IM 2010: Operations Research, Spring 2014

Nonlinear Programming (Part 1)

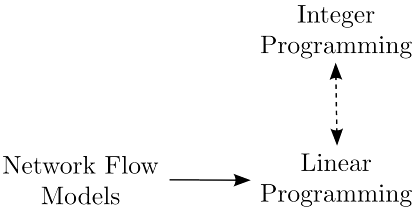
Ling-Chieh Kung

Department of Information Management
National Taiwan University

April 17, 2014

Introduction

- So far we spent most of our time on **Linear Programming**.
  - (Linear) **Integer Programming** complements Linear Programming.
  - **Network Flow Models** are special cases of Linear Programming.



- In these two lectures we introduce **Nonlinear Programming** (NLP).
  - Some functions are no more linear.
  - A generalization of Linear Programming.
  - Single-variate NLP in this week and multi-variate NLP in the next week.

Road map

- **Motivating examples**.
- Convex analysis.
- Solving single-variate NLPs.

Example: pricing a single good

- A retailer buys one product at a unit cost c.
- It chooses a unit retail price p.
- The demand is a function of p: $D(p) = a - bp$.
- How to formulate the problem of finding the profit-maximizing price?
  - Parameters: $a > 0, b > 0, c > 0$.
  - Decision variable: p.
  - Constraint: $p \geq 0$.
  - Formulation:

$$\begin{aligned} \max_{p} \quad & (p - c)(a - bp) \\ \text{s.t.} \quad & p \geq 0 \end{aligned}$$

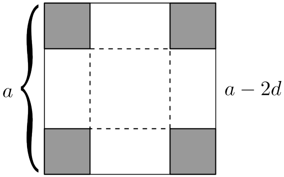
or

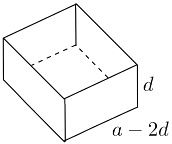
$$\max_{p \geq 0} (p - c)(a - bp).$$

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?

$$\max_{d \in [0, \frac{a}{2}]} (a - 2d)^2 d.$$





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 average Euclidean distance from the cities to the hospital.

$$\min_{x,y} \sum_{i=1}^{n} \sqrt{(x - x_i)^2 + (y - y_i)^2}.$$

- The problem can be formulated as an LP if we are working on Manhattan distances. For Euclidean distances, the formulation must be nonlinear.

Nonlinear Programming

- In all the three examples, the programs are by nature **nonlinear**.
  - Because the trade off can only be modeled in a nonlinear way.
- In general, a **nonlinear program** (NLP) can be formulated as

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

  - $x \in \mathbb{R}^n$: there are n decision variables.
  - There are m constraints.
  - This is an LP if f and g_is are all linear in x.
  - This is an NLP f and g_is are allowed to be nonlinear in x.
- The study of formulating and optimizing NLPs is **Nonlinear Programming** (also abbreviated as NLP).
  - Formulation is easy but optimization is hard.

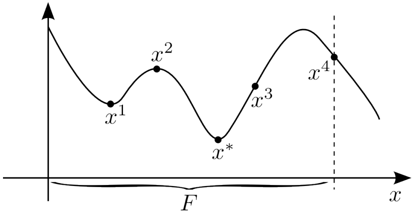
Difficulties of NLP

- Compared with LP, NLP is much more **difficult**.

Observation 1

In an NLP, a local minimum **is not always** *a global minimum.*

- Over the feasible region F, x^1 is a local minimum but not a global minimum. How about other points?



- A greedy search may be trapped at a local minimum.

Difficulties of NLP

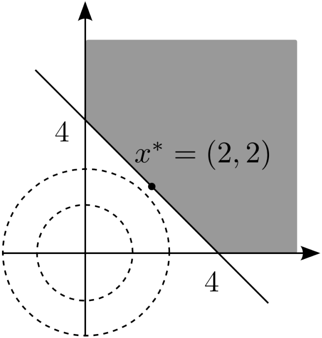
Observation 2

In an NLP which has an optimal solution, there may **exist no** *extreme point optimal solution.*

- For example:

$$\begin{array}{ll} \min_{x_1 \geq 0, x_2 \geq 0} & x_1^2 + x_2^2 \\ \text{s.t.} & x_1 + x_2 \geq 4. \end{array}$$

- The optimal solution $x^* = (2, 2)$ is not an extreme point.
- The two extreme points are not optimal.



