# IM2010: Operations Research
# Nonlinear Programming
# (Chapter 11)

Ling-Chieh Kung

Department of Information Management
National Taiwan University

May 8, 2013

# Road map

- **Motivating examples**.
- Convex programming.
- Solving single-variate NLPs.
- Lagrangian duality and the KKT condition.

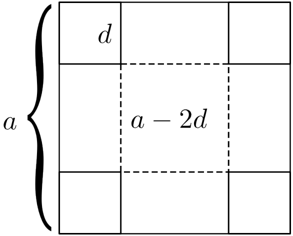
# Example: pricing a single good

- Suppose a retailer purchases one product at a unit cost $c$.
- It chooses a unit retail price $p$ to maximize its total profit.
- The demand is a function of $p$: $D(p) = a - bp$.
- What is the mathematical program that finds the optimal price?
  - Parameters: $a > 0, b > 0, c > 0$.
  - Decision variable: $p$.

$$\begin{array}{rl} \max & (p-c)(a-bp) \\ \text{s.t.} & p \geq 0. \end{array}$$

# Example: folding a piece of paper

- We are given a piece of square paper whose edge length is $a$.
- We want to cut down four small squares, each with edge length $d$, at the four corners.
- We then fold this paper to create a container.
- How to choose $d$ to maximize the volume of the container?




$$\begin{array}{rl} \max & (a-2d)^2 d \\ \text{s.t.} & 0 \leq d \leq \frac{a}{2}. \end{array}$$

# Example: locating a hospital

- In a country, there are $n$ cities, each lies at location $(x_i, y_i)$.
- We want to locate a hospital at location $(x, y)$ to minimize the distance between city 1 (the capital) and the hospital.
- However, we want none of the cities is far from the hospital by distance $d$.

$$
\begin{array}{rl}
\min & \sqrt{(x - x_1)^2 + (y - y_1)^2} \\
\text{s.t.} & \sqrt{(x - x_i)^2 + (y - y_i)^2} \leq d \quad \forall i = 1, ..., n.
\end{array}
$$

# Nonlinear programming

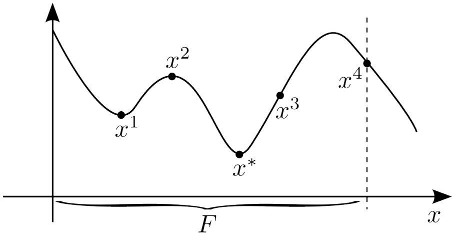
- In all the three examples, the program is by nature **nonlinear**.
- Moreover, it is impossible to linearize these formulation.
  - Because the trade off can only be modeled in a nonlinear way.
- In general, a nonlinear program (NLP) can be formulated as

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq b_i \quad \forall i = 1, ..., m. \end{aligned}$$

  - $x \in \mathbb{R}^n$: there are $n$ decision variables.
  - There are $m$ constraints.
  - This is a nonlinear program unless $f$ and $g_i$s are all linear in $x$.
- The study of optimizing nonlinear programs is nonlinear programming (also abbreviated as NLP).

# Difficulties of nonlinear programming

- Compared with LP, NLP is much more **difficult**.
- Given an NLP, it is possible that **no one** in the world knows how to solve it (i.e., find the global optimum) efficiently. Why?
- Difficulty 1: In an NLP, a local min **may not** be a global min.



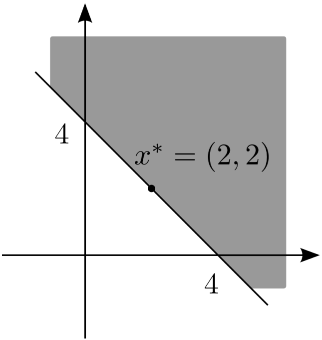

- A greedy search may stop at a local min.

# Difficulties of nonlinear programming

- Difficulty 2: In an NLP which has an optimal solution, there may be **no extreme point** optimal solution.
- For example:

$$\begin{array}{rl} \min & x_1^2 + x_2^2 \\ \text{s.t.} & x_1 + x_2 \geq 4. \end{array}$$

- The optimal solution $x^* = (2, 2)$ is not an extreme point.
- In fact, there is no extreme point.



# Difficulties of nonlinear programming

- For an NLP:
  - What are the conditions that make a local min always a global min?
  - What are the conditions that guarantee an extreme point optimal solution (when there is an optimal solution)?
- To answer these questions, we need convex sets and convex and concave functions.

# Road map

- Motivating examples.
- **Convex programming.**
- Solving single-variate NLPs.
- Lagrangian duality and the KKT condition.

