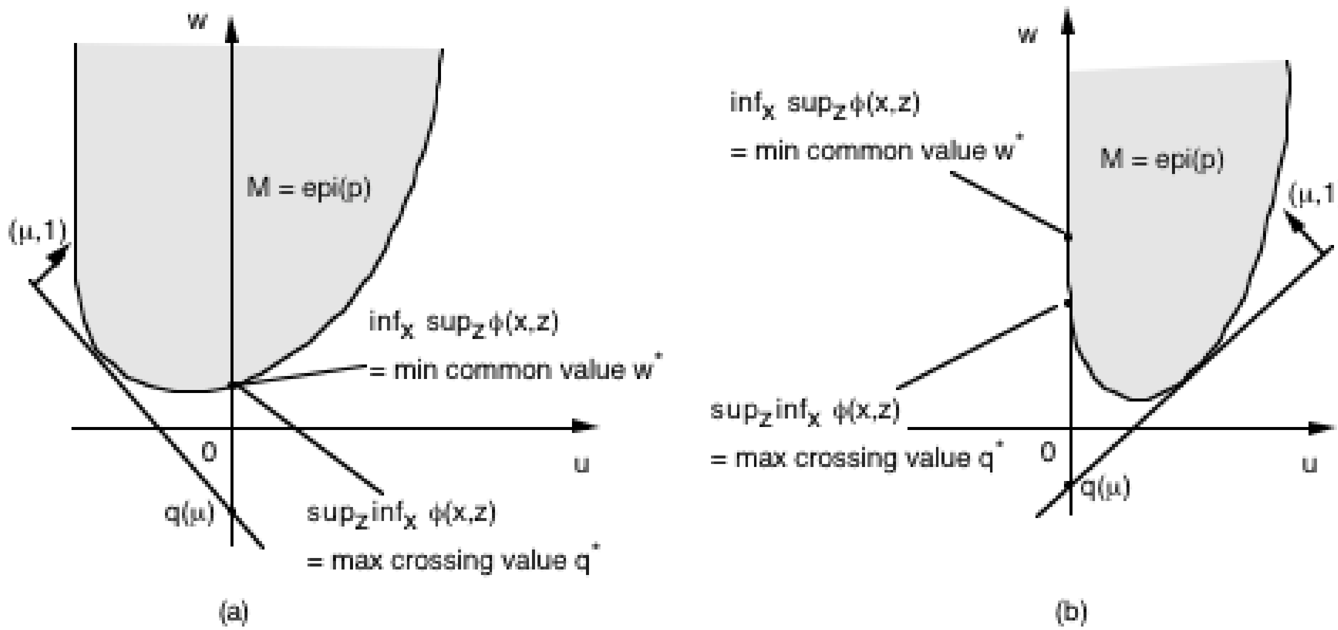
$$p(u) = \inf_x \sup_z \{\Phi(x,z) - u'z\}$$

