CS 2601 Linear and Convex Optimization

14. Dual LP

Bo Jiang

John Hopcroft Center for Computer Science
Shanghai Jiao Tong University

Fall 2022

Outline

- Dual LP
- Interpretation of dual problems
- Weak and strong duality
- Dual LP via Lagragian

Lower bounds in LP

$$\begin{aligned}
\min_{\boldsymbol{x}\in\mathbb{R}^2} \quad & f(\boldsymbol{x}) = x_1 + 2x_2 \\
\text{s.t.} \quad & 2x_1 + x_2 \geq 2 \\
& x_1, x_2 \geq 0
\end{aligned}$$

Any feasible solution $\boldsymbol{x}_0$ gives an upper bound on the optimal value f^*,

$$f^* \leq f(\boldsymbol{x}_0)$$

Can we also get a lower bound f_{LB} on f^*?

$$f^* \geq f_{\text{LB}}$$

Note. A lower bound on f^* is the same as a lower bound on $f(\boldsymbol{x})$ for all feasible $\boldsymbol{x} \in X$, i.e.

$$f^* \geq f_{\text{LB}} \iff f(\boldsymbol{x}) \geq f_{\text{LB}}, \quad \forall \boldsymbol{x} \in X$$

Lower bounds in LP (cont'd)

For any $\mu_1, \mu_2, \mu_3 \geq 0$,

$$
\begin{array}{rlrcrcrl}
\mu_1 \times & [& 2x_1 & + & x_2 & \geq & 2 &] \\
\mu_2 \times & [& x_1 & & & \geq & 0 &] \\
\mu_3 \times & [& & & x_2 & \geq & 0 &] \\
\hline
& & (2\mu_1 + \mu_2)x_1 & + & (\mu_1 + \mu_3)x_2 & \geq & 2\mu_1 =: \psi(\boldsymbol{\mu}) &
\end{array}
$$

We can set $2\mu_1 + \mu_2 = 1$ and $\mu_1 + \mu_3 = 2$ so the LHS becomes f.

Thus

$$f(\boldsymbol{x}) \geq \psi(\boldsymbol{\mu}) = 2\mu_1$$

for any $\boldsymbol{x} \in X$ and any μ_1, μ_2, μ_3 s.t.

$$2\mu_1 + \mu_2 = 1, \quad \mu_1 + \mu_3 = 2, \quad \mu_1, \mu_2, \mu_3 \geq 0$$

In particular, $f^* = \inf_{\boldsymbol{x} \in X} f(\boldsymbol{x}) \geq \psi(\boldsymbol{\mu})$ for such $\boldsymbol{\mu}$.

Lower bounds in LP (cont'd)

The quality of the lower bound $\psi(\boldsymbol{\mu}) = 2\mu_1$ varies for different $\boldsymbol{\mu}$.

- $\psi(0,1,2) = 0$, so

$$0 = \psi(0,1,2) \leq f^* \leq f(1,0) = 1$$

  which also tells us $\boldsymbol{x}_0 = (1,0)^T$ is 1-suboptimal, i.e.

$$f(1,0) - f^* \leq 1$$

- $\psi(\frac{1}{2}, 0, \frac{3}{2}) = 1$, so

$$1 = \psi(\frac{1}{2}, 0, \frac{3}{2}) \leq f^* \leq f(1,0) = 1$$

  which tells us $f^* = 1$ and $\boldsymbol{x}_0 = (1,0)^T$ is the optimal solution!

Dual LP

To get the best lower bound, we maximize over μ_1, μ_2, μ_3,

$$\begin{aligned} \min_{\boldsymbol{x} \in \mathbb{R}^2} \quad & f(\boldsymbol{x}) = x_1 + 2x_2 \\ \text{s.t.} \quad & 2x_1 + x_2 \geq 2 \\ & x_1 \geq 0 \\ & x_2 \geq 0 \end{aligned}$$

primal LP

$$\begin{aligned} \max_{\boldsymbol{\mu} \in \mathbb{R}^3} \quad & \psi(\boldsymbol{\mu}) = 2\mu_1 \\ \text{s.t.} \quad & 2\mu_1 + \mu_2 = 1 \\ & \mu_1 + \mu_3 = 2 \\ & \mu_1 \geq 0 \\ & \mu_2 \geq 0 \\ & \mu_3 \geq 0 \end{aligned}$$

dual LP

The variables μ_1, μ_2, μ_3 are called dual variables.

