# Duality

## ME598/494 Lecture

Max Yi Ren

Department of Mechanical Engineering, Arizona State University

November 22, 2017

# Outline

# Constrained optimization

Consider the constrained optimization problem:

$$\begin{aligned} \max_{\mathbf{x}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{g}(\mathbf{x}) \le \mathbf{0}. \end{aligned} \tag{1}$$

This is equivalent to the following unconstrained problem:

$$\begin{aligned} J(\mathbf{x}) &= \begin{cases} f(\mathbf{x}), & \text{if } \mathbf{g}(\mathbf{x}) \le \mathbf{0} \\ \infty, & \text{otherwise} \end{cases} \\ &= f(\mathbf{x}) + \sum_i I(g_i(\mathbf{x})), \end{aligned} \tag{2}$$

where $I(u)$ is a infinite step function:

$$I(u) = \begin{cases} 0, & \text{if } u \le 0 \\ \infty, & \text{otherwise} \end{cases} \tag{3}$$

But $I(u)$ is non-differentiable and discontinuous.

# Lagrangian

So let us replace $I(u)$ with something nicer, say, $\mu u$ for $\mu \geq 0$. Notice that $\mu u$ is a lower bound on $I(u)$.

This way we get the Lagrange function (Lagrangian):

$$L(\mathbf{x}, \boldsymbol{\mu}) = f(\mathbf{x}) + \boldsymbol{\mu}^T \mathbf{g}(\mathbf{x}). \tag{4}$$

We have

$$\max_{\boldsymbol{\mu} \geq \mathbf{0}} L(\mathbf{x}, \boldsymbol{\mu}) = J(\mathbf{x}). \tag{5}$$

To solve our original problem, we need to solve:

$$\min_{\mathbf{x}} \max_{\boldsymbol{\mu} \geq \mathbf{0}} L(\mathbf{x}, \boldsymbol{\mu}) \tag{6}$$

This problem is hard if our original problem is so. But let us look at another problem:

$$\max_{\boldsymbol{\mu} \geq \mathbf{0}} \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\mu}) = \max_{\boldsymbol{\mu} \geq \mathbf{0}} \Phi(\boldsymbol{\mu}), \tag{7}$$

where $\Phi(\boldsymbol{\mu}) = \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\mu})$. $\Phi(\boldsymbol{\mu})$ is a concave function (the pointwise minimum of affine functions is concave).

# Duality

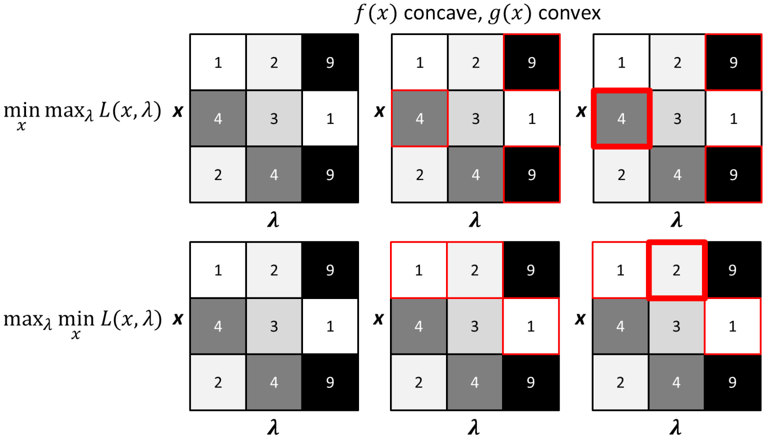



Figure: $\min_{\mathbf{x}} \max_{\boldsymbol{\mu} \geq 0} L(\mathbf{x}, \boldsymbol{\mu})$ and $\max_{\boldsymbol{\mu} \geq 0} \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\mu})$ may have different solutions