$$M = \text{epigraph of } p$$

$\inf_x \sup_z \Phi$ = Min common value

$\sup_z \inf_x \Phi$ = Max crossing value

VISUALIZATION



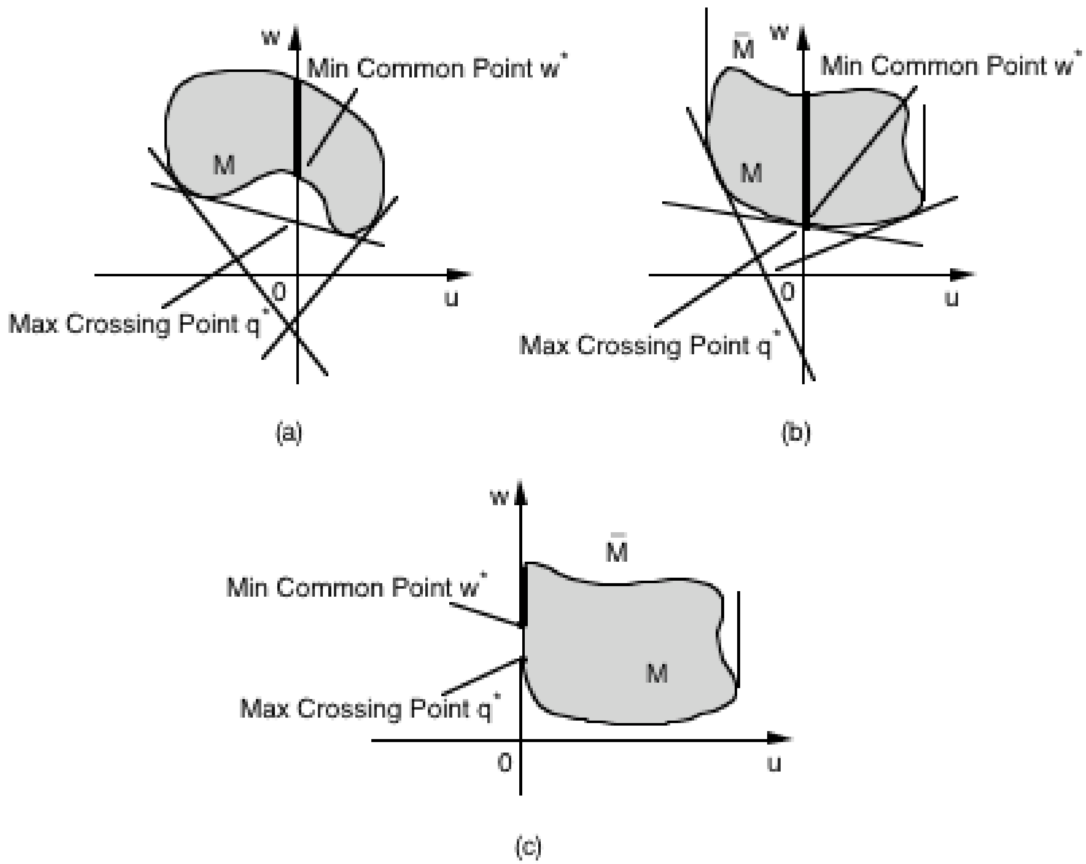
$\inf_x \sup_z \Phi$ = Min common value

$\sup_z \inf_x \Phi$ = Max crossing value

TWO ISSUES IN CONVEX PROGRAMMING AND MINIMAX

- When is there no duality gap (in convex programming), or inf sup = sup inf (in minimax)?
- When does an optimal dual solution exist (in convex programming), or the sup is attained (in minimax)?
- Min common/max crossing framework shows that
  - 1st question is a lower semicontinuity issue
  - 2nd question is an issue of existence of a nonvertical support hyperplane (or subgradient) at the origin
- Further analysis is needed for more specific answers

GRAPHICAL VIEW (ASSUMING CONVEXITY)



- Existence of nonvertical support plane (dual solution)
- No duality gap (semicontinuity issue)

II
UNIFICATION OF EXISTENCE AND NO DUALITY GAP ISSUES

INTERSECTIONS OF NESTED FAMILIES OF CLOSED SETS

- **We will connect two basic problems in optimization**
  - Attainment of a minimum of a function f over a set X
  - Existence of a duality gap
- **The 1st question is a set intersection issue:**

  The set of minima is the intersection of the nonempty level sets $\{x \in X \mid f(x) \leq \gamma\}$
- **The 2nd question is a lower semicontinuity issue:**

  When is the function

  $$p(u) = \inf_x F(x,u)$$

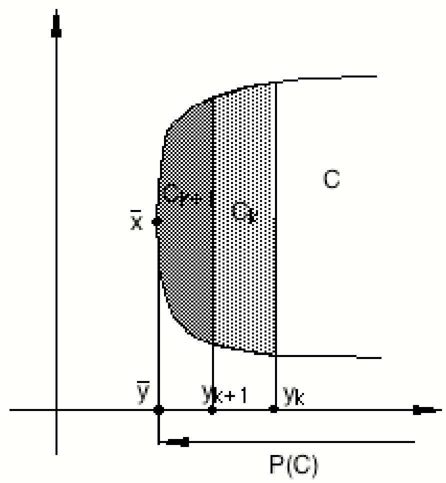
  lower semicontinuous, assuming F(x,u) is convex and lower semicontinuous?