We have one dual variable for each constraint in the primal problem.

Dual LP (cont'd)

Given $\boldsymbol{c} \in \mathbb{R}^n, \boldsymbol{A} \in \mathbb{R}^{k\times n}, \boldsymbol{b} \in \mathbb{R}^k, \boldsymbol{G} \in \mathbb{R}^{m\times n}, \boldsymbol{h} \in \mathbb{R}^m$, consider

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) = \boldsymbol{c}^T\boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & \boldsymbol{G}\boldsymbol{x} \geq \boldsymbol{h} \end{aligned}$$

For $\boldsymbol{\lambda} \in \mathbb{R}^k$, $\boldsymbol{\mu} \in \mathbb{R}^m$ and $\boldsymbol{\mu} \geq \mathbf{0}$,

$$\boldsymbol{\lambda}^T\boldsymbol{A}\boldsymbol{x} + \boldsymbol{\mu}^T\boldsymbol{G}\boldsymbol{x} \geq \boldsymbol{\lambda}^T\boldsymbol{b} + \boldsymbol{\mu}^T\boldsymbol{h} =: \psi(\boldsymbol{\lambda}, \boldsymbol{\mu})$$

If $\boldsymbol{A}^T\boldsymbol{\lambda} + \boldsymbol{G}^T\boldsymbol{\mu} = \boldsymbol{c}$, then we can lower bound f^* by $f^* \geq \psi(\boldsymbol{\lambda}, \boldsymbol{\mu})$.

To maximize the lower bound, solve the following dual LP

$$\begin{aligned} \max_{\boldsymbol{\lambda}, \boldsymbol{\mu}} \quad & \psi(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \boldsymbol{b}^T\boldsymbol{\lambda} + \boldsymbol{h}^T\boldsymbol{\mu} \\ \text{s.t.} \quad & \boldsymbol{A}^T\boldsymbol{\lambda} + \boldsymbol{G}^T\boldsymbol{\mu} = \boldsymbol{c} \\ & \boldsymbol{\mu} \geq \mathbf{0} \end{aligned}$$

Dual LP (cont'd)

It is common to eliminate dual variables corresponding to $\boldsymbol{x} \geq \mathbf{0}$, and call the result the dual LP. Here are some common forms of dual LP,

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \boldsymbol{c}^T \boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & \boldsymbol{x} \geq \mathbf{0} \end{aligned}$$	$$\begin{aligned} \max_{\boldsymbol{y}} \quad & \boldsymbol{b}^T \boldsymbol{y} \\ \text{s.t.} \quad & \boldsymbol{A}^T \boldsymbol{y} \leq \boldsymbol{c} \end{aligned}$$
$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \boldsymbol{c}^T \boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{b} \end{aligned}$$	$$\begin{aligned} \max_{\boldsymbol{y}} \quad & \boldsymbol{b}^T \boldsymbol{y} \\ \text{s.t.} \quad & \boldsymbol{A}^T \boldsymbol{y} = \boldsymbol{c} \\ & \boldsymbol{y} \geq \mathbf{0} \end{aligned}$$
$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \boldsymbol{c}^T \boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{b} \\ & \boldsymbol{x} \geq \mathbf{0} \end{aligned}$$	$$\begin{aligned} \max_{\boldsymbol{y}} \quad & \boldsymbol{b}^T \boldsymbol{y} \\ \text{s.t.} \quad & \boldsymbol{A}^T \boldsymbol{y} \leq \boldsymbol{c} \\ & \boldsymbol{y} \geq \mathbf{0} \end{aligned}$$

Outline

- Dual LP
- Interpretation of dual problems
- Weak and strong duality
- Dual LP via Lagragian

Manufacturing problem

Suppose company A can manufacture n products from m materials.