# Convex sets

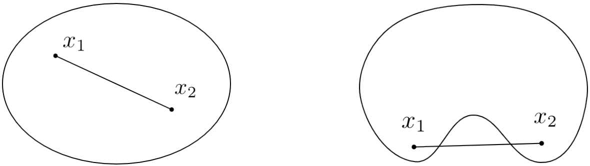
- Recall that we have defined convex sets and functions:

**Definition 1 (Convex sets)**

*A set $F$ is convex if*

$$\lambda x_1 + (1 - \lambda) x_2 \in F$$

*for all $\lambda \in [0, 1]$ and $x_1, x_2 \in F$.*



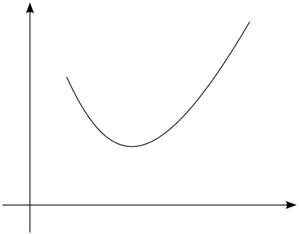
# Convex functions

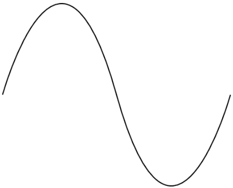
**Definition 2 (Convex functions)**

*A function* $f(\cdot)$ *is convex if*

$$f\big(\lambda x_1 + (1-\lambda)x_2\big) \leq \lambda f(x_1) + (1-\lambda) f(x_2)$$

*for all* $\lambda \in [0,1]$ *and* $x_1, x_2 \in F$.





# Condition for global optimality

- Suppose we minimize a convex function with no constraint, a local minimum is a global minimum.
- When there are constraints, as long as the **feasible region** is also **convex**, the desired property still holds.

**Proposition 1**

*For an NLP* $\min_{x \in F} f(x)$, *if*

- *the feasible region* $F$ *is a convex set and*
- *the objective function* $f$ *is a convex function,*

*a local min is a global min.*

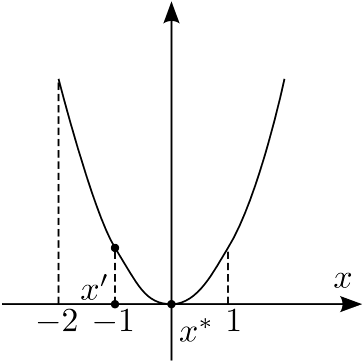
*Proof.* See Proposition 1 in slides "ORSP13_03_BasicsOfLP". □

# Convexity of the feasible region is required

- Consider the following example

$$\begin{array}{rl} \min & x^2 \\ \text{s.t.} & x \in [-2,-1] \cup [0,1]. \end{array}$$

  Note that the feasible region $[-2,-1] \cup [0,1]$ is not convex.
- The local min $x' = -1$ is not a global min. The unique global min is $x^* = 0$.



# Condition for extreme point optimal solutions

- While minimizing a convex function gives us a special property, how about minimizing a concave function?

**Proposition 2**

*For an NLP* $\min_{x \in F} f(x)$, *if*

- *the feasible region* $F$ *is a convex set,*
- *the objective function* $f$ *is a concave function, and*
- *an optimal solution exists,*

*there exists an extreme point optimal solution.*

*Proof.* Beyond the scope of this course. □

# Convex programs

- Between the above two propositions, Proposition 1 is applied more in solving NLPs.
- We give those NLPs that satisfy the condition in Proposition 1 a special name: **convex programs**.

**Definition 3**

*An NLP* $\min_{x \in F} f(x)$ *is a convex program if its feasible region* $F$ *is convex and the objective function* $f$ *is convex over* $F$*.*

**Corollary 1**

*For a convex program, a local min is a global min.*

- Therefore, for convex programs, a **greedy search** finds an optimal solution (if one exists).

# Convex programming

- The field of solving convex programs is **convex programming**.
  - Several optimality conditions have been developed to **analytically** solve convex programs.
  - Many efficient search algorithms have been developed to **numerically** solve convex programs.
  - In particular, the simplex method numerically solve LPs, which are special cases of convex programs.
- In this course, we will only discuss how to analytically solve single-variate convex programs.
- All you need to know are:
  - People **can** solve convex programs.
  - People **cannot** solve general NLPs.

# Road map

- Motivating examples.
- Convex programming.
- **Solving single-variate NLPs**.
- Lagrangian duality and the KKT condition.

# Solving single-variate NLPs

- Here we discuss how to analytically solve single-variate NLPs.
  - "Analytically solving a problem" means to express the solution as a **function** of problem parameters **symbolically**.
- Even though solving problems with only one variable is restrictive, we will see some useful examples in the remaining semester.
- We will focus on **twice differentiable** functions and try to utilize **convexity** (if possible).

