

# **A NEW LOOK AT CONVEX ANALYSIS AND OPTIMIZATION**

**Dimitri Bertsekas  
M.I.T.**

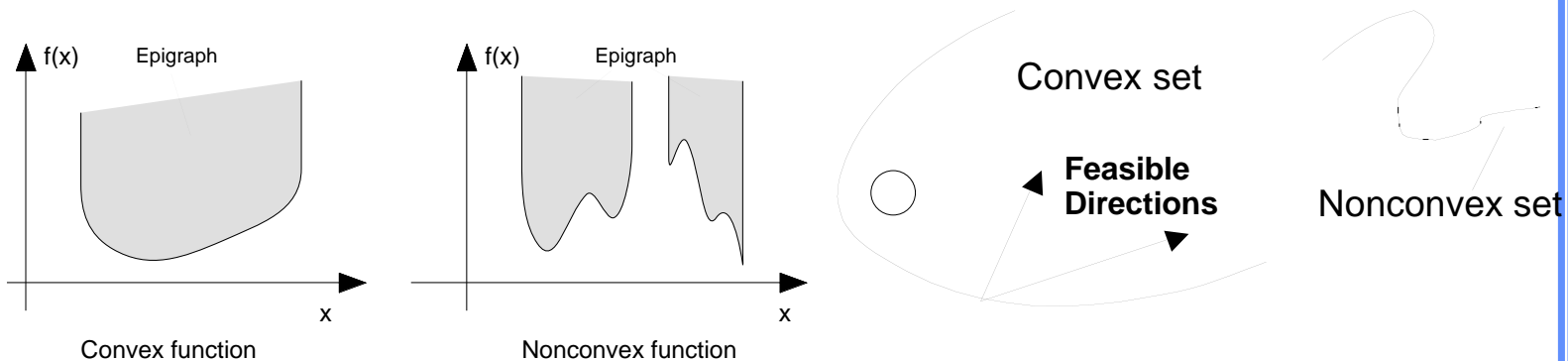
**FEBRUARY 2003**

# OUTLINE

- **Convexity issues in optimization**
- **Historical remarks**
- **Our treatment of the subject**
- **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**

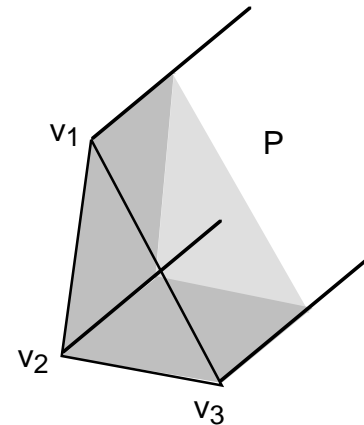
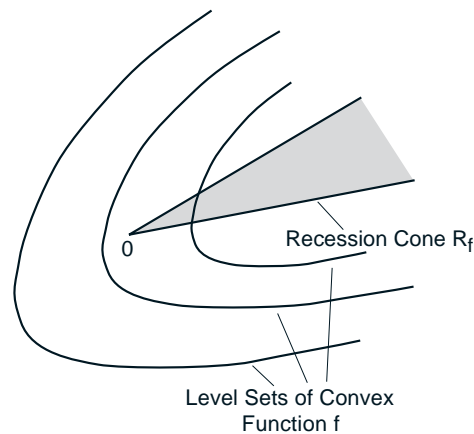
# WHY IS CONVEXITY IMPORTANT IN OPTIMIZATION I