# Duality

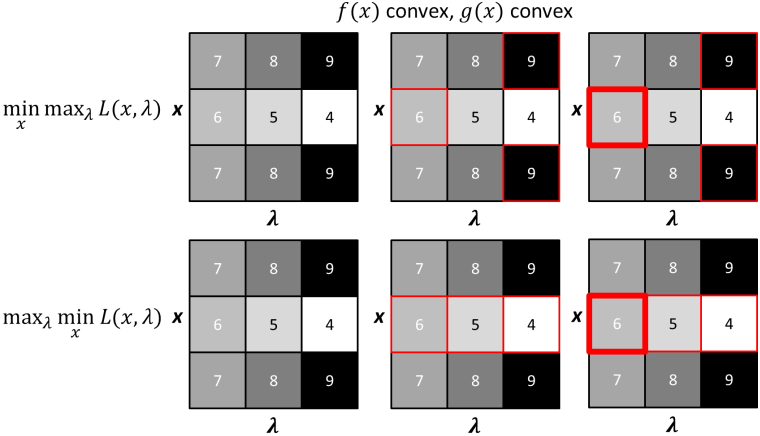



Figure: For convex problems, $\min_{\mathbf{x}} \max_{\boldsymbol{\mu} \geq 0} L(\mathbf{x}, \boldsymbol{\mu})$ and $\max_{\boldsymbol{\mu} \geq 0} \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\mu})$ have the same solutions

# Weak duality

Since $\Phi(\boldsymbol{\mu})$ is concave and $\boldsymbol{\mu} \geq \mathbf{0}$ are linear constraints, maximizing $\Phi(\boldsymbol{\mu})$ is a convex optimization problem, easy! But how does solving this problem help our original problem?

Recall that $\mu u$ is a lower bound on $I(u)$, and thus $L(\mathbf{x}, \boldsymbol{\mu})$ is a lower bound on $J(\mathbf{x})$ for all $\boldsymbol{\mu} \geq \mathbf{0}$. Thus

$$
\begin{aligned}
L(\mathbf{x}, \boldsymbol{\mu}) \leq J(\mathbf{x}) \to \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\mu}) = \mathbf{g}(\boldsymbol{\mu}) \leq \min_{\mathbf{x}} J(\mathbf{x}) = p^* \\
\to d^* = \max_{\boldsymbol{\mu}} \mathbf{g}(\boldsymbol{\mu}) \leq p^*,
\end{aligned}
\tag{8}
$$

where $d^*$ and $p^*$ are the optima of the dual and primal problems respectively.

**Weak duality**: $d^* \leq p^*$.

Duality gives us a relatively easy way to find a lower bound on a hard minimization problem.

# Example

Consider the least-norm solution of linear equations:

$$\begin{aligned} \max_{\mathbf{x}} \quad & \mathbf{x}^T\mathbf{x} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b}. \end{aligned} \tag{9}$$

- Lagrangian is $L(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{x}^T\mathbf{x} + \boldsymbol{\lambda}^T(\mathbf{A}\mathbf{x} - \mathbf{b})$
- $\nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}) = 2\mathbf{x} + \mathbf{A}^T\boldsymbol{\lambda} = \mathbf{0} \to \mathbf{x} = -0.5\mathbf{A}^T\boldsymbol{\lambda}$
- $\Phi(\boldsymbol{\lambda}) = -\frac{1}{4}\boldsymbol{\lambda}^T\mathbf{A}\mathbf{A}^T\boldsymbol{\lambda} - \mathbf{b}^T\boldsymbol{\lambda}$ (concave!)
- solution $\mathbf{x}^* = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{b}$, $\boldsymbol{\lambda}^* = -2(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{b}$

If $\mathbf{A}$ is large, calculating $\mathbf{y} = (\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{b}$ (solving $\mathbf{A}\mathbf{A}^T\mathbf{y} = \mathbf{b}$) requires a cost iteration. When to stop? One can calculate the gap $\mathbf{x}^T\mathbf{x} - \Phi(\boldsymbol{\lambda})$ to see how far away is the current solution. $\boldsymbol{\lambda}^*$ is called a certificate of the optimal solution.

