Chapter 2

Basic Concepts in Minimization Problems

Theory of Non-Linear Programming (NLP)
非線形計画法の理論

Notations

Characteristics of the optimal solutions to mathematical programs

- Formulation and solution of mathematical optimization (minimization) programs.
- Conditions that the optimum solution should have, and its uniqueness.

- Variable vector $\mathbf{x} = (x_1, x_2, .., x_i, .., x_I)^T$
- Objective function $z(\mathbf{x})$
- The j-th constraints $g_j(\mathbf{x}) \geq b_j \quad for\_\ j = 1,2,\ldots J$

Optimization Program (standard form)

Min. (minimize) $z(\mathbf{x})$

sub. to (subject to) $g_j(\mathbf{x}) \geq b_j \qquad for\_\ j = 1,2,\ldots J$

----

example

Min. $z(x_1, x_2) = x_1^2 + 2x_2^2 + 2x_1x_2 - 2x_1 - 4x_2$

sub. to

$-x_1 - x_2 \geq -4 \qquad \cdots \qquad g_1(\mathbf{x}) \geq b_1$

$-x_1 + 2x_2 \geq 2 \qquad \cdots \qquad g_2(\mathbf{x}) \geq b_2$

Standard form

$$g_j(\mathbf{x}) \le b_j \qquad \Rightarrow \qquad -g_j(\mathbf{x}) \ge -b_j$$

$$g_j(\mathbf{x}) = b_j \qquad \Rightarrow \qquad g_j(\mathbf{x}) \ge b_j\,, \quad g_j(\mathbf{x}) \le b_j$$

Feasible solution satisfies all constraints

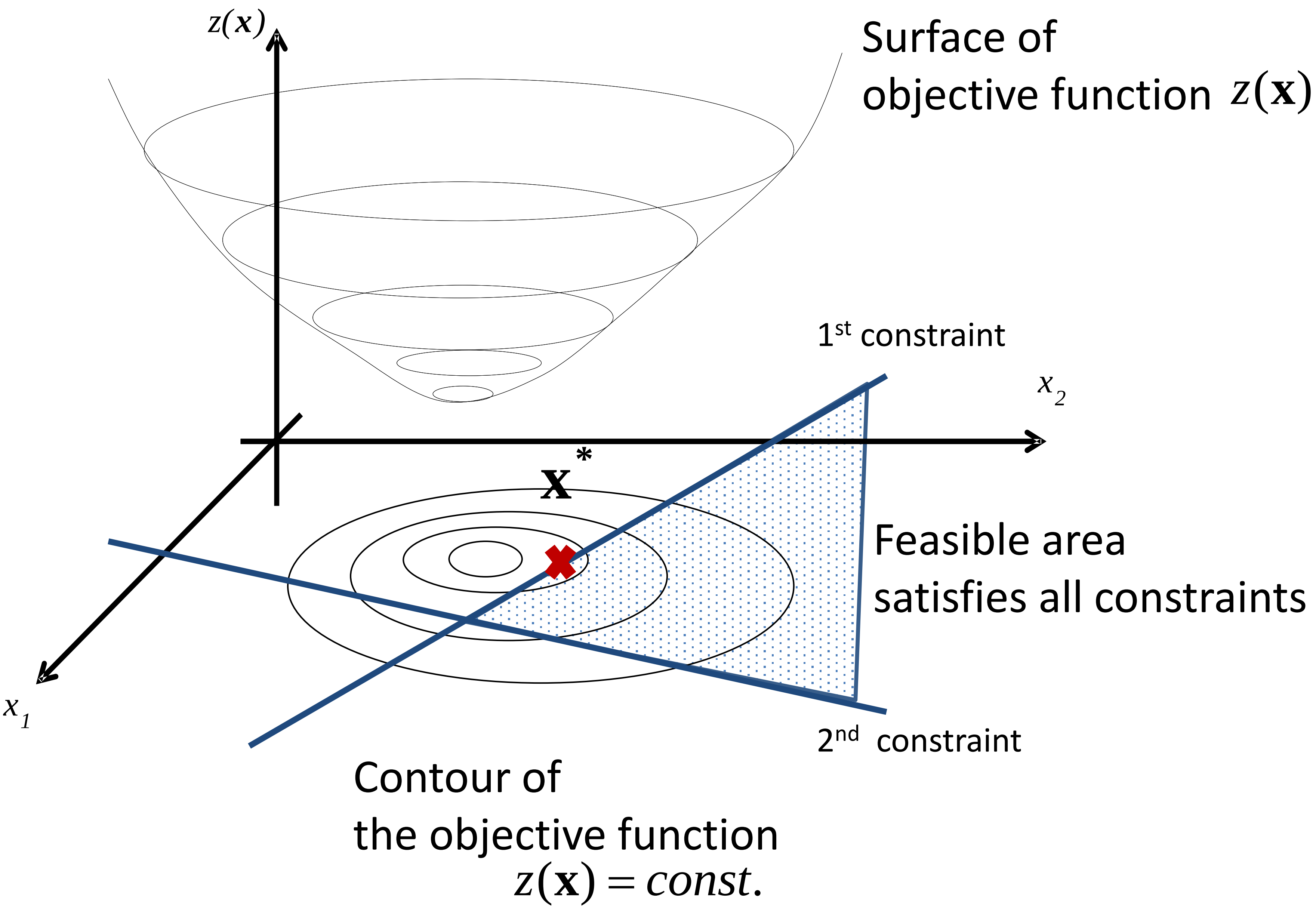
$$g_j(\mathbf{x}) \ge b_j \quad for\_\, j = 1,2,\ldots J$$