- 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



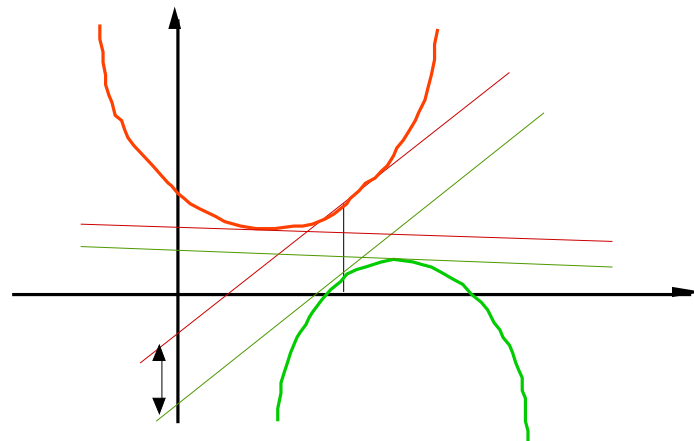
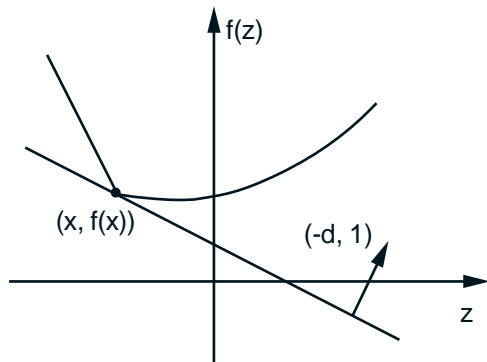
# WHY IS CONVEXITY IMPORTANT IN OPTIMIZATION II

- 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**
  - **Duality: Fenchel, Gale, Kuhn, Tucker**
  - **Optimization-related convexity: Fenchel**
- **60s-70s: Consolidation**
  - **Rockafellar**
- **80s-90s: Extensions to nonconvex optimization and nonsmooth analysis**
  - **Clarke, Mordukovich, Rockafellar-Wets**

# ABOUT THE BOOK

- **Convex Analysis and Optimization, by D. P. Bertsekas, with A. Nedic and A. Ozdaglar (March 2003)**
- **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**

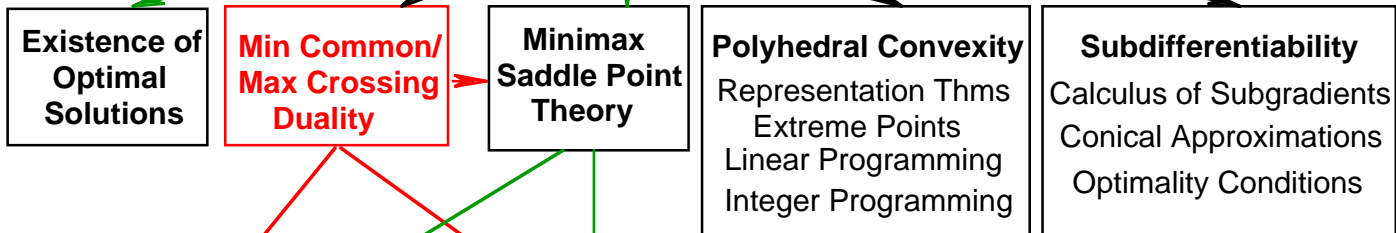
# OUTLINE OF THE BOOK

## Chapter 1

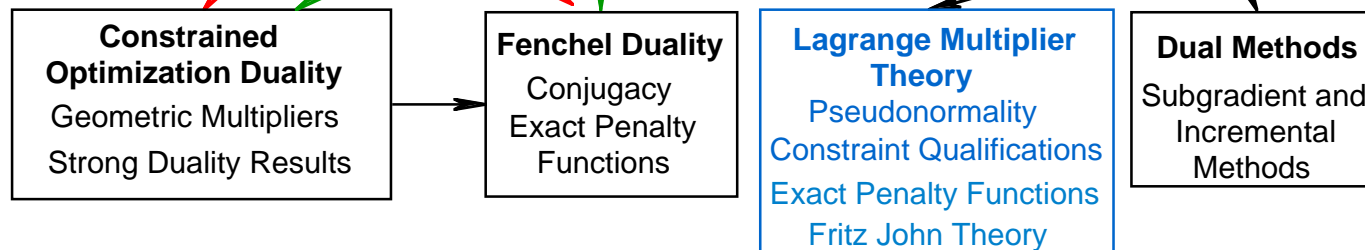
### Basic Convexity Concepts

Convex Sets and Functions  
 Semicontinuity and Epigraphs  
 Convex and Affine Hulls  
 Relative Interior  
 Directions of Recession  
 Closed Set Intersections