# Convexity of twice differentiable functions

- For a general function, we may need to use the definition of convex functions to show its convexity.
- For single-variate twice differentiable functions (i.e., the second-order derivative exists), there are useful properties:

**Proposition 3**

*For a single-variate twice differentiable function $f(x)$:*

- $f$ *is convex in* $[a, b]$ *if* $f''(x) \geq 0$ *for all* $x \in [a, b]$.
- $\bar{x}$ *is a local min only if* $f'(\bar{x}) = 0$.
- *If* $f$ *is convex,* $x^*$ *is a global min if and only if* $f'(x^*) = 0$.

*Proof.* For the first two, see your Calculus textbook. The last one is a combination of the second one and Proposition 1. □

# Convexity of twice differentiable functions

- The condition $f'(x) = 0$ is called the **first order condition** (FOC).
- For all functions, FOC is **necessary** for a local min.
- For convex functions, FOC is also **sufficient** for a global min.

# Example 1

- Now let's apply these properties to solve Example 1

$$\begin{array}{rl} \max & \pi(p) = (p-c)(a-bp) \\ \text{s.t.} & p \geq 0. \end{array}$$

  - The feasible region $[0, \infty)$ is convex.
  - Let's first ignore this constraint.
  - The profit function is concave in $p$:

$$\pi'(p) = a - bp - b(p-c) \quad \text{and} \quad \pi''(p) = -2b < 0.$$

  - An optimal solution $p^*$ satisfies

$$\pi'(p^*) = 0 \Rightarrow a - 2bp^* + bc = 0 \Rightarrow p^* = \frac{a+bc}{2b}.$$

  - As $p^* = \frac{a+bc}{2b}$ is feasible, it is optimal.
  - Does $p^* = \frac{a+bc}{2b}$ make sense?

# Example 2

- Now condition Example 2:

$$\begin{array}{ll} \max & V(d) = (a-2d)^2 d \\ \text{s.t.} & 0 \leq d \leq \frac{a}{2} \end{array}.$$

  - The feasible region $[0, \frac{d}{2}]$ is convex.
  - The volume function $V(d) = 4d^3 - 4ad^2 + a^2 d$ is not concave!
  - However, as long as it is concave over the feasible region, FOC will still be sufficient (if we apply it to only feasible points). Is it?

$$V'(d) = 12d^2 - 8ad + a^2 \quad \text{and} \quad V''(d) = 24d - 8a.$$

    In the feasible region $[0, \frac{a}{2}]$, $V$ is also not concave.
  - What should we do?

# Example 2

- Recall that FOC is always necessary!
- We may find all the points that satisfy FOC and **compare** all those that are feasible.

$$V'(d) = 12d^2 - 8ad + a^2 = 0 \quad \Rightarrow \quad d = \frac{a}{6} \text{ or } \frac{a}{2}.$$

  - As $V\left(\frac{a}{6}\right) > V\left(\frac{a}{2}\right) = 0$, $\frac{a}{6}$ is optimal... ?
- Is this enough?
- As there are constraints, we also need to check the **boundaries**!
  - As both boundary points 0 and $\frac{a}{2}$ result in a zero objective value, $\frac{a}{6}$ is indeed optimal.

# Road map

- Motivating examples.
- Convex programming.
- Solving single-variate NLPs.
- **Lagrangian duality and the KKT condition**.

# Lagrangian relaxation

- Recall that we have learned duality for LP.
- The same idea can be applied to NLPs.
- Consider a **primal** NLP

$$
\begin{aligned}
z^* = \max_{x \in \mathbb{R}^n} \quad & f(x) \\
\text{s.t.} \quad & g_i(x) \leq b_i \quad \forall i = 1, ..., m.
\end{aligned}
$$

- The primal may be difficult:
  - There are many constraints.
  - The primal may be a nonconvex program.

# Lagrangian relaxation

- Instead of solving the primal directly, we may move all the constraints to the objective function:

$$\max_{x \in \mathbb{R}^n} \ f(x) + \sum_{i=1}^{m} \left[ b_i - g_i(x) \right].$$

  - Solving this program is easier but is not helpful. For example, the optimal solution may be infeasible!
  - To avoid violating a constraint $g_i(x) \leq b_i$, we may add a **penalty** $\lambda_i$ to this constraint. These $\lambda_i$s are called **Lagrange multipliers**.
  - This penalty $\lambda_i$ should be nonnegative. Why?
- For $\lambda = (\lambda_1, ..., \lambda_m) \geq 0$, the **Lagrangian relaxation** is

$$L(\lambda) = \max_{x \in \mathbb{R}^n} \ f(x) + \sum_{i=1}^{m} \lambda_i \left[ b_i - g_i(x) \right].$$

# Lagrangian relaxation provides a bound

- Like what we have done in LP duality, the Lagrangian relaxation provides a bound of the primal.