Difficulties of NLP

- **No one** has invented an efficient algorithm for solving general NLPs (i.e., finding a global optimum).
- For an NLP:
  - We want to have a condition that makes a local minimum always a global minimum.
  - We want to have a condition that guarantees an extreme point optimal solution (when there is an optimal solution).
- To answer these questions, we need **convex analysis**.
  - Let's define convex sets and convex and concave functions.
  - Then we define convex programs and show that they have the first desired property.

Road map

- Motivating examples.
- **Convex analysis**.
- Solving single-variate NLPs.

Convex sets

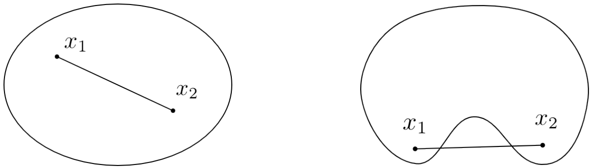
- Let's start by defining **convex sets** and **convex functions**:

Definition 1 (Convex sets)

A set $F \subseteq \mathbb{R}^n$ *is convex if*

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

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



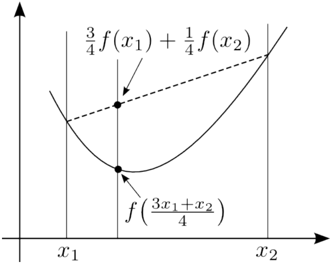
Convex functions

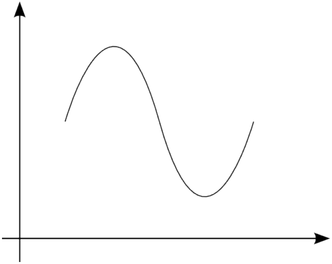
Definition 2 (Convex functions)

For a convex domain $F \subseteq \mathbb{R}^n$, a function $f : \mathbb{R}^n \to \mathbb{R}$ is convex over F 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 $x_1, x_2 \in F$.





Concave functions and some examples

Definition 3 (Concave functions)

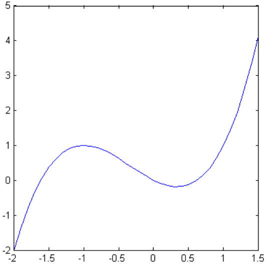
For a convex domain $F \in \mathbb{R}^n$, a function $f : \mathbb{R}^n \to \mathbb{R}$ is **concave** *over F if $-f$ is convex.*

- Convex sets?
  - $X_1 = [10, 20]$.
  - $X_2 = (10, 20)$.
  - $X_3 = \mathbb{N}$.
  - $X_4 = \mathbb{R}$.
  - $X_5 = \{(x, y) \in \mathbb{R}^2 | x^2 + y^2 \leq 4\}$.
  - $X_6 = \{(x, y) \in \mathbb{R}^2 | x^2 + y^2 \geq 4\}$.
- Convex functions?
  - $f_1(x) = x + 2, x \in \mathbb{R}$.
  - $f_2(x) = x^2 + 2, x \in \mathbb{R}$.
  - $f_3(x) = \sin x, x \in [0, 2\pi]$.
  - $f_4(x) = \sin x, x \in [\pi, 2\pi]$.
  - $f_5(x) = \log x, x \in (0, \infty)$.
  - $f_6(x, y) = x^2 + y^2, (x, y) \in \mathbb{R}^2$.

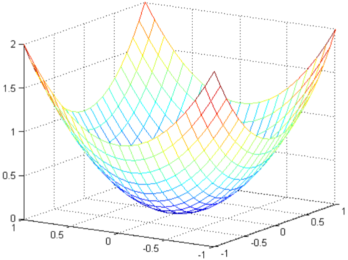
Local v.s. global optima

Proposition 1 (Global optimality of convex functions)

For a convex (concave) function f over a convex domain F, a local minimum (maximum) is a global minimum (maximum).



$f(x) = x^3 + x^2 - x.$

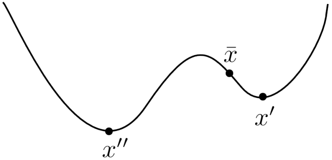


$f(x, y) = x^2 + y^2.$

Local v.s. global optima

Proof. Suppose a local minimum x' is not a global minimum and there exists x'' such that $f(x'') < f(x')$. Consider a small enough $\lambda > 0$ such that $\bar{x} = \lambda x'' + (1-\lambda)x'$ satisfies $f(\bar{x}) > f(x')$. Such $\bar{x}$ exists because x is a local minimum. Now, note that

$$\begin{aligned} f(\bar{x}) &= f(\lambda x'' + (1-\lambda)x') \\ &> f(x') \\ &= \lambda f(x') + (1-\lambda)f(x') \\ &> \lambda f(x'') + (1-\lambda)f(x'), \end{aligned}$$



which violates the fact that $f(\cdot)$ is convex. Therefore, by contradiction, the local minimum x must be a global minimum. □