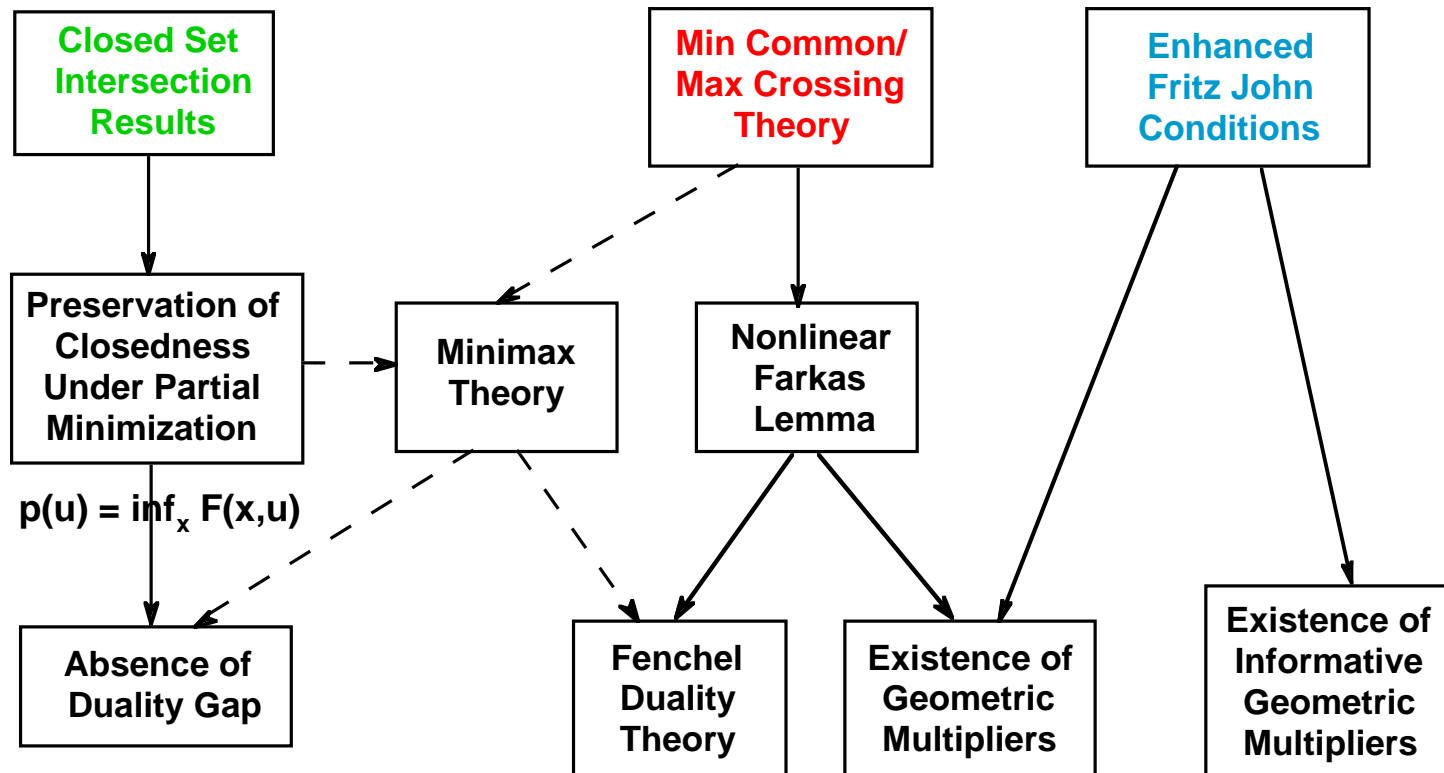
## Chapters 2-4



## Chapters 5-8

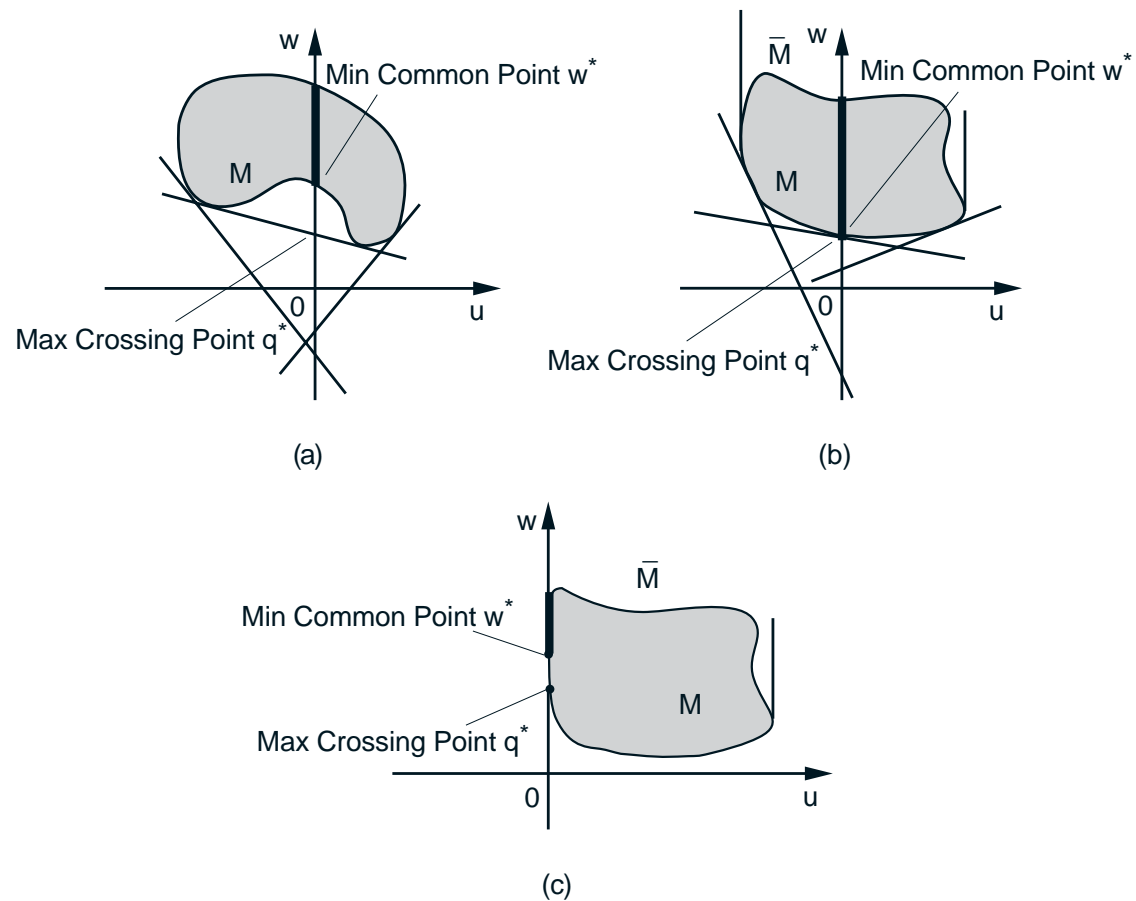


# OUTLINE OF DUALITY ANALYSIS



# I MIN COMMON/MAX CROSSING DUALITY

# GEOMETRICAL VIEW OF DUALITY



# APPROACH

- **Prove theorems about the geometry of  $M$** 
  - Conditions on  $M$  guarantee that  $w^* = q^*$
  - Conditions on  $M$  that guarantee existence of a max crossing hyperplane
- **Choose  $M$  to reduce the constrained optimization problem, the minimax problem, and others, to special cases of the min common/max crossing framework**
- **Specialize the min common/max crossing theorems to duality and minimax theorems**

