11 Lagrangian duality

Our exposition follows [13] and [9].

11.1 Lagrangian relaxation and Lagrangian dual

We want to solve the problem

$$z = \max\{c^T x : x \geq \mathbf{0},\ Ax \leq b,\ Dx \leq d,\ x \in \mathbb{Z}^n\}. \tag{IP}$$

Sometimes it may be the case that some constraints are "easier" to handle (in our case these are the $Ax \leq b$) and some are "complicating" ($Dx \leq d$). Then we may like to get rid of the $Dx \leq d$ and instead make them part of the objective function:

$$z(u) = \max\{c^T x + u^T(d - Dx) : x \geq \mathbf{0},\ Ax \leq b,\ x \in \mathbb{Z}^n\}, \tag{LR(u)}$$

for $u \in \mathbb{R}^m$, $u \geq \mathbf{0}$, where m is the number of rows of D.

Each problem (LR(u)) is called a **Lagrangian relaxation** of (IP); the coefficients u_i are called **Lagrange multipliers**. Indeed, (LR(u)) is a relaxation of (IP):

Proposition 11.1 *Let $u \in \mathbb{R}^m$, $u \geq \mathbf{0}$. Then the feasible region of* (IP) *is a subset of the feasible region of* (LR(u)). *Moreover, if z is the optimal value of* (IP) *and $z(u)$ is the optimal value of* (LR(u)), *then $z(u) \geq z$.*

Proof. Inclusion of feasible regions is obvious. If x is a feasible solution of (IP), then $c^T x + u^T(d - Dx) \geq c^T x$; this implies the second claim. ∎

In this way, for every u the Lagrangian relaxation (LR(u)) provides an upper bound on z. To find the best Lagrangian upper bound, we would like to compute

$$z_{\text{LD}} = \min\{z(u) : u \geq \mathbf{0},\ u \in \mathbb{R}^m\}. \tag{LD}$$

Problem (LD) is called the **Lagrangian dual**.

Now we are left with three issues:

(1) How do we compute $z(u)$?
(2) How do we compute z_{LD}?
(3) How good is the upper bound provided by z_{LD}?

11.2 How to compute $z(u)$?

Example 11.1 The uncapacitated facility location problem of Assignment 1 leads to the IP

$$\min \quad \sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij}x_{ij} + \sum_{i=1}^{n} o_i y_i \tag{11.1a}$$

$$\text{s.t.} \quad \sum_{i=1}^{n} x_{ij} = d_j, \qquad \forall j = 1,\dots,m \tag{11.1b}$$

$$x_{ij} \leq y_i d_j, \qquad \forall i = 1,\dots,n,\ \forall j = 1,\dots,m \tag{11.1c}$$

$$x_{ij} \geq 0, \qquad \forall i = 1,\dots,n,\ \forall j = 1,\dots,m \tag{11.1d}$$

$$y_j \in \{0,1\}. \qquad \forall j = 1,\dots,m \tag{11.1e}$$

We dualize the demand constraints (11.1b):

$$\begin{aligned} z(u) = \min \quad & \sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij}x_{ij} + \sum_{i=1}^{n} o_i y_i + \sum_{j=1}^{m} u_j\Big(d_j - \sum_{i=1}^{n} x_{ij}\Big) && (11.2a)\\ \text{s.t.} \quad & x_{ij} \le y_i d_j, && (11.2b)\\ & x_{ij} \ge 0, && (11.2c)\\ & y_j \in \{0,1\}, && (11.2d) \end{aligned}$$

which in turn yields

$$z(u) = u^T d + \sum_{i=1}^{n} z_i(u),$$

where

$$z_i(u) = \min\Big\{\sum_{j=1}^{m}(c_{ij} - u_j)x_{ij} + o_i y_i : x_{ij} \le y_i d_j,\ x_{ij} \ge 0,\ y_i \in \{0,1\} \text{ for all } i, j\Big\}.$$

Now it is easy to compute $z_i(u)$. If $y_i = 0$, then all $x_{ij} = 0$ and the objective value is 0. If on the other hand $y_i = 1$, then to minimize the sum, set $x_{ij} = d_j$ if $c_{ij} < u_j$ and set $x_{ij} = 0$ if $c_{ij} \ge u_j$. Hence $z_i(u) = \min\big\{0, o_i + \sum_{j=1}^{m} \min\{0, (c_{ij} - u_j)d_j\}\big\}$.