PRESERVATION OF SEMICONTINUITY UNDER PARTIAL MINIMIZATION

- 2nd question also involves set intersection
- Key observation: For $p(u) = \inf_x F(x,u)$, we have

$$\text{Closure}\big(P(\text{epi}(F))\big) \supset \text{epi}(p) \supset P\big(\text{epi}(F)\big)$$

where $P(\cdot)$ is projection on the space of u. So if projection preserves closedness, F is l.s.c.



Given C, when is P(C) closed?

If y_k is a sequence in P(C) that converges to $\bar{y}$, we must show that the intersection of the C_k is nonempty

UNIFIED TREATMENT OF EXISTENCE OF SOLUTIONS AND DUALITY GAP ISSUES

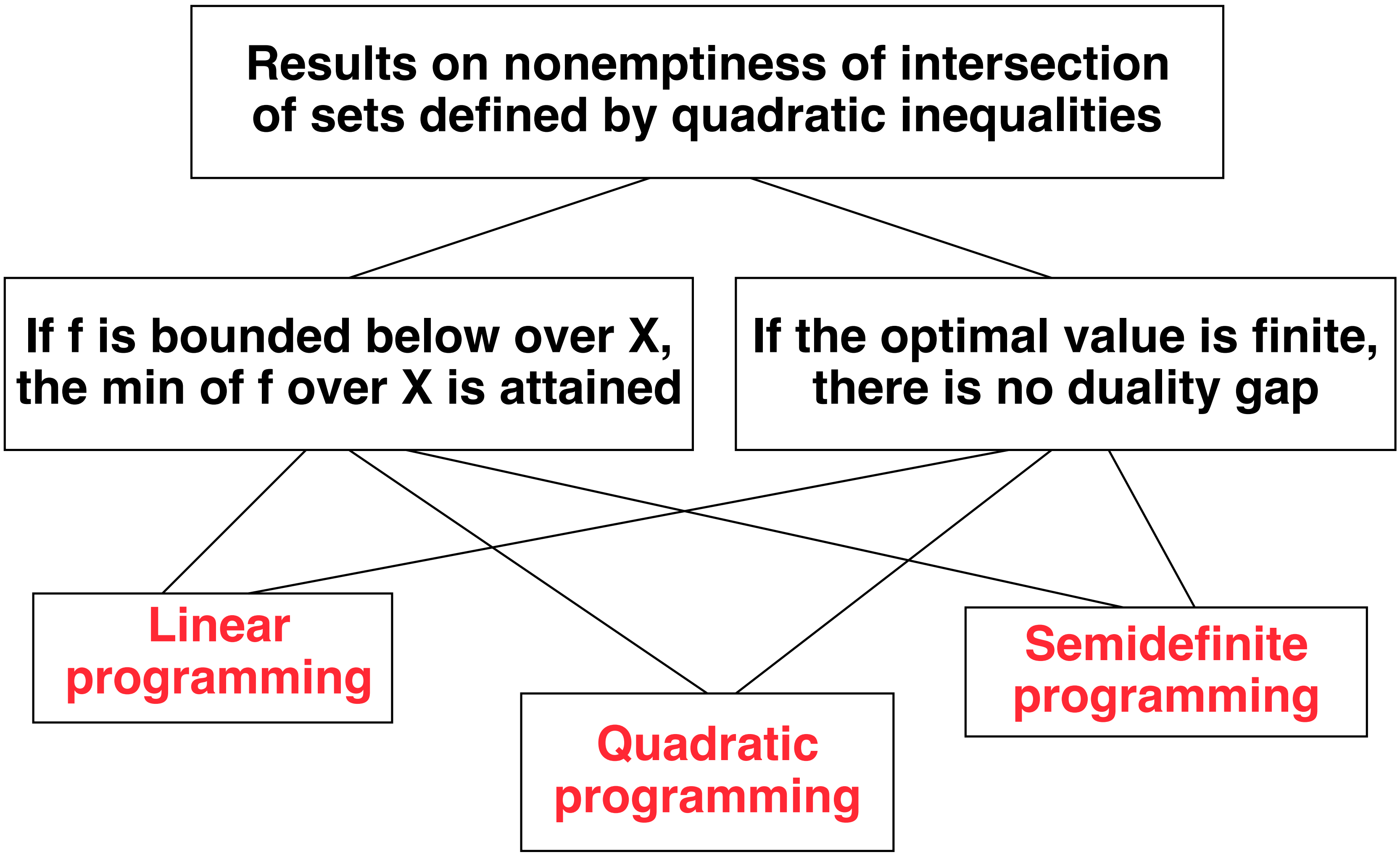
Results on nonemptiness of intersection of a nested family of closed sets (use of directions of recession)