**Proposition 4**

*$L(\lambda) \geq z^*$ if $\lambda_i \geq 0$ for all $i = 1, ..., m$.*

*Proof.* We have

$$\begin{aligned} z^* &= \max_{x \in \mathbb{R}^n} \left\{ f(x) \middle| g_i(x) \leq b_i \ \forall i = 1, ..., m \right\} \\ &\leq \max_{x \in \mathbb{R}^n} \left\{ f(x) + \sum_{i=1}^m \lambda_i [b_i - g_i(x)] \middle| g_i(x) \leq b_i \ \forall i = 1, ..., m \right\} \\ &\leq \max_{x \in \mathbb{R}^n} \left\{ f(x) + \sum_{i=1}^m \lambda_i [b_i - g_i(x)] \right\} = L(\lambda), \end{aligned}$$

where the first inequality relies on $\lambda \geq 0$. □

# Lagrangian duality

- For a given $\lambda \geq 0$, the Lagrangian relaxation provides an upper bound of the primal.
- It is natural to search for the $\lambda$ that results in the **lowest** upper bound. This defines the **Lagrangian dual program**:

$$w^* = \min_{\lambda \geq 0} L(\lambda)$$

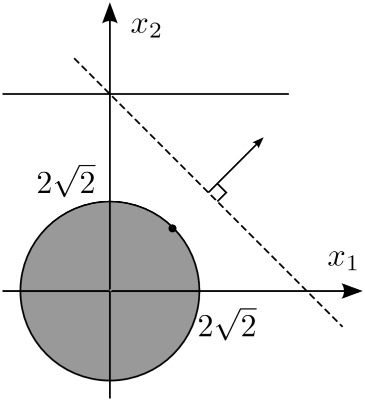
- As $L(\lambda) \geq z^*$ for all $\lambda \geq 0$, certainly $w^* \geq z^*$.
  - Examples exist show that $w^* > z^*$ for some NLPs.
  - It can be shown that $w^* = z^*$ for all convex programs (under some mild conditions).

# Example 1

- Consider the following example

$$\begin{aligned} z^* = \max \quad & x_1 + x_2 \\ \text{s.t.} \quad & x_1^2 + x_2^2 \leq 8 \\ & x_2 \leq 6. \end{aligned}$$

- For this primal program, the optimal solution is $x^* = (2, 2)$.
- What is the Lagrangian dual?



# Example 1

- Lagrangian relaxation:

$$L(\lambda) = \max_{x \in \mathbb{R}^2} \; x_1 + x_2 + \lambda_1(8 - x_1^2 - x_2^2) + \lambda_2(6 - x_2)$$

  for all $\lambda = (\lambda_1, \lambda_2) \geq 0$.
- Some examples:
  - $L(1,2) = \max_{x \in \mathbb{R}^2} \; -x_1^2 + x_1 - x_2^2 - x_2 + 20 = 20.5.$
  - $L(1,0) = \max_{x \in \mathbb{R}^2} \; -x_1^2 + x_1 - x_2^2 - x_2 + 8 = 8.5.$
  - $L(0,1) = \max_{x \in \mathbb{R}^2} \; x_1 + 6 = \infty.$

# Example 1

- Let's express $L(\lambda)$ as a function of $\lambda$ only:

$$L(\lambda) = \max_{x \in \mathbb{R}^2} \; -\lambda_1 x_1^2 + x_1 - \lambda_1 x_2^2 + (1 - \lambda_2) x_2 + 8\lambda_1 + 6\lambda_2.$$

- The optimal $x$ is $x_1^* = \frac{1}{2\lambda_1}$ and $x_2^* = \frac{1-\lambda_2}{2\lambda_1}$.
- So we plug in $x_1^*$ and $x_2^*$ back to the above program and obtain

$$L(\lambda) = \frac{1}{4\lambda_1} + \frac{(1-\lambda_2)^2}{4\lambda_1} + 8\lambda_1 + 6\lambda_2.$$

- The Lagrangian dual $\min_{\lambda \geq 0} L(\lambda)$ is thus

$$w^* = \min_{\lambda \geq 0} \; \frac{1}{4\lambda_1} + \frac{(1-\lambda_2)^2}{4\lambda_1} + 8\lambda_1 + 6\lambda_2,$$

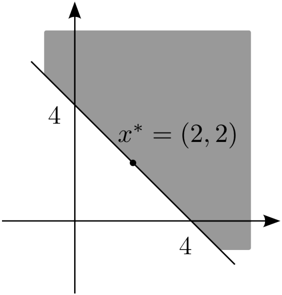
which is another NLP.

