Lecture X
Introduction to Nonlinear Programming

A. Introduction

- Consider the **generalized nonlinear programming [NLP] problem**:

$$\min_{\mathbf{x}} \; f(\mathbf{x})$$

subject to: $\mathbf{x} \in \mathbf{S}$

$$\mathbf{S} = \{\mathbf{x} \mid \mathbf{g}(\mathbf{x}) \le \mathbf{0}\}$$

where $f(\cdot)$ and $\mathbf{g}(\cdot)$ are differentiable;

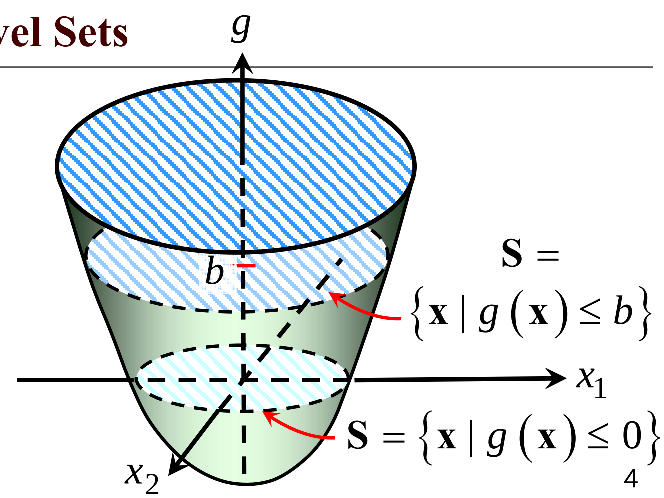
$\mathbf{x} \in E^n;\; f : E^n \to E^1;\; \mathbf{g} : E^n \to E^m$

Equality Constraints

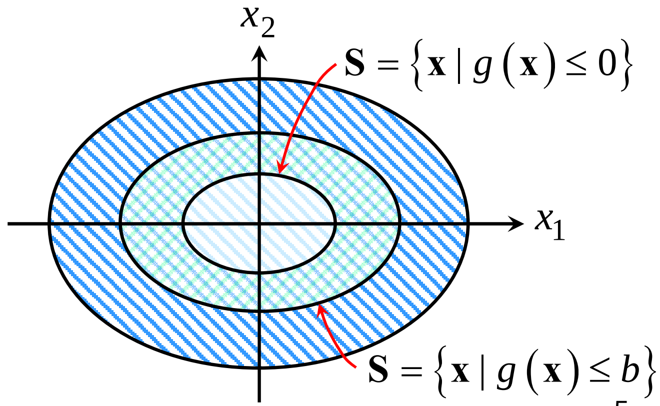
- Instead of starting with NLP theory for equality constraints, we will start with inequalities and derive the former:

$$\mathbf{h}(\mathbf{x}) = \mathbf{0} \Leftrightarrow \begin{cases} \mathbf{h}(\mathbf{x}) \le \mathbf{0} \\ -\mathbf{h}(\mathbf{x}) \le \mathbf{0} \end{cases}$$

Level Sets



Level Sets (cont.)



Active Constraints

- Define:

  $$\mathsf{A}(\mathbf{x}) = \{i \mid g_i(\mathbf{x}) = 0, i = 1, \ldots, m\}$$

- Constraint is **active** if:

  if $i \in \mathsf{A}(\mathbf{x})$ for **given** $\mathbf{x}$

- Constraint is **inactive** if:

  if $i \notin \mathsf{A}(\mathbf{x})$ for **given** $\mathbf{x}$

Tangent Plane

- Suppose some given $\mathbf{x}^*$ lies on boundary (i.e., at least 1 constraint is active at $\mathbf{x}^*$).

  Then, a tangent hyperplane (or **tangent plane**) @ $\mathbf{x}^*$ is the set of all $\mathbf{x}$ such that:

  $$\mathbf{a}^T \mathbf{x} = c$$

Tangent Plane (cont.)

where

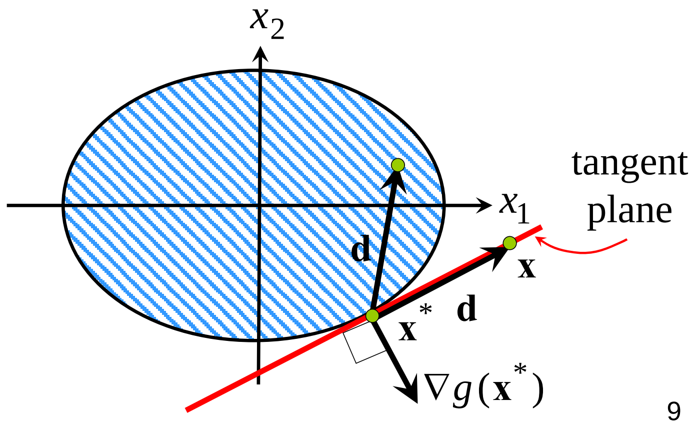
$$\mathbf{a} = \nabla g_i(\mathbf{x}^*) \text{ for each } i \in \mathsf{A}(\mathbf{x}^*)$$

$$c = \nabla g_i(\mathbf{x}^*)^T \mathbf{x}^*$$

i.e.,

$$\underbrace{\nabla g_i(\mathbf{x}^*)^T}_{\mathbf{a}^T} \mathbf{x} = \underbrace{\nabla g_i(\mathbf{x}^*)^T \mathbf{x}^*}_{c}$$

Tangent Plane (cont.)



