Lecture 24 18 April 2013

Recall from last time:

- Necessary Condition of optimality for unconstrained optimization.
  Suppose f is a continuous, differentiable function in n variables, and $\bar{x}$ is an optimal solution to:

$$\min_{x \in \mathbb{R}^n} f(x). \quad \Big\} \text{(NLP)}$$

  Then, $\bar{x}$ must necessarily satisfy:

$$\nabla f(\bar{x}) = 0.$$

- Remark: However, if x satisfy $\nabla f(x) = 0$, x might not be an optimal solution for NLP. x is a local minimum/maximum, or an inflection pt.

- Defn (convex function)
  A ~~continuous differentiable~~ function f (in n variables) is convex if:
  for any 2 points a, b
  and any λ where $0 \le \lambda \le 1$,
  the following inequality is satisfied:

$$f(\lambda a + (1-\lambda) b) \le \lambda f(a) + (1-\lambda) f(b).$$

- Thm: Suppose f is a continuous, differentiable, and convex function. Then,
  $\bar{x}$ is an opt solution to
  if and only if $\nabla f(\bar{x}) = 0$.

- Remark: if f is convex, then any x that satisfies $\nabla f(x) = 0$ is an optimal solution to (NLP).

- Remark: There are convex functions that are not differenticable. However, if f is a differentiable function, then f is convex if and only if:

$$\nabla^2 f(x) \text{ has nonnegative eigenvalues }, \forall x.$$

$$(\text{or: } f''(x) \geq 0 \quad \text{for all } x, \quad \text{if } n=1).$$

Unconstrained Optimization

$$\left.\begin{aligned} \text{Min} \quad & f(x) \\ \text{S.t.} \quad & g_i(x) \geq 0 \qquad i=1,\dots,p \\ & h_j(x) = 0 \qquad j=1,\dots,q. \end{aligned}\right\} (NLP).$$

Assume that: f, g_i, h_j are continuous and differentiable functions in $x_1, \dots, x_n$.

Ex:

$$\begin{aligned} \text{Min} \quad & x_1^2 + x_2^2 \\ \text{S.T.} \quad & x_1^2 + x_2 \geq 4 \\ & x_1 + 4x_2 \geq 13. \end{aligned}$$

$$\therefore \quad \begin{aligned} \text{Min} \quad & f(x) \\ \text{S.t.} \quad & g_1(x) \geq 0 \\ & g_2(x) \geq 0 \end{aligned}$$

where
$$f(x) = x_1^2 + x_2^2$$
$$\left.\begin{aligned} g_1(x) &= x_1^2 + x_2 - 4 \\ g_2(x) &= x_1 + 4x_2 - 13. \end{aligned}\right\} g(x) = \begin{pmatrix} g_1(x) \\ g_2(x) \end{pmatrix} = \begin{pmatrix} x_1^2 + x_2 - 4 \\ x_1 + 4x_2 - 13 \end{pmatrix}.$$

- Let us try to use what we know: how to deal with <u>unconstrained problems</u>.

We will change our constrained problem into an unconstrained problem as follows:

① Fix $y = (y_1, y_2)$ where $y_1 \geq 0$, $y_2 \geq 0$.

② $\displaystyle\min_{x \in \mathbb{R}^n} L_y(x)$ $\}$ (*)

where $L_y(x) := f(x) - y_1 g_1(x) - y_2 g_2(x)$.

$\therefore L_y(x) = (x_1^2 + x_2^2) - y_1(x_1^2 + x_2 - 4) - y_2(x_1 + 4x_2 - 13)$.

<u>Observations</u>:

- We now have an unconstrained problem:
$$\min_{x \in \mathbb{R}^n} L_y(x), \quad \}$$
(for which we have a necessary condition to check if a solution is optimal or not)

→ Show pictures to illustrate that opt soln ~~dep~~ depends on y.

- Since (*) is unconstrained, its optimal solution could be infeasible for the original problem, (NLP).
  <u>Ex</u>: $y_1 = y_2 = 0$.
  Then $L_{(0,0)}(x) = x_1^2 + x_2^2$.
  $\therefore (x_1, x_2) = (0,0)$ is optimal for (*) but not feasible for (NLP).
- Suppose, however, that we choose $y_1 > 0$, $y_2 > 0$.
  If $x = (x_1, x_2)$ violates the first constraint, then
  $$g_1(x) = x_1^2 + x_2 - 4 < 0.$$
  $$\therefore \quad -y_1 g_1(x) > 0.$$
  $\therefore$ We can think of $-y_i g_i(x)$ as a <u>penalty</u> term.

→ We would like to minimize $L_y(x)$, but if the original constraints are violated, then the penalty terms make $L_y(x)$ larger.