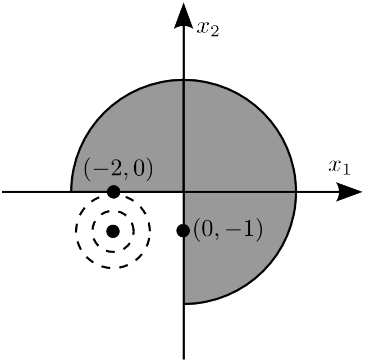
Convexity of the feasible region is required

- Consider the following example

$$\begin{array}{ll} \min\limits_{x \in \mathbb{R}^2} & (x_1+2)^2 + (x_2+1)^2 \\ \text{s.t.} & x_1^2 + x_2^2 \leq 9 \\ & x_1 \geq 0 \text{ or } x_2 \geq 0. \end{array}$$

  Note that the feasible region is not convex.

- The local minimum $(0, -1)$ is not a global minimum. The unique global minimum is $(-2, 0)$.



Extreme points and optimal solutions

- Now we know if we minimize a convex function over a convex feasible region, a local minimum is a global minimum.
- What may happen if we **minimize a concave function**?
- One "goes down" on a concave function if she moves "towards its boundary".
- We thus have the following proposition:

Proposition 2

For any concave function that has a global minimum over a convex feasible region, there exists a global minimum that is an extreme point.

Proof. Beyond the scope of this course. □

Special case: LP

- Now we know when we minimize $f(\cdot)$ over a convex feasible region F:
  - If $f(\cdot)$ is **convex**, search for a **local minimum**.
  - If $f(\cdot)$ is **concave**, search among the **extreme points** of F.
- For any LP, we have both!

Proposition 3

The feasible region of an LP is convex.

Proof. First, note that the feasible region of an LP is the intersection of several half spaces (each one is determined by an inequality constraint) and hyperplanes (each one is determined by an equality constraint). It is trivial to show that half spaces and hyperplanes are always convex. It then remains to show that the intersection of convex sets are convex, which is left as an exercise. □

Special case: LP

Proposition 4

A linear function $f : \mathbb{R}^n \to \mathbb{R}$ is both convex and concave.

Proof. To show that a function f is convex and concave, we need to show that $f(\lambda x^1 + (1-\lambda)x^2) = \lambda f(x^1) + (1-\lambda)f(x^2)$, which is exactly the separability of linear functions: Let $f(x) = c^T x + b$ be a linear function, $c \in \mathbb{R}^n, b \in \mathbb{R}$, then

$$f\Big(\lambda x^1 + (1-\lambda)x^2\Big) = c^T\Big(\lambda x^1 + (1-\lambda)x^2\Big) + b$$
$$= \lambda(c^T x^1 + b) + (1-\lambda)(c^T x^2 + b) = \lambda f(x^1) + (1-\lambda)f(x^2).$$

Therefore, a linear function is both convex and concave. □

- To solve an LP, use a **greedy search** focusing on **extreme points**.
- This is exactly the simplex method.

Convex Programming

- Consider a general NLP

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

- If the feasible region $F = \{x \in \mathbb{R}^n | g_i(x) \leq b_i \forall i = 1, ..., m\}$ is convex and f is convex over F, a local minimum is a global minimum.
- In this case, the NLP is called a **convex program** (CP).

Definition 4 (Convex programs)

An NLP is a CP if its feasible region is convex and its objective function is convex over the feasible region.

- Efficient algorithms exist for solving CPs.
- The subject of formulating and solving CPs is **Convex Programming**.

A sufficient condition for CP

- When is an NLP a CP?

Proposition 5

For an NLP

$$\min_{x \in \mathbb{R}^n} \left\{ f(x) \middle| g_i(x) \leq b_i \forall i = 1, ..., m \right\},$$

if f and g_is are all convex functions, the NLP is a CP.