Tangent Plane (cont.)

- Note that ∇g and any $\mathbf{d} = \mathbf{x} - \mathbf{x}^*$ (for any $\mathbf{x}$ contained in the tangent plant) are **orthogonal:**

  $$\nabla g(\mathbf{x}^*)^T \left[\mathbf{x} - \mathbf{x}^* \right] = 0 \text{ , or}$$

  $$\nabla g(\mathbf{x}^*)^T \mathbf{d} = 0$$

  $\forall \mathbf{d}$ emanating from $\mathbf{x}^*$ and lying in the tangent plane.

Feasible Directions

- Notice that those $\mathbf{d}$ such that:

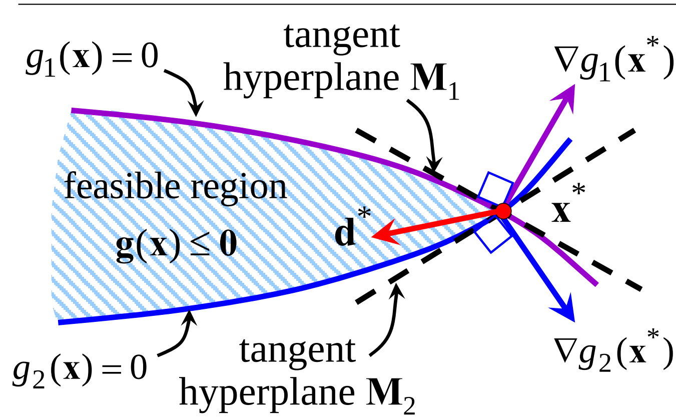
  $$\nabla g(\mathbf{x}^*)^T \mathbf{d} < 0$$

  are **feasible directions** (if g nonlinear).
- For 2 active constraints, tangent planes are:

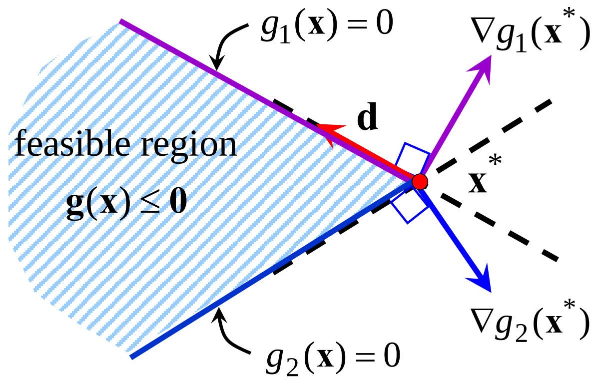
  $$\mathbf{M}_1 = \left\{ \mathbf{x} \mid \nabla g_1(\mathbf{x}^*)^T \left[\mathbf{x} - \mathbf{x}^* \right] = 0 \right\}$$

  $$\mathbf{M}_2 = \left\{ \mathbf{x} \mid \nabla g_2(\mathbf{x}^*)^T \left[\mathbf{x} - \mathbf{x}^* \right] = 0 \right\}$$

Feasible Directions (cont.)



Linear Case



Linear Case (cont.)

- In the linear case, the feasible directions satisfy:

$$\nabla g_1(\mathbf{x}^*)^T \mathbf{d} \le 0$$

$$\nabla g_2(\mathbf{x}^*)^T \mathbf{d} \le 0$$

B. Constraint Qualification

- The set of all feasible directions and all limits of feasible directions from some pt. $\mathbf{x}^*$ can be expressed as:

$$\mathbf{D}(\mathbf{x}^*) = \left\{ \mathbf{d} \mid \nabla g_i(\mathbf{x}^*)^T \mathbf{d} \le \mathbf{0} \ ; \ \forall i \in \mathrm{A}(\mathbf{x}^*) \right\}$$

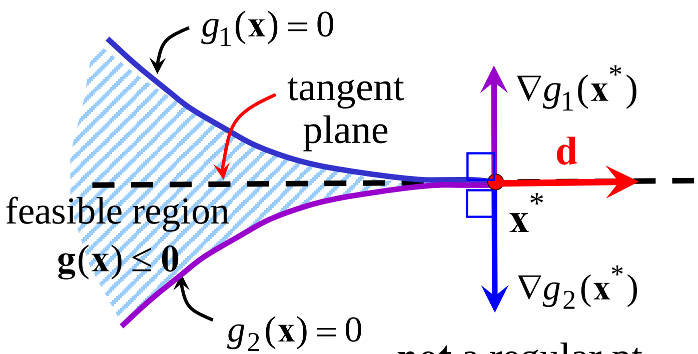
as long as the following **qualification** is made: the vectors $\nabla g_i(\mathbf{x}^*)$ are **linearly independent** $\forall i \in \mathrm{A}(\mathbf{x}^*)$ (active).

