Project Paper

- Due on **Tuesday, 3/15 @ 11:59 pm** (this is a **<u>HARD</u> deadline**)
- The **project paper** will be
  - **8 pages <u>maximum</u>** (double–column)
  - should include:
    - abstract
    - introduction
    - description of the project (methods)
    - experiments
    - conclusion
    - bibliography (this does not count toward the 8 pages)
- **LaTeX style files** are available on Canvas under the Project Module

ECE 271B – Winter 2022

Duality

Disclaimer:
This class will be recorded
and made available to students asynchronously.

Manuela Vasconcelos

ECE Department, UCSD

Optimization

- **goal**: find **maximum** or **minimum** of a function

- **Definition:** given functions f, $g_i, i = 1, \dots, r$ and $h_i, i = 1, \dots, m$ defined on some domain $\Omega \in \mathbb{R}^n$

$$\begin{aligned} \min_{\mathbf{w}} \quad & f(\mathbf{w}), \mathbf{w} \in \Omega \\ \text{subject to} \quad & g_i(\mathbf{w}) \leq 0, \forall i \\ & h_i(\mathbf{w}) = 0, \forall i \end{aligned}$$

- for compactness,
  we write $g(\mathbf{w}) \leq 0$ instead of $g_i(\mathbf{w}) \leq 0, \forall i$ and similarly $h(\mathbf{w}) = 0$

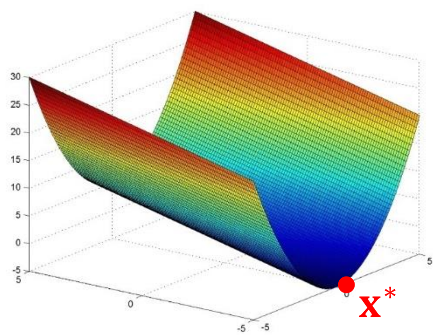
- we derived necessary and sufficient conditions for (local) **optimality**
  - in the **absence of constraints** (<u>unconstrained</u>)
  - with <u>equality</u> constraints <u>only</u>
  - with <u>equality</u> and <u>inequality constraints</u>

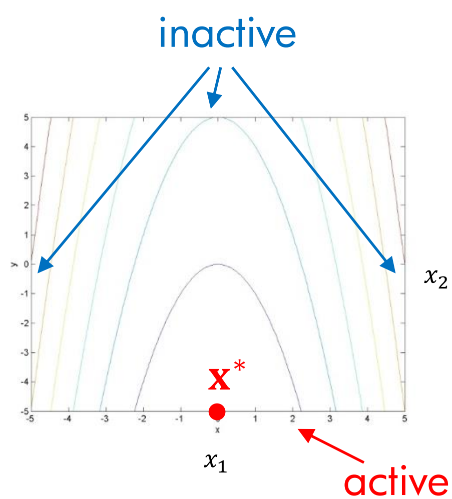
Inequality Constraints

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} f(\mathbf{x}) \quad \text{subject to} \quad h(\mathbf{x}) = 0, g(\mathbf{x}) \leq 0$$

- we start by defining the set $A(\mathbf{x})$ of <u>**active**</u> inequality constraints

$$A(\mathbf{x}) = \{ j \mid g_j(\mathbf{x}) = 0 \}$$





- inactive constraints do <u>not</u> do **anything**
- active constraints are <u>**equalities**</u>

- the ones that **matter** are those which are <u>**active**</u>, and these are <u>equalities</u>

Constrained Optimization

- hence, the problem

$$\boxed{\mathbf{x}^* = \arg\min_{\mathbf{x}} f(\mathbf{x}) \quad \text{subject to} \quad h(\mathbf{x}) = 0, g(\mathbf{x}) \leq 0}$$

- is equivalent to

$$\boxed{\mathbf{x}^* = \arg\min_{\mathbf{x}} f(\mathbf{x}) \quad \text{subject to} \quad h(\mathbf{x}) = 0, g_i(\mathbf{x}) = 0, \forall i \in A(\mathbf{x}^*)}$$

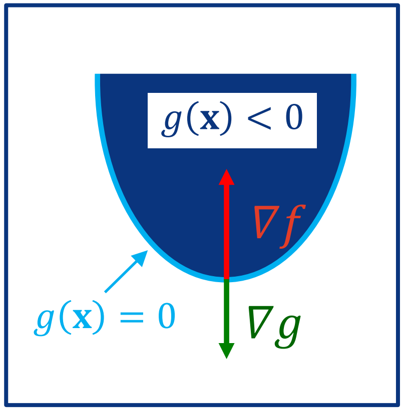
- this is a problem with equality constraints:

there must be a $\boldsymbol{\lambda}^*$ and $\boldsymbol{\mu}^*$, such that

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(\mathbf{x}^*) + \sum_{j=1}^{r} \mu_j^* \nabla g_j(\mathbf{x}^*) = 0$$

with $\mu_j^* = 0,\ j \notin A(\mathbf{x}^*)$

finally, we need $\mu_j^* \geq 0,\ \forall j$, to guarantee this →



The KKT Conditions

- **Theorem:** for the problem

$$\boxed{\mathbf{x}^* = \arg\min_{\mathbf{x}} f(\mathbf{x}) \quad \text{subject to} \quad h(\mathbf{x}) = 0, g(\mathbf{x}) \leq 0}$$

$\mathbf{x}^*$ is a local minimum if and only if there exist $\boldsymbol{\lambda}^*$ and $\boldsymbol{\mu}^*$ such that

i) $$\boxed{\nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(\mathbf{x}^*) + \sum_{j=1}^{r} \mu_j^* \nabla g_j(\mathbf{x}^*) = 0}$$

these conditions would be the same if all constraints were equalities

ii) $\boxed{\mu_j^* \geq 0, \ \forall j}$ condition on <u>all</u> inequality constraints

iii) $\boxed{\mu_j^* = 0, \ \forall j \notin A(\mathbf{x}^*)}$ this condition eliminates <u>inactive</u> constraints

iv) $h(\mathbf{x}^*) = 0$

v) $$\mathbf{y}^T \nabla \left[\nabla f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(\mathbf{x}) + \sum_{j=1}^{r} \mu_j^* \nabla g_j(\mathbf{x}) \right]_{\mathbf{x}=\mathbf{x}^*} \mathbf{y} \geq 0, \forall \mathbf{y} \in V(\mathbf{x}^*)$$

where $V(\mathbf{x}^*) = \left\{ \mathbf{y} \,\middle|\, \nabla h_i^T(\mathbf{x}^*)\mathbf{y} = 0, \forall i \text{ and } \nabla g_j^T(\mathbf{x}^*)\mathbf{y} = 0, \forall j \in A(\mathbf{x}^*) \right\}$

Geometric Interpretation

- first, we consider the case <u>without</u> **equality constraints**

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} f(\mathbf{x}) \quad \text{subject to} \quad g(\mathbf{x}) \le 0$$

- from the KKT conditions, the solution satisfies

i) $\nabla[L(\mathbf{x}^*, \boldsymbol{\mu}^*)] = 0$

ii) $\mu_j^* \ge 0, \ \forall j$

iii) $\mu_j^* = 0, \forall j \notin A(\mathbf{x}^*)$

→

this implies that

$\mu_j^* g_j(\mathbf{x}^*) = 0, \forall j$

active: $g_j(\mathbf{x}^*) = 0$

inactive: $\mu_j^* = 0$

with

$$L(\mathbf{x}^*, \boldsymbol{\mu}^*) = f(\mathbf{x}^*) + \sum_{j=1}^{r} \mu_j^* g_j(\mathbf{x}^*)$$

which is equivalent to

$$L^* = \min_{\mathbf{x}}[L(\mathbf{x}, \boldsymbol{\mu}^*)] = \min_{\mathbf{x}}[f(\mathbf{x}) + (\boldsymbol{\mu}^*)^T g(\mathbf{x})]$$

with $\mu_j^* \ge 0, \forall j$ and $\mu_j^* = 0, \forall j \notin A(\mathbf{x}^*)$

Geometric Interpretation

$$L^* = \min_{\mathbf{x}}[L(\mathbf{x},\boldsymbol{\mu}^*)] = \min_{\mathbf{x}}[f(\mathbf{x}) + (\boldsymbol{\mu}^*)^T g(\mathbf{x})]$$

with $\mu_j^* \geq 0,\ \forall j$ and $\mu_j^* = 0, \forall j \notin A(\mathbf{x}^*)$

