

# LECTURE 7: CONSTRAINED OPTIMIZATION – SENSITIVITY ANALYSIS AND DUALITY

---

1. Basic concepts
2. Sensitivity analysis
3. Duality theory

# Sensitivity analysis

- Consider NLP with equality constraints:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{s.t.} & h(x) = 0 \\ & x \in E^n. \end{array}$$

Let  $x^*$  be a local minimizer with a Lagrange vector  $\lambda^* \in E^m$ .

If the right-hand-side is perturbed by a vector  $c \in E^m$ ,

- (i) How will the solution  $x^*(c)$  be perturbed?
- (ii) How will the optimal value  $f(x^*(c))$  be perturbed?
- (iii) How is the change associated with  $\lambda^*$ ?

# Basic Idea of Implicit Functions

Suppose we have a set of  $m$  equations in  $n$  variables

$$u_i(x) = 0, \quad i = 1, 2, \dots, m.$$

If we fix  $n - m$  variables, can the equations be solved for the remaining  $m$  variables?

In other words, for  $m$  selected variables  $x_1, x_2, \dots, x_m$ , can they be expressed implicitly as a function of the remaining variables  $x_{m+1}, \dots, x_n$  in the form of

$$x_i = \phi_i(x_{m+1}, \dots, x_n), \quad i = 1, \dots, m?$$

# Example

For LP problem with  $Ax = b$ , we let  $x = \begin{bmatrix} x_B \\ \text{---} \\ x_N \end{bmatrix}$

and  $A = [B \mid N]$ ,

$$\text{then } [B \mid N] \begin{bmatrix} x_B \\ \text{---} \\ x_N \end{bmatrix} = b$$

$$\Rightarrow Bx_B + Nx_N = b$$

$$\Rightarrow Bx_B = b - Nx_N$$

$$\Rightarrow x_B = B^{-1}(b - Nx_N)$$

Therefore, the  $m$  basic variables are functions of the  $n - m$  nonbasic variables!

# Implicit function theorem

- Theorem:

Let  $\bar{x} \in E^n$  such that

- (i)  $u_i(\bar{x}) = 0$ ,  $i = 1, 2, \dots, m$ .
- (ii)  $u_i \in C^p$  (for some  $p \geq 1$ ) in a neighborhood of  $\bar{x}$ .
- (iii) The Jacobian matrix

$$J = \begin{bmatrix} \frac{\partial u_1(\bar{x})}{\partial x_1} & \cdots & \frac{\partial u_1(\bar{x})}{\partial x_m} \\ \vdots & & \vdots \\ \frac{\partial u_m(\bar{x})}{\partial x_1} & \cdots & \frac{\partial u_m(\bar{x})}{\partial x_m} \end{bmatrix}_{m \times m}$$

is nonsingular.

Then  $\exists$  a neighborhood of

$\bar{x}_{-m} = (\bar{x}_{m+1}, \dots, \bar{x}_n) \in E^{n-m}$  s.t. for

$x_{-m} = (x_{m+1}, \dots, x_n) \in N(\bar{x}_{-m})$ ,  $\exists$

functions  $\phi_i(x_{-m})$ ,  $i = 1, 2, \dots, m$ , with

- (i)  $\phi_i \in C^p(N(\bar{x}_{-m}))$ .
- (ii)  $\bar{x}_i = \phi_i(\bar{x}_{-m})$   $i = 1, 2, \dots, m$ .
- (iii)  $u_i(\phi_1(x_{-m}), \dots, \phi_m(x_{-m}), x_{-m}) = 0$ ,  $i = 1, 2, \dots, m$ .

# Sensitivity theorem

Let  $f$  and  $h_i$  ( $i = 1, 2, \dots, m$ )  $\in C^2$ . Consider the family of problems associated with  $c \in E^m$  :

$$(P) \quad \begin{array}{ll} \text{minimize} & f(x) \\ \text{s. t.} & h(x) = c \\ & x \in E^n. \end{array}$$

Suppose that

- (i)  $x^*$  is a local solution to the problem with  $c = 0$ .
- (ii)  $x^*$  is a regular point.
- (iii)  $x^*$  and the associated Lagrange vector  $\lambda^*$  satisfy the 2nd-order sufficient conditions for a strict, local minimum solution.