Optimum (minimum) solution $\mathbf{x}^*$

$$z(\mathbf{x}^*) \le z(\mathbf{x})$$ for any feasible solutions,

and

$$g_j(\mathbf{x}^*) \ge b_j \quad for\_\, j = 1,2,\ldots J$$


$z(\mathbf{x})$
Surface of
objective function $z(\mathbf{x})$
1st constraint
x_2
$\mathbf{x}^*$
Feasible area
satisfies all constraints
x_1
2nd constraint
Contour of
the objective function
$z(\mathbf{x}) = const.$

Assumptions on Non-linear Optimization

- Existence: at least one feasible solution
- Finite minimum
- Continuity: objective function and constraints are continuously differentiable

cf. discrete (combinatorial) optimization

- Single variable minimization without constraints
- Multivariable minimization without constraints
- Multivariable minimization with constraints
- Some special cases (but frequently used in UE analysis)

2.1_Unconstrained Minimization Program with One Variable

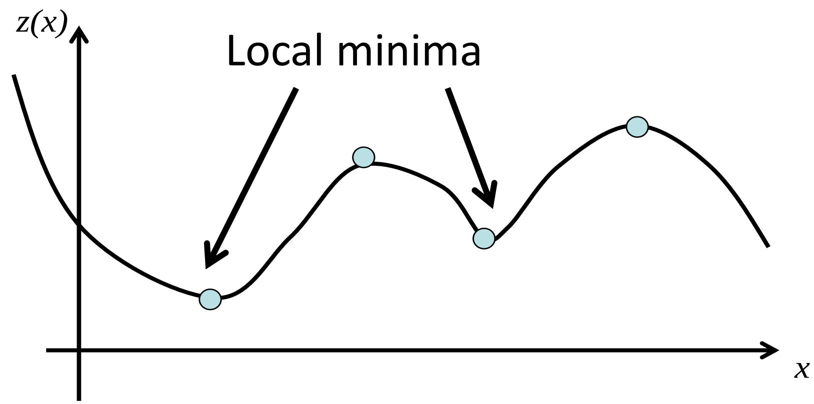
$$Min. \quad z(x)$$

- Necessary condition (first order condition) ; if z(x) has a minimum at x=x*, the derivative of z(x=x*) equals zero.

$$dz(x^*)/dx = 0$$

- Stationary points;