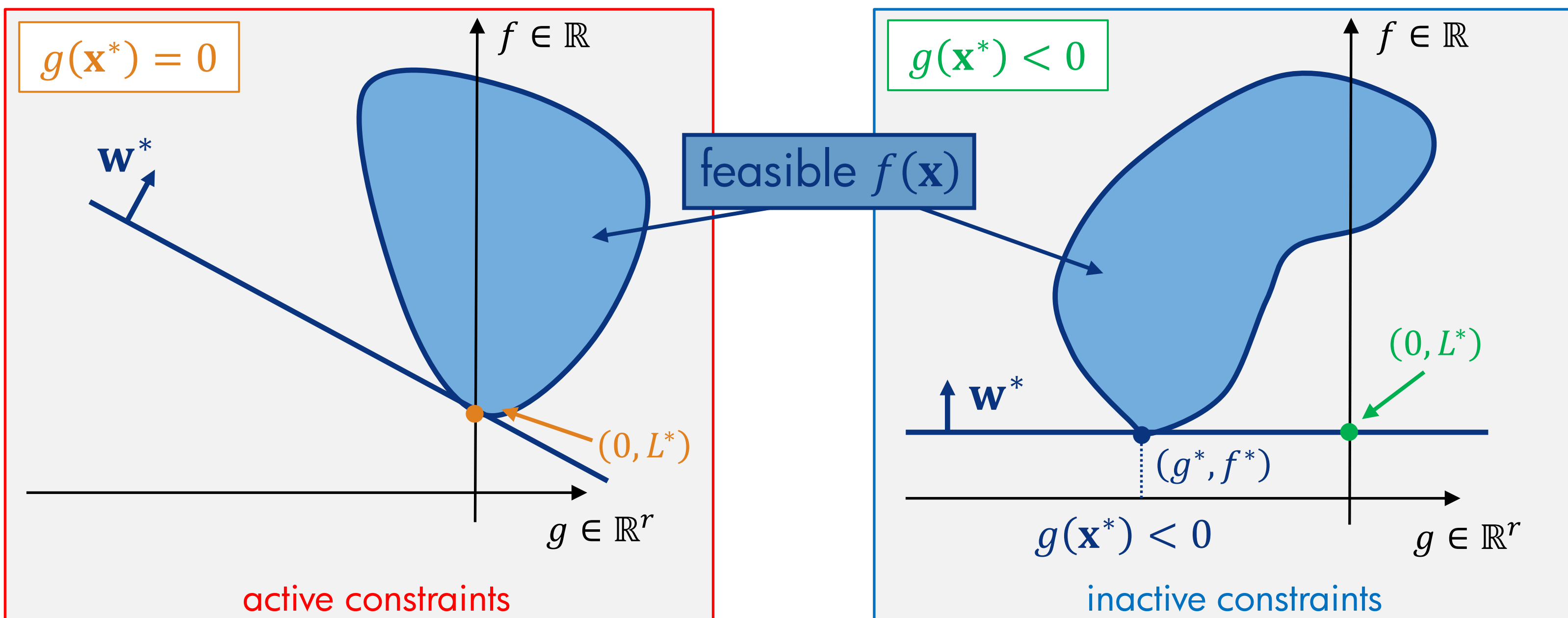
- is equivalent to
  - $\mathbf{x} = \mathbf{x}^* \Rightarrow (\mathbf{w}^*)^T\mathbf{z} - b = 0$
  - $\mathbf{x} \neq \mathbf{x}^* \Rightarrow (\mathbf{w}^*)^T\mathbf{z} - b \geq 0$

plane in z −space
normal $\mathbf{w}^*$, bias b

$\mathbf{z}$ is in half−space
pointed to by $\mathbf{w}^*$

$$b = L^* \qquad \mathbf{w}^* = \begin{bmatrix} 1 \\ \boldsymbol{\mu}^* \end{bmatrix} \qquad \mathbf{z} = \begin{bmatrix} f(\mathbf{x}) \\ g(\mathbf{x}) \end{bmatrix}$$

and can be visualized as



$$L^* = \min_{\mathbf{x}}[L(\mathbf{x}, \boldsymbol{\mu}^*)] = \min_{\mathbf{x}}[f(\mathbf{x}) + (\boldsymbol{\mu}^*)^T g(\mathbf{x})]$$
with $\mu_j^* \geq 0,\ \forall j$ and $\mu_j^* = 0, \forall j \notin A(\mathbf{x}^*)$

Duality

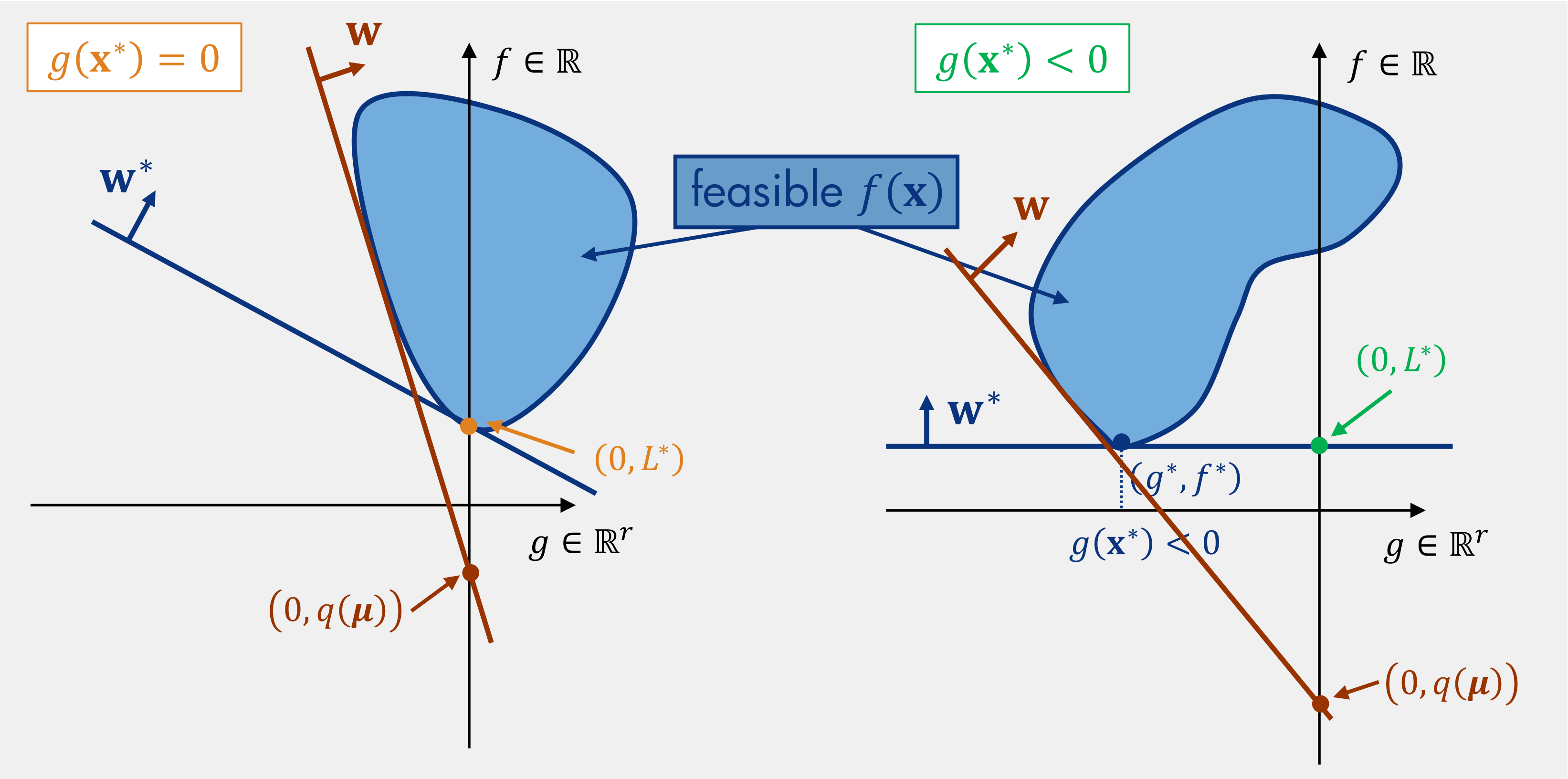
- we solve **instead**

$q(\boldsymbol{\mu})$ – (Lagrangian) dual function

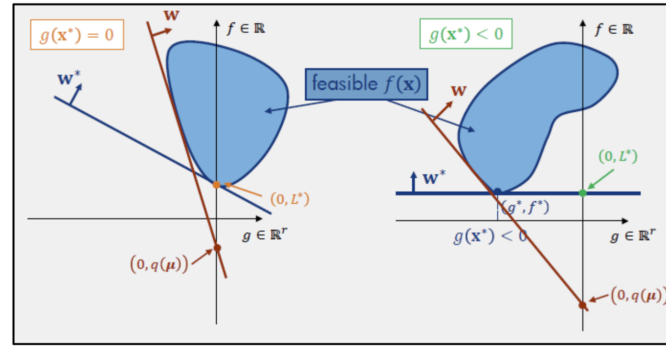
$$q(\boldsymbol{\mu}) = \min_{\mathbf{x}}[L(\mathbf{x}, \boldsymbol{\mu})] = \min_{\mathbf{x}}[f(\mathbf{x}) + \boldsymbol{\mu}^T g(\mathbf{x})]$$
with $\boldsymbol{\mu} \geq \mathbf{0}$

$$b = q(\boldsymbol{\mu}) \qquad \mathbf{w} = \begin{bmatrix} 1 \\ \boldsymbol{\mu} \end{bmatrix} \qquad \mathbf{z} = \begin{bmatrix} f(\mathbf{x}) \\ g(\mathbf{x}) \end{bmatrix}$$

