B6.3 Integer Programming

Raphael Hauser

Mathematical Institute
University Of Oxford

Michaelmas Term 2018

The Strength of Lagrangian Dual Bounds

Let us consider the IP

$$
\begin{aligned}
\text{(IP)} \quad z = \max\, & c^{\mathsf{T}} x \\
\text{s.t. } & D_1 x \leq d_1, \\
& D_2 x = d_2, \\
& x \in \mathscr{X} = \{x \in \mathbb{R}^n : Ax \leq a,\ x \geq 0,\ x \in \mathbb{Z}^n\},
\end{aligned}
$$

with Lagrangian relaxation

$$
\text{(IP}(u)) \quad z(u) = \max\{c^{\mathsf{T}} x + u^{\mathsf{T}}(d - Dx) : x \in \mathscr{X}\},
$$

and let us compare the dual bounds associated with the Lagrangian Dual and the LP relaxation of (IP),

$$
\text{(LD)} \quad w_{LD} = \min_u \{z(u) : u \in \mathbb{R}^m,\ u_i \geq 0\ (i \in \mathcal{I})\},
$$

$$
\text{(LP)} \quad w_{LP} = \max_x \{c^{\mathsf{T}} x : D_1 x \leq d_1,\ D_2 x = d_2,\ Ax \leq a,\ x \geq 0\}.
$$

Theorem (Characterisation of Lagrangian dual bound)

The Lagrangian dual bound is characterised as follows,

$$
\begin{aligned}
(\mathrm{LD}') \quad w_{LD} = \max_{x} \; & c^{\mathsf{T}} x \\
\text{s.t. } & Dx \le d \\
& x \in \operatorname{conv}(\mathscr{X}).
\end{aligned}
$$

(See Lecture 12.)

Theorem (Lagrangian dual and LP relaxation)

The Lagrangian dual bound is at least as tight as the LP relaxation bound,

$$z \le w_{LD} \le w_{LP}.$$

If $\{x \in \mathbb{R}^n : Ax \le a, \, x \ge 0\}$ is an ideal formulation of $\mathscr{X}$, then $w_{LD} = w_{LP}$.

Proof. $\mathscr{X} \subset \{x \in \mathbb{R}^n : Ax \le a, \, x \ge 0\}$ implies $\operatorname{conv}(\mathscr{X}) \subseteq \{x \in \mathbb{R}^n : Ax \le a, \, x \ge 0\}$. Hence, Theorem (Characterisation of Lagrangian dual) shows that (LP) is a relaxation of (LD).

Moreover, if $\{x \in \mathbb{R}^n : Ax \le a, \, x \ge 0\}$ is an ideal formulation of $\mathscr{X}$, then

$$\{x \in \mathbb{R}^n : Ax \le a, \, x \ge 0\} = \operatorname{conv}(\mathscr{X}),$$

so that (LD) and (LP) coincide.

In the latter situation the Lagrangian Dual offers an alternative to solving the LP relaxation directly in cases where this is too costly.

Choosing a Lagrangian Dual

Many problems have several reasonable Lagrangian Duals. In this case it is worthwhile thinking about the advantages and disadvantages of the different formulations before starting any calculations.

Example (Generalised assignment problem (GAP))

Consider the *generalised assignment problem*

$$
\begin{aligned}
\text{(IP)} \quad z = \max & \sum_{j=1}^{n} \sum_{i=1}^{m} c_{ij} x_{ij} \\
\text{s.t.} & \sum_{j=1}^{n} x_{ij} \leq 1 \quad (i = 1, \dots, m), \\
& \sum_{i=1}^{m} a_{ij} x_{ij} \leq b_j \quad (j = 1, \dots, n), \\
& x \in \{0, 1\}^{m \times n}.
\end{aligned}
$$

In this case we have multiple choices of a Lagrangian dual:

Example (GAP continued)

1. Dualising both sets of constraints

$$(\mathrm{IP}(u)) \quad z(u) = \max_{x} \sum_{j=1}^{n} \sum_{i=1}^{m} (c_{ij} - u_i - a_{ij} v_j) x_{ij} + \sum_{i=1}^{m} u_i + \sum_{j=1}^{n} v_j b_j$$
$$\text{s.t. } x \in \{0,1\}^{m \times n}.$$

It is certainly easy to solve this IP, as the remaining feasible set $\mathscr{X} = \{0,1\}^{m \times n}$ has the ideal formulation $\{x \in \mathbb{R}^m : 0 \leq x_i \leq 1 \; \forall\, i\}$. By Theorem 2, solving the Lagrangian dual $\min_{u \geq 0} z(u)$ yields the same result as the LP relaxation of (IP). However, it might still be better to solve the Lagrangian dual, as $(\mathrm{IP}(u))$ can be solved by inspection:

$$x^*_{ij} = \begin{cases} 1 & \text{if } c_{ij} - u_i - a_{ij} v_j > 0 \\ 0 & \text{otherwise.} \end{cases}$$

2. Dualising only the second set of constraints The advantages and disadvantages of this relaxation are similar to the first case.