$$dz(x)/dx = 0$$

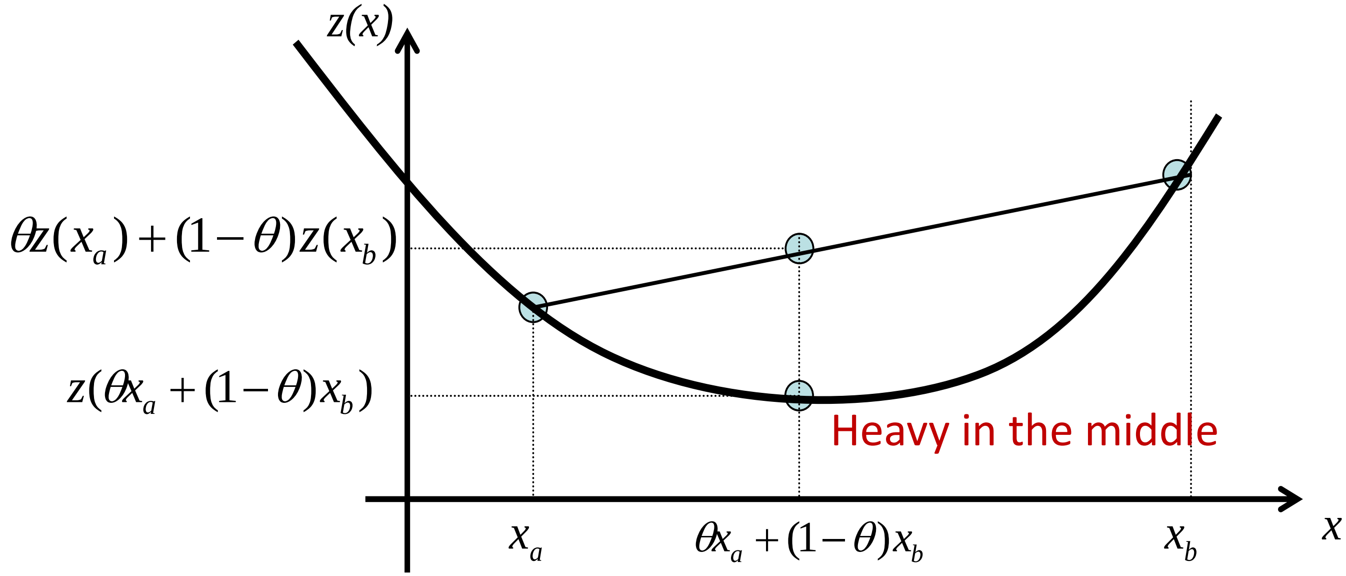


- If a function is *ditonic*, its stationary point is a global minimum.

Strictly Convex

- A sufficient condition for a stationary point to be a local minimum is for the function to be strictly convex in the vicinity of the point.
- <u>Strictly convexity</u>; a line segment connecting any two points of the function lies entirely above the function.

$$z(\theta x_a + (1-\theta)x_b) < \theta z(x_a) + (1-\theta)z(x_b) \qquad 0 < \theta < 1$$



Uniqueness Condition

- If a function is twice differentiable, ‘strictly convex’ is equal to that the second derivative is positive.

$$d^2 z(x) / dx^2 > 0$$

- If the function is strictly convex, a local minimum point is the unique global minimum.

- For strictly convex function, the necessary condition becomes the necessary and sufficient condition.

Find the global minimum solution of the following problem.

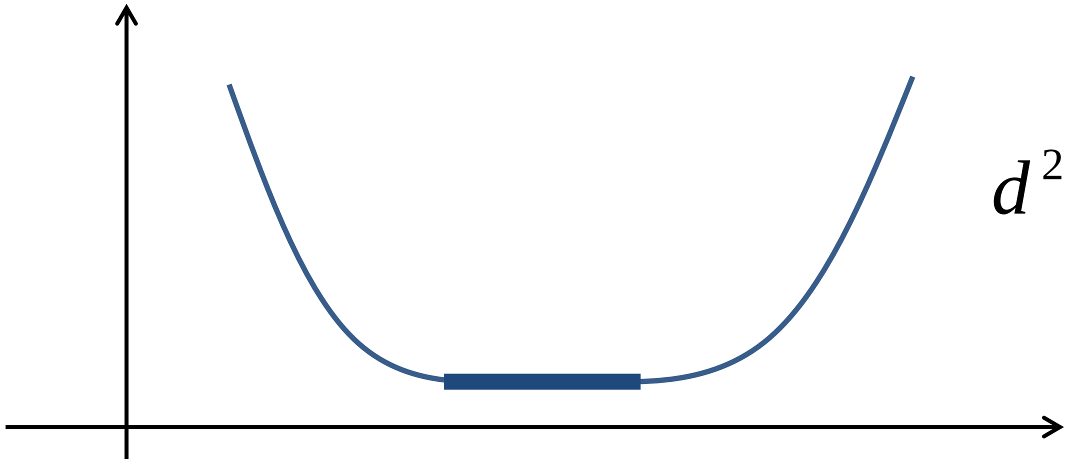
$$Min.\_\ z(x) = x^2 + 2x - 2$$

(Strictly) Convex

- (strictly) Convex; a line segment connecting any two points of the function never lies below the function.

$$z(\theta x_a + (1-\theta) x_b) \leq \theta z(x_a) + (1-\theta) z(x_b) \qquad 0 < \theta < 1$$

- There may be a flat bottom, and the minimum solution is not necessarily unique.



$$d^2 z(x) / dx^2 \geq 0$$

2.2.1_Unconstrained Minimization Program with Multiple Variables

$$Min. \quad z(\mathbf{x})$$

$$\mathbf{x} = (x_1, x_2, ..., x_i, ..., x_I)^T$$

- Gradient vector is the vector of partial derivatives.

$$\nabla z(\mathbf{x}) = \left(\partial z(\mathbf{x}) / \partial x_1, \partial z(\mathbf{x}) / \partial x_2, ..., \partial z(\mathbf{x}) / \partial x_I\right)^T$$

Nabla

- The first-order condition (necessary condition) for a minimum at x=x* is that the gradient of z(x) vanish at x*.