- <u>same</u> picture with L^* replaced by $q(\boldsymbol{\mu})$ and $\boldsymbol{\mu}^*$ replaced by $\boldsymbol{\mu}$



Duality



- note that
  - $q(\boldsymbol{\mu}) \leq L^* = f^*$
  - if we keep increasing $q(\boldsymbol{\mu})$, we will get $q(\boldsymbol{\mu}) = L^*$
  - we cannot go beyond L^*

- this is exactly the definition of the **dual problem**

$$\max_{\boldsymbol{\mu} \geq \mathbf{0}} q(\boldsymbol{\mu}) \qquad q(\boldsymbol{\mu}) = \min_{\mathbf{x}} [L(\mathbf{x}, \boldsymbol{\mu})] = \min_{\mathbf{x}} [f(\mathbf{x}) + \boldsymbol{\mu}^T g(\mathbf{x})]$$

- **note**:
  - $q(\boldsymbol{\mu})$ may go to $-\infty$ for some $\boldsymbol{\mu}$, which means that there is <u>no</u> Lagrange multiplier (plane would be vertical)
  - this is avoided by introducing the **constraint**

$$\boldsymbol{\mu} \in D_q = \{\boldsymbol{\mu} \mid q(\boldsymbol{\mu}) > -\infty\}$$

primal problem:
$$\mathbf{x}^* = \arg\min_{\mathbf{x}} f(\mathbf{x}) \quad \text{subject to} \quad g(\mathbf{x}) \le 0$$

Duality

- we therefore have a **two–step** recipe to find the optimal solution

1. for any $\boldsymbol{\mu}$, solve

$$\boxed{q(\boldsymbol{\mu}) = \min_{\mathbf{x}}[L(\mathbf{x}, \boldsymbol{\mu})] = \min_{\mathbf{x}}[f(\mathbf{x}) + \boldsymbol{\mu}^T g(\mathbf{x})]}$$

2. then solve

$$\boxed{\max_{\boldsymbol{\mu} \ge \mathbf{0},\, \boldsymbol{\mu} \in D_q} q(\boldsymbol{\mu})} \qquad D_q = \{\boldsymbol{\mu} \mid q(\boldsymbol{\mu}) > -\infty\}$$

- 1. is similar to the Lagrangian of an equality constraint problem but **easier** because we do not need to solve for $\boldsymbol{\mu}$
- 2. is called the **dual problem**
- one of the reasons why this is **interesting** is that 2. turns out to be quite manageable (we will see why)

Equality Constraints

- so far, we have **disregard** them
- what about

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} f(\mathbf{x}) \quad \text{subject to} \quad h(\mathbf{x}) = 0, g(\mathbf{x}) \leq 0$$

- intuitively, nothing should change, since

$$h(\mathbf{x}) = 0 \Leftrightarrow \{h(\mathbf{x}) \leq 0 \text{ and } -h(\mathbf{x}) \leq 0\}$$

- i.e., each equality is the same as two inequalities

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} f(\mathbf{x}) \quad \text{subject to} \quad h(\mathbf{x}) \leq 0, -h(\mathbf{x}) \leq 0, g(\mathbf{x}) \leq 0$$

- this has Lagrangian

$$L(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\alpha}^+, \boldsymbol{\alpha}^-) = f(\mathbf{x}) + \sum_{i=1}^{r} \mu_i g_i(\mathbf{x}) + \sum_{i=1}^{m} \alpha_i^+ h_i(\mathbf{x}) - \sum_{i=1}^{m} \alpha_i^- h_i(\mathbf{x})$$

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} f(\mathbf{x}) \quad \text{subject to} \quad h(\mathbf{x}) = 0, g(\mathbf{x}) \le 0$$

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} f(\mathbf{x}) \quad \text{subject to} \quad h(\mathbf{x}) \le 0, -h(\mathbf{x}) \le 0, g(\mathbf{x}) \le 0$$

Equality Constraints

- which is equivalent to

$$L(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\alpha}^+, \boldsymbol{\alpha}^-) = f(\mathbf{x}) + \sum_{i=1}^{r} \mu_i g_i(\mathbf{x}) + \sum_{i=1}^{m} \alpha_i^+ h_i(\mathbf{x}) - \sum_{i=1}^{m} \alpha_i^- h_i(\mathbf{x})$$

$$= f(\mathbf{x}) + \sum_{i=1}^{r} \mu_i g_i(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i h_i(\mathbf{x}) \quad \text{with} \quad \lambda_i = \alpha_i^+ - \alpha_i^-$$

- i.e., it is basically the **same**, but $\boxed{\lambda_i \text{ do } \textbf{not} \text{ have to be } \ge 0}$

- in summary, $(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*)$ is a Lagrange multiplier if $\boxed{\boldsymbol{\mu}^* \ge 0}$ and

$$\boxed{f^* = \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*)}$$

- the **dual** is

$$\boxed{\max_{(\boldsymbol{\mu} \ge \mathbf{0}, \boldsymbol{\lambda} \in \mathbb{R}^m) \in D_q} q(\boldsymbol{\mu}, \boldsymbol{\lambda})}$$

$$D_q = \{\boldsymbol{\mu}, \boldsymbol{\lambda} \mid q(\boldsymbol{\mu}, \boldsymbol{\lambda}) > -\infty\}$$

$$\boxed{q(\boldsymbol{\mu}, \boldsymbol{\lambda}) = \min_{\mathbf{x}} [L(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda})] \quad \text{with } \boldsymbol{\mu} \ge 0, \boldsymbol{\lambda} \in \mathbb{R}^m}$$

Back to Duality

- last class, we proved

  Theorem: D_q is a convex set and $q(\boldsymbol{\mu})$ is concave on D_q.

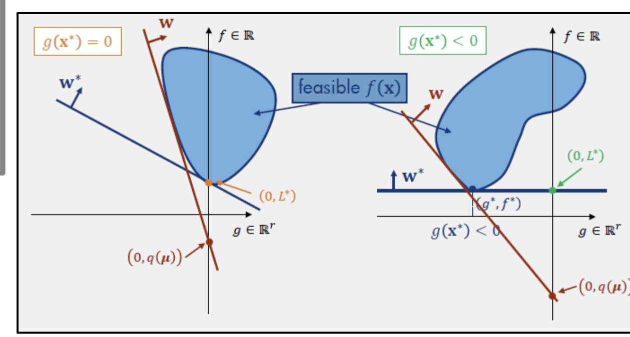
- note that the dual is always concave, irrespective of the primal optimization problem

- very appealing result since convex optimization problems are among the easiest to solve

- **Theorem:** (weak duality) it is always true that

  $$\boxed{q^* \leq f^*}$$

weak duality:
maximum of the dual is never larger than the minimum of the primal



Duality Gap

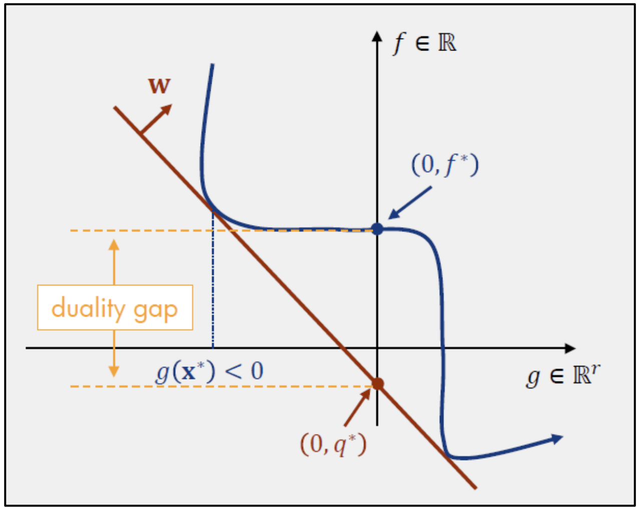
- we say that

  - if $q^* = f^*$, there is <u>no</u> **duality gap**
  - otherwise, there is a **duality gap**

- the **duality gap** <u>constrains</u> the existence of Lagrange multipliers

- **Theorem:**
  - if there is <u>**no**</u> **duality gap**, the set of Lagrange multipliers is the set of optimal dual solutions;
  - if there is a **duality gap**, there are <u>**no**</u> Lagrange multipliers.