- Manufacturing one unit of product i needs a_{ij} units of material j
- The unit price of product i is b_i
- The inventory of material j is c_j

where $a_{ij} > 0$, $b_i > 0$, $c_j > 0$.

Let y_i denote the amount of product i manufactured. To maximize its revenue, the company solves the following LP,

$$
\begin{aligned}
\max_{\boldsymbol{y}} \quad & \sum_{i=1}^{n} b_i y_i \\
\text{s.t.} \quad & \sum_{i=1}^{n} y_i a_{ij} \le c_j, \forall j = 1, 2, \ldots, m \\
& y_i \ge 0, \quad i = 1, 2, \ldots, n
\end{aligned}
$$

Manufacturing problem (cont'd)

Now suppose company B offers to buy the raw materials at the price of x_j per unit of material j.

The equivalent offer for one unit of product i is $\sum_{j=1}^{m} a_{ij}x_j$. Company A will accept the offer only if

$$\sum_{j=1}^{m} a_{ij}x_j \geq b_i, \quad i = 1, 2, \ldots, n$$

To minimizes its cost, company B solves the dual LP,

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \sum_{j=1}^{m} c_j x_j \\ \text{s.t.} \quad & \sum_{j=1}^{m} a_{ij}x_j \geq b_i, \forall i = 1, 2, \ldots, n \\ & x_j \geq 0, \quad j = 1, 2, \ldots, m \end{aligned}$$

Optimal transport problem

Recall the optimal transport problem in §1,

$$
\begin{aligned}
\min_{(x_{ij})} \quad & \sum_{i=1}^{n} \sum_{j=1}^{m} c_{ij} x_{ij} \\
\text{s.t.} \quad & \sum_{i=1}^{n} x_{ij} = b_j \qquad \text{for} \quad j = 1, 2, \ldots, m \\
& \sum_{j=1}^{m} x_{ij} \le a_i \qquad \text{for} \quad i = 1, 2, \ldots, n \\
& x_{ij} \ge 0 \qquad\qquad \text{for} \quad i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, m
\end{aligned}
$$

- x_{ij}: quantity shipped from warehouse i to customer j
- c_{ij}: unit shipping cost from warehouse i to customer j
- a_i: inventory at warehouse i
- b_j: demand at customer j

Optimal transport problem (cont'd)

Now instead of actually shipping the products, you decide to fulfill the orders by trading with another seller of the same product, who

- buys your stock at warehouse i at the unit price μ_i
- delivers the order of customer j at the unit price λ_j

The cost of "sending" one unit from i to j is $\lambda_j - \mu_i$. The deal will be competitive if

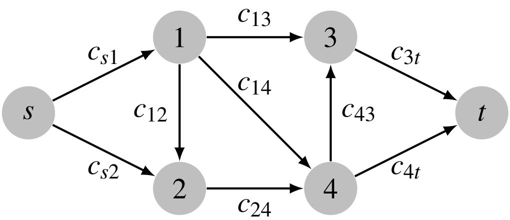
$$\lambda_j - \mu_i \leq c_{ij}$$

To maximize his profit, the other seller solves the dual LP,

$$\begin{aligned}
\max_{(\lambda_j),(\mu_i)} \quad & \sum_{j=1}^{m} b_j \lambda_j - \sum_{i=1}^{n} a_i \mu_i \\
\text{s.t.} \quad & \lambda_j - \mu_i \leq c_{ij} && \text{for} \quad i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, m \\
& \mu_i \geq 0 && \text{for} \quad i = 1, 2, \ldots, n
\end{aligned}$$

Maximum flow problem

Consider a directed graph $G = (V, E)$, each edge $(i, j) \in E$ of which has a capacity $c_{ij} > 0$. There are two special nodes: a source $s \in V$ and a sink $t \in V$, $s \neq t$. Assume G is acyclic and $(s, t) \notin E$ for simplicity.