$$\nabla z(\mathbf{x}^*) = \mathbf{0} \quad \Leftrightarrow \quad \partial z(\mathbf{x}^*) / \partial x_i = 0 \quad for\_i = 1,2,..., I$$

Simultaneous equations

Convexity of Multivariate Function

- Strictly convexity; a line segment connecting any two points of the function lies entirely above the function.

$$z(\theta \mathbf{x}_a + (1-\theta)\mathbf{x}_b) < \theta z(\mathbf{x}_a) + (1-\theta) z(\mathbf{x}_b)$$

- Hessian Matrix; matrix of the second derivatives

$$H(\mathbf{x}) = \nabla^2 z(\mathbf{x}) = \begin{pmatrix} \partial^2 z(\mathbf{x})/\partial x_1^2 & \partial^2 z(\mathbf{x})/\partial x_1 \partial x_2 & .. & \partial^2 z(\mathbf{x})/\partial x_1 \partial x_I \\ \partial^2 z(\mathbf{x})/\partial x_2 \partial x_1 & \partial^2 z(\mathbf{x})/\partial x_2^2 & .. & : \\ : & .. & .. & : \\ \partial^2 z(\mathbf{x})/\partial x_I \partial x_1 & .. & .. & \partial^2 z(\mathbf{x})/\partial x_I^2 \end{pmatrix}$$

- When Hessian is ‘positive definite’, the function is strictly convex.
- When z(x) is strictly convex, the first-order condition becomes the necessary and sufficient condition.

Positive Definite → Linear Algebra

- A matrix H is positive definite if one of three conditions are satisfied.

1. All eigenvalues are positive,
2. A quadratic form is positive. $\mathbf{h}H\mathbf{h}^T > 0 \quad for\_\mathbf{h} \neq \mathbf{0}$
3. All minor determinants are positive,

$$\begin{vmatrix} a_{11} \end{vmatrix} > 0\,, \quad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0\,, \quad 2\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} > 0\,, \quad \ldots.., \quad \begin{vmatrix} a_{11} & .. & : & a_{1n} \\ .. & .. & : & .. \\ .. & .. & a_{ij} & .. \\ a_{n1} & .. & : & a_{nn} \end{vmatrix} > 0$$

Any diagonal matrix with positive element is positive definite.
This can be applied to the objective function of UE problem.

Exercise

$$z(x_1, x_2) = x_1^2 + 2x_2^2 + 2x_1x_2 - 2x_1 - 4x_2$$

1. Show gradient vector of z(x).
2. Show Hessian matrix of z(x).
3. Obtain all eigenvalues of H.
4. Examine the quadratic form for H.
5. Calculate all minor determinants of H.
6. Discuss the convexity of z(x).
7. Find the optimum solution of Min. z(x).

Application: Multiple Regression Analysis

- N sets of samples are observed for three variables X,Y and Z. A linear relation can be assumed. Estimate the parameters (α, β, γ) by multiple regression analysis.

$$z_i = \alpha + \beta x_i + \gamma y_i + \varepsilon_i$$

$$i = 1,2,\ldots,N$$

$$\varepsilon_i\_random\_error$$

標本番号	X	Y	Z
1	x_1	y_1	z_1
2	x_2	y_2	z_2
:	:	:	:
:	:	:	:
N	x_N	y_N	z_N

Minimize $S = \sum_{i=1}^{N} \varepsilon_i^2 = \sum_{i=1}^{N} \{z_i - (\alpha + \beta x_i + \gamma y_i)\}^2$

S is a function of parameters S(α,β,γ)

- When S is strictly convex*, the optimum solution satisfies

$$\nabla S(\alpha, \beta, \gamma) = \mathbf{0}$$

- Linear simultaneous equations

$$\begin{pmatrix} N & \sum_{i=1}^{N} x_i & \sum_{i=1}^{N} y_i \\ \sum_{i=1}^{N} x_i & \sum_{i=1}^{N} x_i^2 & \sum_{i=1}^{N} x_i y_i \\ \sum_{i=1}^{N} y_i & \sum_{i=1}^{N} x_i y_i & \sum_{i=1}^{N} y_i^2 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{N} z_i \\ \sum_{i=1}^{N} x_i z_i \\ \sum_{i=1}^{N} y_i z_i \end{pmatrix}$$

*) Almost always 'yes', but extremely strong correlation among variables may disturb this.