# CONVEX PROGRAMMING DUALITY

- Primal problem:

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

- Dual problem:

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

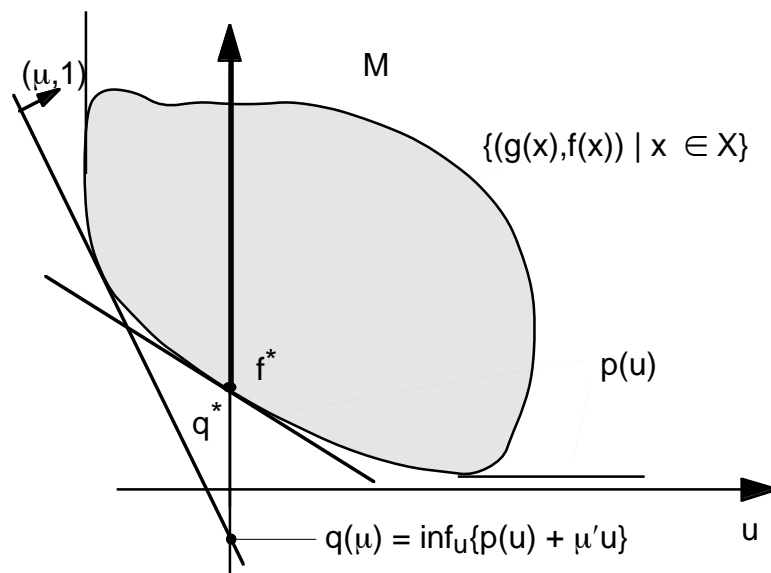
where the dual function is

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

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

$$M = \text{epi}(p), \quad p(u) = \inf_{x \in X, g_j(x) \leq u_j} f(x)$$

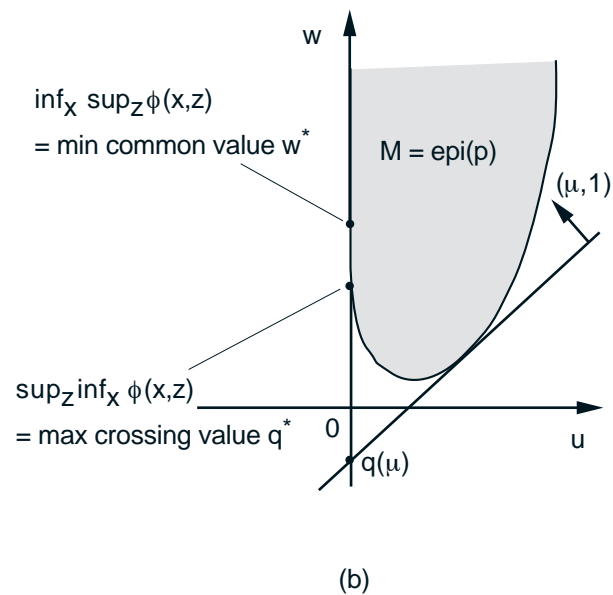
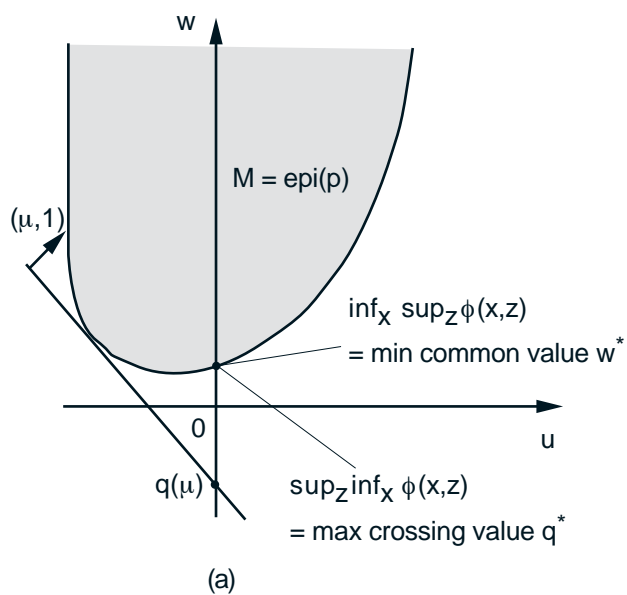
# VISUALIZATION



