Reformulations of Integer Programs

July, 2012

Outline

1. Definitions and Notation
2. Lagrangean relaxation
3. Dantzig-Wolfe reformulations

Premise

This lecture is largely based on the chapter by **F. Vanderbeck and L. Wolsey. Reformulation and Decomposition of Integer Programs.** In M. Junnger, Th.M. Liebling, D. Naddef, G.L. Nemhauser, W.R. Pulleyblank, G. Reinelt, G. Rinaldi, and L.A. Wolsey, editors, 50 Years of Integer Programming 1958 – 2008. Springer, Berlin, 2010.

Definitions and Notation

Given an initial formulation of an IP, knowledge of problem structure can be used to obtain improved problem formulations and more effective algorithms.

- (IP) $\min\{cx : x \in X\}$ where X is a discrete solution set that can be modeled as the set of integer points satisfying a set of linear inequalities.
- $X = P \cap \mathbb{Z}^n$ *with* $P = \{x \in \mathbb{R}^n_+ : Ax \geq a\}$

Definition

A polyhedron $P \subseteq \mathbb{R}^n$ *is the intersection of a finite number of half-spaces. There exists* $A \in \mathbb{R}^{m \times n}$, $a \in \mathbb{R}^m$ *such that* $P = \{x \in \mathbb{R}^n : Ax \geq a\}$.

Definition

A polyhedron P *is a formulation for* X *if* $X = P \cap \mathbb{Z}^n$

Definitions and Notation

Definition

Given $X \subseteq \mathbb{R}^n$, the convex hull of X, denoted $conv(X)$, is the smallest closed convex set containing X.

Definition

An extended formulation for a polyhedron $P \subseteq \mathbb{R}^n$ is a polyhedron $Q = \{(x, w) \in \mathbb{R}^{n+p} : Gx + Hw \geq d\}$ such that $P = proj_x(Q)$.

Definition

Given a non-empty polyhedron $P \subseteq \mathbb{R}^n$, i) $x \in P$ is an extreme point of P if $x = \lambda x_1 + (1 - \lambda)x_2,\ 0 < \lambda < 1,\ x_1, x_2 \in P$ implies that $x = x1 = x2$.
ii) r is a ray of P if $r \neq 0$ and $x \in P$ implies $x + \mu r \in P$ for all $\mu \in \mathbb{R}_+$.
iii) r is an extreme ray of P if r is a ray of P and $r = \mu_1 r_1 + \mu_2 r_2$, $\mu \in \mathbb{R}^2_+ \setminus \{0\}$, r_1, r_2 rays of P implies $r_1 = \alpha r_2$ for some $\alpha > 0$.

Minkowski Theorem

Theorem

(Minkowski) Every polyhedron $P = \{x \in \mathbb{R}^n : Ax \geq a\}$ can be represented in the form

$P = \{x \in \mathbb{R}^n : x = \sum_{g \in G} \lambda_g x^g + \sum_{r \in \mathbb{R}} \mu_r v_r,\ \sum_{g \in G} \lambda_g = 1,\ \lambda \in \mathbb{R}_+^{|G|},\ \mu \in \mathbb{R}_+^{|R|}\}$

where $\{x^g\}_{g \in G}$ are the extreme points of P and $\{v^r\}_{r \in R}$ the extreme rays of P.

Definition

An extended formulation for an IP set $X \subseteq \mathbb{Z}^n$ is a polyhedron $Q \subseteq \mathbb{R}^{n+p}$ such that $X = proj_x(Q) \cap \mathbb{Z}^n$.

Definition

An extended formulation $Q \subseteq \mathbb{R}^{n+p}$ for an IP set $X \subseteq \mathbb{Z}^n$ is tight if $proj_x(Q) = conv(X)$.

equivalent of Minkowski Theorem for IPs

Definition

An extended IP-formulation for an IP set $X \in \mathbb{Z}^n$ *is a set* $Q_I = \{(x, w1, w2) \in \mathbb{R}^n \times \mathbb{Z}^{p_1} \times \mathbb{R}^{p_2} : Gx + H^1 w^1 + H^2 w^2 \geq b\}$ *such that* $X = proj_x Q_I$.