Constraint Qualification (cont.)

- Points where CQ holds are called **regular points**.
- In the following example, CQ not satisfied at $\mathbf{x}^*$ since $\nabla g_1(\mathbf{x}^*)$ and $\nabla g_2(\mathbf{x}^*)$ are linearly dependent and $\mathbf{d}$ is **not** a feasible direction, but $\mathbf{d}$ still satisfies:

$$\nabla \mathbf{g}(\mathbf{x}^*)^T \mathbf{d} \le 0$$

Constraint Qualification (cont.)



not a regular pt.

Constraint Qualification (cont.)

- CQ is automatically satisfied if:

1. All constraints **linear**
2. All constraints defined by **convex** functions with nonempty interiors

C. Farkas' Lemma

Given $\mathbf{c}\in E^n$, $\mathbf{A}=\left(\mathbf{a}^1,\ldots,\mathbf{a}^m\right)^T$ where

$\mathbf{a}^i\in E^n$ for $i=1,\ldots,m$,

then $\mathbf{c}^T\mathbf{x}\ge 0\ \ \forall\mathbf{x}\ \ni\ \mathbf{A}\mathbf{x}\le\mathbf{0}$

$\Leftrightarrow\ \exists\boldsymbol{\lambda}\ge\mathbf{0}\ \ni\ \mathbf{c}^T+\boldsymbol{\lambda}^T\mathbf{A}=\mathbf{0}$

or $\exists\lambda_i$ s.t. $\sum_{i=1}^{m}\lambda_i\mathbf{a}^i=-\mathbf{c}$

proof:

$(\Rightarrow)$ Set up the following linear programming problem:

$$\min\ \mathbf{c}^T\mathbf{x}$$

$$\text{s.t.}\quad \left.\begin{array}{l}\mathbf{A}\mathbf{x}\le\mathbf{0}\\ \mathbf{c}^T\mathbf{x}\ge 0\end{array}\right\}\quad \begin{array}{l}-\mathbf{A}\mathbf{x}\ge\mathbf{0}\\ \mathbf{c}^T\mathbf{x}\ge 0\end{array}$$

$$\mathbf{x}\ \text{free}$$

Proof (cont.):

$(\Rightarrow)$ An obvious optimal solution is $\mathbf{x}^* = \mathbf{0}$.
Therefore, since the primal has a finite optimum, then so does the dual:

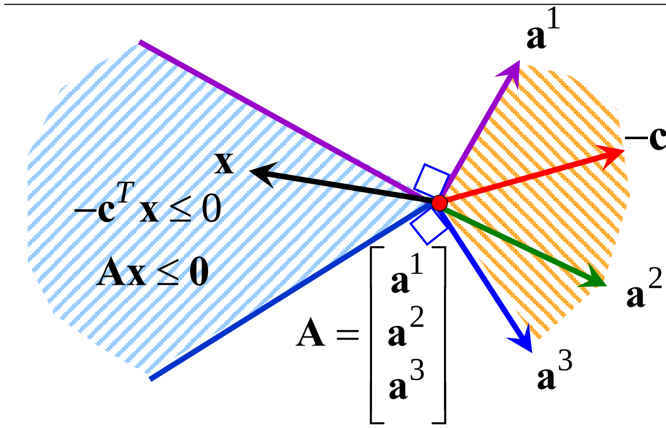
$$\max \ \boldsymbol{\lambda}^T \cdot \mathbf{0} + \upsilon \cdot 0$$

$$\text{s.t.} \quad \begin{bmatrix} \boldsymbol{\lambda}^T & \upsilon \end{bmatrix} \begin{bmatrix} -\mathbf{A} \\ \mathbf{c}^T \end{bmatrix} = \mathbf{c}^T$$

$$\boldsymbol{\lambda}, \upsilon \geq \mathbf{0}$$

$$\textit{or} \quad -\boldsymbol{\lambda}^T \mathbf{A} + \upsilon \, \mathbf{c}^T = \mathbf{c}^T$$

Geometric Interpretation



Proof (cont.):

$(\Rightarrow)$ Notice that any feasible solution is also optimal. Set $\upsilon^* = 0$, then

$$-\boldsymbol{\lambda}^T \mathbf{A} = \mathbf{c}^T$$

$(\Leftarrow)$ $\mathbf{c}^T + \boldsymbol{\lambda}^T \mathbf{A} = \mathbf{0}$

$$\mathbf{c}^T \mathbf{x} + \underbrace{\underbrace{\boldsymbol{\lambda}^T}_{\geq 0} \underbrace{\mathbf{A}\mathbf{x}}_{\leq 0}}_{\leq 0} = 0 \ \Rightarrow \mathbf{c}^T \mathbf{x} \geq 0 \,\|$$