A flow $f = \{f_{ij} : (i, j) \in E\}$ on G is an assignment of weights to edges that satisfies

- capacity constraint: $0 \leq f_{ij} \leq c_{ij}$, $\forall (i, j) \in E$
- flow conservation:

$$\sum_{(i,k)\in E} f_{ik} = \sum_{(k,j)\in E} f_{kj}, \quad \forall k \in V \setminus \{s, t\}$$

Maximum flow problem (cont'd)

Max-flow problem. Maximize the value $|f|$ of flow f,

$$\begin{aligned}
\max_{f} \quad & |f| \triangleq \sum_{(s,j)\in E} f_{sj} \\
\text{s. t.} \quad & 0 \le f_{ij} \le c_{ij}, && \forall (i,j) \in E \\
& \sum_{(i,k)\in E} f_{ik} = \sum_{(k,j)\in E} f_{kj}, && \forall k \in V \setminus \{s,t\}
\end{aligned}$$

To find the dual, introduce $a_{ij}, b_{ij} \ge 0$ for $(i,j) \in E$, x_k for $k \in V \setminus \{s,t\}$,

$$\begin{array}{llrcll}
a_{ij}\times & [& -f_{ij} & \le & 0 &] \quad \forall (i,j) \in E \\
b_{ij}\times & [& f_{ij} & \le & c_{ij} &] \quad \forall (i,j) \in E \\
x_k\times & [& \sum_{(i,k)\in E} f_{ik} - \sum_{(k,j)\in E} f_{kj} & = & 0 &] \quad \forall k \in V \setminus \{s,t\} \\
\hline
\end{array}$$

$$\sum_{(i,j)\in E} (-a_{ij}f_{ij} + b_{ij}f_{ij}) + \sum_{k\in V\setminus\{s,t\}} x_k \left(\sum_{(i,k)\in E} f_{ik} - \sum_{(k,j)\in E} f_{kj} \right) \le \sum_{(i,j)\in E} c_{ij}b_{ij}$$

Maximum flow problem (cont'd)

Renaming the dummy variable k to i and j,

$$\sum_{(i,j)\in E}(b_{ij}f_{ij} - a_{ij}f_{ij}) + \sum_{\substack{(i,j)\in E\\ j\neq t}} x_j f_{ij} - \sum_{\substack{(i,j)\in E\\ i\neq s}} x_i f_{ij} \leq \sum_{(i,j)\in E} c_{ij}b_{ij}$$

Matching the coefficients for f_{ij} in the objective,

$$\begin{array}{llllll} b_{sj} - a_{sj} & +x_j & & = & 1 & \forall (s,j) \in E \\ b_{it} - a_{it} & & -x_i & = & 0 & \forall (i,t) \in E \\ b_{ij} - a_{ij} & +x_j & -x_i & = & 0 & \forall (i,j) \in E,\, i \neq s, j \neq t \end{array}$$

and defining $x_s = 1$, $x_t = 0$, we obtain the dual LP,

$$\begin{aligned} \min_{a,b,x} \quad & \sum_{(i,j)\in E} c_{ij}b_{ij} \\ \text{s.t.} \quad & b_{ij} - a_{ij} + x_j - x_i = 0, \quad && \forall (i,j) \in E \\ & a_{ij} \geq 0, b_{ij} \geq 0, && \forall (i,j) \in E \end{aligned}$$

Maximum flow problem (cont'd)

Partial minimization over all a_{ij} yields the equivalent dual LP,

$$\begin{aligned}
\min_{b,x} \quad & \sum_{(i,j)\in E} c_{ij} b_{ij} \\
\text{s.t.} \quad & b_{ij} + x_j - x_i \geq 0, && \forall (i,j) \in E \\
& b_{ij} \geq 0, && \forall (i,j) \in E
\end{aligned}$$

This is a relaxation of the following integer programming (IP) problem

$$\begin{aligned}
\min_{b,x} \quad & \sum_{(i,j)\in E} c_{ij} b_{ij} \\
\text{s.t.} \quad & b_{ij} + x_j - x_i \geq 0, && \forall (i,j) \in E \\
& b_{ij} \in \{0,1\}, && \forall (i,j) \in E \\
& x_i \in \{0,1\}, && \forall i \in V \setminus \{s,t\}
\end{aligned}$$