# 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  $\Phi$
- Min common/max crossing framework:
  - $M = \text{epigraph of the function}$
  - $$p(u) = \inf_x \sup_z \{ \Phi(x,z) - u'z \}$$
- $\inf_x \sup_z \Phi = \text{Min common value}$
- $\sup_z \inf_x \Phi = \text{Max crossing value}$

# VISUALIZATION



## 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

**II**  
**UNIFICATION OF EXISTENCE  
AND NO DUALITY GAP ISSUES**

# INTERSECTIONS OF NESTED FAMILIES OF CLOSED SETS

- Two basic problems in convex 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

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

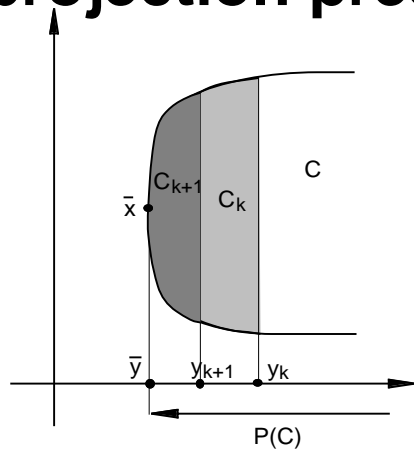
lower semicontinuous, assuming  $F(x,u)$  is convex and lower semicontinuous?

# PRESERVATION OF SEMICONTINUITY UNDER PARTIAL MINIMIZATION

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

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

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

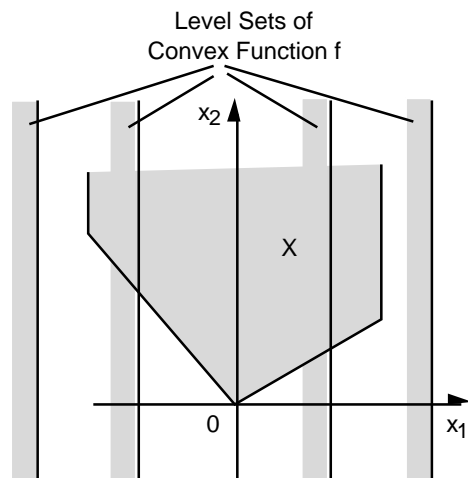
No duality gap results  
In convex programming

$$\inf \sup \Phi = \sup \inf \Phi$$

Existence of minima of  
f over X

# THE ROLE OF POLYHEDRAL ASSUMPTIONS: AN EXAMPLE

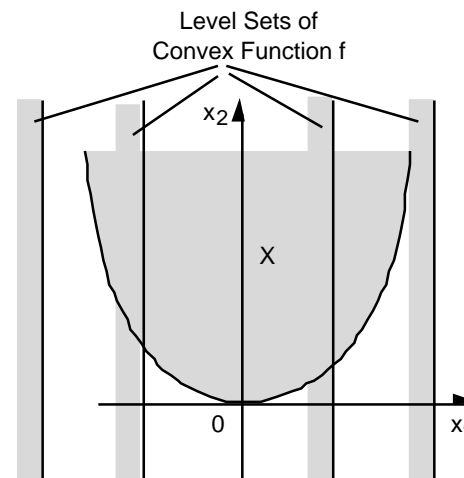
Polyhedral Constraint Set  $X$



**Minimum Attained**

$$f(x) = e^{x_1}$$

Nonpolyhedral Constraint Set  $X$



**Minimum not Attained**

**The min is attained if  $X$  is polyhedral, and  $f$  is constant along common directions of recession of  $f$  and  $X$**