D. Necessary Conditions

- Recall Theorem II-1 (1^{st} order necessary conditions) Given $\mathbf{x}^*$ is a local minimum of $f \in \mathbf{C}^1$ over $\mathbf{X}$, then: $\nabla f(\mathbf{x}^*)^T \mathbf{d} \geq 0$ for any feasible **d.**
  We now want to update this theorem for feasible directions (and limits of feasible directions) $\mathbf{d} \in \mathbf{D}(\mathbf{x}^*)$

Theorem 1

- If $\mathbf{x}^*$ is a local minimum and the constraint qualification (CQ) holds at $\mathbf{x}^*$, then
$$\nabla f(\mathbf{x}^*)^T \mathbf{d} \geq 0$$
$$\forall \mathbf{d} \in \mathbf{D}(\mathbf{x}^*)$$

Proof:
From CQ, the vectors $\nabla g_i(\mathbf{x}^*)$ are assumed linearly independent.

25

Proof (cont.):

We can directly apply Farkas' Lemma by letting
$$\mathbf{c} = \nabla f(\mathbf{x}^*)$$
$$\mathbf{a}^i = \nabla g_i(\mathbf{x}^*)^T \quad \text{(row vector)}$$
$$\mathbf{x} = \mathbf{d}$$

Therefore $\exists \lambda_i \geq 0$ not all zero such that :
$$\underbrace{\nabla f(\mathbf{x}^*)}_{\mathbf{c}} + \sum_{i \in \mathrm{A}(\mathbf{x}^*)} \lambda_i \underbrace{\nabla g_i(\mathbf{x}^*)}_{\mathbf{a}^{iT}} = \mathbf{0}$$

26

Proof (cont.):

or
$$\nabla f(\mathbf{x}^*)^T \mathbf{d} + \sum_{i \in \mathrm{A}(\mathbf{x}^*)} \lambda_i \nabla g_i(\mathbf{x}^*)^T \mathbf{d} = 0$$

Notice then that…
$$\nabla f(\mathbf{x}^*)^T \mathbf{d} \geq 0$$
$$\forall \mathbf{d} \text{ s.t. } \nabla g_i(\mathbf{x}^*)^T \mathbf{d} \leq 0 \qquad ||$$

27

E. Karush-Kuhn-Tucker Theorem

- **Theorem 2:**

IF $\mathbf{x}^*$ is a local minimum and regular point,

THEN $\exists \lambda_i \geq 0$ [not all 0] $\forall i \in \mathrm{A}(\mathbf{x}^*)$ s.t.
$$\nabla f(\mathbf{x}^*) + \sum_{i \in \mathrm{A}(\mathbf{x}^*)} \lambda_i \nabla g_i(\mathbf{x}^*) = \mathbf{0}$$
or
$$-\nabla f(\mathbf{x}^*) = \sum_{i \in \mathrm{A}(\mathbf{x}^*)} \lambda_i \nabla g_i(\mathbf{x}^*)$$

Proof: Follows from Farkas' Lemma

28

KKT Conditions

- **Theorem 3: IF $\mathbf{x}^*$** local min. and reg. pt., $\exists \lambda_i^* \geq 0\,[i = 1, \ldots, m]$ such that:
  1) $g_i(\mathbf{x}^*) \leq 0\ (i = 1, \ldots, m)$ **[Feasibility]**
  2) $\lambda_i^* \cdot g_i(\mathbf{x}^*) = 0\ (i = 1, \ldots, m)$
     [Complementary Slackness]
  3) $\nabla f(\mathbf{x}^*) + \boldsymbol{\lambda}^{*T} \cdot \nabla \mathbf{g}(\mathbf{x}^*) = \mathbf{0}$
     [Stationarity Conditions]

29

KKT Conditions (cont.)

Proof:

From Theorem 2:

$$\nabla f(\mathbf{x}^*) + \sum_{i \in A(\mathbf{x}^*)} \lambda_i^* \nabla g_i(\mathbf{x}^*) = \mathbf{0}$$

Let $\lambda_i^* = 0 \quad \forall i \notin A(\mathbf{x}^*)$

Then

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(\mathbf{x}^*) = \mathbf{0}$$

30

KKT Conditions (cont.)