→ We can think of y_i as the "<u>price</u>" ~~per unit~~ that we have to pay "per unit" that the ith constraint is violated.

In general: Given the constrained problem:

$$\left.\begin{array}{lll} \text{Min } f(x) & & \\ \text{s.t. } g_i(x) \geq 0 & i=1,\dots,p. \\ \quad\ \ h_j(x) = 0 & j=1,\dots,q. \end{array}\right\} (NLP),$$

The function:

$$\left(L_{y,z}(x) := f(x) - y_1 g_1(x) - \dots - y_p g_p(x) - z_1 h_1(x) - \dots - z_q h_q(x) \right)$$

$$L_{y,z}(x) := f(x) - \sum_{i=1}^{p} y_i g_i(x) - \sum_{j=1}^{q} z_j h_j(x)$$

is called the Lagrangian.
The y_i's and z_j's are called shadow prices or Lagrange multipliers.

Remarks: • So far, it is not clear how solving for $L_{y,z}(x)$'s optimal solution gives us an optimal solution for the original problem, (NLP), because we don't know what values of y, z are appropriate to choose.

EX ① $y_1=0$, $y_2=0$, $(x_1, x_2) = (0,0)$ is opt for $L_y(x)$, but not feasible for (NLP).

② $y_1=0.5$, $y_2=0.5$.
Claim: $(x_1, x_2) = (0.5, 1.25)$ is optimal for $L_{(0.5,0.5)}(x)$:
Check that $\nabla L_{0.5,0.5}(0.5, 1.25) = 0$:

$\nabla L_{0.5,0.5}(0.5, 1.25) = ($

- However, if we are given a solution $\bar{x}$, feasible for (NLP), the following theorem provides a way to check if $\bar{x}$ is an opt solution for (NLP).

<u>Thm</u> (Necessary conditions of optimality for constrained optimization problems)

Suppose we are solving

$$\left.\begin{array}{lll} \min & f(x) & \\ \text{s.t.} & g_i(x) \geq 0 & i=1,\dots,p \\ & h_j(x) = 0 & j=1,\dots,q, \end{array}\right\} \text{(NLP)}$$

where f, g_i, h_j are continuous and differentiable.
If $\bar{x}$ is an opt solution for (NLP),
then, there exists <u>Lagrange multipliers</u>

$$\begin{array}{ll} y_i \geq 0 & i=1,\dots,p \\ z_j & j=1,\dots,q, \end{array}$$

satisfying:

1. whenever $g_i(\bar{x}) < 0$ then $y_i = 0$.
(but if $g_i(\bar{x}) = 0$ then $y_i \geq 0$).

2. $\nabla L_{y,z}(\bar{x}) = 0$
(where $\nabla L_{y,z}(\bar{x}) = \nabla f(\bar{x}) - \sum_{i=1}^{p} y_i \nabla g_i(\bar{x}) - \sum_{j=1}^{q} z_j \nabla h_j(\bar{x})$)

<u>Remarks</u>:
- Condition 1 looks like "complementary slackness" for LP
- Condition 2 looks like necessary opt condition for unconstrained problems.

$\boxed{\text{Ask as Q}}$ →

- The theorem gives us a way to check if $\bar{x}$ is optimal or not:
  - First, check that $\bar{x}$ is <u>feasible</u> for (NLP)
  - Then, compute $\nabla L_{y,z}(\bar{x})$
  - Solve for y and z such that $\nabla L_{y,z}(\bar{x}) = 0$.
  - For the resulting values of y and z, make sure that $y_i \geq 0$, and that if $g_i(\bar{x}) < 0$, then $y_i = 0$
  - If no such y, z exists, then $\bar{x}$ is not optimal

Given $\bar{x}$, how ~~would~~ might we use the thm to check if $\bar{x}$ is optimal?
("there exists" is very 'abstract')

EX: From our example,

$$L_y(x) = (x_1^2 + x_2^2) - y_1(x_1^2 + x_2 - 4) - y_2(x_1 + 4x_2 - 13)$$

$$\nabla L_y(x) = \begin{pmatrix} 2x_1 - 2y_1 x_1 - y_2 \\ 2x_2 - y_1 - 4y_2 \end{pmatrix}$$

iclicker → Check if $\bar{x} = (10, 20.25)$ is optimal for (NLP).

① Is $\bar{x}$ feasible?

$g_1(\bar{x}) = x_1^2 + x_2 - 4 = 100 + 20.25 - 4 = 116.25 > 0$ ✓

$g_2(\bar{x}) = x_1 + 4x_2 - 13 = 10 + 81 - 13 = 78 > 0$ ✓.

$\therefore \bar{x}$ is feasible.