# THE ROLE OF QUADRATIC FUNCTIONS

Results on nonemptiness of intersection of sets defined by quadratic inequalities

If  $f$  is bounded below over  $X$ , the min of  $f$  over  $X$  is attained

If the optimal value is finite, there is no duality gap

**Linear programming**

**Quadratic programming**

**Semidefinite programming**

# III LAGRANGE MULTIPLIER THEORY / PSEUDONORMALITY

# LAGRANGE MULTIPLIERS

- Problem (smooth, nonconvex):

$$\text{Min } f(x)$$

$$\text{subject to } x \in X, \quad h_i(x) = 0, \quad i = 1, \dots, m$$

- Necessary condition for optimality of  $x^*$  (case  $X = \mathbb{R}^n$ ):  
Under some “constraint qualification”, we have

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

for some Lagrange multipliers  $\lambda_i$

- Basic analytical issue: **What is the fundamental structure of the constraint set that guarantees the existence of a Lagrange multiplier?**
- Standard constraint qualifications (case  $X = \mathbb{R}^n$ ):
  - 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  $\lambda_i$ , not all 0, such that

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

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

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

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

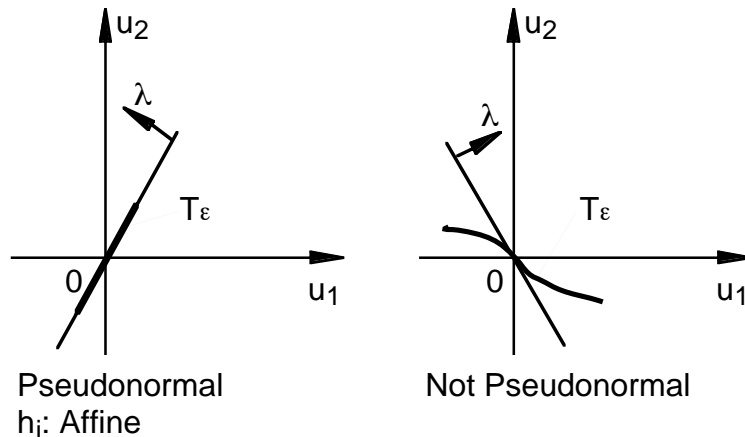
**NOTE:** If  $\mu_0 > 0$ , the  $\lambda_i$  are Lagrange multipliers with a sensitivity property (we call them **informative**)

# PSEUDONORMALITY

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

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

- Pseudonormality at  $x^*$  guarantees that, if  $x^*$  is optimal, we can take  $\mu_0 = 1$  in the F-J conditions (so there exists an informative Lagrange multiplier)