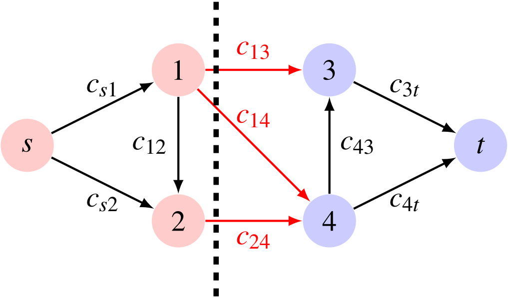
which describes the minimum cut (min-cut) problem.

Minimum cut problem

An s-t cut of G is a partition of the vertex set V into $S \subset V$ and $\bar{S} = V \setminus S$ s.t. $s \in S$ and $t \in \bar{S}$.

The capacity of a cut $(S, \bar{S})$ is

$$c(S, \bar{S}) = \sum_{\substack{(i,j) \in E \\ i \in S,\, j \in \bar{S}}} c_{ij}$$



The min-cut problem is

$$\min_{(S,\bar{S}) \text{ is an } s\text{-}t \text{ cut}} c(S, \bar{S})$$

In the integer programming formulation,

- $x_i = 1$ for $i \in S$, and $x_i = 0$ for $i \in \bar{S}$
- $b_{ij} = 1$ if $i \in S, j \in \bar{S}$, and $b_{ij} = 0$ otherwise (use $b_{ij} \geq x_i - x_j$)

Outline

- Dual LP
- Interpretation of dual problems
- Weak and strong duality
- Dual LP via Lagragian

Weak and strong duality

Consider an LP and its dual. WLOG, consider the inequality form,

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \boldsymbol{c}^T \boldsymbol{x} \\ \text{s.t.} \quad & A\boldsymbol{x} \geq \boldsymbol{b} \end{aligned} \qquad \text{(P)}$$

$$\begin{aligned} \max_{\boldsymbol{y}} \quad & \boldsymbol{b}^T \boldsymbol{y} \\ \text{s.t.} \quad & A^T \boldsymbol{y} = \boldsymbol{c} \\ & \boldsymbol{y} \geq \boldsymbol{0} \end{aligned} \qquad \text{(D)}$$

Weak duality. If $\boldsymbol{x}$ is primal feasible and $\boldsymbol{y}$ is dual feasible, then

$$\boldsymbol{c}^T \boldsymbol{x} \geq \boldsymbol{b}^T \boldsymbol{y}$$

Proof.

$$\boldsymbol{c}^T \boldsymbol{x} = (A^T \boldsymbol{y})^T \boldsymbol{x} = \boldsymbol{y}^T A \boldsymbol{x} \geq \boldsymbol{y}^T \boldsymbol{b}$$

Strong duality. If either (P) or (D) has a finite optimal value, so does the other, the optimal solutions $\boldsymbol{x}^*$ and $\boldsymbol{y}^*$ exist, and $\boldsymbol{c}^T \boldsymbol{x}^* = \boldsymbol{b}^T \boldsymbol{y}^*$.

Proof. Will follow from Slater's Theorem. Show weaker form next.

Strong duality

Lemma. If all (equality and inequality) constraint functions are affine, then the KKT conditions hold at a local minimum.

Note. We do not assume the local minimum is a regular point.

Theorem. If (P) has an optimal solution, then (D) also has an optimal solution and strong duality holds.

Proof. Let $\boldsymbol{x}^*$ be the optimal solution of (P). By the KKT conditions, there exist Lagrange multipliers $\boldsymbol{\mu}^*$ s.t.