No duality gap results In convex programming

inf sup Φ = sup inf Φ

Existence of minima of f over X

THE ROLE OF QUADRATIC FUNCTIONS



III
LAGRANGE MULTIPLIER THEORY / PSEUDONORMALITY

LAGRANGE MULTIPLIERS

- Problem (smooth, nonconvex):

  Min $f(x)$

  subject to $x \in X$, $h_i(x)=0$, $i = 1,\ldots,m$

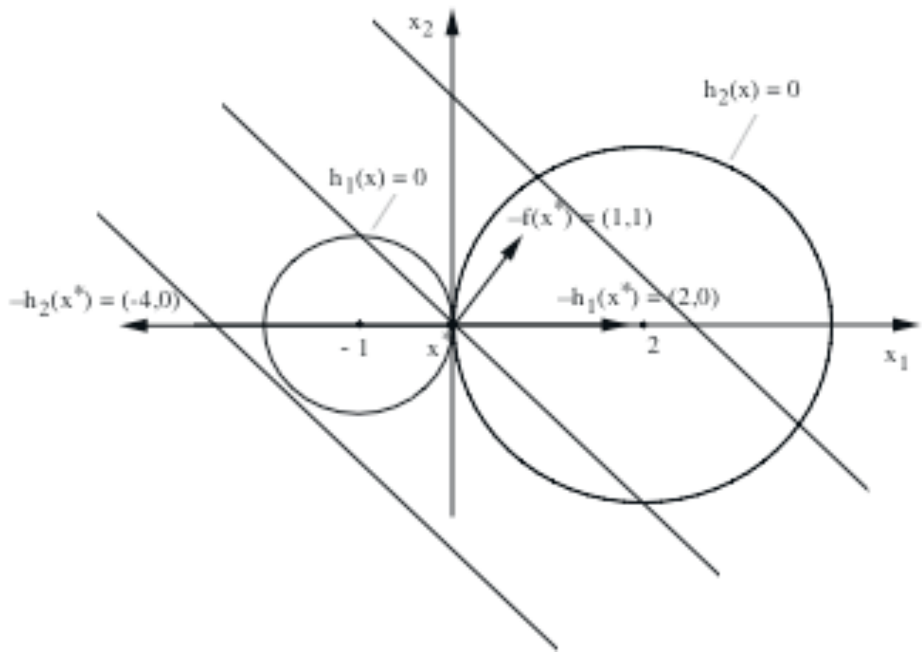
- Necessary condition for optimality of x^* (case $X = R^n$): Under some "constraint qualification", we have

  $$\nabla f(x^*) + \Sigma_i \lambda_i \nabla h_i(x^*) = 0$$

  for some Lagrange multipliers λ_i

- Basic analytical issue: What is the fundamental structure of the constraint set that guarantees the existence of a Lagrange multiplier?