- last class, we discussed a **dual problem** that has a **solution**, but for which there is <u>no</u> Lagrange multiplier →



duality is <u>only</u> interesting when there is <u>no</u> **duality gap**

Duality Gap

- the KKT theorem assures a **local minimum** only when there is a set of Lagrange multipliers that satisfies the KKT conditions

- this is impossible if there is a **duality gap**

- **when is this the case?**
  - as far as I know, this is still an open question
  - there are various results which characterize the existence of solutions for **certain** classes of problems
  - the bulk of the results are for the case of **convex programming problems**

- recall: the problem is **convex** if the function f is convex and the constraints h and g are convex

Duality Gap

- the following theorems are relevant
  (**note:** proofs are hard, not particularly insightful, therefore omitted)

- **Theorem:** (strong duality) Consider the problem

$$\boxed{\mathbf{x}^* = \arg\min_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}) \quad \text{subject to} \quad g(\mathbf{x}) \leq 0}$$

where $\mathcal{X}, f$, and g_i are <u>**all**</u> **convex**, the optimal value f^* is finite, and there is a vector $\bar{\mathbf{x}}$ such that

$$\boxed{g_j(\bar{\mathbf{x}}) < 0, \forall j} \quad (*)$$

Then, there is <u>at least</u> **one** Lagrange multiplier vector and there is <u>**no**</u> duality gap.

- i.e., **convex problems** <u>have **dual**</u> as long as (*) holds!

$D_q = \{\boldsymbol{\mu}, \boldsymbol{\lambda} | q(\boldsymbol{\mu}, \boldsymbol{\lambda}) > -\infty\}$

Duality Gap

- the condition $\boxed{g_j(\bar{\mathbf{x}}) < 0, \forall j}$ is needed to guarantee that there <u>are</u> Lagrange multipliers
  - consider the following example that **<u>violates</u> the condition**

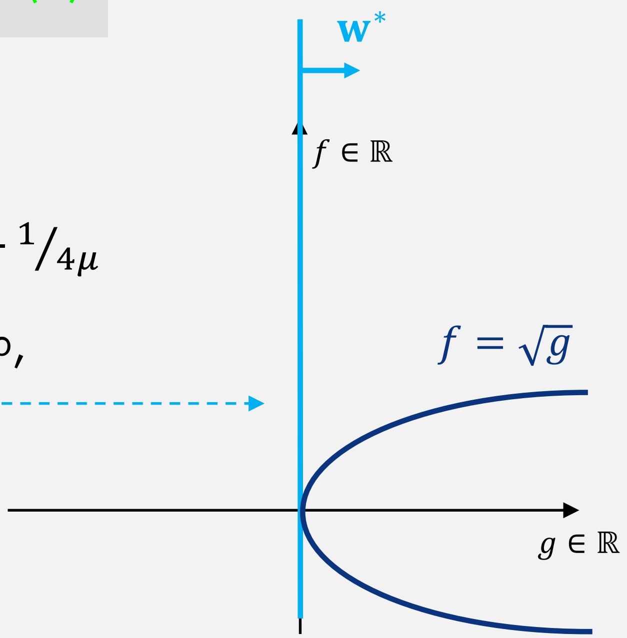
$$\min_{x \in \mathbb{R}} f(x) = x \quad \text{subject to} \quad g(x) = x^2 \leq 0 \qquad (*)$$

  - the solution of

$$q(\mu) = \min_x [f(x) + \mu g(x)], \forall \mu \geq 0$$

  occurs at $x^* = -{}^{1}/_{2\mu}$ and has value $q(\mu) = -{}^{1}/_{4\mu}$

  - since $\mu \geq 0$, $q(\mu)$ converges to zero as $\mu \to \infty$, i.e., a vertical supporting plane
  - geometrically, this <u>cannot</u> happen for finite μ since the first coordinate of $\mathbf{w}^*$ is 1



  - there is <u>no</u> Lagrange multiplier since

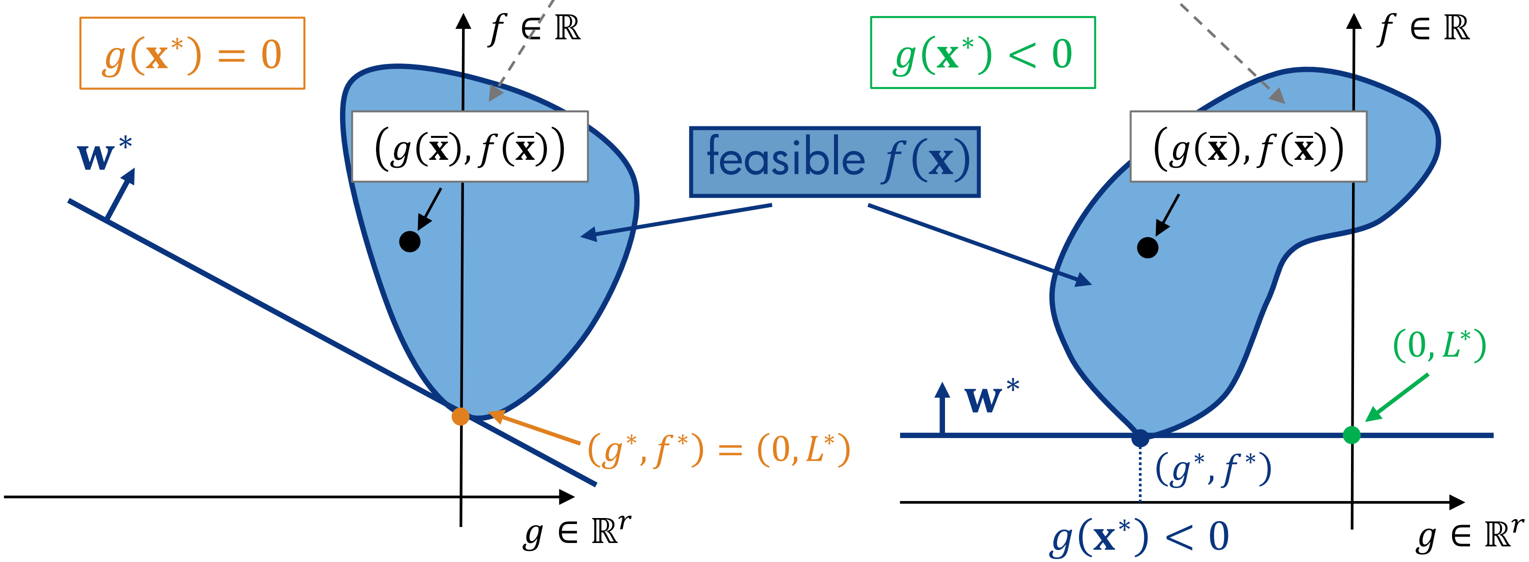
$\boxed{\max_{\boldsymbol{\mu} \geq \mathbf{0}} q(\boldsymbol{\mu})}$ has no solution, even though $x^* = 0$ is a solution of $(*)$!

Duality Gap

- The condition

$$g_j(\bar{\mathbf{x}}) < 0, \forall j$$

guarantees that this (see example) **never** happens



Duality Gap

- there is also a slightly **more general** result when the constraints are linear

- **Theorem:** (strong duality) Consider the problem

$$\mathbf{x}^* = \arg\min_{\mathbf{x}\in\mathcal{X}} f(\mathbf{x}) \quad \text{subject to} \quad \mathbf{e}_j^T\mathbf{x} - d_j \leq 0 \qquad (**)$$

  where $\mathcal{X}$ and f are convex, and the optimal value f^* is finite. Then, there is at least one Lagrange multiplier vector and there is **no** duality gap.

- **Corollary:** if, in addition to $(**)$, f is **linear** and $\mathcal{X}$ polyhedral, then there is **no** duality gap.