# Example 2

- Consider the primal

$$z^* = \min \quad x_1^2 + x_2^2$$
$$\text{s.t.} \quad x_1 + x_2 \geq 4$$

whose optimal solution is $x^* = (2, 2)$ with objective value $z^* = 8$.




- Lagrangian relaxation with $\lambda \geq 0$ (why nonnegative):

$$L(\lambda) = \min_{x \in \mathbb{R}^2} x_1^2 + x_2^2 + \lambda(4 - x_1 - x_2)$$
$$= 4\lambda + \min_{x \in \mathbb{R}^2} x_1^2 - \lambda x_1 + x_2^2 - \lambda x_2 = 4\lambda + \frac{x^2}{2}.$$

  - Note that $x_1^* = x_2^* = \frac{\lambda}{2}$ are optimal to the subprogram.

# Example 2

- Lagrangian duality:

$$w^* = \max_{\lambda \geq 0} \; L(\lambda) = 4\lambda - \frac{\lambda^2}{2}.$$

- Note that this is a convex program!
- As $L''(\lambda) = -1 < 0$, we apply FOC:

$$L'(\lambda^*) = 4 - \lambda^* = 0 \quad \Rightarrow \quad \lambda^* = 4.$$

  As $\lambda^*$ is feasible, it is optimal.
- The optimal dual objective value $w^* = 8 = z^*$.
- Moreover, the dual optimal solution allows us to find the primal optimal solution:

$$x_1^* = \frac{\lambda}{2} = 2 \quad \text{and} \quad x_2^* = \frac{\lambda}{2} = 2.$$

# From dual to primal

- Solving the Lagrangian dual may allow us to solve the primal.

**Proposition 5**

*For a "regular" convex program, solving the Lagrangian duality results in a primal optimal solution.*

*Proof.* Beyond the scope of this course. □

- We need some mild conditions to make a convex program "regular". While we omit those conditions in this course, all NLPs you see in this course are "regular".
- For a nonconvex program, this is not true!

# The KKT condition

- Now we present an optimality condition for general NLPs to close this session.

**Proposition 6 (KKT condition)**

*For a "regular" nonlinear program*

$$\begin{array}{rl} \max & f(x) \\ \text{s.t.} & g_i(x) \leq b_i \quad \forall i = 1, ..., m, \end{array}$$

*if $\bar{x}$ is a local max, then there exists $\lambda \in \mathbb{R}^m$ such that*

- $g_i(\bar{x}) \leq b_i$ *for all* $i = 1, ..., m$,
- $\lambda \geq 0$, *and*
- $\nabla f(\bar{x}) = \sum_{i=1}^m \lambda_i \nabla g_i(\bar{x})$,
- $\lambda_i g_i(\bar{x}) = 0$ *for all* $i = 1, ..., m$.

# The KKT condition

- For a multi-variate function $f(x)$ where $x \in \mathbb{R}^m$,

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} & \cdots & \frac{\partial f(x)}{\partial x_n} \end{bmatrix}^T$$

  is the gradient of $f$.
- Remarks for the KKT condition:
  - Condition 1 means $\bar{x}$ must be feasible.
  - Condition 2 means the Lagrangian multipliers should be penalties.
  - Condition 3 means the objective function in the Lagrangian relaxation satisfies the first order condition.
  - Condition 4 means "if the constraint is not binding at $\bar{x}$, the corresponding shadow price must be 0."
- Anyway, this will not appear in homework or exams.

# The story of the KKT condition

- About the discovery of this condition:
  - Harold W. Kuhn and Albert W. Tucker are two very famous mathematicians and economists.
  - In 1951, they together published a paper stating the KKT condition, which was called the Kuhn-Tucker condition at that time.
  - However, later scholars found that a master student William Karush has proved this condition in his master thesis in 1939.
  - Starting from then, the condition is called the KKT condition.
- Two things we may learn from this story:
  - Do not underestimate what we are doing.
  - Sadly, what you are reading (the KKT condition) was discovered 70 years ago, and we cannot even put it in your homework and exam...
- One final remark: The KKT condition is sufficient for convex programs.