Proof (cont.):

…and

$$\lambda_i^* \cancelto{0}{g_i}(\mathbf{x}^*) = 0 \quad \forall i \in A(\mathbf{x}^*)$$

$$\cancelto{0}{\lambda_i^*} g_i(\mathbf{x}^*) = 0 \quad \forall i \notin A(\mathbf{x}^*) \;\|$$

31

Lagrange Multipliers

- The λ_i^* are **Lagrange multipliers**:
- The KKT conditions can be expressed concisely in terms of the **Lagrangian function**:

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda}^T \cdot \mathbf{g}(\mathbf{x})$$

$$= f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i \cdot g_i(\mathbf{x})$$

32

Restatement of KKT Conditions

KKT Conditions : **IF** $\mathbf{x}^*$ a local minimum, and a regular pt., **THEN** $\exists\, \boldsymbol{\lambda}^* \geq \mathbf{0}$ (not all zero) such that:

1') $\nabla_\lambda L(\mathbf{x}^*, \boldsymbol{\lambda}^*) \leq \mathbf{0}$ **[feasibility]**

2') $L(\mathbf{x}^*, \boldsymbol{\lambda}^*) = f(\mathbf{x}^*)$ **[comp. slack]**

3') $\nabla_x L(\mathbf{x}^*, \boldsymbol{\lambda}^*) = \mathbf{0}$ **[stationarity]**

Geometric Interpretation of KKT Conditions

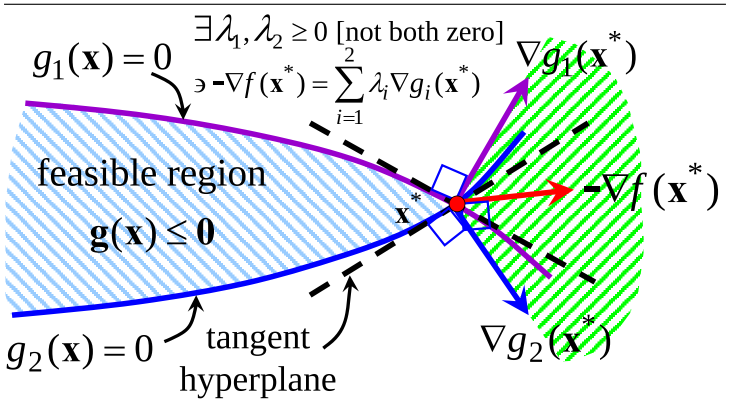


Illustration of Nonoptimal Point

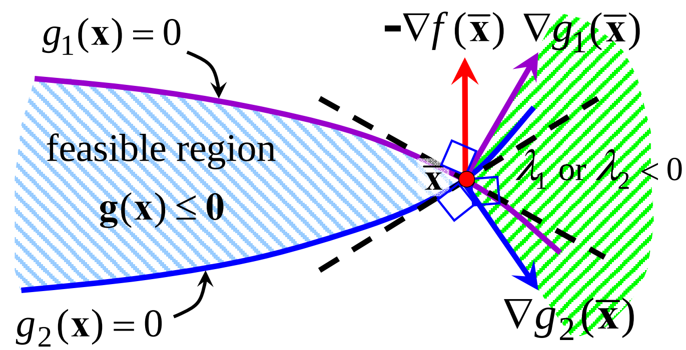
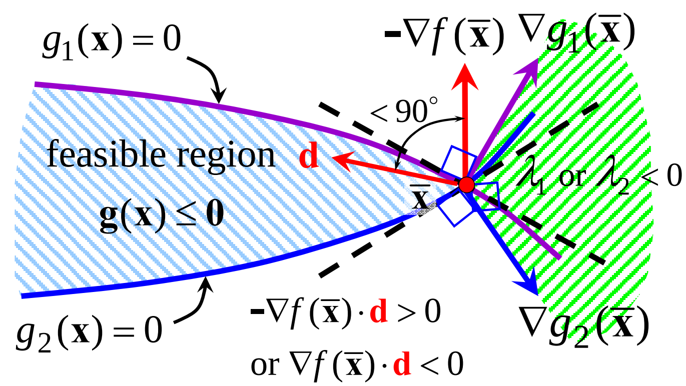


Illustration of Nonoptimal Point



Complementary Slackness

- Complementary slackness:

$$\lambda_i^* g_i(\mathbf{x}^*) = 0 \quad \forall i \in \mathsf{A}(\mathbf{x}^*)$$
$$\lambda_i^* g_i(\mathbf{x}^*) = 0 \quad \forall i \notin \mathsf{A}(\mathbf{x}^*)$$

- For any constraint which is slack (inactive) at $\mathbf{x}^*$, its $\lambda_i^* = 0$ $\left[\text{i.e., } g_i(x^*) < 0\right]$
- For any constraint which is tight (active) at $\mathbf{x}^*$, its $\lambda_i^* \geq 0$ $\left[\text{i.e., } g_i(x^*) = 0\right]$

F. Economic Interpretation of Lagrange Mulipliers

$$\lambda_i^* = -\frac{\partial f}{\partial b_i} \Big| b_i = g_i(\mathbf{x}^*) = 0$$

Proof:

Define a more general problem:

$$\min_{\mathbf{x}} \quad f(\mathbf{x})$$
$$\text{s.t.:} \quad \mathbf{g}(\mathbf{x}) \leq \mathbf{b}$$

Economic Interpretation (cont.)