- these problems are called linear programming problems

Linear Programming

- consider the problem

$$\boxed{\mathbf{x}^* = \arg\min_{\mathbf{x} \geq \mathbf{0}} \mathbf{c}^T\mathbf{x} \quad \text{subject to} \quad \mathbf{A}\mathbf{x} - \mathbf{b} \leq \mathbf{0}}$$

- the dual function is

$$\begin{aligned} q(\boldsymbol{\mu}) &= \min_{\mathbf{x}\geq\mathbf{0}}\{\mathbf{c}^T\mathbf{x} + \boldsymbol{\mu}^T(\mathbf{A}\mathbf{x} - \mathbf{b})\} \\ &= \min_{\mathbf{x}\geq\mathbf{0}}\{(\mathbf{c}^T + \boldsymbol{\mu}^T\mathbf{A})\mathbf{x} - \boldsymbol{\mu}^T\mathbf{b}\} \\ &= \min_{\mathbf{x}\geq\mathbf{0}}\{\textstyle\sum_i (c_i + (\boldsymbol{\mu}^T\mathbf{A})_i)\,\mathbf{x}_i - \sum_i \mu_i b_i\} \end{aligned}$$

- **note**: if, $c_i + (\boldsymbol{\mu}^T\mathbf{A})_i < 0, \forall i$, we can make $q(\boldsymbol{\mu}) = -\infty$ by making $\mathbf{x}_i$ arbitrarily large. So, to have a solution, we **need**

$$\boxed{c_i + (\boldsymbol{\mu}^T\mathbf{A})_i \geq 0, \forall i}$$

$\mathbf{x}^* = \arg\min_{\mathbf{x}\geq 0} \mathbf{c}^T\mathbf{x}$ subject to $\mathbf{A}\mathbf{x} - \mathbf{b} \leq 0$

Linear Programming

- $$q(\boldsymbol{\mu}) = \min_{\mathbf{x}\geq 0}\{\textstyle\sum_i [c_i + (\boldsymbol{\mu}^T\mathbf{A})_i]\,\mathbf{x}_i - \sum_i \mu_i b_i\}$$

  when $\boxed{c_i + (\boldsymbol{\mu}^T\mathbf{A})_i \geq 0, \forall i}$ the minimum is at $\mathbf{x}^* = \mathbf{0}$ and $\boxed{q^* = -\boldsymbol{\mu}^T\mathbf{b}}$

- this leads to

primal	dual
$\min_{\mathbf{x}} \mathbf{c}^T\mathbf{x}$	$\max_{\boldsymbol{\mu}}(-\boldsymbol{\mu}^T\mathbf{b})$
s.t. $\mathbf{A}\mathbf{x} \leq \mathbf{b}$	s.t. $\boldsymbol{\mu}^T\mathbf{A} \geq -\mathbf{c}^T$
$\mathbf{x} \geq \mathbf{0}$	$\boldsymbol{\mu} \geq \mathbf{0}$

switching to $-\mathbf{b}$ and $-\mathbf{A}$

primal	dual
$\min_{\mathbf{x}} \mathbf{c}^T\mathbf{x}$	$\max_{\boldsymbol{\mu}} \boldsymbol{\mu}^T\mathbf{b}$
s.t. $\mathbf{A}\mathbf{x} \geq \mathbf{b}$	s.t. $\boldsymbol{\mu}^T\mathbf{A} \leq \mathbf{c}^T$
$\mathbf{x} \geq \mathbf{0}$	$\boldsymbol{\mu} \geq \mathbf{0}$

- which is the standard form of duality for linear programming problems
- the dual can be obtained with a **simple recipe**

Linear Programming

primal	dual
$$\min_{\mathbf{x}} \mathbf{c}^T\mathbf{x}$$	$$\max_{\boldsymbol{\mu}} \boldsymbol{\mu}^T\mathbf{b}$$
s.t. $\mathbf{A}\mathbf{x} \geq \mathbf{b}$	s.t. $\boldsymbol{\mu}^T\mathbf{A} \leq \mathbf{c}^T$
$\mathbf{x} \geq \mathbf{0}$	$\boldsymbol{\mu} \geq \mathbf{0}$

- this gives us a recipe for primal to dual conversion for linear programming problems

1. interchange $\mathbf{x}$ with $\boldsymbol{\mu}^T$ and $\mathbf{b}$ with $\mathbf{c}^T$
2. reverse the constraint inequalities
3. maximize instead of minimizing

- the dual of a linear programming problem is **trivial** to obtain!

Linear Programming

- can be applied to <u>any</u> problem, e.g. with <u>equality constraints</u>

$$\min_{\mathbf{x}} \mathbf{c}^T\mathbf{x} \quad \text{s.t. } \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}$$

$$\Leftrightarrow \quad \min_{\mathbf{x}} \mathbf{c}^T\mathbf{x} \quad \text{s.t. } \mathbf{A}\mathbf{x} \geq \mathbf{b}, -\mathbf{A}\mathbf{x} \geq -\mathbf{b}, \mathbf{x} \geq \mathbf{0}$$

$$\Leftrightarrow \quad \min_{\mathbf{x}} \mathbf{c}^T\mathbf{x} \quad \text{s.t. } \begin{bmatrix} \mathbf{A} \\ -\mathbf{A} \end{bmatrix} \mathbf{x} \geq \begin{bmatrix} \mathbf{b} \\ -\mathbf{b} \end{bmatrix}, \quad \mathbf{x} \geq \mathbf{0}$$

1. interchange $\mathbf{x}$ with $\boldsymbol{\mu}^T$ and $\mathbf{b}$ with $\mathbf{c}^T$
2. reverse the inequalities
3. maximize

and the dual is

$$\max_{\boldsymbol{\mu}} \ (\boldsymbol{\mu}_1^T\mathbf{b} - \boldsymbol{\mu}_2^T\mathbf{b}) \quad \text{s.t. } [\boldsymbol{\mu}_1^T \quad \boldsymbol{\mu}_2^T] \begin{bmatrix} \mathbf{A} \\ -\mathbf{A} \end{bmatrix} \leq \mathbf{c}^T, \quad \boldsymbol{\mu}_1^T \geq \mathbf{0}, \ \boldsymbol{\mu}_2^T \geq \mathbf{0}$$

$$\Leftrightarrow \quad \max_{\boldsymbol{\mu}} \ (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{b} \quad \text{s.t. } (\boldsymbol{\mu}_1^T - \boldsymbol{\mu}_2^T)\mathbf{A} \leq \mathbf{c}^T, \quad \boldsymbol{\mu}_1^T \geq \mathbf{0}, \ \boldsymbol{\mu}_2^T \geq \mathbf{0}$$

$$\Leftrightarrow \quad \max_{\boldsymbol{\mu}} \ \boldsymbol{\mu}^T\boldsymbol{b} \quad \text{s.t. } \boldsymbol{\mu}^T\mathbf{A} \leq \mathbf{c}^T \quad (\boldsymbol{\mu} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$$

- this has a <u>nice</u> geometric interpretation

Linear Programming Example

1. interchange $\mathbf{x}$ with $\boldsymbol{\mu}^T$ and $\mathbf{b}$ with $\mathbf{c}^T$
2. reverse the inequalities
3. maximize

primal

$$\min_{\mathbf{x}} \ \mathbf{c}^T\mathbf{x}$$
$$\text{s.t.} \ \mathbf{A}\mathbf{x} = \mathbf{b}$$
$$\mathbf{x} \geq \mathbf{0}$$

dual

$$\max_{\boldsymbol{\mu}} \ \boldsymbol{\mu}^T\mathbf{b}$$
$$\text{s.t.} \ \boldsymbol{\mu}^T\mathbf{A} \leq \mathbf{c}^T$$

- for the example

$$\min_{\mathbf{x}} \ (12x_1 + 12x_2 + 2x_3 + 4x_4)$$
$$\text{s.t.} \quad 3x_1 + x_2 - 2x_3 + x_4 = 2$$
$$x_1 + 3x_2 - x_4 = 2$$
$$\mathbf{x} \geq \mathbf{0}$$

$$\mathbf{A} = \begin{bmatrix} 3 & 1 & -2 & 1 \\ 1 & 3 & 0 & -1 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \quad \mathbf{c} = \begin{bmatrix} 12 \\ 12 \\ 2 \\ 4 \end{bmatrix}$$

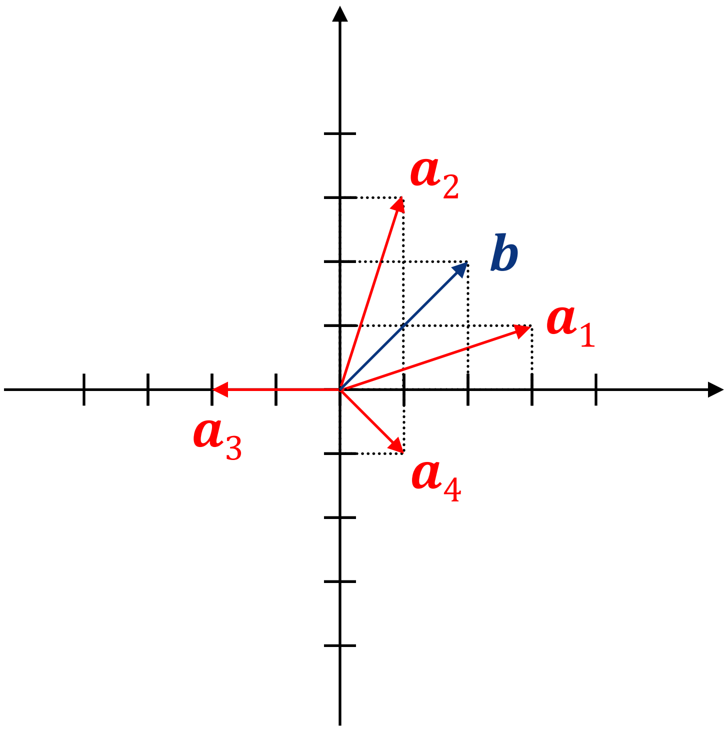
the dual is

$$\max_{\boldsymbol{\mu}} \ (2\mu_1 + 2\mu_2)$$
$$\text{s.t.} \quad 3\mu_1 + \mu_2 \leq 12$$
$$\mu_1 + 3\mu_2 \leq 12$$
$$-2\mu_1 \leq 2$$
$$\mu_1 - \mu_2 \leq 4$$