A data matrix and a vector make a simpler equation.

$$D = \begin{pmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ \vdots & \vdots & \vdots \\ 1 & x_N & y_N \end{pmatrix} \qquad \mathbf{z} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_N \end{pmatrix} \qquad \mathbf{b} = \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}$$

Linear equation becomes;

$$D^T D\mathbf{b} = D^T \mathbf{z}$$

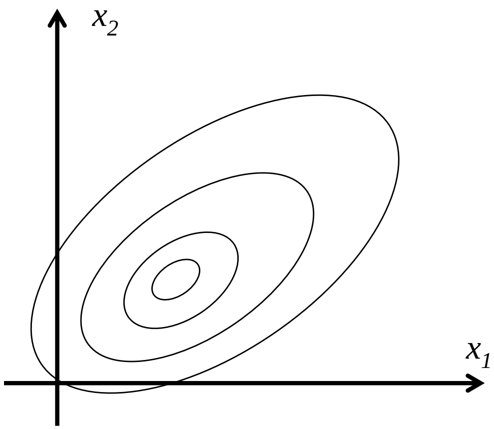
Parameter vector is estimated as;

$$\mathbf{b} = [D^T D]^{-1} D^T \mathbf{z}$$

2.2.2_Constrained Minimization Programs

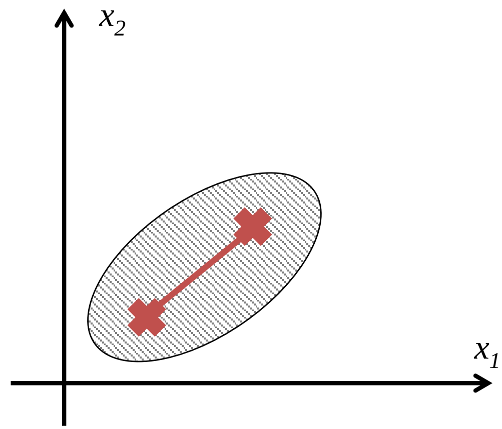
- A minimization program with multiple constraints.

- A feasible area is convex if $g_j(x)$ is concave.
- A common area surrounded by multiple convex area is also convex.



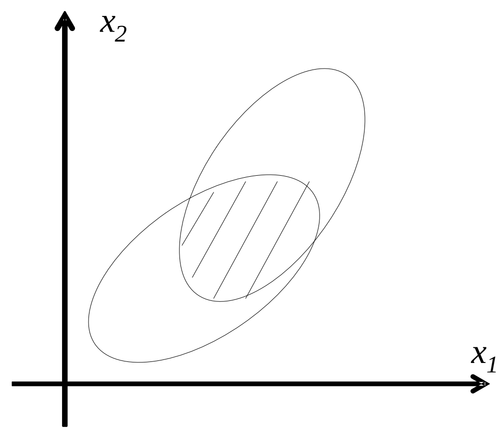
Contour of a constraint

$$g_j(\mathbf{x}) = const.$$



Convex area for

$$g_j(\mathbf{x}) \geq b_j$$



Common area is convex

A line segment connecting any two points in the area never goes out of the area. If all constraints are linear, feasible area is convex.

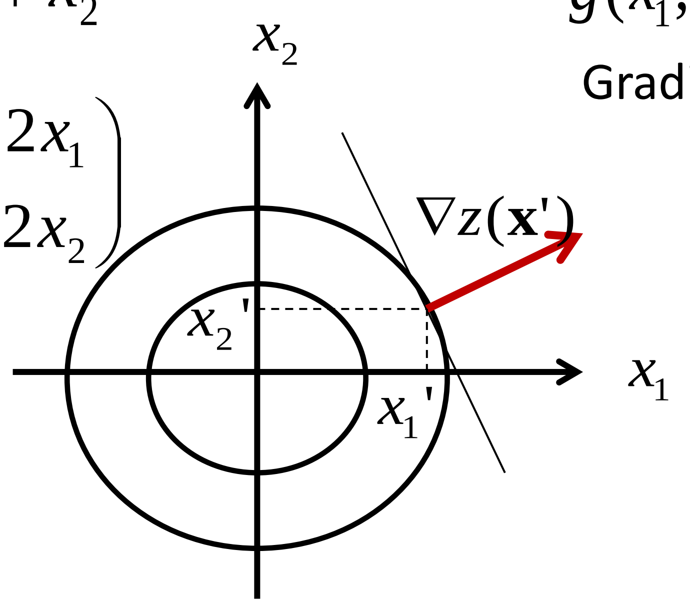
Effective Constraints

- Some of the constraints bind the optimum solution. They are called as effective. $g_j(\mathbf{x}^*) = b_j$
- Gradient vectors of effective constraints $\nabla g_j(\mathbf{x}^*)$ and the gradient vector of objective function $\nabla z(\mathbf{x}^*)$ are balanced at optimum.