Proof (cont.):

So…

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda}^T \cdot \left[\mathbf{g}(\mathbf{x}) - \mathbf{b}\right]$$

Suppose $\mathbf{x}^*(\mathbf{b})$ solves the above problem (i.e., solution depends on given **b**). Applying Condition 3' of KKT conditions:

Economic Interpretation (cont.)

Proof (cont.):

$$\frac{\partial L}{\partial x_j} = \frac{\partial f(\mathbf{x}^*(\mathbf{b}))}{\partial x_j} + \sum_{i=1}^{m} \lambda_i^* \frac{\partial g_i(\mathbf{x}^*(\mathbf{b}))}{\partial x_j} = 0$$

For all active constraints, we have…

$$g_i(\mathbf{x}^*(\mathbf{b})) = b_i$$

Economic Interpretation (cont.)

Proof (cont.):

So…

$$\frac{\partial g_i(\mathbf{x}^*(\mathbf{b}))}{\partial b_k} = 1 \quad \text{for } i = k$$

$$\frac{\partial g_i(\mathbf{x}^*(\mathbf{b}))}{\partial b_k} = 0 \quad \text{for all } i \neq k$$

For all inactive constraints:

$$\lambda_i^* = 0 \quad \left[\forall i \notin \mathbf{A}(\mathbf{x}^*)\right]$$

41

Economic Interpretation (cont.)

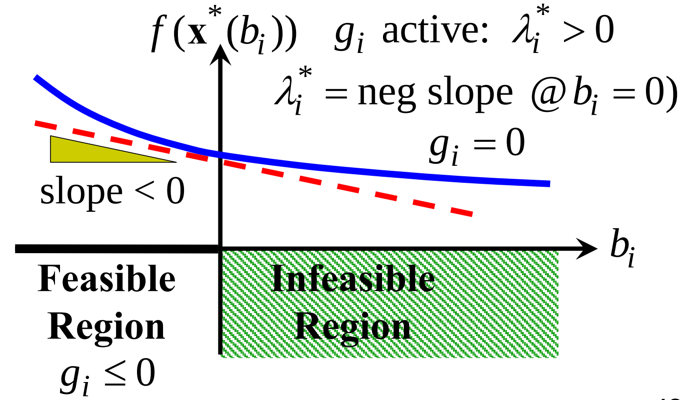
Proof (cont.):

So…

$$\frac{\partial f}{\partial b_k}\frac{\partial b_k}{\partial x_j} + \lambda_k^* \overset{=1}{\frac{\partial g_k}{\partial b_k}}\frac{\partial b_k}{\partial x_j} + \sum_{\substack{i=1\\ i\neq k}}^{m} \lambda_i^* \overset{0}{\frac{\partial g_i}{\partial b_k}}\frac{\partial b_k}{\partial x_j} = 0 \quad ||$$

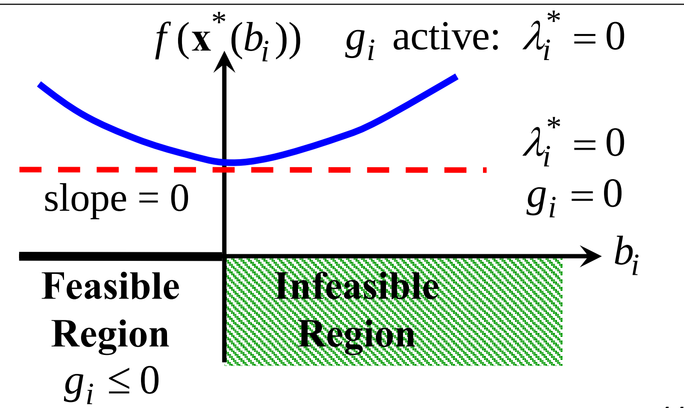
42

Graphical Interpretation



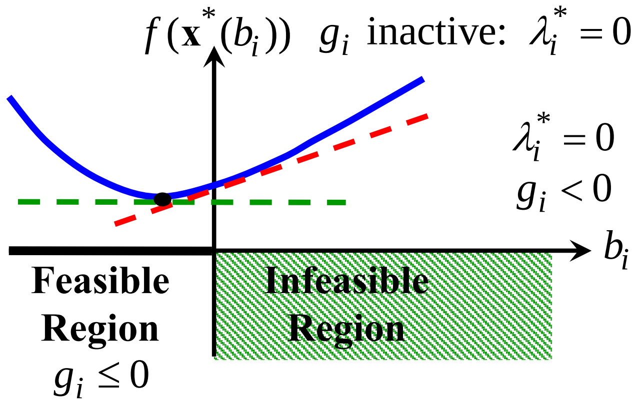
43

Graphical Interpretation (cont.)



44

Graphical Interpretation (cont.)