# Strong duality

For convex problems that satisfy the Slater's constraint qualification, strong duality holds, i.e., $d^* = p^*$.

Slater's constraint qualification: There exists $\mathbf{x}$ that is strictly feasible, i.e., $\mathbf{g}(\mathbf{x}) < \mathbf{0}$ and $\mathbf{h}(\mathbf{x}) = \mathbf{0}$.

Example: The following problem is nonconvex for non-positive definite $\mathbf{A}$:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{x}^T \mathbf{A} \mathbf{x} + 2\mathbf{b}^T \mathbf{x} \\ \text{subject to} \quad & \mathbf{x}^T \mathbf{x} \leq 1. \end{aligned} \tag{10}$$

But it has strong duality, thus one can solve the following equivalent convex problem:

$$\begin{aligned} \max_{t,\lambda} \quad & -t - \lambda \\ \text{subject to} \quad & [\mathbf{A} + \lambda \mathbf{I}, \mathbf{b}; \mathbf{b}^T, t] \succeq 0. \end{aligned} \tag{11}$$

# Economic interpretation of duality

Consider the LP problem:

$$
\begin{aligned}
\max_{x_1,\cdot,x_n} \quad & c_1x_1 + \cdots + c_nx_n \\
\text{subject to} \quad & a_{11}x_1 + \cdots + a_{1n}x_n \leq b_1 \\
& \qquad \ldots \\
& a_{m1}x_1 + \cdots + a_{mn}x_n \leq b_m \\
& x_1, \cdots, x_n \geq 0.
\end{aligned}
\tag{12}
$$

The economic interpretation is:

- $n$ economic activities, $m$ resources
- $c_j$ is revenue per unit of activity $j$
- $b_i$ is maximum availability of resource $i$
- $a_ij$ is consumption of resource $i$ per unit of activity $j$

# Lagrangian of LP

The Lagrange function of the LP is:

$$
\begin{aligned}
L(\mathbf{x}, \boldsymbol{\mu}, \tilde{\boldsymbol{\mu}}) &= \mathbf{c}^T \mathbf{x} + \boldsymbol{\mu}^T (\mathbf{A}\mathbf{x} - \mathbf{b}) - \tilde{\boldsymbol{\mu}}^T \mathbf{x} \\
&= -\boldsymbol{\mu}^T \mathbf{b} + (\mathbf{c}^T + \boldsymbol{\mu}_1^T \mathbf{A} - \tilde{\boldsymbol{\mu}}^T)\mathbf{x}
\end{aligned}
\tag{13}
$$

The dual function is

$$
\begin{aligned}
\Phi(\boldsymbol{\mu}, \tilde{\boldsymbol{\mu}}) &= \inf_{\mathbf{x} \geq \mathbf{0}} -\boldsymbol{\mu}^T \mathbf{b} + (\mathbf{c}^T + \boldsymbol{\mu}^T \mathbf{A} - \tilde{\boldsymbol{\mu}}^T)\mathbf{x} \\
&= \begin{cases} -\boldsymbol{\mu}^T \mathbf{b} & \mathbf{c} + \mathbf{A}^T \boldsymbol{\mu} - \tilde{\boldsymbol{\mu}} = 0 \\ -\infty & \text{otherwise} \end{cases}
\end{aligned}
\tag{14}
$$

Note that $\Phi(\boldsymbol{\mu}, \tilde{\boldsymbol{\mu}})$ is concave.