Example 11.2 Recall the symmetric traveling salesman problem from Chapter 1. It led to the IP

$$\begin{aligned} \min \quad & \sum_{ij\in E} w_{ij}x_{ij} && && (11.3a)\\ \text{s.t.} \quad & \sum_{ij\in E} x_{ij} = 2, && \forall i \in V && (11.3b)\\ & \sum_{\substack{ij\in E\\ i,j\in S}} x_{ij} \le |S| - 1, && \forall S \subset V : 2 \le |S| \le |V| - 1 && (11.3c)\\ & x_{ij} \in \{0,1\}. && \forall ij \in E && (11.3d) \end{aligned}$$

In fact, as

$$|S| - \sum_{\substack{i,j\in S\\ ij\in E}} x_{ij} = \frac{1}{2}\sum_{i\in S}\sum_{ij\in E} x_{ij} - \sum_{\substack{i,j\in S\\ ij\in E}} x_{ij} = \frac{1}{2}\sum_{\substack{ij\in E\\ i\in S,\ j\notin S}} x_{ij},$$

we have

$$|S| - \sum_{\substack{i,j\in S\\ ij\in E}} x_{ij} = |V \setminus S| - \sum_{\substack{i,j\in V\setminus S\\ ij\in E}} x_{ij},$$

and so half of the constraints (11.3c) are redundant. Thus we may replace (11.3c) with

$$\sum_{\substack{ij\in E\\ i,j\in S}} x_{ij} \le |S| - 1, \qquad \forall S \subset V : 2 \le |S| \le |V| - 1,\ 1 \notin S. \qquad (11.3e)$$

Now we dualize the degree constraints (11.3b), leaving the degree constraint on the vertex 0 in place:

$$z(u) = \min \sum_{ij\in E}(w_{ij} - u_i - u_j)x_{ij} + 2\sum_{i\in V} u_i \tag{11.4a}$$

$$\text{s.t.} \quad \sum_{1j\in E} x_{1j} = 2, \tag{11.4b}$$

$$\sum_{\substack{ij\in E\\ i,j\in S}} x_{ij} \le |S| - 1, \quad \forall S \subset V : 2 \le |S| \le |V| - 1,\ 1 \notin S \tag{11.4c}$$

$$\sum_{ij\in E} x_{ij} = n, \tag{11.4d}$$

$$x_{ij} \in \{0, 1\}. \quad \forall ij \in E \tag{11.4e}$$

A feasible solution to (11.4), interpreted as a set of edges or a subgraph T of G, satisfies: 1. The degree of vertex 1 in T is 2. 2. The subgraph of T induced by $\{2, 3, \ldots, n\}$ is a tree. Such a subgraph T of G is called a **1-tree**. Finding a minimum-weight 1-tree (with weight $w_{ij} - u_i - u_j$ on edge ij) is easy. Solving the Lagrangian dual $z_{\mathrm{LD}} = \max\{z(u) : u \in \mathbb{R}^n\}$ is less easy; it was first approached by Held & Karp (1970–71), who were able to solve TSP instances too large for all other approaches of the time.

11.3 How good is the upper bound?

Sometimes it is the best possible:

Theorem 11.2 *If $u \ge \mathbf{0}$, $x(u)$ is an optimal solution to* (LR(u))*, $Dx(u) \le d$ and $u^T \cdot (d - Dx(u)) = 0$, then $x(u)$ is an optimal solution to* (IP)*.*

Proof. If z is the optimal value of (IP), then $c^T x(u) \le z$, since by assumption $x(u)$ is feasible for (IP). But

$$c^T x(u) = c^T x(u) + u^T(d - Dx(u)) = z(u) \ge z_{\mathrm{LD}} \ge z.$$

Therefore $z = z_{\mathrm{LD}} = z(u)$. ∎

In general, though, we have:

Theorem 11.3 *Let $Q = \{x \in \mathbb{Z}^n : x \ge \mathbf{0},\ Ax \le b\}$. Then $z_{\mathrm{LD}} = \max\{c^T x : Dx \le d,\ x \in \operatorname{conv} Q\}$.*

Proof.
By definition,

$$z(u) = \max\{c^T x + u^T(d - Dx) : x \in Q\} = \max\{c^T x + u^T(d - Dx) : x \in \operatorname{conv} Q\}.$$

Hence if $Q = \emptyset$, then $z_{\mathrm{LD}} = -\infty = \max\{c^T x : Dx \le d,\ x \in \operatorname{conv} Q\}$. Otherwise, by Theorem 3.16, $\operatorname{conv} Q = \operatorname{conv}\{v_1, \ldots, v_s\} + \operatorname{cone}\{r_1, \ldots, r_t\}$. In other words, $\operatorname{conv} Q$ has vertices $v_1, \ldots, v_s$ and extremal rays $r_1, \ldots, r_t$.