G. KKT Conditions for Equality Constraints

Consider:

$$\min_{\mathbf{x}} \ f(\mathbf{x})$$

$$\text{s.t.:} \quad \mathbf{h}(\mathbf{x}) = \mathbf{0} \ \text{ or } \begin{cases} \mathbf{h}(\mathbf{x}) \le \mathbf{0} \\ -\mathbf{h}(\mathbf{x}) \le \mathbf{0} \end{cases}$$

$$L(\mathbf{x}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_1) = f(\mathbf{x}) + \boldsymbol{\lambda}_1^T \mathbf{h}(\mathbf{x}) + \boldsymbol{\lambda}_2^T [-\mathbf{h}(\mathbf{x})]$$

$$= f(\mathbf{x}) + [\underbrace{\boldsymbol{\lambda}_1^T}_{\ge 0} - \underbrace{\boldsymbol{\lambda}_2^T}_{\ge 0}] \cdot \mathbf{h}(\mathbf{x}) = f(\mathbf{x}) + \underbrace{\boldsymbol{\lambda}^T}_{\text{free}} \mathbf{h}(\mathbf{x})$$

Equality Constraints (cont.)

$\exists \, \boldsymbol{\lambda}^* \in E^m$ (free) such that:

1") $\nabla_\lambda L(\mathbf{x}^*, \boldsymbol{\lambda}^*) = \mathbf{0}$

2") $\nabla_x L(\mathbf{x}^*, \boldsymbol{\lambda}^*) = \mathbf{0}$

H. Sufficient Conditions

- **Theorem 4:** Assume f and $\mathbf{g}$ are:
  i. differentiable
  ii. convex*

  Then if the KKT conditions (1-3) or (1’-3’) are satisfied for some $\mathbf{x}^*$, then that $\mathbf{x}^*$ must be the **global optimum**.

* f pseudo convex; g_i quasiconvex

Sufficient Conditions (cont.)

Proof:

$L(\mathbf{x}, \boldsymbol{\lambda}^*)$ must be convex w.r.t. $\mathbf{x}$:

Therefore, if Condition 3' is satisfied where

$$\nabla_x L(\mathbf{x}^*, \boldsymbol{\lambda}^*) = 0$$

then by convexity:

$$L(\mathbf{x}^*, \boldsymbol{\lambda}^*) \le L(\mathbf{x}, \boldsymbol{\lambda}^*) \quad \forall \mathbf{x} \in E^n$$

49

Sufficient Conditions (cont.)

Proof (cont.):

Therefore…

$$f(\mathbf{x}^*) + \overbrace{\boldsymbol{\lambda}^{*T}\mathbf{g}(\mathbf{x}^*)}^{=0}$$

$$\le f(\mathbf{x}) + \overbrace{\boldsymbol{\lambda}^{*T}\mathbf{g}(\mathbf{x})}^{\le 0}$$

$$\Rightarrow f(\mathbf{x}^*) \le f(\mathbf{x}) \quad \forall \mathbf{x} \ni \mathbf{g}(\mathbf{x}) \le \mathbf{0} \quad \|$$

50

I. Example

$$\min \quad (x_1 - 1)^2 + (x_2 - 2)^2$$

subject to:

$$-x_1 + x_2 = 1$$
$$x_1 + x_2 \le 2$$
$$x_1, x_2 \ge 0$$

51

Example (cont.)

- To apply KKT conditions, we will temporarily **relax** nonnegativity constraints and then see if they are satisfied anyway.
  - these are normally included in Lagrangian function with associated Lagrange multipliers (unlike simplex method where nonnegativity holds)

52

Example (cont.)

- KKT conditions:

  $\exists\, \mathbf{x}$ and $\exists \lambda_1$ free ; $\lambda_2 \geq 0$ such that:

  $$1)\quad -x_1 + x_2 = 1$$
  $$x_1 + x_2 \leq 2$$
  $$2)\quad \lambda_2 \left[x_1 + x_2 - 2\right] = 0$$
  $$3)\quad \frac{\partial L}{\partial x_1} = 0 = 2\left(x_1 - 1\right) - \lambda_1 + \lambda_2$$
  $$\frac{\partial L}{\partial x_2} = 0 = 2\left(x_2 - 2\right) + \lambda_1 + \lambda_2$$

Example (cont.)

Assume:

$x_1 + x_2 = 2$ 2) OK!

From 1): $x_1^* = \frac{1}{2} \quad x_2^* = \frac{3}{2}$

From 3): $2\left[\frac{1}{2} - 1\right] - \lambda_1 + \lambda_2 = 0$

$$2\left[\frac{3}{2} - 2\right] + \lambda_1 + \lambda_2 = 0$$

$\lambda_1^* = 0 \quad \lambda_2^* = 1 \ (\geq 0)$ OK!

Example (cont.)