1. $\boldsymbol{c} - \boldsymbol{A}^T\boldsymbol{\mu}^* = \boldsymbol{0}$ (stationarity)
2. $\boldsymbol{\mu}^* \geq \boldsymbol{0}$ (nonnegativity)
3. $(\boldsymbol{\mu}^*)^T(\boldsymbol{A}\boldsymbol{x}^* - \boldsymbol{b}) = 0$ (complementary slackness)

By 1 and 2, $\boldsymbol{\mu}^*$ is feasible for (D). By 1 and 3,

$$\boldsymbol{c}^T\boldsymbol{x}^* = (\boldsymbol{A}\boldsymbol{\mu}^*)^T\boldsymbol{x}^* = \boldsymbol{b}^T\boldsymbol{\mu}^*$$

By weak duality, $\boldsymbol{\mu}^*$ is optimal for (D) and hence strong duality holds.

Example

Recall the following pair of primal and dual LP,

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^2} \quad & f(\boldsymbol{x}) = x_1 + 2x_2 \\ \text{s.t.} \quad & 2x_1 + x_2 \geq 2 \\ & \boldsymbol{x} \geq \mathbf{0} \end{aligned}$$

$$\begin{aligned} \max_{\boldsymbol{\mu}\in\mathbb{R}^3} \quad & \psi(\boldsymbol{\mu}) = 2\mu_1 \\ \text{s.t.} \quad & 2\mu_1 + \mu_2 = 1 \\ & \mu_1 + \mu_3 = 2 \\ & \boldsymbol{\mu} \geq \mathbf{0} \end{aligned}$$

- The primal LP can be solved graphically, $\boldsymbol{x}^* = (1,0)^T, f^* = 1$.
- The dual is equivalent to

$$\begin{aligned} \max_{\mu_1} \quad & 2\mu_1 \\ \text{s.t.} \quad & 2\mu_1 \leq 1 \\ & \mu_1 \leq 2 \\ & \mu_1 \geq 0 \end{aligned}$$

So $\boldsymbol{\mu}^* = (\frac{1}{2}, 0, \frac{3}{2})^T$, $\psi^* = 1 = f^*$.

Max-flow min-cut theorem

Theorem. The max-flow value is equal to the min-cut capacity.

Proof. The max-flow problem is feasible, since the flow with $f_{ij} = 0$ for all $(i, j) \in E$ is feasible. The primal optimal value is finite,

$$0 \leq |f| \leq \sum_{(i,j) \in E} c_{ij} < +\infty$$

By strong duality, the max-flow value is equal to the optimal value of the LP relaxation of the min-cut problem.

The theorem then follows from the following

Lemma. The min-cut capacity is equal to the optimal value of the LP relaxation.

Max-flow min-cut theorem (cont'd)

Proof of lemma. We refer to the IP formulation as (IP) and its LP relaxation as (LP). Let c_{IP} and c_{LP} be their optimal values. We show $c_{LP} = c_{IP}$.

$c_{LP} \le c_{IP}$. A feasible solution of IP is also feasible for LP, so $c_{LP} \le c_{IP}$.

$c_{IP} \le c_{LP}$. We show there exists an s-t cut with capacity $\le c_{LP}$ using the probabilistic method.

- Let $(x_i^*), (b_{ij}^*)$ be an optimal solution of (LP), so $c_{LP} = \sum_{(i,j)} c_{ij} b_{ij}^*$.
- Let $U \sim$ uniform$(0,1)$ be a uniform random variable on $(0,1)$
- Let $S_U = \{i : U < x_i^*\}$. Then $(S_U, \bar{S}_U)$ is a (random) s-t cut, and

$$\begin{aligned}\mathbb{E}_U[c(S_U,\bar{S}_U)] &= \mathbb{E}_U\left[\sum c_{ij}\mathbb{1}\{i \in S_U, j \notin S_U\}\right] = \sum c_{ij}\Pr[i \in S_U, j \notin S_U] \\ &= \sum c_{ij}\Pr[x_j^* \le U < x_i^*] \le \sum c_{ij}(x_i^* - x_j^*)^+ = \sum c_{ij} b_{ij}^*\end{aligned}$$

- For some $u \in (0,1)$, the cut $(S_u, \bar{S}_u)$ has capacity $c(S_u, \bar{S}_u) \le c_{LP}$