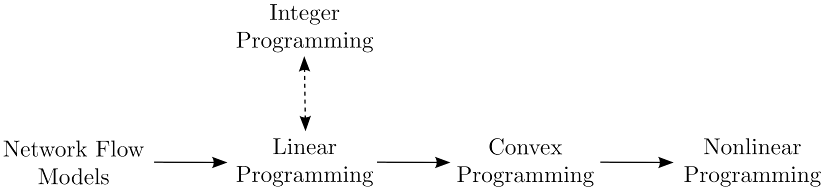
Proof. We only need to prove that the feasible region is convex, which is implied if $F_i = \{x \in \mathbb{R}^n | g_i(x) \leq b_i\}$ is convex for all i. For two points x_1, $x_2 \in F_i$ and an arbitrary $\lambda \in [0, 1]$, we have

$$\begin{aligned} g(\lambda x_1 + (1-\lambda)x_2) &\leq \lambda g(x_1) + (1-\lambda)g(x_2) \\ &\leq \lambda b_i + (1-\lambda)b_i = b_i, \end{aligned}$$

which implies that F_i is convex. Repeating this argument for all i completes the proof. □

Convex programming

- Now we have a larger relationship map:



- In this course, we will only discuss how to analytically solve NLPs.
  - Analytical solutions are the foundations for managerial insights.
  - We will not discuss algorithms for solving NLPs.
- All you need to know are:
  - People **can** efficiently solve CPs.
  - People **cannot** efficiently solve general NLPs.

Road map

- Motivating examples.
- Convex analysis.
- **Solving single-variate NLPs**.

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 6

For a twice differentiable function $f : \mathbb{R} \to \mathbb{R}$ over an interval (a, b):

- *f is convex over (a, b) if and only if $f''(x) \geq 0$ for all $x \in (a, b)$.*
- *$\bar{x}$ is a local minimum over (a, b) only if $f'(\bar{x}) = 0$.*
- *If f is convex over (a, b), x^* is a global minimum over (a, b) 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 the convexity of f. □

- Note that the two boundary points may need special considerations.

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 minimum.
  - For convex functions, FOC is also **sufficient** for a global minimum.
- To solve an NLP, convexity is the key.

Example 1: a retailer's pricing problem

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

$$\max_{p \geq 0} \; \pi(p) = (p - c)(a - bp).$$

  - 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 **unconstrained optimal solution** p^* satisfies

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

  - As $p^* = \frac{a+bc}{2b} > 0$ is **feasible**, it is optimal.
- $p^* = \frac{a+bc}{2b}$ is an **analytical solution**.

Example 1: economic interpretations

- For the retailer's pricing problem

$$\pi^* = \max_{p \geq 0} \ \pi(p) = (p - c)(a - bp),$$

  the optimal retail price is $p^* = \frac{a+bc}{2b}$. $\pi^* = \pi(p^*) = \frac{(a-bc)^2}{4b}$.
- Does p^* make sense?
  - p^* goes up when a goes up.
  - p^* goes down when b goes up.
  - p^* goes up when c goes up.
- Does π^* make sense?
  - π^* goes up when a goes up.
  - π^* goes down when c goes up.
  - What happens when b goes up?
- Any condition on a, b, and c for the solution to be reasonable?

Example 2: folding paper

- Now condition Example 2:

$$\max_{d \in [0, \frac{a}{2}]} V(d) = (a - 2d)^2 d.$$

  - 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?

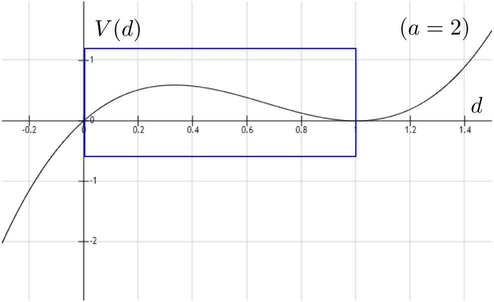
$$V'(d) = 12d^2 - 8ad + a^2 \quad \text{and} \quad V''(d) = 24d - 8a.$$

  In the feasible region $[0, \frac{a}{2}]$, V is also not concave.

- What should we do?

Example 2: graphical illustration

- Let's depict $V(d) = 4d^3 - 4ad^2 + a^2d$:



  - The reflection point (at which $V''(d) = 24d - 8a = 0$) is $\frac{a}{3}$.
  - When $a = 2$, this is $\frac{2}{3}$. $V(d)$ is not concave over $[\frac{2}{3}, \infty)$.

Example 2: solving the problem

- 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 "over $(0, \frac{d}{2})$".
  - We may verify that $\frac{a}{6}$ and $\frac{a}{2}$ are local maximum and local minimum:

$$V''\left(\frac{a}{6}\right) = 24\left(\frac{a}{6}\right) - 8a = -4a < 0 \quad \text{and} \quad V''\left(\frac{a}{2}\right) = 4a > 0.$$

- 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.
- Do $d^* = \frac{a}{6}$ and $V(d^*) = \frac{2a^3}{27}$ make sense?