EXAMPLE WITH NO LAGRANGE MULTIPLIERS



- **Standard constraint qualifications (case $X = R^n$):**
  - **The gradients $\nabla h_i(x^*)$ are linearly independent**
  - **The functions h_i are affine**

ENHANCED FRITZ JOHN CONDITIONS

If x^* is optimal, there exist $\mu_0 \geq 0$ and λ_i, not all 0, such that

$$\mu_0 \nabla f(x^*) + \Sigma_i \lambda_i \nabla h_i(x^*) = 0$$

and a sequence $\{x^k\}$ with $x^k \to x^*$ and such that

$$f(x^k) < f(x^*) \text{ for all } k,$$

$$\lambda_i h_i(x^k) > 0 \quad \text{for all } i \text{ with } \lambda_i \neq 0 \quad \text{and all } k$$

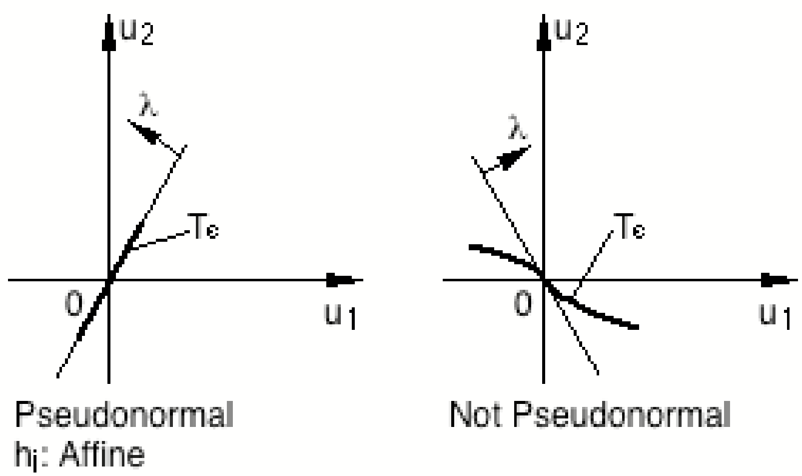
NOTE: If $\mu_0 > 0$, the λ_i are Lagrange multipliers with a special sensitivity property (they indicate the direction in which constraints must be violated to effect cost improvement)

PSEUDONORMALITY

- A feasible point x^* is pseudonormal if one cannot find λ_i and a sequence $\{x^k\}$ with $x^k \to x^*$ such that

$$\Sigma_i \lambda_i \nabla h_i(x^*) = 0, \qquad \Sigma_i \lambda_i h_i(x^k) > 0 \quad \text{for all } k$$

- Pseudonormality at x^* ==> $\mu_0 = 1$ in the F-J conditions (so there exists a "special" Lagrange multiplier)



To visualize:

Map an ε -ball around x^* onto the constraint space $T_\varepsilon = \{h(x) \mid \|x-x^*\| < \varepsilon\}$