Linear Programming Example: Primal

- $$\begin{aligned} \min_{\mathbf{x}} \ & (12x_1 + 12x_2 + 2x_3 + 4x_4) \\ \text{s.t.} \quad & 3x_1 + x_2 - 2x_3 + x_4 = 2 \\ & x_1 + 3x_2 - x_4 = 2 \\ & \mathbf{x} \geq \mathbf{0} \end{aligned}$$

$\mathbf{A} = \begin{bmatrix} 3 & 1 & -2 & 1 \\ 1 & 3 & 0 & -1 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$



- solution is a linear combination of

$$\boldsymbol{a}_1 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \boldsymbol{a}_2 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \boldsymbol{a}_3 = \begin{bmatrix} -2 \\ 0 \end{bmatrix}, \boldsymbol{a}_4 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

that adds up to

$$\mathbf{b} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

- it is <u>not</u> obvious what it is
- **what about the dual?**

Linear Programming Example: Dual

- $$\max_{\boldsymbol{\mu}} \ (2\mu_1 + 2\mu_2)$$
  $$\text{s.t.} \quad 3\mu_1 + \mu_2 \le 12 \quad (1)$$
  $$\mu_1 + 3\mu_2 \le 12 \quad (2)$$
  $$-2\mu_1 \le 2 \quad (3)$$
  $$\mu_1 - \mu_2 \le 4 \quad (4)$$

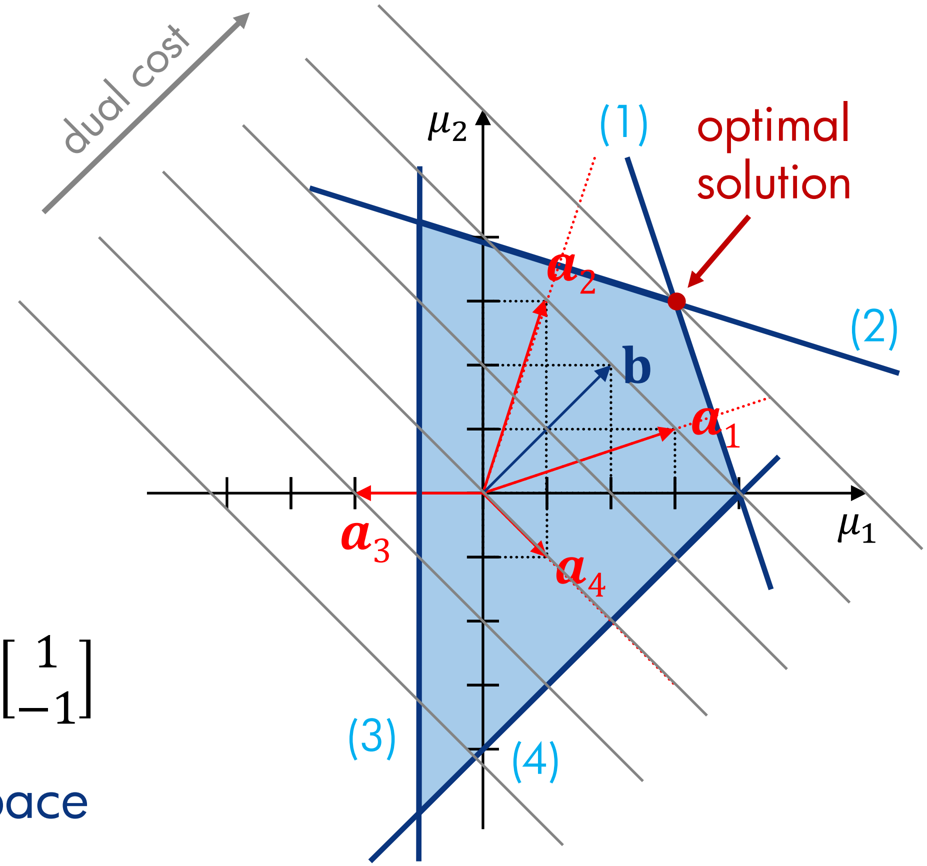
- vectors

  $$\boldsymbol{a}_1 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \boldsymbol{a}_2 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \boldsymbol{a}_3 = \begin{bmatrix} -2 \\ 0 \end{bmatrix}, \boldsymbol{a}_4 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

  are normal to planes in (μ_1, μ_2) space

- the bias of each plane is set by $\mathbf{c} = (12, 12, 2, 4)^T$ and defines a half−space where the solution must be

- solution can be obtained by **inspection**



Linear Programming Example: Dual

$$\max_{\boldsymbol{\mu}} \ (2\mu_1 + 2\mu_2)$$
$$\text{s.t.} \quad 3\mu_1 + \mu_2 \leq 12, \quad \mu_1 + 3\mu_2 \leq 12, \quad -2\mu_1 \leq 2, \quad \mu_1 - \mu_2 \leq 4$$

- noting that only constraints (1) and (2) are active

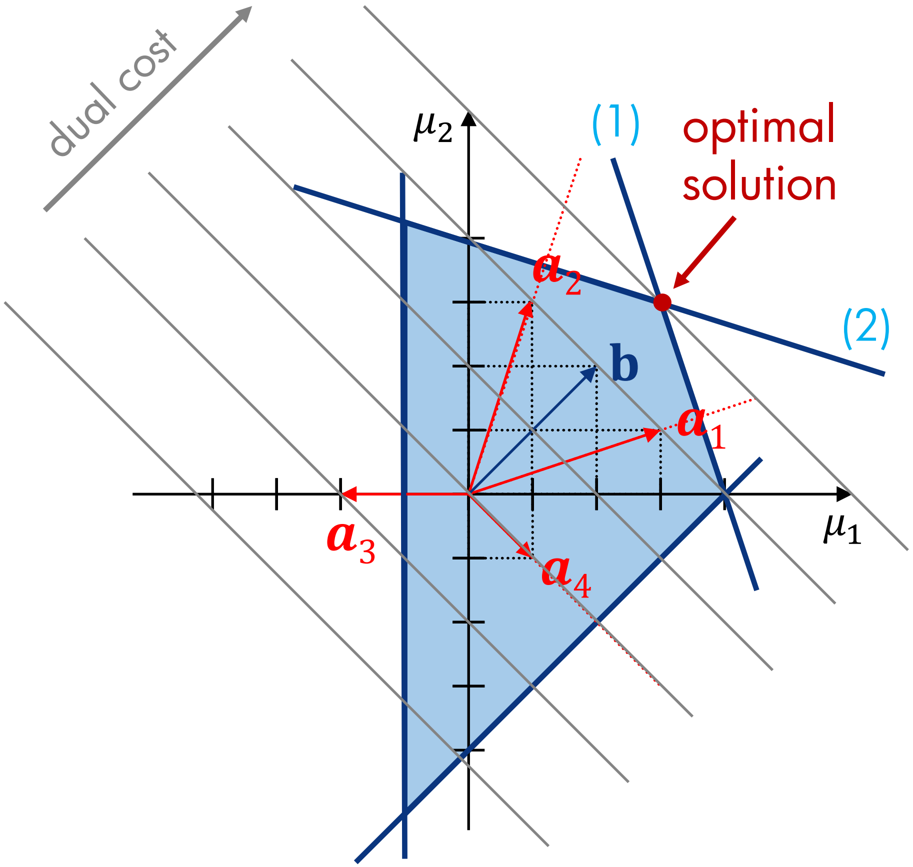
$$\begin{cases} 3\mu_1 + \mu_2 = 12 \\ \mu_1 + 3\mu_2 = 12 \end{cases} \Leftrightarrow \begin{cases} \mu_1 = 3 \\ \mu_2 = 3 \end{cases}$$