Theorem

Every IP set $X = \{x \in \mathbb{Z}^n : Ax \geq a\}$ *can be represented in the form* $X = proj_x(Q_I)$, *where*

$$Q_I = \{(x, \lambda, \mu) \in \mathbb{R}^n \times \mathbb{Z}_+^{|G|} \times \mathbb{Z}_+^{|R|} : \\ x = \sum_{g \in G} \lambda_g x^g + \sum_{r \in R} \mu_r v_r, \ \sum_{g \in G} \lambda_g = 1\}$$

where $\{x^g\}_{g \in G}$ *is a finite set of integer points in* X *and* $\{v^r\}_{r \in R}$ *the extreme rays of* $conv(X)$ *(scaled to integrality).*

Decompositions

Consider a (minimization) problem (IP) $\min\{cx : x \in X\}$ with the property that a subset of the constraints of X defines a set $Z \subset X$ over which optimization easier.

$$(IP) \quad \min\{cx : Dx \geq d, Bx \geq b, x \in \mathbb{Z}_+^n\}$$

where the constraints $Bx \geq b$ define a set $Z = \{x \in \mathbb{Z}_+^n : Bx \geq b\}$ that is "tractable".

In many cases B has a block diagonal structure, i.e., $Z = Z^1 \times Z^2 \times \ldots Z^K$ an IP has the form:

$$\begin{array}{lccccccccc}
\min & c^1x^1 & + & c^2x^2 & + & \ldots & + & c^nx^n & \\
 & D^1x^1 & + & D^2x^2 & + & \ldots & + & D^nx^n & \geq d \\
 & B^1x^1 & + & & & & & & \geq b^1 \\
 & & + & B^2x^2 & & & & & \geq b^2 \\
 & & & & & \ldots & & & \\
 & & + & & & & & B^Kx^K & \geq b^K \\
 & x^1 \in \mathbb{Z}_+^{n_1}, & & x^2 \in \mathbb{Z}_+^{n_2}, & & \ldots & x^K \in \mathbb{Z}_+^{n_K}. & &
\end{array}$$

Decompositions

In the case of identical sub-problems we have $D^K = D,\ B^K = B,\ c^K = c,\ Z^K = Z^*$ for all K. We can define the aggregate variables $y = \sum_{k=1}^{K} x^k$ and the problem is:

$$(IP) \quad \min\{cy : Dy \geq d,\ y = \sum_{k=1}^{K} x^k,\ x^k \in Z^*,\ k = 1, \ldots, K\}$$

Lagrangean relaxation

The idea is to turn the difficult constraints $Dx \geq d$ into constraints that can be violated at a non-negative price π.

Definition

Lagrangean sub-problem:
$L(\pi) = \min_x \{cx + \pi(d - Dx) : Bx \geq b, x \in \mathbb{Z}^n_+\}$.

For any $\pi \geq 0$, $L(\pi)$ defines a lower bound on the optimal value z of IP. Indeed the optimal solution x^* satisfies: $cx^* \geq cx^* + \pi(d - Dx^*) \geq L(\pi)$. The problem of maximizing this bound over vectors $\pi \geq 0$ is known as the Lagrangean dual:

Definition

Lagrangean dual: $z_{LD} = \max_{\pi \geq 0} L(\pi) = \max_{\pi \geq 0} \min_{x \in Z} \{cx + \pi(d - Dx)\}$

Lagrangean relaxation

Assuming that Z is non-empty and bounded, the Lagrangean sub-problem achieves its optimum at an extreme point $x^t \in conv(Z)$, so one can write

$$z_{LD} = \max_{\pi \geq 0} \min_{t=1,\ldots,T} cx^t + \pi(d - Dx^t),$$

where $\{x^t\}_{t=1,\ldots,T}$ is the set of extreme points of $conv(Z)$. Introducing an additional variable σ representing a lower bound on $(c - \pi D)x^t$, we get:

$$\begin{aligned} z_{LD} = \max \pi d + \sigma & \\ \pi D x^t + \sigma \leq cx^t & \quad t = 1, \ldots, T \\ \pi \geq 0,\ \sigma \in \mathbb{R} & \end{aligned}$$