INFORMATIVE LAGRANGE MULTIPLIERS

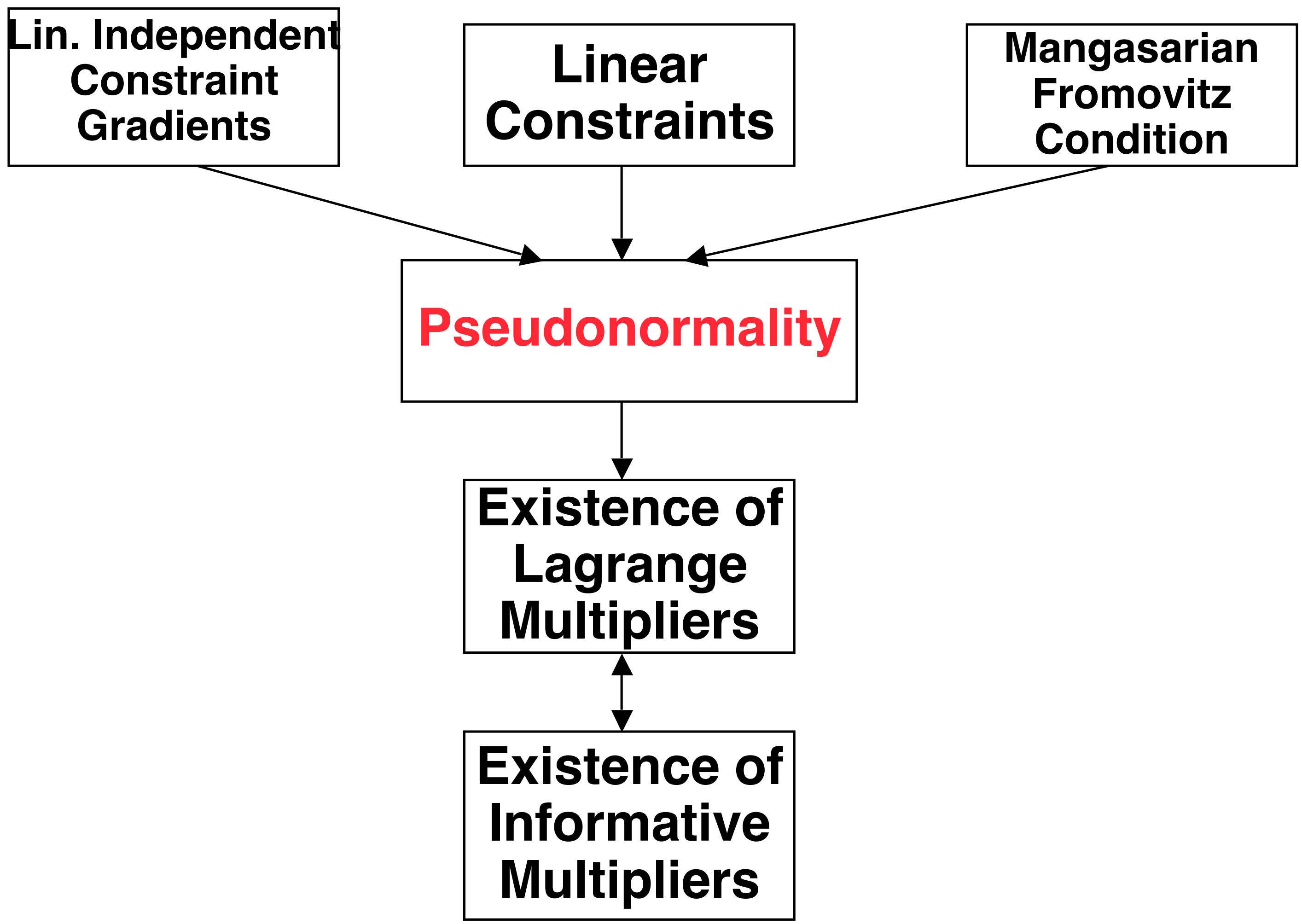
- **The Lagrange multipliers obtained from the enhanced Fritz John conditions have a special sensitivity property:**

  They indicate the constraints to violate in order to improve the cost

- **We call such multipliers informative**

- **Proposition: If there exists at least one Lagrange multiplier vector, there exists one that is informative**