Example (GAP continued)

3. Dualising only the first set of constraints

$$(\mathrm{IP}(u)) \quad z(u) = \max_x \sum_{j=1}^{n}\sum_{i=1}^{m}(c_{ij} - u_i)x_{ij} + \sum_{i=1}^{m} u_i$$
$$\text{s.t.} \sum_{i=1}^{m} a_{ij}x_{ij} \leq b_j \quad (j = 1, \dots, n),$$
$$x \in \{0,1\}^{m\times n}.$$

Here the remaining feasible set $\mathscr{X} = \left\{x \in \{0,1\}^{m\times n} : \sum_i a_{ij}x_{ij} \leq b_j,\, j = 1, \dots, n\right\}$ is not of the "easy" type, as

$$\mathrm{conv}(\mathscr{X}) \subset \left\{x \in \mathbb{R}^n : 0 \leq x_i \leq 1,\, i = 1, \dots, m,\, \sum_i a_{ij}x_{ij} \leq b_j\, j = 1, \dots, n\right\}$$

is generally a strict inclusion. Consequently, w_{LD} may be a strictly tighter bound than the bound obtained from the LP relaxation.

The Lagrangian dual is more difficult to solve, but (IP(u)) decouples into blocks $(\{x_{ij} : i = 1, \dots, m\})_{j=1}^{n}$ and can be parallelised.

Solving the Lagrangian Dual

By the definition of a Lagrangian Relaxation, the map

$$u \mapsto z(u) = \max\{c^{\mathsf{T}} x + u^{\mathsf{T}}(d - Dx) : x \in \mathscr{X}\}$$

is a piecewise linear function of u. Therefore, $z(u)$ is a convex function on $\mathscr{D} := \{u \in \mathbb{R}^m : u_i \geq 0, (i \in \mathcal{I})\}$ by virtue of the following lemma:

Lemma (Pointwise maximum of convex functions)

Let $\{f_i(u) : \mathscr{D} \to \mathbb{R} \big| i \in \mathcal{N}\}$ be a set of convex functions defined on a convex domain $\mathscr{D}$. Then

$$u \mapsto \max_{i \in \mathcal{N}} f_i(u)$$

is a convex function on $\mathscr{D}$.

Proof. For all $u_1, u_2 \in \mathscr{D}$ and $\lambda \in [0, 1]$,

$$\max_i f_i\left(\lambda u_1 + (1-\lambda)u_2\right) \leq \max_i \left(\lambda f_i(u_1) + (1-\lambda) f_i(u_2)\right) \leq \lambda \max_i f_i(u_1) + (1-\lambda) \max_i f_i(u_2).$$

Note that $z(u)$ is not differentiable at breakpoints u where $\arg\max\{c^{\mathsf{T}}x + u^{\mathsf{T}}(d - Dx) : x \in \mathscr{X}\}$ contains more than one point.

Lemma (Gradient characterisation of convex functions)

Let $\mathscr{D} \subseteq \mathbb{R}^m$ be a convex domain and $f : \mathscr{D} \to \mathbb{R}$ a convex function with gradient $\gamma = \nabla f(u)$ at $u \in \mathscr{D}$. Then the first order Taylor approximation is a lower bounding function:

$$f(u) + \gamma^{\mathsf{T}}(v - u) \leq f(v), \quad (v \in \mathscr{D}).$$

Proof. By definition, f is convex if for all $u, v \in \mathscr{D}$ and $\lambda \in [0, 1]$,

$$f(\lambda v + (1 - \lambda)u) \leq \lambda f(v) + (1 - \lambda)f(u).$$

Therefore,

$$f(u) + \frac{f(u + \lambda(v - u)) - f(u)}{\lambda} \leq f(v),$$

and taking the limit $\lambda \to 0$ yields the result.