- hence, the dual problem

$$\begin{aligned} \max_{\boldsymbol{\mu}} \ & (2\mu_1 + 2\mu_2) \\ \text{s.t.} \quad & 3\mu_1 + \mu_2 \leq 12 \quad (1) \\ & \mu_1 + 3\mu_2 \leq 12 \quad (2) \\ & -2\mu_1 \leq 2 \quad (3) \\ & \mu_1 - \mu_2 \leq 4 \quad (4) \end{aligned}$$

has the same solution as

$$\begin{aligned} \max_{\boldsymbol{\mu}} \ & (2\mu_1 + 2\mu_2) \\ \text{s.t.} \quad & 3\mu_1 + \mu_2 \leq 12 \quad (1) \\ & \mu_1 + 3\mu_2 \leq 12 \quad (2) \end{aligned}$$



Linear Programming Example: Dual

- and using duality again

$$\mathbf{A} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \quad \mathbf{c} = \begin{bmatrix} 12 \\ 12 \end{bmatrix}$$

$$\min_{\mathbf{x}} \ \mathbf{c}^T\mathbf{x} \quad \text{s.t. } \mathbf{A}\mathbf{x} = \mathbf{b}, \ \mathbf{x} \geq \mathbf{0}$$

$$\min_{\mathbf{x}} \ (12x_1 + 12x_2) \quad \text{s.t.} \quad 3x_1 + x_2 = 2, \ x_1 + 3x_2 = 2, \ \mathbf{x} \geq \mathbf{0}$$

$$\max_{\boldsymbol{\mu}} \ (2\mu_1 + 2\mu_2) \quad \text{s.t.} \quad 3\mu_1 + \mu_2 \leq 12, \ \mu_1 + 3\mu_2 \leq 12$$

$$\max_{\boldsymbol{\mu}} \ \boldsymbol{\mu}^T\mathbf{b} \quad \text{s.t. } \boldsymbol{\mu}^T\mathbf{A} \leq \mathbf{c}^T$$

- the basis vectors for the primal solution are

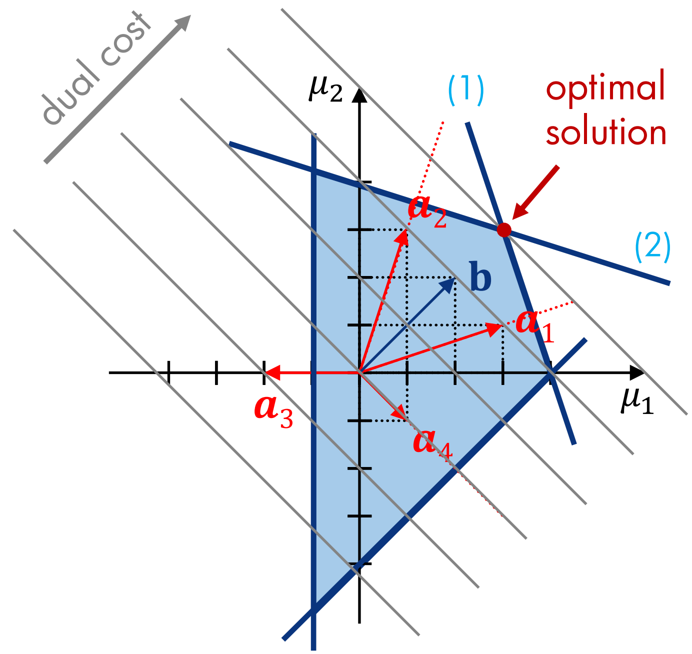
$$\boldsymbol{a}_1 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \boldsymbol{a}_2 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

and add to $\mathbf{b} = (2,2)^T$ when

$$x_1 = x_2 = 1/2$$

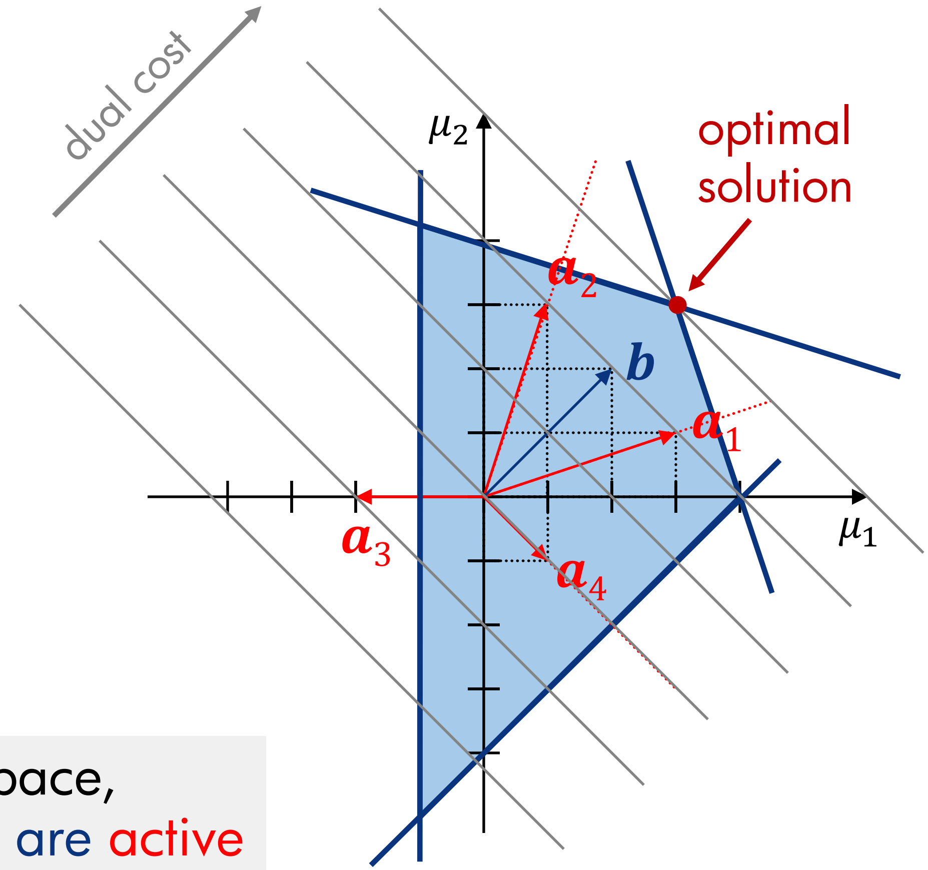
- hence, the optimal solution is

$$\mathbf{x}^* = (1/2\,, 1/2\,, 0,0)^T$$



Notes

- by **using the dual**
  1. we were able to solve the problem with **minimal** (none?) computation
  2. we quickly **identified** what constraints are active

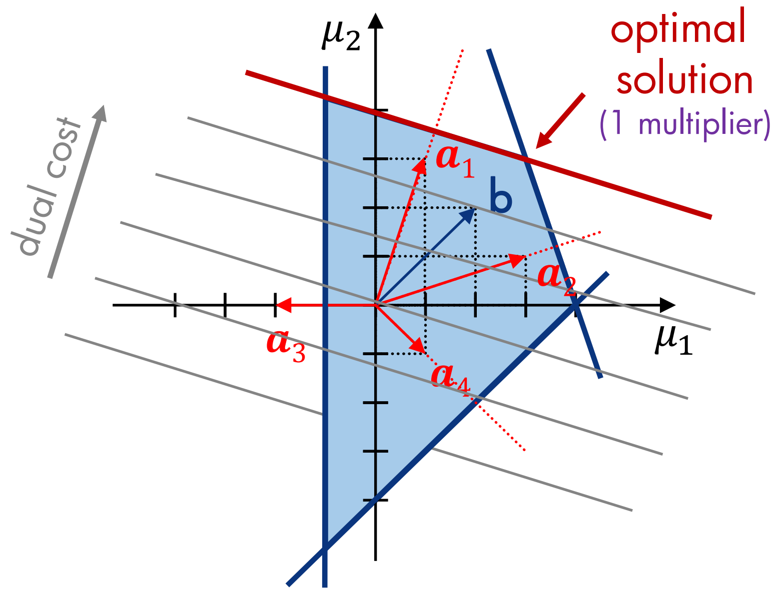


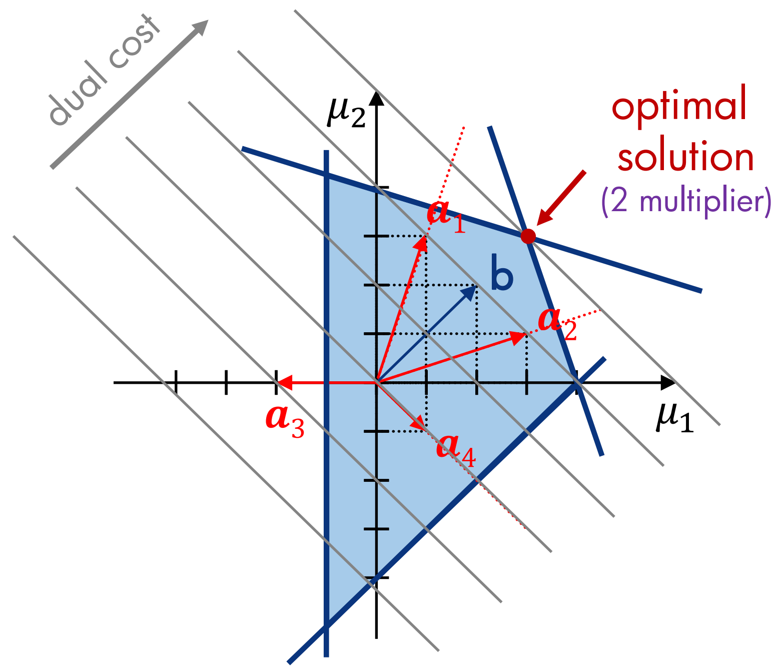
- property 2 is **always true**:
  - at any given region of the space, only a **few** of the constraints are active
  - by taking the remaining Lagrange multipliers to zero, the dual solution **automatically** identifies those