THE STRUCTURE OF THE THEORY FOR $X = R^n$



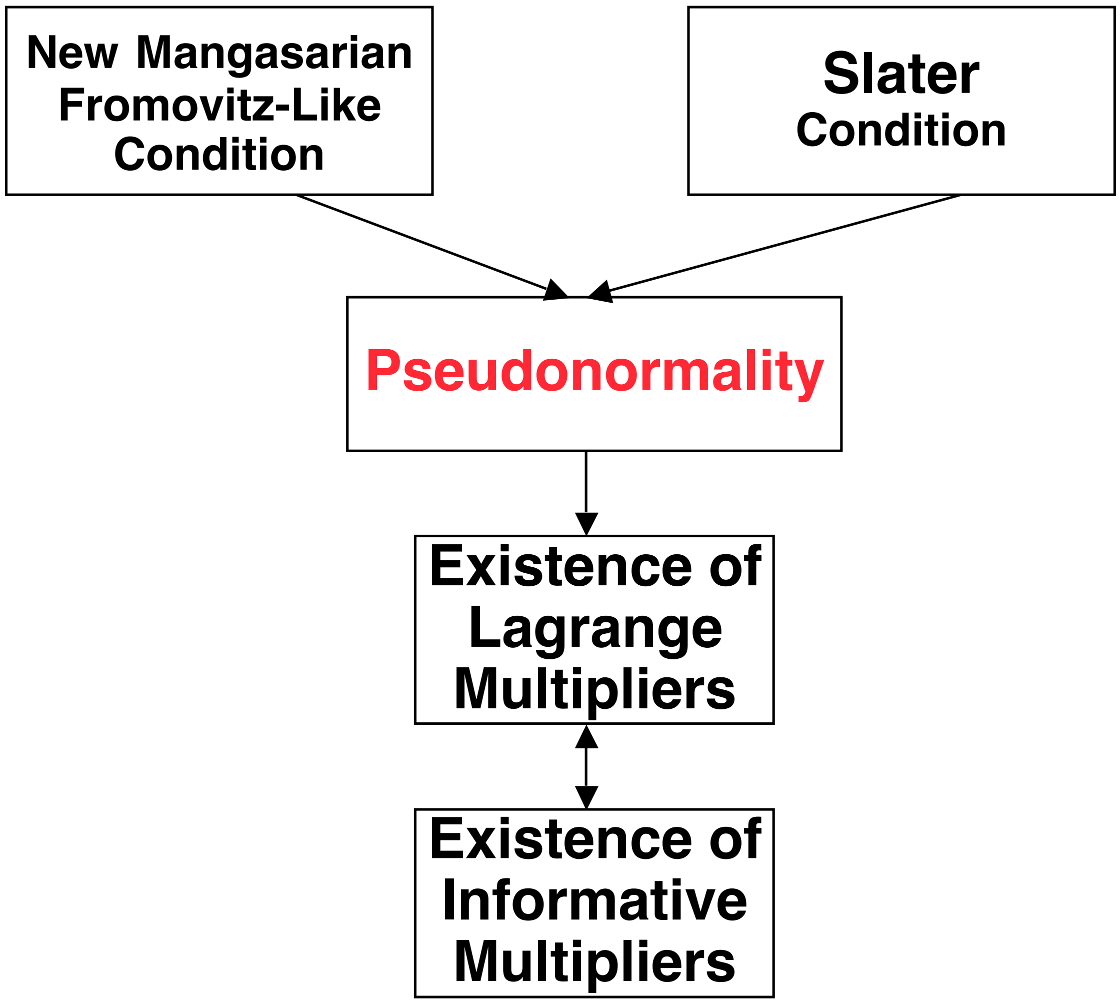
EXTENSIONS/CONNECTIONS TO NONSMOOTH ANALYSIS

- **F-J conditions for an additional constraint $x \in X$**
- **The stationarity condition becomes**

$$-\left(\mu_0 \nabla f(x^*) + \Sigma_I \lambda_i \nabla h_i(x^*)\right) \in (\text{normal cone of } X \text{ at } x^*)$$

- **X is called regular at x* if the normal cone is equal to the polar of its tangent cone at x* (example: X convex)**
- **If X is not regular at x*, the Lagrangian may have negative slope along some feasible directions**
- **Regularity is the fault line beyond which there is no satisfactory Lagrange multiplier theory**

THE STRUCTURE OF THE THEORY FOR X: REGULAR



EXTENSIONS

- Enhanced Fritz John conditions and pseudonormality for convex problems, when existence of a primal optimal solution is not assumed
- Connection of pseudonormality and exact penalty functions
- Connection of pseudonormality and the classical notion of quasiregularity

CONCLUDING REMARKS

- Optimization is becoming a universal tool in applications
- Convexity is the “soul” of optimization
- Geometry is the “soul” of convexity
- Very few simple geometric ideas are sufficient to unify/clarify most of convex optimization
- Theoretical research on convexity is still active