Then for every  $c$  in a sufficiently small neighborhood of 0, there is an  $x^*(c)$ , depending continuously on  $c$ , such that  $x^*(0) = x^*$  and  $x^*(c)$  is a local minimum of (P). Moreover,

$$\nabla_c f(x^*(c)) \Big|_{c=0} = -(\lambda^*)^T.$$

# Proof

Consider the following system of  $n + m$  equations in  $n + 2m$  variables  $(x, \lambda, c)$ :

$$(*) \quad \begin{aligned} \nabla f(x) + \lambda^T \nabla h(x) &= 0 \\ h(x) - c &= 0. \end{aligned}$$

Previous results tell us there exists a solution  $(x^*, \lambda^*)$  at  $c = 0$ . We would like to represent the  $n + m$  variables  $(x, \lambda)$  in terms of the remaining  $m$  variables  $c$ .

The Jacobian matrix of the above system at  $(x^*, \lambda^*)$  is

$$J = \begin{bmatrix} L(x^*) & \nabla h(x^*)^T \\ \nabla h(x^*) & 0 \end{bmatrix}.$$

Because  $x^*$  is a regular point and  $L(x^*)$  is positive definite on  $T(x^*)$ ,  $J$  is nonsingular.

By the Implicit Function Theorem,  $\exists$  a solution  $(x^*(c), \lambda^*(c))$  to the system  $(*)$  which is  $C^2$  in a small neighborhood of  $c = 0$  with  $x^*(0) = x^*$ ,  $\lambda^*(0) = \lambda^*$ .

Using assumption (iii), we know  $x^*(c)$  is a local minimizer of (P).

# Proof - continue

Using the chain rule, we have

$$\nabla_c f(x^*(c)) \Big|_{c=0} = \nabla_x f(x^*) \nabla_c x^*(0)$$

and

$$\nabla_c h(x^*(c)) \Big|_{c=0} = \nabla_x h(x^*) \nabla_c x^*(0).$$

But  $h(x^*(c)) - c = 0$  implies that

$$\nabla_c h(x^*(c)) \Big|_{c=0} = I.$$

Notice that  $\nabla_x f(x^*) + (\lambda^*)^T \nabla h_x(x^*) = 0$ .

Hence

$$\begin{aligned} \nabla_c f(x^*(c)) \Big|_{c=0} &= \nabla_x f(x^*) \nabla_c x^*(0) \\ &= -(\lambda^*)^T \nabla h_x(x^*) \nabla_c x^*(0) \\ &= -(\lambda^*)^T \nabla_c h(x^*(c)) \Big|_{c=0} \\ &= -(\lambda^*)^T. \end{aligned}$$



# Corollary

- NLP with equality and inequality constraints

Let  $f, g_j, h_i \in C^2$ . Consider the family of problems:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{(NLP)} \quad \text{s.t.} & h(x) = c \\ & g(x) \leq d. \end{array}$$

Suppose that for  $c = 0$ ,  $d = 0$ , there exists a local solution  $x^*$  that is a regular point and that, together with the associated Lagrange vectors  $\lambda^* \in E^m$ ,  $\mu^* \in E_+^p$ , satisfying the 2nd-order sufficient conditions.

Assume further that no active inequality constraint is degenerate.

Then  $\forall (c, d) \in E^{m+p}$  in a sufficiently small neighborhood of  $(0, 0)$ , there exists a solution  $x^*(c, d)$ , depending continuously on  $(c, d)$ , s.t.  $x^*(0, 0) = x^*$  and  $x^*(c, d)$  is a local minimum of problem (NLP). Moreover,

$$\nabla_c f(x^*(c, d)) \Big|_{(0,0)} = -(\lambda^*)^T$$

$$\nabla_d f(x^*(c, d)) \Big|_{(0,0)} = -(\mu^*)^T.$$

# Observations

1. Consider a manufacturer who is facing a problem with inequality constraints:

$$\begin{aligned} \text{Minimize} \quad & f(x) \\ \text{s. t.} \quad & g(x) \leq 0 \\ & x \in X. \end{aligned}$$

Take  $f(x) \longleftrightarrow$  production cost,  
 $g(x) \leq 0 \longleftrightarrow$  resource constraints,  
 $X \longleftrightarrow$  set of interests.

Then  $\nabla_d f(x(d))|_{d=0} = -\mu^T$  may mean that a unit incremental in resource  $j$  will cause a marginal cost of  $-\mu_j$  in production.

Since  $-\mu_j \leq 0$ , this is a cost reduction. Hence the manufacturer is willing to pay a price of  $\mu_j$  ( $\geq 0$ ) to acquire this extra resource.