Wasserstein distance

In the optimal transport problem, let $\boldsymbol{x}_i$ be the location of warehouse i, and $\boldsymbol{y}_j$ be the location of customer j. Suppose

$$\sum_{i=1}^{n} a_i = 1 = \sum_{j=1}^{m} b_j$$

Note a, b can be interpreted as two discrete probability distributions,

$$\Pr[\boldsymbol{X} = \boldsymbol{x}_i] = a_i, \quad \Pr[\boldsymbol{Y} = \boldsymbol{y}_j] = b_j$$

Now suppose the shipping cost c_{ij} is determined by the distance between $\boldsymbol{x}_i$ and $\boldsymbol{y}_j$,

$$c_{ij} = d(\boldsymbol{x}_i, \boldsymbol{y}_j)$$

e.g. $d(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|_2$.

The (first) Wasserstein distance $W_1(a, b)$ between the two distributions a and b is the optimal value of the optimal transport problem.

Wasserstein distance (cont'd)

Note $x_{ij} = a_i b_j$ is feasible (what's the probabilistic interpretation?), and the optimal cost is bounded and hence finite. By strong duality, $W_1(a, b)$ is equal to the dual optimal value.

Note the dual optimal solution satisfies

$$\lambda_j - \mu_i = c_{ij}$$

There exists h s.t. $\lambda_j = h(\boldsymbol{y}_j)$ and $\mu_i = h(\boldsymbol{x}_i)$.

- If $\boldsymbol{x}_i = \boldsymbol{y}_j$, then $\lambda_j - \mu_i = c_{ij} = d(\boldsymbol{x}_i, \boldsymbol{y}_j) = 0$, and $\lambda_j = \mu_i$
- We can specify the values of h at distinct points

The dual is equivalent to

$$\begin{aligned} \max_{h} \quad & \sum_{j=1}^{m} b_j h(\boldsymbol{y}_j) - \sum_{i=1}^{n} a_i h(\boldsymbol{x}_i) \\ \text{s.t.} \quad & h(\boldsymbol{y}_j) - h(\boldsymbol{x}_i) \le d(\boldsymbol{x}_i, \boldsymbol{y}_j), \quad \forall i, j \\ & h(\boldsymbol{x}_i) \ge 0, \quad \forall i \end{aligned}$$

Wasserstein distance (cont'd)

We can

- remove the constraints $h(\boldsymbol{x}_i) \geq 0$, since adding a constant to h doesn't change the optimal value
- add the constraints $h(\boldsymbol{y}_j) - h(\boldsymbol{x}_i) \geq -d(\boldsymbol{x}_i, \boldsymbol{y}_j)$, since the optimal solution satisfies $h(\boldsymbol{y}_j) - h(\boldsymbol{x}_i) = d(\boldsymbol{x}_i, \boldsymbol{y}_j)$

Thus the dual is equivalent to

$$\begin{aligned} \max_h \quad & \sum_{j=1}^m b_j h(\boldsymbol{y}_j) - \sum_{i=1}^n a_i h(\boldsymbol{x}_i) = \mathbb{E}_{\boldsymbol{Y}\sim b}\, h(\boldsymbol{Y}) - \mathbb{E}_{\boldsymbol{X}\sim a}\, h(\boldsymbol{X}) \\ \mathrm{s.\,t.} \quad & |h(\boldsymbol{y}_j) - h(\boldsymbol{x}_i)| \leq d(\boldsymbol{x}_i, \boldsymbol{y}_j), \quad \forall i, j \end{aligned}$$

The condition $|h(\boldsymbol{y}) - h(\boldsymbol{x})| \leq d(\boldsymbol{x}, \boldsymbol{y})$ simply means h is 1-Lipschitz, so

$$W_1(a, b) = \max_{h \text{ is } 1\text{-Lipschitz}} \mathbb{E}_{\boldsymbol{Y}\sim b}\, h(\boldsymbol{Y}) - \mathbb{E}_{\boldsymbol{X}\sim a}\, h(\boldsymbol{X})$$

Note. The general case is given by the Kantorovich-Rubinstein duality.