Lagrangean relaxation

Taking the dual:

$$z_{LD} = \max \pi d + \sigma$$
$$\pi D x^t + \sigma \leq c x^t \quad \forall t$$
$$\pi \geq 0, \ \sigma \in \mathbb{R}$$

$$z_{LD} = \min \sum_{t=1}^{T} (cx^t)\lambda_t$$
$$\sum_{t=1}^{T} (Dx^t)\lambda_t \geq d$$
$$\sum_{t=1}^{T} \lambda_t = 1$$
$$\lambda_t \geq 0, \ \forall t.$$

Theorem

Lagrangean duality: $z_{LD} = \min\{cx : \ Dx \geq d, x \in conv(Z)\}$

$(conv(Z) = \{x = \sum_{t=1}^{T} x^t \lambda_t : \sum_{t=1}^{T} \lambda_t = 1, \lambda_t \geq 0, t = 1, \ldots, T\})$

Lagrangean relaxation - sub-gradient algorithm

The sub-gradient algorithm for the Lagrangean dual in form:
$z_{LD} = \max_{\pi \geq 0} \min_{t=1,\dots,T} \{cx^t + \pi(d - Dx^t)\}$

- Initialize $\pi_0 = 0, t = 1$.
- At iteration t, solve the Lagrangean subproblem to obtain the dual bound $L(\pi^t) = \min\{cx + \pi^t(dDx)\}$ and an optimal solution x^t.
- Compute the violation of the dualized constraints $(d - Dx^t)$; this provides a "sub-gradient" that can be used to modify the dual variables.
- Update the dual solution using $\pi^{t+1} = max\{0, \pi^t + \epsilon^t(d - Dx^t)\}$ where ϵ_t is an appropriately chosen step-size.
- If $t \leq \tau$, increment t and iterate.

Lagrangean relaxation - sub-gradient algorithm

Usually a normalized step-size is used:

$$\epsilon_t = \frac{\alpha_t}{||d - Dx^t||} \tag{1}$$

When the α_t form a divergent series: $\alpha_t \to 0$ and $\sum_t \alpha_t \to \infty$ convergence of the sub-gradient method is guaranteed.

Dantzig-Wolfe reformulations

Consider problem IP : $\min\{cx : Dx \geq d,\ x \in Z\}$, and assume Z is bounded. From Minkowski Theorem we have:

Definition

Dantzig-Wolfe reformulation: convexification approach DW_c

$$\begin{aligned} \min \sum_{g \in G^c} (cx^g)\lambda_g \\ \sum_{g \in G^c} (Dx^g)\lambda_g \geq d \\ \sum_{g \in G^c} \lambda_g = 1 \\ x = \sum_{g \in G^c} x^g \lambda_g \in \mathbb{Z}^n \\ \lambda_g \geq 0,\ g \in G^c \end{aligned}$$

where $\{x^g\}_{g \in G^c}$ are the extreme points of $conv(Z)$.

Dantzig-Wolfe reformulations

From the Minkowski Theorem for IPs we have:

Definition

Dantzig-Wolfe reformulation: discretization approach DW_d

$$\min \sum_{g \in G^d} (cx^g)\lambda_g$$

$$\sum_{g \in G^d} (Dx^g)\lambda_g \geq d$$

$$\sum_{g \in G^d} \lambda_g = 1$$

$$\lambda_g \in \{0,1\},\ g \in G^d$$

where $\{x^g\}_{g \in G^d}$ are all the points of Z.

In general the extreme points of $conv(Z)$ are a subset of the points of Z.The two approaches are equivalent when considering LP relaxations.

Dantzig-Wolfe reformulations

- The Lagrangean dual problem is the same as the LP relaxation of DW_c;
- $z_{LP}^{DW_c} = z_{LP}^{DW_d} = \min\{cx : Dx \geq d,\ x \in conv(Z)\} = z_{LD}$.