Definition (Extension of a convex function)

Let $\mathscr{D} \subseteq \mathbb{R}^m$ be a convex domain and $f : \mathscr{D} \to \mathbb{R}$ a convex function. We extend f to a *proper convex function* defined on $\mathbb{R}^m$ by setting $f(u) := +\infty$ for $u \notin \mathscr{D}$. The extension satisfies satisfies

$$f(\lambda u + (1 - \lambda)v) \leq \lambda f(u) + (1 - \lambda)f(v), \quad (u, v \in \mathbb{R}^m, \ \lambda \in [0, 1]).$$

Motivated by Lemma (Gradient characterisation of convex functions), the notion of gradient can be generalised to non-differentiable points of convex functions:

Definition (Subgradient and Subdifferential)

Let $u \in \mathbb{R}^{\boldsymbol{m}}$ and let $f : \mathbb{R}^{\boldsymbol{m}} \to \mathbb{R}$ be a proper convex function. A *subgradient* of f at $u \in \mathbb{R}^{\boldsymbol{m}}$ is a vector $\gamma \in \mathbb{R}^{\boldsymbol{m}}$ such that

$$f(u) + \gamma^{\mathsf{T}}(v - u) \leq f(v), \quad (v \in \mathbb{R}^{\boldsymbol{m}}).$$

The *subdifferential* $\partial f(u)$ of f at u is the set of subgradients of f at u.

Proposition (Properties of the subdifferential)

- *If f is differentiable at u, then $\partial f(u) = \{\nabla f(u)\}$ is a singleton containing only the gradient.*
- *$\partial f(u)$ is a convex set.*
- *$u^* \in \arg\min f(u)$ if and only if $\vec{0} \in \partial f(u^*)$.*
- *Let $z(u) := \max\{c^{\mathsf{T}}x + u^{\mathsf{T}}(d - Dx) : x \in \mathscr{X}\}$. Then*

  $$\partial z(u) = \operatorname{conv}\left(\left\{d - Dx^* : x^* \in \arg\max\{c^{\mathsf{T}}x + u^{\mathsf{T}}(d - Dx) : x \in \mathscr{X}\}\right\}\right),$$

  where $\arg\max$ is the set of all maximisers.

Proof. See problem sheet.

Algorithm (Subgradient Algorithm for Solving (LD))

Initialise:
  choose $u \in \mathbb{R}^m$ with $u_i \geq 0$, $(i \in \mathcal{I})$;
  $X^* := \arg\max\{c^\mathrm{T} x + u^\mathrm{T}(d - Dx) : x \in \mathscr{X}\}$, $V := \{d - Dx^* : x^* \in X^*\}$;
while $\vec{0} \notin \mathrm{conv}(V)$ **do**
  choose $v \in V$, $\mu > 0$;
  if $i \in \mathcal{I}$ **then**
    $u_i := \max(u_i - \mu v_i, 0)$;
  else
    $u_i := u_i - \mu v_i$;
  end
  $X^* := \arg\max\{c^\mathrm{T} x + u^\mathrm{T}(d - Dx) : x \in \mathscr{X}\}$, $V := \{d - Dx^* : x^* \in X^*\}$;
end

Notes:

- In each iteration of the main loop the Lagrange multiplier vector is improved by correcting it in a direction that makes the objective function $z(u)$ *decrease*.
- Note the built-in safeguard mechanism that prevents individual components of the updated u to become negative for $i \in \mathcal{I}$.
- The termination criterion of the main loop can be evaluated by solving an LP (see problem sheet).
- The choice of step length μ requires further discussion.

Let $u^{[k]}$, $v^{[k]}$ and μ_k be the values of u, v and μ in the k-th iteration of the main loop. Let $z_k := z(u^{[k]})$, $U^* := \arg\min_u z(u)$, $z^* := \min_u z(u)$ and $\text{dist}(u^{[1]}, U^*) := \min_{u^* \in U^*} \|u^{[1]} - u^*\|_2$.

Lemma (Convergence of the subgradient algorithm)

Suppose $\|v^{[k]}\|_2 \leq G$ for all k. Then

$$\min_{i \in [1,k]} z_i - z^* \leq \frac{\text{dist}(u^{[1]}, U^*) + G^2 \sum_{i=1}^k \mu_i^2}{2 \sum_{i=1}^k \mu_i}.$$

Proof. For any $u^* \in U^*$,

$$\begin{aligned}
\|u^{[k+1]} - u^*\|_2^2 &\leq \|u^{[k]} - \mu_k v^{[k]} - u^*\|_2^2 \quad \text{(see problem sheet)} \\
&= \|u^{[k]} - u^*\|_2^2 - 2\mu_k v^{[k]\,\mathsf{T}}(u^{[k]} - u^*) + \mu_k^2 \|v^{[k]}\|_2^2 \\
&\leq \|u^{[k]} - u^*\|_2^2 - 2\mu_k (z_k - z^*) + \mu_k^2 G^2 \quad \text{(since } v^{[k]} \text{ is a subgradient).}
\end{aligned}$$