Outline

- Dual LP
- Interpretation of dual problems
- Weak and strong duality
- Dual LP via Lagragian

Dual LP via Lagrangian

The Lagrangian for the general LP on slide 6 is

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \boldsymbol{c}^T\boldsymbol{x} - \boldsymbol{\lambda}^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) - \boldsymbol{\mu}^T(\boldsymbol{G}\boldsymbol{x} - \boldsymbol{h})$$

If $\boldsymbol{\mu} \geq \mathbf{0}$ and $\boldsymbol{x} \in X$, i.e. $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ and $\boldsymbol{G}\boldsymbol{x} \geq \boldsymbol{h}$, then

$$f(\boldsymbol{x}) = \boldsymbol{c}^T\boldsymbol{x} \geq \boldsymbol{c}^T\boldsymbol{x} - \underbrace{\boldsymbol{\lambda}^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})}_{=0} - \underbrace{\boldsymbol{\mu}^T(\boldsymbol{G}\boldsymbol{x} - \boldsymbol{h})}_{\geq 0} = \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$$

Taking the infimum over $\boldsymbol{x} \in X$ first and then relaxing the constraint,

$$f^* = \inf_{x \in X} f(\boldsymbol{x}) \geq \inf_{\boldsymbol{x} \in X} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \geq \inf_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) =: \phi(\boldsymbol{\lambda}, \boldsymbol{\mu})$$

To maximize the lower bound, solve the following dual problem

$$\begin{array}{cl} \max\limits_{\boldsymbol{\lambda}, \boldsymbol{\mu}} & \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) \\ \text{s.t.} & \boldsymbol{\mu} \geq \mathbf{0} \end{array}$$

Dual LP via Lagrangian (cont'd)

Note

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = (\boldsymbol{c} - \boldsymbol{A}^T \boldsymbol{\lambda} - \boldsymbol{G}^T \boldsymbol{\mu})^T \boldsymbol{x} + \boldsymbol{b}^T \boldsymbol{\lambda} + \boldsymbol{h}^T \boldsymbol{\mu}.$$

An affine function is bounded below iff the coefficient for $\boldsymbol{x}$ is zero[1].

The dual problem

$$\begin{aligned} \max_{\boldsymbol{\lambda}, \boldsymbol{\mu}} \quad & \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \begin{cases} \boldsymbol{b}^T \boldsymbol{\lambda} + \boldsymbol{h}^T \boldsymbol{\mu}, & \text{if } \boldsymbol{c} - \boldsymbol{A}^T \boldsymbol{\lambda} - \boldsymbol{G}^T \boldsymbol{\mu} = \boldsymbol{0} \\ -\infty & \text{otherwise} \end{cases} \\ \text{s.t.} \quad & \boldsymbol{\mu} \geq \boldsymbol{0} \end{aligned}$$

which is equivalent to the dual LP

$$\begin{aligned} \max_{\boldsymbol{\lambda}, \boldsymbol{\mu}} \quad & \psi(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \boldsymbol{b}^T \boldsymbol{\lambda} + \boldsymbol{h}^T \boldsymbol{\mu} \\ \text{s.t.} \quad & \boldsymbol{A}^T \boldsymbol{\lambda} + \boldsymbol{G}^T \boldsymbol{\mu} = \boldsymbol{c} \\ & \boldsymbol{\mu} \geq \boldsymbol{0} \end{aligned}$$

---

[1] Consider $f(\boldsymbol{x}) = \boldsymbol{a}^T \boldsymbol{x} + c$. If $\boldsymbol{a} = \boldsymbol{0}$, then $\inf_{\boldsymbol{x}} f(\boldsymbol{x}) = c$. If $\boldsymbol{a} \neq \boldsymbol{0}$, letting $\boldsymbol{x} = -t\boldsymbol{a}$ and $t \to +\infty$ yields $\inf_{\boldsymbol{x}} f(\boldsymbol{x}) \leq -t\|\boldsymbol{a}\|^2 + c \to -\infty$.