If $(c^T - u^T D)r_j > 0$ for some j, then $\max\{c^T x + u^T(d - Dx) : x \in \operatorname{conv} Q\} = +\infty$. Otherwise $\max\{c^T x + u^T(d - Dx) : x \in \operatorname{conv} Q\} = c^T v_k + u^T(d - Dv_k)$ for some k. Thus

$$\begin{aligned}
z_{\mathrm{LD}} &= \min\bigl\{\max\{c^T v_k + u^T(d - Dv_k) : k = 1, \dots, s\} : u \geq \mathbf{0}, \\
&\qquad\qquad (c^T - u^T D)r_j \leq 0 \text{ for all } j = 1, \dots, t\bigr\} \\
&= \min\{w : w + u^T(Dv_k - d) \geq c^T v_k \text{ for } k = 1, \dots, s; \\
&\qquad\qquad u^T D r_j \geq c^T r_j \text{ for } j = 1, \dots, t;\ u \geq \mathbf{0}\} \qquad (11.5) \\
&= \max\Bigl\{c^T\Bigl(\sum_{k=1}^{s} y_k v_k + \sum_{j=1}^{t} z_j r_j\Bigr) : \sum_{k=1}^{s} y_k = 0;\ D\Bigl(\sum_{k=1}^{s} y_k v_k + \sum_{j=1}^{t} z_j r_j\Bigr) \leq d \sum_{k=1}^{s} y_k; \\
&\qquad\qquad y_k, z_j \geq 0 \text{ for all } k, j\Bigr\} \qquad (11.6) \\
&= \max\{c^T x : x \in \operatorname{conv} Q,\ Dx \leq d\}.
\end{aligned}$$

∎

Note. So z_{LD} can be computed by solving one of the dual LPs (11.5) and (11.6).

Corollary 11.4 *If $\{x \in \mathbb{R}^n : x \geq \mathbf{0},\ Ax \leq b\}$ is integral, then z_{LD} is equal to the value of the LP relaxation of* (IP).

Note. In the STSP, the corollary implies that we can potentially solve the LP relaxation with exponentially many constraints by setting weights so as to maximize the weight of a minimum-weight 1-tree, without explicitly treating the many constraints.

Corollary 11.5 *The value $z(u)$ is finite if and only if $u^T D r_j \geq c^T r_j$ for all $j = 1, \dots, t$. The vectors u satisfying these conditions form a polyhedron, over which $z(u)$ is convex and piecewise linear.*

Proof. On this polyhedron, $z(u) = \max\{c^T v_k + u^T(d - Dv_k) : k = 1, \dots, t\}$; the maximum of finitely many affine functions is always piecewise linear and convex. ∎

11.4 How to compute z_{LD}?

The function $z(u)$ is piecewise linear and convex, but in general not differentiable. As a generalization of the gradient method for differentiable convex functions, Held & Karp (1970) proposed the **subgradient method**.

For a convex function $z : \mathbb{R}^m \to \mathbb{R}$, a **subgradient** at u is a vector $s \in \mathbb{R}^m$ such that $z(v) \geq s^T(v - u)$ for all $v \in \mathbb{R}^m$.

Lemma 11.6 *Let $u_0 \in \mathbb{R}^m$; let $x^*(u_0)$ be an optimal solution to*

$$z(u_0) = \max\{c^T x + u_0^T(d - Dx) : x \in Q\}$$

with Q defined as in Theorem 11.3. Then $d - Dx^*(u_0)$ is a subgradient of $z(u)$ at u_0.

Algorithm 11.1 (Subgradient method for the Lagrangian dual)
Given a starting point u and step lengths $\{\mu_t : t = 0, 1, \dots\}$.
1. Set $u^0 := u$, $t := 0$.
2. Find an optimal solution $x^*(u^t)$ to $\max\{c^T x + u_0^T(d - Dx) : x \in Q\}$.
3. Set $u^{t+1} := \max\big\{u^t - \mu_t\big(d - Dx^*(u^t)\big), 0\big\}$.
4. If stopping criterion fulfilled, stop. Otherwise set $t := t + 1$ and go to **2.**

There are still several degrees of freedom. How should the starting point be selected? (One might go with $u = \mathbf{0}$.) What are good step lengths? What is a good stopping criterion?

As for the step lengths, it may be proved that if $\mu_k \to \infty$ and $\sum_{i=0}^{k} \mu_t \to \infty$ as $k \to \infty$, then $z(u^t) \to z_{\text{LD}}$ as $t \to \infty$. In practice, the convergence may be too slow. Similarly $z(u^t) \to z_{\text{LD}}$ as $t \to \infty$ if $\mu_t = \mu_0 \lambda^t$ for some $\lambda < 1$, provided μ_0 and λ are sufficiently large. This may lead to faster convergence, but sometimes the geometric series tends to zero too rapidly and u^t converges before reaching an optimal point.

The method is often terminated before the solution z_{LD} is reached. The result may be used as an upper bound in some branch-and-bound algorithm; as we saw in the STSP case, it might be easier to compute than directly solving the LP relaxation.

Note. In fact, if (LR(u)) can be solved in polynomial time, then so can be (LD).