$z(x_1, x_2) = x_1^2 + x_2^2$

Gradient vector

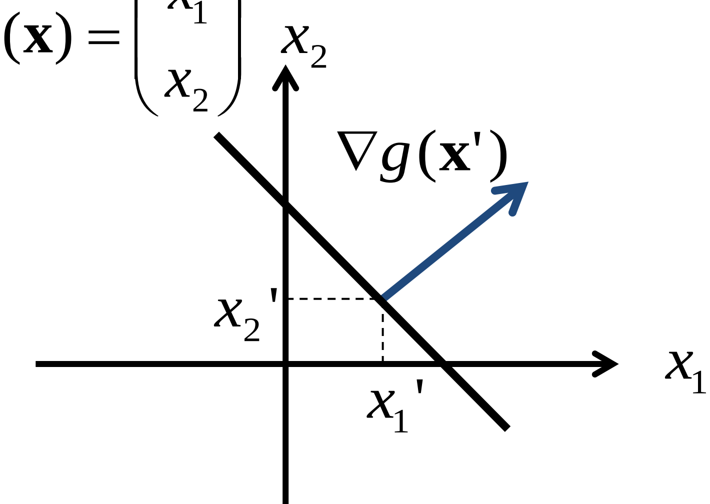
$$\nabla z(\mathbf{x}) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix}$$



$g(x_1, x_2) = x_1 + x_2 - 2 \geq 0$

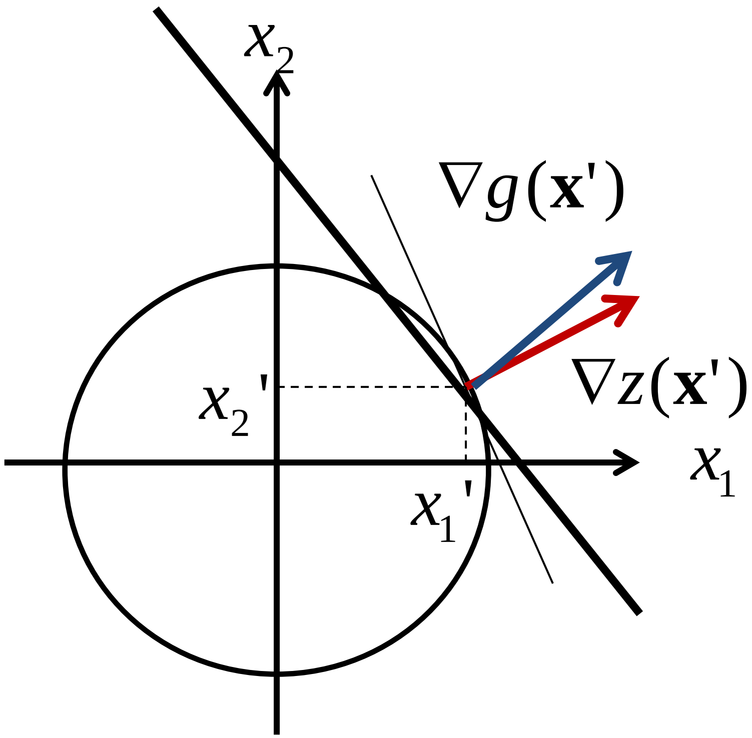
Gradient vector

$$\nabla g(\mathbf{x}) = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$



At optimum, two gradient vectors should have the same direction.

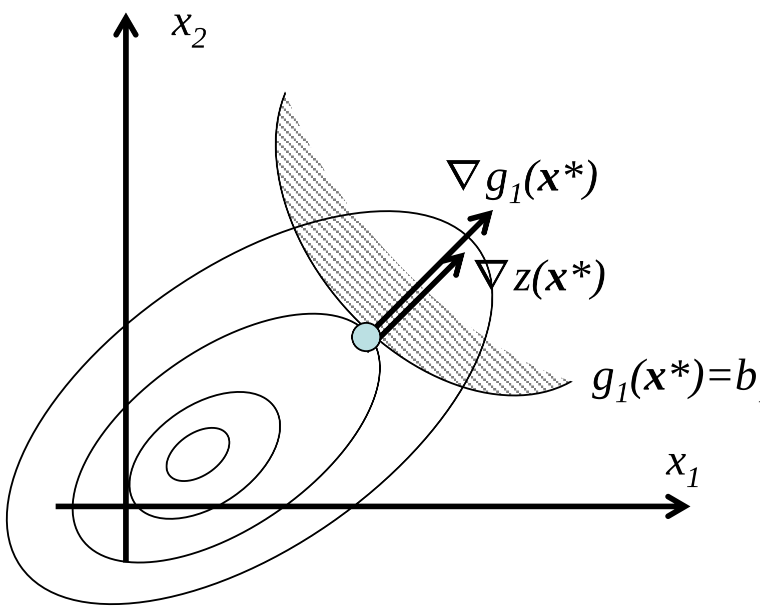
$$g(x_1, x_2) = x_1 + x_2 - 2 = 0$$



On the constraint, but not optimum.