② Find y s.t. $\nabla L_y(\bar{x}) = \vec{0}$

$$\nabla L_y(\bar{x}) = \begin{pmatrix} 20 - 20y_1 - y_2 \\ 40.5 - y_1 - 4y_2 \end{pmatrix}.$$

$\therefore$ $0 = 20 - 20y_1 - y_2$. $\times 4$ | $0 = 80 - 80y_1 - 4y_2$.

$0 = 40.5 - y_1 - 4y_2$. $\times 1$ | $0 = 40.5 - y_1 - 4y_2$ −

$0 = 39.5 - 79y_1$

$\therefore y_1 = \frac{39.5}{79} = 0.5$.

$\therefore y_2 = 20 - 10 = 10$.

③ However, since $g_1(\bar{x}) > 0$, y_1 should have been 0

$g_2(\bar{x}) > 0$, y_2 should have been 0.

$\therefore \bar{x}$ is not optimal for (NLP).

iclicker → Check if $\bar{x} = (1, 3)$ is opt for (NLP):

① If $\bar{x}$ feasible?

$g_1(\bar{x}) = 1 + 3 - 4 = 0$. ✓

$g_2(\bar{x}) = 1 + 4(3) - 13 = 0$ ✓

② $\nabla L_y(\bar{x}) = \begin{pmatrix} 2 - 2y_1 - y_2 \\ 6 - y_1 - 4y_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ | $2 - 2y_1 - y_2 = 0$, $6 - y_1 - 4y_2 = 0$

$\therefore$ $2 - 2y_1 - y_2 = 0$, $12 - 2y_1 - 8y_2 = 0$ − $\} \rightarrow -10 + 7y_2 = 0$. $y_2 = 10/7$.

$y_1 = 6 - 4(\frac{10}{7}) = 2/7$.

So, $\bar{x} = (1, 3)$ is optimal!

Intuition behind this theorem: ← skip to next page?

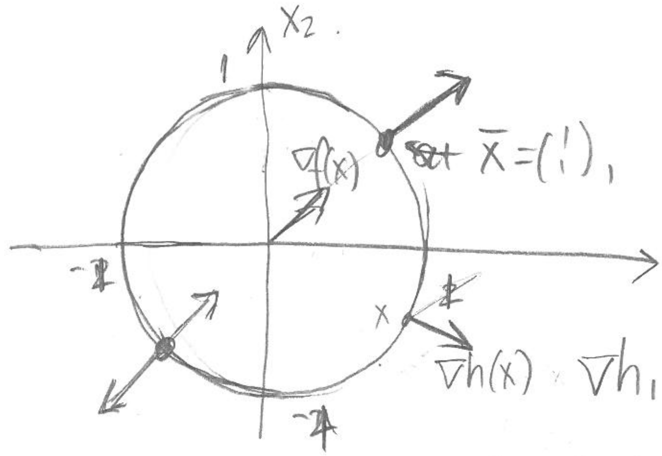
- First consider a problem with just the equality constraint:

$$\begin{aligned} &\text{Min } f(x) \\ &\text{s.t. } h_j(x) = 0 \qquad j = 1, \dots, q. \end{aligned}$$

Ex. $\text{Min } x_1 + x_2$ ← $f(x) = x_1 + x_2$

s.t. $x_1^2 + x_2^2 = 1.$ ← $h_1(x) = x_1^2 + x_2^2 - 1 = 0$

Circle of radius 1.



at $\bar{x} = \begin{pmatrix}1\\1\end{pmatrix}$, $\nabla f(x) = \begin{pmatrix}1\\1\end{pmatrix} = \frac{1}{2}\nabla h_1(x)$.

$\nabla h_1(x) = \begin{pmatrix}2x_1\\2x_2\end{pmatrix}$.

$\nabla f(x) = \begin{pmatrix}1\\1\end{pmatrix}$.

Ex

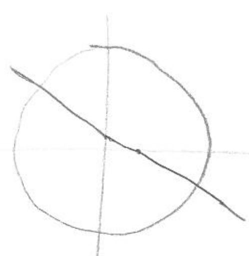
$$\begin{aligned} &\text{Min } x_1 + x_2 \\ &\text{s.t. } x_1^2 + x_2^2 \leq 1 \\ &\quad +2x_1 + x_2 \geq 0.5 \end{aligned}$$

$g_1(x) = -x_1^2 - x_2 + 1$

$g_2(x) = 2x_1 + x_2 - 0.5.$

$\nabla g_1(x) = \begin{pmatrix}-2x_1\\-1\end{pmatrix}$

$\nabla g_2(x) = \begin{pmatrix}2\\1\end{pmatrix}.$



$\text{Min } c^T x$

S.t. $Ax \geq b$

$x \geq 0$

y

$\cancel{y^T A x = y^T b}$

Satisfies: $A^T y \leq c$.