By recursion, $\|u^{[k+1]} - u^*\|_2^2 \leq \|u^{[1]} - u^*\|_2^2 - 2\sum_{i=1}^k \mu_i (z_i - z^*) + \sum_{i=1}^k \mu_i^2 G^2$, and hence,

$$\begin{aligned}
2\left(\sum_{i=1}^k \mu_i\right) \times \left(\min_{i \in [1,k]} z_i - z^*\right) &\leq 2\sum_{i=1}^k \mu_i (z_i - z^*) + \|u^{[k+1]} - u^*\|_2^2 \\
&\leq \|u^{[1]} - u^*\|_2^2 + \sum_{i=1}^k \mu_i^2 G^2.
\end{aligned}$$

Theorem (Step length choice in subgradient algorithm)

i) *If* $\sum_k \mu_k \to \infty$ *and* $\sum_k \mu_k^2 \to 0$ *as* $k \to \infty$, *then* $z(u^{[k]}) \to w_{LD}$.

ii) *If* $\sum_k \mu_k \to \infty$ *and* $\mu_k \to 0$ *as* $k \to \infty$, *then* $z(u^{[k]}) \to w_{LD}$.

iii) *If* $\mu_k = \mu_0 \rho^k$ *for some fixed* $\rho \in (0,1)$ *for* μ_0 *sufficiently large and* ρ *sufficiently close to* 1, *then* $z(u^{[k]}) \to w_{LD}$.

iv) *if* $\overline{w} \geq w_{LD}$ *and*

$$\mu_k = \frac{\varepsilon_k \times (z(u^{[k]}) - \overline{w})}{\|v^{[k]}\|^2},$$

where $\varepsilon_k \in (0,2)$ *for all* k, *then either* $z(u^{[k]}) \to w_{LD}$ *for* $k \to \infty$, *or else* $\overline{w} \geq z(u^{[k]}) \geq w_{LD}$ *occurs for some finite* k.

Step length choice iv) gives the most useful step lengths in practice, but note that for μ_k to be positive, we need an upper bound $\overline{w} \in (w_{LD}, z(u^{[k]}))$. In practical applications such a bound is not available explicitly.

Note however:

- Lower bounds $\underline{w}$ of w_{LD} are available by ways of using heuristics that produce primal feasible solutions.
- If the bound $\overline{w}$ in Rule iv) is chosen too low, μ_k is positive but possibly too large. If $z(u^{[k+1]}) < z(u^{[k]})$, this does not pose a problem, as descent is achieved and the point $u^{[k+1]}$ can be accepted as the next iterate.
- If $z(u^{[k+1]}) \geq z(u^{[k]})$, the step μ_k took the iterate to a point where the objective function $z(u)$ increases again. The guess of $\overline{w}$ then needs to be increased to reduce the step length μ_k.

Algorithm (Practical Subgradient Algorithm for Solving (LD))

Initialise:

fix $\varepsilon \in (0, 2)$, choose $u \in \mathbb{R}^m$ with $u_i \geq 0$, $(i \in \mathcal{I})$;

$X^* := \arg\max\{c^T x + u^T(d - Dx) : x \in \mathscr{X}\}$, $V := \{d - Dx^* : x^* \in X^*\}$;

find $x \in \mathscr{F} = \mathscr{X} \cap \{D_1 x \leq d_1, D_2 x = d_2\}$;

set $\overline{w} := \underline{w} := c^T x$;

while $\vec{0} \notin \mathrm{conv}(V)$ **do**

- choose $v \in V$;
- $z^+ := +\infty$;
- **while** $z^+ \geq z(u)$ **do**
  - $\overline{w} := \frac{z(u)+\overline{w}}{2}$; `// this is our guess of a suitable dual bound`
  - $\mu := \frac{\varepsilon(z(u)-\overline{w})}{\|v\|^2}$;
  - **if** $i \in \mathcal{I}$ **then**
    - $u_i^+ := \max(u_i - \mu v_i, 0)$; `// compute candidate updates`
  - **else**
    - $u_i^+ := u_i - \mu v_i$; `// compute candidate updates`
  - **end**
  - $z^+ := z(u^+)$; `// evaluate candidate updates`
- **end**
- $u := u^+$; `// accept candidate updates as actual updates`
- $\overline{w} := \underline{w}$;
- $X^* := \arg\max\{c^T x + u^T(d - Dx) : x \in \mathscr{X}\}$, $V := \{d - Dx^* : x^* \in X^*\}$;

end