# The dual problem

We can formulate the dual problem as

$$\begin{aligned} \max_{\boldsymbol{\mu}} \quad & -\mathbf{b}^T\boldsymbol{\mu} \\ \text{subject to:} \quad & \mathbf{A}^T\boldsymbol{\mu} + \mathbf{c} \geq \mathbf{0} \\ & \boldsymbol{\mu} \geq \mathbf{0} \end{aligned} \tag{15}$$

The economic interpretation of the dual **variables** is as follows:

- Let $\mathbf{x}^*$ be the optimal for the primal, and $\boldsymbol{\mu}^*$ be the optimal for the dual, then $\mathbf{c}^T\mathbf{x} = -\mathbf{b}^T\boldsymbol{\mu}$
- $-\mathbf{c}^T\mathbf{x}$ is the maximal revenue, $\mathbf{b}^T\boldsymbol{\mu}$ is the summation of "availability of resource" times "revenue per unit of resource".
- $-\boldsymbol{\mu}$ is the *dual price of resource*.
- For any non-optimal $x$ and $\boldsymbol{\mu}$, we have $-\mathbf{c}^T\mathbf{x} < \mathbf{b}^T\boldsymbol{\mu}$, i.e., at non-optimal solutions, the profit does not fully utilize the resources.

# The dual problem

We can formulate the dual problem as

$$\begin{aligned} \max_{\boldsymbol{\mu}} \quad & -\mathbf{b}^T\boldsymbol{\mu} \\ \text{subject to:} \quad & \mathbf{A}^T\boldsymbol{\mu} + \mathbf{c} \geq \mathbf{0} \\ & \boldsymbol{\mu} \geq \mathbf{0} \end{aligned} \tag{16}$$

The economic interpretation of the dual **constraints** is as follows:

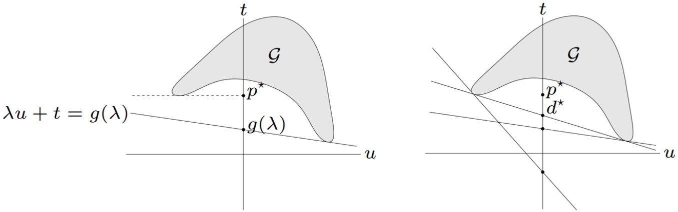
- Each row $j$ of $\mathbf{A}^T\boldsymbol{\mu} + \mathbf{c} \geq \mathbf{0}$ says that at any solution, the profit of activity $j$ should be lower than the summation of the resource price for that activity.
- At the optimal solution, the two should match.
- If row $j$ of $\mathbf{A}^T\boldsymbol{\mu} + \mathbf{c}$ is less than $\mathbf{0}$, then we should do more of activity $j$.

# Geometric interpretation

For simplicity, consider a problem with a single inequality constraint $g(\mathbf{x}) \leq 0$.

Interpretation of dual function:

$$\Phi(\mu) = \inf_{(u,t)\in\mathcal{G}} (t + \mu u), \text{ where } \mathcal{G} = \{(f(\mathbf{x}), g(\mathbf{x}))|\mathbf{x} \in \mathcal{D}\} \tag{17}$$




- $\mu u + t = \Phi(\mu)$ is supporting hyperplane to $\mathcal{G}$.
- hyperplane intersects $t$-axis at $t = \Phi(\mu)$.
- for convex problems, $\mathcal{G}$ is convex.

# Examples - Support Vector Machine

SVM: Find the maximum margin hyperplane that cuts between data with labels $y = 1$ and $y = -1$.

The primal problem (simplified):

$$\begin{aligned} \min_{\mathbf{w},b} \quad & \frac{1}{2}\mathbf{w}^T\mathbf{w} \\ \text{subject to:} \quad & (\mathbf{w}^T\mathbf{x}_i + b)y_i \geq 1, \forall i. \end{aligned} \tag{18}$$

The dual problem

$$\begin{aligned} \max_{\boldsymbol{\alpha}\geq\mathbf{0}} \quad & -\frac{1}{2}\boldsymbol{\alpha}\mathbf{X}\mathbf{X}^T\boldsymbol{\alpha} + \mathbf{1}^T\boldsymbol{\alpha} \\ \text{subject to:} \quad & \mathbf{y}^T\boldsymbol{\alpha} = 0. \end{aligned} \tag{19}$$

The solution only depends on the inner products $\mathbf{X}\mathbf{X}^T$. We can introduce other inner products instead of the vector product.

Algorithms based on the dual problem can solve the SVM problem faster when sample size is smaller than the number of features.