2. The resource provider may have a profit function based on the price he/she sets for the manufacturer. An associated problem becomes

$$\begin{aligned} \text{Max} \quad & \phi(\mu) \\ \text{s. t.} \quad & \mu \in E_+^p, \end{aligned}$$

where  $\phi(\mu)$  is the profit function of price.

3. For a given price  $\mu \geq 0$ , the manufacturer will try to

$$\min_{x \in X} \{f(x) + \mu^T g(x)\}.$$

4. The provider may want to associate the profit function with the relation

$$\phi(\mu) \triangleq \min_{x \in X} \{f(x) + \mu^T g(x)\}, \quad \mu \geq 0,$$

and try to maximize it.

5. This will be done through the concept of “local duality”.

# Duality theory

- Motivation of Local Duality

1. Start with a problem with equality constraints:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{s.t.} & h(x) = 0 \end{array}$$

where  $x \in E^n$ ,  $h(x) \in E^m$  and  $f, h \in C^2$ .

2. Let  $x^*$  be a regular, local minimum point. Then  $\exists \lambda^* \in E^m$  such that

$$\nabla f(x^*) + (\lambda^*)^T \nabla h(x^*) = 0$$

and

$$L(x^*) = F(x^*) + (\lambda^*)^T H(x^*)$$

which is positive semidefinite on the tangent subspace

$$T(x^*) = \{y \in E^n \mid \nabla h(x^*)y = 0\}.$$

# Local duality

3. For “local convexity,” assume that  $L(x^*)$  is positive definite on a sufficiently small  $N_1(x^*)$ .

Then  $\ell(x) \triangleq f(x) + (\lambda^*)^T h(x)$  is locally convex at  $x^*$  with  $x^*$  being a strict, local minimizer over  $N_1(x^*)$ .

4. By the Implicit Function Theorem, we know there exists a  $C^2$  function  $x \triangleq x^*(\lambda)$  with  $x^* = x^*(\lambda^*)$  that provides a local minimizer of  $\ell(x)$  near  $x^*$  when  $\lambda$  is near  $\lambda^*$ , say  $\lambda \in N_2(\lambda^*)$ .

5. Given  $\lambda \in N_2(\lambda^*)$ , we define a dual function:

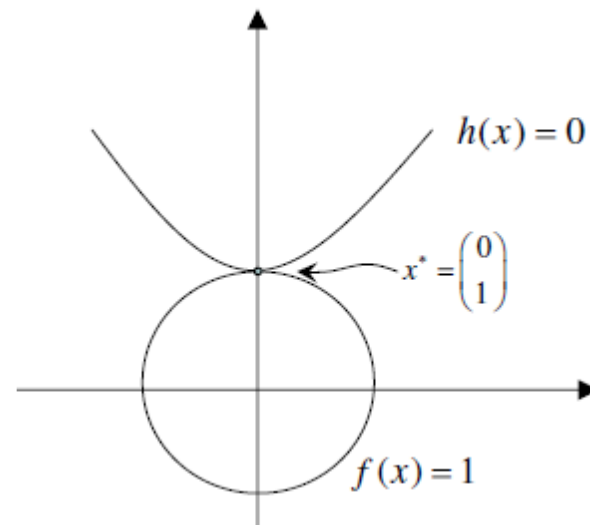
$$\phi(\lambda) \triangleq \underset{\substack{\text{minimum} \\ (\text{infimum}) \\ x \in N_1(x^*)}}{[f(x) + \lambda^T h(x)]},$$

Then  $\phi(\lambda)$  is a concave function over  $N_2(\lambda^*)$ .

6. We show that the original constrained local minimization problem is equivalent to the unconstrained local maximization problem of “maximize  $\phi(\lambda)$ ”.
7. Once the equivalence relation is established, all unconstrained optimization theory and solution methods become applicable.