- property 1:
  - dual much **simpler** whenever # of constraints ≪ # variables

Notes

- in general, on <u>linear programming</u> problems, we can have
  1. solution is **one entire constraint** (1 multiplier)
  2. solution is at the **intersection of two constraints** (2 multipliers)
  3. **more multipliers** only if several constraints intersect at <u>single</u> point





Quadratic Programming

- consider the problem

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \left\{ \frac{1}{2} \mathbf{x}^T \mathbf{Q}\, \mathbf{x} - \mathbf{b}^T \mathbf{x} \right\} \quad \text{subject to} \quad \mathbf{A}\mathbf{x} \leq \mathbf{c}$$

  where $\mathbf{Q}$ is positive–definite

- this is a convex problem with linear constraints and has <u>no</u> duality gap
- the dual problem is

$$q^* = \max_{\boldsymbol{\alpha} \geq \mathbf{0}} \left\{ \min_{\mathbf{x}} \left[\frac{1}{2} \mathbf{x}^T \mathbf{Q}\, \mathbf{x} - \mathbf{b}^T \mathbf{x} + \boldsymbol{\alpha}^T (\mathbf{A}\mathbf{x} - \mathbf{c}) \right] \right\}$$

- setting gradient w.r.t. $\mathbf{x}$ to zero, we obtain

$$\mathbf{Q}\, \mathbf{x} - \mathbf{b} + \mathbf{A}^T \boldsymbol{\alpha} = \mathbf{0} \quad \Leftrightarrow \quad \mathbf{x} = \mathbf{Q}^{-1} (\mathbf{b} - \mathbf{A}^T \boldsymbol{\alpha})$$

Quadratic Programming

$$q^* = \max_{\boldsymbol{\alpha}\geq\mathbf{0}}\left\{\min_{\mathbf{x}}\left[\frac{1}{2}\mathbf{x}^T\mathbf{Q}\,\mathbf{x} - \mathbf{b}^T\mathbf{x} + \boldsymbol{\alpha}^T(\mathbf{A}\mathbf{x} - \mathbf{c})\right]\right\}$$

- and

$$\mathbf{x} = \mathbf{Q}^{-1}(\mathbf{b} - \mathbf{A}^T\boldsymbol{\alpha})$$

$$\mathbf{x}^T = (\mathbf{b}^T - \boldsymbol{\alpha}^T\mathbf{A})\mathbf{Q}^{-1}$$

$$q^* = \max_{\boldsymbol{\alpha}\geq\mathbf{0}}\left\{\min_{\mathbf{x}}\left[\frac{1}{2}\mathbf{x}^T\mathbf{Q}\,\mathbf{x} - \mathbf{b}^T\mathbf{x} + \boldsymbol{\alpha}^T(\mathbf{A}\mathbf{x} - \mathbf{c})\right]\right\}$$

$$= \max_{\boldsymbol{\alpha}\geq\mathbf{0}}\left\{\frac{1}{2}(\mathbf{b}^T - \boldsymbol{\alpha}^T\mathbf{A})\mathbf{Q}^{-1}\mathbf{Q}\mathbf{Q}^{-1}(\mathbf{b} - \mathbf{A}^T\boldsymbol{\alpha}) - \mathbf{b}^T\mathbf{Q}^{-1}(\mathbf{b} - \mathbf{A}^T\boldsymbol{\alpha}) + \boldsymbol{\alpha}^T(\mathbf{A}\mathbf{Q}^{-1}(\mathbf{b} - \mathbf{A}^T\boldsymbol{\alpha}) - \mathbf{c})\right\}$$

$$= \max_{\boldsymbol{\alpha}\geq\mathbf{0}}\left\{\frac{1}{2}(\mathbf{b}^T - \boldsymbol{\alpha}^T\mathbf{A})\mathbf{Q}^{-1}(\mathbf{b} - \mathbf{A}^T\boldsymbol{\alpha}) - \mathbf{b}^T\mathbf{Q}^{-1}(\mathbf{b} - \mathbf{A}^T\boldsymbol{\alpha}) + \boldsymbol{\alpha}^T\mathbf{A}\mathbf{Q}^{-1}(\mathbf{b} - \mathbf{A}^T\boldsymbol{\alpha}) - \boldsymbol{\alpha}^T\mathbf{c}\right\}$$

$$= \max_{\boldsymbol{\alpha}\geq\mathbf{0}}\left\{\frac{1}{2}(\mathbf{b}^T - \boldsymbol{\alpha}^T\mathbf{A})\mathbf{Q}^{-1}(\mathbf{b} - \mathbf{A}^T\boldsymbol{\alpha}) - (\mathbf{b}^T - \boldsymbol{\alpha}^T\mathbf{A})\mathbf{Q}^{-1}(\mathbf{b} - \mathbf{A}^T\boldsymbol{\alpha}) - \boldsymbol{\alpha}^T\mathbf{c}\right\}$$

$$= \max_{\boldsymbol{\alpha}\geq\mathbf{0}}\left\{-\frac{1}{2}(\mathbf{b}^T - \boldsymbol{\alpha}^T\mathbf{A})\mathbf{Q}^{-1}(\mathbf{b} - \mathbf{A}^T\boldsymbol{\alpha}) - \boldsymbol{\alpha}^T\mathbf{c}\right\}$$

$$= \max_{\boldsymbol{\alpha}\geq\mathbf{0}}\left\{-\frac{1}{2}\boldsymbol{\alpha}^T\mathbf{A}\mathbf{Q}^{-1}\mathbf{A}^T\boldsymbol{\alpha} + \boldsymbol{\alpha}^T\mathbf{A}\mathbf{Q}^{-1}\mathbf{b} - \boldsymbol{\alpha}^T\mathbf{c}\right\}$$

$$= \max_{\boldsymbol{\alpha}\geq\mathbf{0}}\left\{-\frac{1}{2}\boldsymbol{\alpha}^T\mathbf{A}\mathbf{Q}^{-1}\mathbf{A}^T\boldsymbol{\alpha} + \boldsymbol{\alpha}^T(\mathbf{A}\mathbf{Q}^{-1}\mathbf{b} - \mathbf{c})\right\}$$

Quadratic Programming

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \left\{ \frac{1}{2}\mathbf{x}^T\mathbf{Q}\,\mathbf{x} - \mathbf{b}^T\mathbf{x} \right\} \quad \text{subject to} \quad \mathbf{Ax} \le \mathbf{c}$$

$$q^* = \max_{\boldsymbol{\alpha}\ge\mathbf{0}} \left\{ -\frac{1}{2}\boldsymbol{\alpha}^T\mathrm{AQ}^{-1}\mathrm{A}^T\boldsymbol{\alpha} + \boldsymbol{\alpha}^T(\mathrm{AQ}^{-1}\mathrm{b} - \mathrm{c}) \right\}$$

- hence, the dual problem is of the form

$$q^* = \max_{\alpha\ge\mathbf{0}} \left\{ -\frac{1}{2}\boldsymbol{\alpha}^T\mathbf{P}\boldsymbol{\alpha} + \boldsymbol{\alpha}^T\mathbf{d} \right\}$$

with

$$\mathbf{P} = \mathbf{AQ}^{-1}\mathbf{A}^T$$
$$\mathbf{d} = \mathbf{AQ}^{-1}\boldsymbol{b} - \mathbf{c}$$

- note that, like the primal, the dual is a quadratic problem
- the **advantage** is that the constraints are now much simpler
- this is the optimization problem defined by the support vector machine
- next class will talk more about this