When there is block diagonal structure, the DW_d reformulation is:

$$\min \sum_{k=1}^{K} \sum_{g \in G_k^d} (c^k x^g)\lambda_{kg} : \sum_{k=1}^{K} \sum_{g \in G_k^d} (D^k x^g)\lambda_{kg} \geq d; \sum_{g \in G_k^d} \lambda_{kg} = 1,$$
$$\lambda_{kg} \in \{0,1\},\ k = 1,\ldots,K,\ g \in G_k^d\}$$

where $Z^k = \{x^g\}_{g \in G_k^d}$ for all k, with $x^k = \sum_{g \in G_k^d} x^g \lambda_{kg} \in Z^k$.

Dantzig-Wolfe reformulations

When the subproblems are identical, in order to avoid symmetry, we use an aggregate variable $\nu_g = \sum_{k=1}^{K} \lambda_{kg}$. Defining $Z^* = Z^1 = \ldots = Z^K$ and $Z^* = \{x^g\}_{g \in G^*}$ we get:

$$\min \sum_{g \in G^*} (cx^g)\nu_g$$
$$\sum_{g \in G^*} (Dx^g)\nu_g \geq d$$
$$\sum_{g \in G^*} \nu_g = K$$
$$\nu_g \in \mathbb{Z},\ g \in G^*$$

where ν_g is the number of copies of x^g in the solution. By projecting ν on the original space we can only have the aggregate variables $y = \sum_{k=1}^{K} x^k = \sum_{k=1}^{K} \sum_{g \in G^*} x^g \lambda_{kg} = \sum_{g \in G^*} x^g \nu_g$.

Dantzig-Wolfe reformulation of the cutting stock model

We use *integer* variables $x_{ij},\ i = 1, \ldots, m,\ j = 1, \ldots, n$, denoting the number of copies of items of class i inserted into bin j; and *binary* variables $y_j,\ j = 1, \ldots, n$, taking value 1 when bin j is used.

$$\min \sum_{j=1}^{n} y_j$$

$$\sum_{j=1}^{n} x_{ij} = d_i, \qquad i = 1, \ldots, m$$

$$\sum_{i=1}^{m} w_i x_{ij} \leq C y_j, \qquad j = 1, \ldots, n$$

$$x_{ij} \ \textit{integer},\ y_j \in \{0, 1\}, \quad i = 1, \ldots, m;\ j = 1, \ldots, n$$

Solving the Dantzig-Wolfe LP relaxation by Column Generation

Consider the LP relaxation of DW_c or DW_d (master problem, MP) and suppose that only a subset $\{x^g\}_{g \in G'}$, with $G' \subset G$, are known. This is denoted as restricted master problem (RMP).

$$(MP)$$

$$\min \sum_{g \in G} (cx^g)\lambda_g$$

$$\sum_{g \in G} (Dx^g)\lambda_g \geq d$$

$$\sum_{g \in G} \lambda_g = 1$$

$$\lambda_g \geq 0, \;\; g \in G$$

The dual of the MP reads:

$$\max \pi d + \sigma$$

$$\pi D x^g + \sigma \leq cx^g, \quad g \in G$$

$$\pi \geq 0, \;\; \sigma \in \mathbb{R}$$

Let λ' and (π', σ') be a primal-dual solution of the RMP.

Solving the Dantzig-Wolfe LP relaxation by Column Generation

- Given a dual solution (π', σ') to the RMP, the reduced cost of the column associated with x^g is $cx^g - \pi' D x^g - \sigma'$.
- Let $\zeta = min_{g \in G}(cx^g - \pi' D x^g) = min_{x \in Z}(c - \pi' D)x$. Pricing can be carried out implicitly by solving a single integer program over the set Z.
- RMP is solved when $\zeta - \sigma' = 0$, i.e., when there is no column with negative reduced cost (eq., all dual constraints are satisfied).
- The pricing problem is equivalent to the Lagrangean sub-problem: $L(\pi') = \min_{x \in Z}\{cx + \pi'(d - Dx)\}$; hence, each pricing step provides a Lagrangean dual bound (by adding $\pi' d$).
- Equivalently, (π', ζ) is a feasible solution of the dual of the RMP, and therefore $\pi' d + \zeta$ gives a lower bound on z_{RMP}.