# Example 1

$$\begin{aligned} \text{Minimize} \quad & f(x) = x_1^2 + x_2^2 \\ \text{s. t.} \quad & h(x) = x_1^2 - x_2 + 1 = 0 \end{aligned}$$



$$\text{For } x^* = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

$$\nabla f(x^*) + 2\nabla h(x^*) = 0,$$

$$\lambda^* = 2, \quad f(x^*) = 1.$$

$$\begin{aligned} \phi(\lambda) &= \inf_x [f(x) + \lambda h(x)] \\ &= \inf_x \underbrace{[x_1^2 + x_2^2 + \lambda(x_1^2 - x_2 + 1)]}_{\text{convex locally as } \lambda > -1}. \end{aligned}$$

Take partial derivatives,

$$2x_1 + 2\lambda x_1 = 0 \Rightarrow x_1 = 0$$

$$2x_2 - \lambda = 0 \Rightarrow x_2 = \frac{\lambda}{2}.$$

Hence  $\phi(\lambda) = 0^2 + (\frac{\lambda}{2})^2 + \lambda(0^2 - \frac{\lambda}{2} + 1)$

$$= \frac{\lambda^2}{4} - \frac{\lambda^2}{2} + \lambda$$

$$= \underbrace{-\frac{\lambda^2}{4}}_{\text{concave function to be maximized}} + \lambda.$$

concave function to be maximized

Take derivative,  $-\frac{\lambda}{2} + 1 = 0 \Rightarrow \lambda^* = 2$  and  $\phi(\lambda^*) = 1 = f(x^*)$ .

# Example 2

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

s. t.

$$h(x) = x_1^2 + x_2^2 - 1 = 0$$

For  $x^* = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ ,

$$\nabla f(x^*) + \frac{1}{2} \nabla h(x^*) = 0,$$

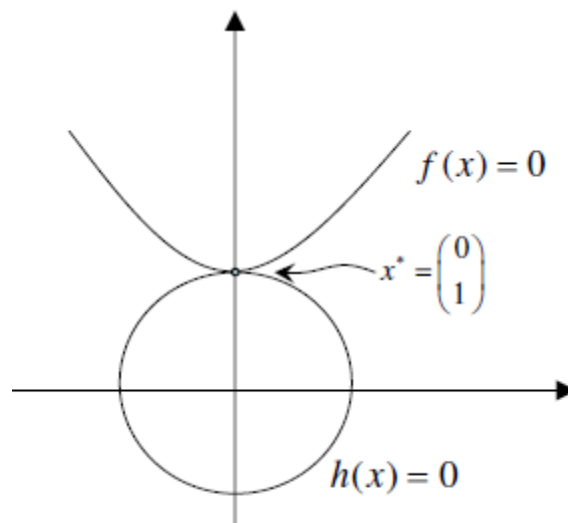
$$\lambda^* = \frac{1}{2}, \quad f(x^*) = 0.$$

$$\begin{aligned} \phi(\lambda) &= \inf_x [x_1^2 - x_2 + 1 + \lambda(x_1^2 + x_2^2 - 1)] \\ &= \inf_x \underbrace{[(1 + \lambda)x_1^2 + \lambda x_2^2 - x_2 + 1 - \lambda]}_{\text{convex as } \lambda > 0}. \end{aligned}$$

Take partial derivatives,

$$2(1 + \lambda)x_1 = 0 \quad \Rightarrow \quad x_1 = 0$$

$$2\lambda x_2 - 1 = 0 \quad \Rightarrow \quad x_2 = \frac{1}{2\lambda}.$$



$$\begin{aligned} \phi(\lambda) &= [0 + \lambda(\frac{1}{2\lambda})^2 - \frac{1}{2\lambda} + 1 - \lambda] \\ &= -\frac{1}{4\lambda} + 1 - \lambda \quad (\text{over a neighborhood of } \lambda^* = \frac{1}{2}). \end{aligned}$$

Maximize  $\phi(\lambda)$  over the neighborhood

$$\Rightarrow \frac{1}{4\lambda^2} - 1 = 0 \quad \Rightarrow \quad \lambda^* = \frac{1}{2}.$$

$$\phi(\lambda^*) = 0 = f(x^*).$$

# Observation 1

1. In our “local convexity” setting,

$$\phi(\lambda) = f(x^*(\lambda)) + \lambda^T h(x^*(\lambda)).$$

Hence

$$\begin{aligned}\nabla \phi(\lambda) &= \nabla f(x^*(\lambda)) \nabla x^*(\lambda) \\ &\quad + \lambda^T \nabla h(x^*(\lambda)) \nabla x^*(\lambda) + h(x^*(\lambda))^T \\ &= \underbrace{[\nabla f(x^*(\lambda)) + \lambda^T \nabla h(x^*(\lambda))]}_{=0} \nabla x^*(\lambda) + h(x^*(\lambda))^T \\ &= h(x^*(\lambda))^T.\end{aligned}$$

This means that given  $\bar{\lambda}$ , we find  $x^*(\bar{\lambda})$  by minimizing  $\{f(x) + \bar{\lambda}h(x)\}$  around  $x^*$ , then  $h(x^*(\bar{\lambda}))^T$  provides the gradient information of  $\phi$  at  $\bar{\lambda}$ , even we don't know the explicit form of  $\phi(\lambda)$ .