- For this problem, the KKT conditions are both necessary and sufficient: $g_i(\mathbf{x})$ linear

  $$f(x) = (x_1 - 1)^2 + (x_2 - 2)^2$$
  $$\frac{\partial f}{\partial x_1} = 2(x_1 - 1) \qquad \frac{\partial f}{\partial x_2} = 2(x_1 - 2)$$
  $$\mathbf{H} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \qquad \det \mathbf{H} = 4 > 0$$

J. Analytical KKT Solutions

- Suppose we have the following problem:

  $$\min \quad \mathbf{c}^T\mathbf{x} + \tfrac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x}$$
  $$s.t. \quad \mathbf{A}\mathbf{x} = \mathbf{b}$$

  where the $\mathbf{x}$ variables are free

  $$L(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{c}^T\mathbf{x} + \tfrac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} + \boldsymbol{\lambda}^T\left[-\mathbf{A}\mathbf{x} + \mathbf{b}\right]$$

Analytical Solution (cont.)

- KKT conditions:

  $\exists$ **x** and $\exists\boldsymbol{\lambda}$ free such that:

  1) $\mathbf{A}\mathbf{x} = \mathbf{b}$

  2) $\mathbf{c} + \mathbf{Q}\mathbf{x} - \mathbf{A}^T\boldsymbol{\lambda} = \mathbf{0}$

  $$\mathbf{x} = \mathbf{Q}^{-1}\left[\mathbf{A}^T\boldsymbol{\lambda} - \mathbf{c}\right]$$

  $$\mathbf{A}\left[\mathbf{Q}^{-1}\left[\mathbf{A}^T\boldsymbol{\lambda} - \mathbf{c}\right]\right] = \mathbf{b}$$

57

Analytical Solution (cont.)

$$\left[\mathbf{A}\mathbf{Q}^{-1}\mathbf{A}^T\right]\boldsymbol{\lambda} - \mathbf{A}\mathbf{Q}^{-1}\mathbf{c} = \mathbf{b}$$

$$\boldsymbol{\lambda} = \left[\mathbf{A}\mathbf{Q}^{-1}\mathbf{A}^T\right]^{-1}\left[\mathbf{A}\mathbf{Q}^{-1}\mathbf{c} + \mathbf{b}\right]$$

$$\mathbf{x} = \mathbf{Q}^{-1}\left[\mathbf{A}^T\left[\mathbf{A}\mathbf{Q}^{-1}\mathbf{A}^T\right]^{-1}\left[\mathbf{A}\mathbf{Q}^{-1}\mathbf{c} + \mathbf{b}\right] - \mathbf{c}\right]$$

58

Affine Scaling Optimal Direction

- Let us relate this analytical solution to $\mathbf{c}_p$ calculated for the affine scaling algorithm.

$$\min \left\| -\hat{\mathbf{c}}\mathbf{D} - \mathbf{d} \right\|^2$$

$$\text{s.t.} \quad \mathbf{B}\mathbf{d} = \mathbf{0}$$

[different **c** - call it $\hat{\mathbf{c}}$]

59

Affine Scaling Direction (cont.)

- Objective function as quadratic form:

$$\min \cancelto{0}{\left[\hat{\mathbf{c}}^T\mathbf{D}\right]^T}\left[\hat{\mathbf{c}}^T\mathbf{D}\right] + 2\left[\hat{\mathbf{c}}^T\mathbf{D}\right]^T \mathbf{d}$$
$$+\, \mathbf{d}^T\mathbf{I}\mathbf{d}$$

- Comparing with general KKT solution:

$$\mathbf{x} = \mathbf{d} \quad \mathbf{Q} = 2\mathbf{I} \quad \mathbf{c} = 2\left[\hat{\mathbf{c}}^T\mathbf{D}\right]$$

$$\mathbf{b} = \mathbf{0} \quad \mathbf{A} = \mathbf{B}$$

60

Affine Scaling Direction (cont.)

$$\mathbf{x} = \tfrac{1}{2}\mathbf{I}\left\{\mathbf{B}^T\left[\mathbf{B}\tfrac{1}{2}\mathbf{I}\mathbf{B}^T\right]^{-1}\left[\mathbf{B}\tfrac{1}{2}\mathbf{I}\cdot 2\hat{\mathbf{c}}^T\mathbf{D}\right]\right.$$

$$\left.-2\hat{\mathbf{c}}^T\mathbf{D}\right\}$$

$$\mathbf{x} = \tfrac{1}{2}\left\{\mathbf{B}^T 2\left[\mathbf{B}\mathbf{B}^T\right]^{-1}\left[\mathbf{B}\hat{\mathbf{c}}^T\mathbf{D} - 2\cdot\mathbf{I}\;\hat{\mathbf{c}}^T\mathbf{D}\right]\right\}$$

$$\mathbf{x} = \mathbf{c}_p = \left\{\mathbf{B}^T\left[\mathbf{B}\mathbf{B}^T\right]^{-1}\mathbf{B} - \mathbf{I}\right\}\underbrace{\hat{\mathbf{c}}^T\mathbf{D}}_{\mathbf{D}\hat{\mathbf{c}}} \qquad \text{since } \mathbf{D}=\mathbf{D}^T$$

61