If the constraint is effective, there exists a positive variable u.

$$\nabla z(\mathbf{x}^*) = u_1 \nabla g_1(\mathbf{x}^*)$$



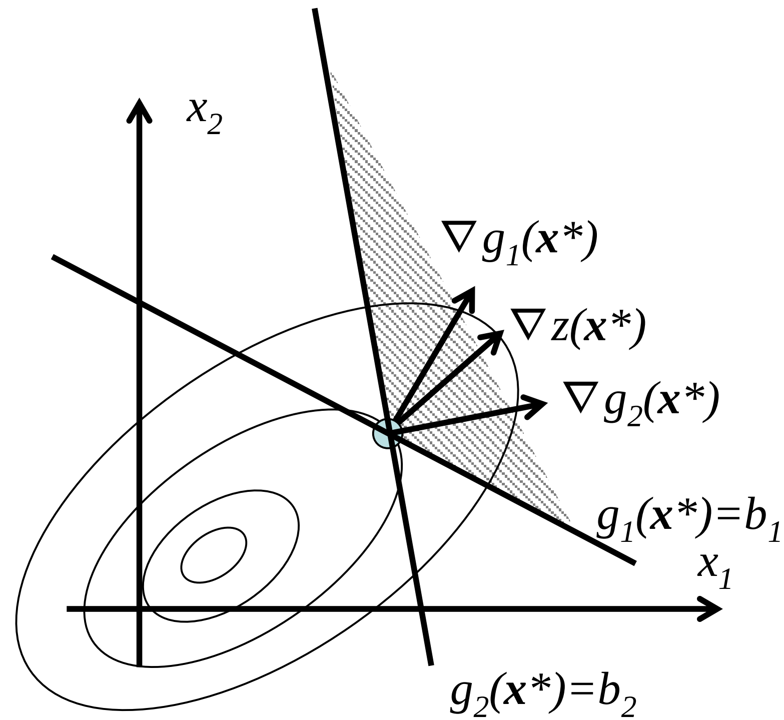
When two constraints are effective,

- Gradient vector of the objective function stays within a <u>cone</u> of two gradient vectors of the effective constraints.

- Two positive variables u1 and u2 exist that satisfy

$$\nabla z(\mathbf{x}^*) = u_1 \nabla g_1(\mathbf{x}^*) + u_2 \nabla g_2(\mathbf{x}^*)$$

$$\begin{pmatrix} \partial z(\mathbf{x}^*)/\partial x_1 \\ \partial z(\mathbf{x}^*)/\partial x_2 \end{pmatrix} = u_1 \begin{pmatrix} \partial g_1(\mathbf{x}^*)/\partial x_1 \\ \partial g_1(\mathbf{x}^*)/\partial x_2 \end{pmatrix} + u_2 \begin{pmatrix} \partial g_2(\mathbf{x}^*)/\partial x_1 \\ \partial g_2(\mathbf{x}^*)/\partial x_2 \end{pmatrix}$$



Kuhn-Tucker Conditions

- The first-order condition is generally written as;

$$\partial z(\mathbf{x}^*)/\partial x_i = \sum_j u_j[\partial g_j(\mathbf{x}^*)/\partial x_i] \qquad \text{for all} \quad i = 1,2,\ldots,I$$

$$u_j \geq 0, \quad \underline{u_j[b_j - g_j(\mathbf{x}^*)] = 0}, \qquad g_j(\mathbf{x}^*) \geq b_j, \quad \text{for all} \quad j = 1,2,..,J$$

If the j-th constraint is effective, $u_j > 0$ and $b_j - g_j(x^*) = 0$
If the constraint is not effective, $u_j = 0$.

The auxiliary variable u_j is named as dual variable as well as *Lagrange multiplier*.

The underlined part is known as the complementary slackness condition.

Exercise

1. Show the Kuhn-Tucker conditions for the following minimization problem.

$$\min. \qquad z(x_1, x_2) = x_1^2 + 2x_2^2 + 2x_1x_2 - 2x_1 - 4x_2$$

$$sub.to$$

$$x_1 + x_2 \geq 2$$

2. Find optimum solution by using KT conditions.

3. Compare the solutions if the constraint is changed by;

$$x_1 + x_2 \geq -2$$

2.3.1_Nonnegativity Constraints

- The first-order conditions for one dimensional case.