## Observation 2

$$\begin{aligned} 2. \quad & \nabla\phi(\lambda) = h(x^*(\lambda))^T \\ & \Rightarrow \Phi(\lambda) = \nabla h(x^*(\lambda))\nabla x^*(\lambda). \end{aligned}$$

$$\begin{aligned} \text{Also,} \quad & \nabla f(x^*(\lambda)) + \lambda^T \nabla h(x^*(\lambda)) = 0 \\ & \Rightarrow L(x^*(\lambda), \lambda)\nabla x^*(\lambda) + \nabla h(x^*(\lambda))^T = 0. \end{aligned}$$

Hence

$$\Phi(\lambda) = -\nabla h(x^*(\lambda))L^{-1}(x^*(\lambda), \lambda)\nabla h(x^*(\lambda))^T.$$

Since  $L^{-1}(x^*(\lambda), \lambda)$  is positive definite and  $\nabla h(x^*(\lambda))$  is of full rank near  $x^*$ , we know  $\Phi(\lambda)$  is negative definite. This gives us “local concavity” on the dual side.



# Observation 3

3. Observation 1 says

$$\nabla\phi(\lambda^*) = h(x^*)^T = 0.$$

Observation 2 says

$\Phi(\lambda)$  is negative definite in  $N_2(\lambda^*)$ .

Hence

$\lambda^*$  is a maximizer of  $\phi(\lambda)$  over  $N_2(\lambda^*)$ .

# Local duality theorem

- Theorem: Let  $x^*$  be a regular, local minimum point of the problem

$$(P) \quad \begin{array}{ll} \text{minimizer} & f(x) \\ \text{s.t.} & h(x) = 0 \end{array}$$

with corresponding value  $v^*$  and Lagrange multipliers  $\lambda^*$ . If the Hessian of the Lagrangian

$$L(x^*) = F(x^*) + (\lambda^*)^T H(x^*)$$

is positive definite, then the dual problem

$$(D) \quad \text{maximize } \{ \phi(\lambda) \triangleq \underset{(\text{infimum})}{\text{minimize}} [f(x) + \lambda h(x)] \}$$

has a local solution at  $\lambda^*$  with corresponding value  $v^*$  and has  $x^*$  as the point corresponding to  $\lambda^*$  in the definition of  $\phi$ .

# Observations

1. The dual problem (D) is a max-min problem.

2. (Inequality Constraints)

The result can be easily extended to the problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{(P)} \quad \text{s.t.} & h(x) = 0 \\ & g(x) \leq 0. \end{array}$$

with  $f, h, g \in C^2$ .

In this case, the local convexity assumption requires that

$$L(x^*) \triangleq F(x^*) + (\lambda^*)^T H(x^*) + (\mu^*)^T G(x^*)$$

is positive definite on  $N_1(x^*)$ .

For  $\lambda$  and  $\mu \geq 0$  near  $\lambda^*$  and  $\mu^*$ , we define

$$\phi(\lambda, \mu) = \min_{x \in N_1(x^*)} [f(x) + \lambda^T h(x) + \mu^T g(x)].$$

Then  $(\lambda^*, \mu^*)$  is a local maximizer of  $\phi(\lambda, \mu)$  with  $\lambda \in E^n$  and  $\mu \in E_+^p$ .

# Observations

## 3. (Convex Duality)

When  $f, g$  are convex and  $h$  is affine, then problem (P) is a convex program. Hence a local optimizer becomes a global optimizer. Moreover, the Lagrangian  $f(x) + \lambda^T h(x) + \mu^T g(x)$  is convex for any  $\lambda \in E^m$  and  $\mu \in E_+^p$ , and  $\phi(\lambda, \mu)$  is concave.

## 4. (Partial Duality)

It is not necessary to include the Lagrangian multipliers of all the constraints in defining the dual function. When the local convexity assumption holds, local duality can be defined with respect to any subset of functional constraints. For problem (P), we may consider

$$\phi(\lambda) = \min_{g(x) \leq 0} \{f(x) + \lambda^T h(x)\}$$

as a partial dual.