Column generation algorithm

- Initialize primal and dual bounds $PB = +\infty$, $DB = -\infty$. Generate a subset of points x^g so that the RMP is feasible.
- Iteratively, solve the RMP over the current set of columns; if the primal solution defines an integer solution, update PB. If $PB = DB$, stop.
- Solve the pricing problem $\zeta = min_{x \in Z}(c - \pi' D)x$. Let x be an optimal solution. If $\zeta - \sigma' = 0$, set $DB = z_{RMP}$ and stop; otherwise, add x to G' and include the associated column in the RMP.
- Compute the dual bound: $L(\pi) = \pi d + \zeta$; update $DB = max\{DB, L(\pi)\}$. If $PB = DB$, stop.

Solving the Dantzig-Wolfe LP relaxation by Column Generation

- When problem IP has a block diagonal structure, the dual constraints are: $\pi D^k x^g + \sigma^k \leq c^k x^g,\ g \in G^k,\ k = 1, \ldots, K$ and the k^{th} subproblem is $\zeta_k = min_{x \in Z^k}(c^k - \pi' D^k)x.$ The dual lower bound is of the form $\pi' d + \sum_{k=1}^{K} \zeta^k$.
- When the K subproblems are identical, this bound takes the form $\pi' d + K\zeta$.

Stabilization of column generation

Convergence of column generation can be very slow, mainly because of:

- primal degeneracy;
- oscillations in the values of the dual variables π.

$$
\begin{aligned}
&[P]\\
\min\ & cx\\
Ax &= b\\
x &\geq 0
\end{aligned}
\qquad\qquad
\begin{aligned}
&[D]\\
\max\ & b\pi\\
\pi A &\leq c
\end{aligned}
$$

Stabilization of column generation

One way to overcome degeneracy is to perturb $[P]$ by adding bounded surplus and slack variables:

$$[P_\epsilon] \quad \min_{x,y_-,y_+\geq 0} \{cx : Ax - y_- + y_+ = b,\ y_- \leq \epsilon_-,\ y_+ \leq \epsilon_+\}$$

In addition, it is possible to reduce dual variables oscillations by narrowing the domain of the dual problem $[D]$:

$$[P_\delta] \quad \min_{x,y_-,y_+\geq 0} \{cx + \delta y_- + \delta y_+ : Ax - y_- + y_+ = b\}$$

This is equivalent to impose in the dual that: $-\delta \leq \pi \leq \delta$.

Stabilization of column generation

A possible way of merging the two approaches was proposed by du Merle et al. (DAM 194, 1999):

$$[\tilde{P}]$$

$$\begin{aligned} \min cx + \delta y_- + \delta y_+ \\ Ax - y_- + y_+ = b \\ y_- \leq \epsilon_- \\ y_+ \leq \epsilon_+ \\ x, y_-, y_+ \geq 0 \end{aligned}$$

$$[\tilde{D}]$$

$$\begin{aligned} \max b\pi - \epsilon_- w -_- \epsilon_+ w_+ \\ \pi A \leq c \\ -\pi - w_- \leq -\delta_- \\ \pi - w_+ \leq \delta_+ \\ w_-, w_+ \geq 0 \end{aligned}$$

In $[\tilde{P}]$, y are slack and surplus variables with bounds ϵ, and they are penalized in the objective function by δ.
In $[\tilde{D}]$, we have $\delta_- - w_- \leq \pi \leq \delta_+ + w_+$, thus we penalize π when they lie outside the interval $[\delta_-, \delta_+]$.
$[\tilde{P}] = [P]$ when either $\epsilon_- = \epsilon_+ = 0$, or $\delta_- < \pi < \delta_+$.