$$\min. \quad z(x), \quad sub.to \quad x \geq 0$$

- If non-negativity constraint is not binding,

$$x^* > 0 \qquad dz(x^*)/dx = 0$$

- If the constraint is effective,

$$x^* = 0 \qquad dz(x^*)/dx \geq 0$$

- Both cases are written as a whole;

$$x^*\left[dz(x^*)/dx\right] = 0 \quad and \quad dz(x^*)/dx \geq 0$$

- Multidimensional case,

$$\min. \quad z(\mathbf{x}), \quad sub.to \quad x_i \geq 0 \quad for\ all \quad i$$

- Either positive area,

$$x_i^* > 0 \quad and \quad \partial z(\mathbf{x}^*)/\partial x_i = 0$$

- Or on the boundary of the feasible region,

$$x_i^* = 0 \quad and \quad \partial z(\mathbf{x}^*)/\partial x_i \geq 0$$

- Accordingly the first-order conditions can be;

$$x_i^* \left[\partial z(\mathbf{x}^*)/\partial x_i\right] = 0 \quad and \quad \partial z(\mathbf{x}^*)/\partial x_i \geq 0 \quad for\ all \quad i$$

2.3.2_Linear Equality Constraints

$$\min. \quad z(\mathbf{x})$$

constant

$$sub.to$$

$$\sum_i h_{ij} x_i = b_j \quad for \quad j$$

Note that

$$\partial g_j(\mathbf{x}) / \partial x_i = h_{ij}$$

Kuhn-Tucker conditions are;

$$\begin{cases} \partial z(\mathbf{x}^*) / \partial x_i = \sum_j u_j h_{ij} & for \quad i \\ \sum_i h_{ij} x_i^* = b_j & for \quad j \end{cases}$$

All constraints are binding and the complementary slackness conditions are automatically satisfied.

Lagrangian Multiplier Method

Objective function and constraints are combined with dual variable known as Lagrangian multipliers.

$$L(\mathbf{x},\mathbf{u}) = z(\mathbf{x}) + \sum_j u_j [b_j - \sum_i h_{ij} x_i]$$

The first-order conditions of this unconstrained program are

$$\begin{cases} \partial L(\mathbf{x}^*,\mathbf{u}^*) / \partial x_i = 0 & \textit{for} \quad i = 1,2,\ldots I \\ \partial L(\mathbf{x}^*,\mathbf{u}^*) / \partial u_j = 0 & \textit{for} \quad j = 1,2,\ldots J \end{cases}$$

This is equivalent to Kuhn-Tucker conditions.

$u_j{}^* > 0$ means $\sum_i h_{ij} x_i \geq b_j$

$u_j{}^* < 0$ means $\sum_i h_{ij} x_i \leq b_j$

2.3.4_Nonnegativity and Linear Equality Constraints

$$\min. \quad z(\mathbf{x})$$

$$sub.to$$

$$\sum_i h_{ij} x_i = b_j \qquad for\_j$$

$$x_i \geq 0 \qquad for\_i$$

Lagrangian function with non-negativity constraints is

$$\min. \quad L(\mathbf{x},\mathbf{u}) = z(\mathbf{x}) + \sum_j u_j [b_j - \sum_i h_{ij} x_i]$$

$$sub.to$$

$$x_i \geq 0 \qquad for \qquad i = 1,2,..,I$$

The first-order conditions for the non-negativity constraints case are

$$x_i^* > 0 \quad \textit{and} \quad \partial L(\mathbf{x}^*, \mathbf{u}^*) / \partial x_i = 0$$

or

$$x_i^* = 0 \quad \textit{and} \quad \partial L(\mathbf{x}^*, \mathbf{u}^*) / \partial x_i \geq 0$$

and

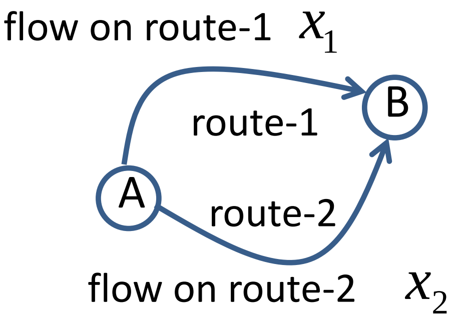
$$\partial L(\mathbf{x}^*, \mathbf{u}^*) / \partial u_j = 0 \quad \textit{and} \quad x_i^* \geq 0$$

original constraints

Example

Two cities A and B are connected by two routes.
The amount of emission from each route is a function of traffic volume of each route.

$$\begin{cases} E_1(x_1) = 10x_1 + x_1^2/2 \\ E_2(x_2) = x_2^2 \end{cases}$$



We would like to find the optimal route flows which can minimize total emissions.
When total traffic flow between two cities is given as 10, formulate the optimum assignment problem, and find the optimum solution. How about the total traffic is given as 4?

It is not necessary to consider the route choice of the drivers.