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

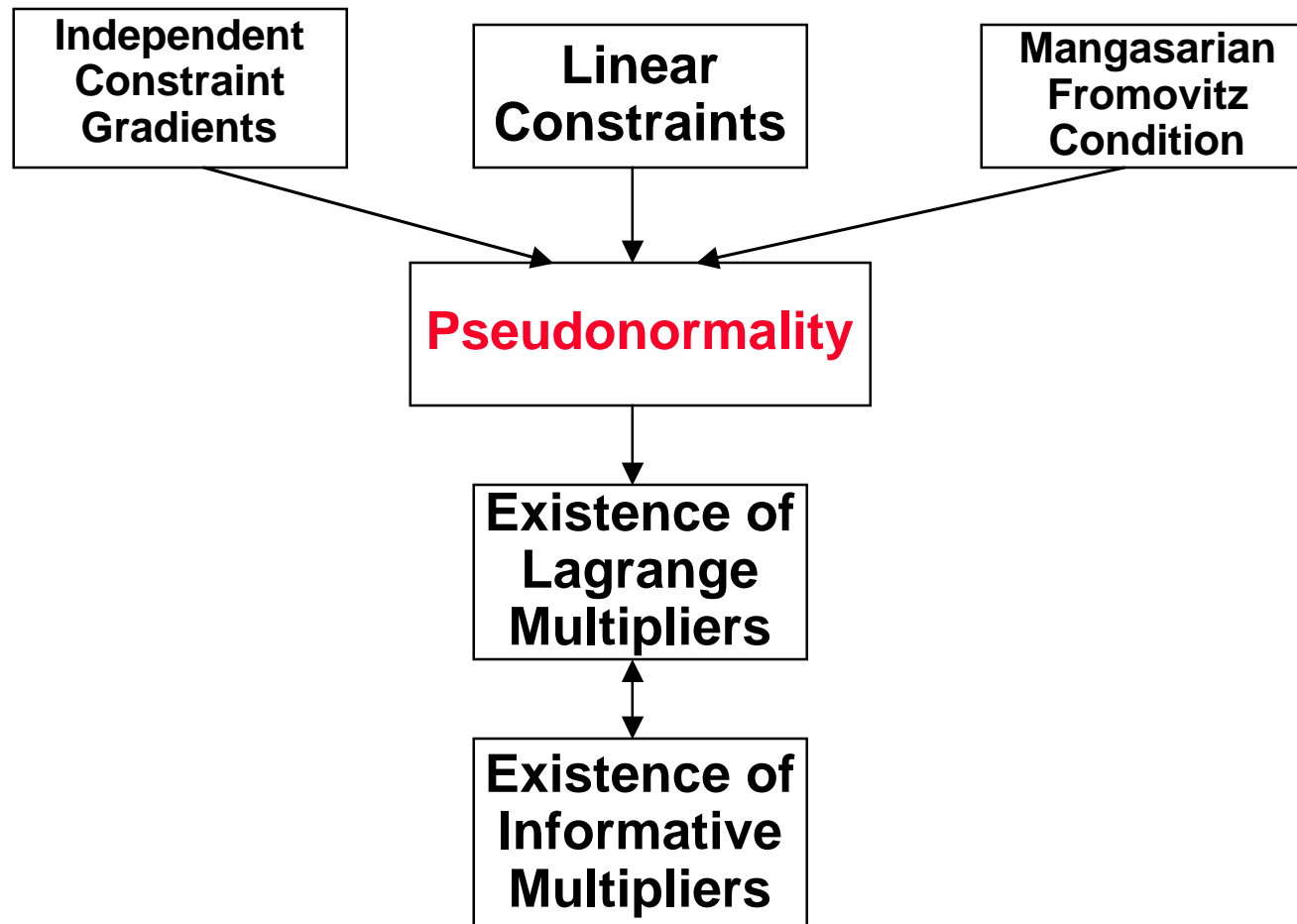
# INFORMATIVE LAGRANGE MULTIPLIERS

- 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

# EXTENSIONS/CONNECTIONS TO NONSMOOTH ANALYSIS

- F-J conditions for an additional constraint  $x \in X$
- The stationarity condition becomes
  - $(\nabla f(x^*) + \sum_i \lambda_i \nabla h_i(x^*)) \in$  (normal cone of  $X$  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 $X = \mathbb{R}^n$



# THE STRUCTURE OF THE THEORY FOR X: REGULAR

New Mangasarian  
Fromovitz-Like  
Condition

Slater  
Condition

**Pseudonormality**

Existence of  
Lagrange  
Multipliers

Existence of  
Informative  
Multipliers

# 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**