Large-scale Integer Linear Programming

https://bicmr.pku.edu.cn/~wenzw/bigdata2024.html

Outline

Lagrangian Relaxation

- Consider the integer programming problem

$$\begin{aligned} \max \quad & c^\top x, \\ \text{s.t.} \quad & Ax \le b,\ Dx \le d, \\ & x \in \mathbb{Z}^n, \end{aligned} \tag{1}$$

  and assume that A, D, b, c, d have integer entries.

- Let Z_{IP} the optimal cost and let

$$X = \{x \in \mathbb{Z}^n \mid Dx \le d\}. \tag{2}$$

  We assume that optimizing over the set X can be done efficiently.

- Let $\lambda \ge 0$ be a vector of dual variables. We introduce the problem

$$\begin{aligned} \max \quad & c^\top x + \lambda^\top (b - Ax), \\ \text{s.t.} \quad & x \in X, \end{aligned} \tag{3}$$

  and denote its optimal cost by $Z(\lambda)$.

Lagrangian Relaxation

Lemma

If the problem (1) *has an optimal solution and if* $\lambda \geq 0$, *then* $Z(\lambda) \geq Z_{IP}$

- **Proof:** Let x^* denote an optimal solution to (1).
  Then, $b - Ax^* \geq 0$ and, therefore

$$c^\top x^* + \lambda^\top (b - Ax^*) \geq c^\top x^* = Z_{IP}.$$

  Since $x^* \in X$,

$$Z(\lambda) \geq c^\top x^* + \lambda^\top (b - Ax^*) \geq c^\top x^* = Z_{IP}.$$

- Problem (3) provides an upper bound to (1). It is natural to consider the tightest such bound.

Lagrangian Dual

- We introduce the problem

$$\min \; Z(\lambda), \quad \text{s.t.} \;\; \lambda \geq 0. \tag{4}$$

- We will refer to problem (4) as the Lagrangian dual. Let

$$Z_D = \min_{\lambda \geq 0} \; Z(\lambda).$$

- Suppose $X = \{x^1, \cdots, x^m\}$. Then $Z(\lambda)$ can be written as

$$Z(\lambda) = \max_{i=1,\cdots,m} (c^\top x^i + \lambda^\top (b - Ax^i)).$$

- The function $Z(\lambda)$ is convex and piecewise linear.
- Computing Z_D can be recast as a linear programming problem with a very large number of constraints.

Weak Duality

Theorem (Weak Duality)

We have $Z_D = \min_{\lambda \geq 0} \; Z(\lambda) \geq Z_{IP}$.

- The previous theorem represents the weak duality theory of integer programming.
- Unlike linear programming, integer programming does not have a strong duality theory. It is possible to have $Z_D > Z_{IP}$.
- The procedure of obtaining bounds for integer programming problems by calculating Z_D is called *Lagrangian relaxation*.

Strength of the Lagrangian Dual

Theorem

The optimal value Z_D of the Lagrangian dual is equal to the optimal cost of the following linear programming problem:

$$\begin{aligned} \max \quad & c^\top x, \\ \textit{s.t.} \quad & Ax \le b, x \in \mathrm{conv}(X). \end{aligned} \tag{5}$$

where $\mathrm{conv}(X)$ be the convex hull of the set $X = \{x \in \mathbb{Z}^n \mid Dx \le d\}$.

Proof:

$$Z(\lambda) = \max_{x \in X} \quad (c^\top x + \lambda^\top (b - Ax)).$$

- The optimal cost remains same if we allow convex combinations of the elements of X.

$$Z(\lambda) = \max_{x \in \mathrm{conv}(X)} \quad (c^\top x + \lambda^\top (b - Ax)).$$

Proof

- By definition, we have

$$Z_D = \min_{\lambda \geq 0} Z(\lambda) = \min_{\lambda \geq 0} \max_{x \in \mathrm{conv}(X)} (c^\top x + \lambda^\top (b - Ax)).$$

- Let $\{v^k, k \in K\}$ be the extreme points, and $\{r^j, j \in J\}$ be the complete set of extreme rays of $\mathrm{conv}(X)$.
- Then, for any fixed λ, we have

$$Z(\lambda) = \begin{cases} +\infty, & \exists\, j \in J, (c^\top - \lambda^\top A) r^j > 0, \\ \max_{k \in K} (c^\top v^k + \lambda^\top (b - Av^k)), & \text{otherwise.} \end{cases} \tag{6}$$

Proof

- According to (6), the Lagrangian dual is equivalent to and has the same optimal value as the problem

$$\begin{aligned} \min_{\lambda \geq 0} \quad & \max_{k \in K} (c^\top v^k + \lambda^\top (b - Av^k)), \\ \text{s.t.} \quad & (c^\top - \lambda^\top A) r^j \leq 0, \quad j \in J. \end{aligned} \tag{7}$$

- Problem (7) is equivalent to the linear programming problem

$$\begin{aligned} \min_{\lambda \geq 0} \quad & y, \\ \text{s.t.} \quad & y + \lambda^\top (Av^k - b) \geq c^\top v^k, \quad k \in K, \\ & \lambda^\top A r^j \geq c^\top r^j, \quad j \in J. \end{aligned} \tag{8}$$

Proof

- Taking the linear programming dual of problem (8), and using strong duality, Z_D is equal to the optimal cost of the problem

$$\begin{aligned} \max \quad & c^\top \left(\sum_{k \in K} \alpha_k v^k + \sum_{j \in J} \beta_j r^j \right), \\ \text{s.t.} \quad & A \left(\sum_{k \in K} \alpha_k v^k + \sum_{j \in J} \beta_j r^j \right) \le b, \\ & \sum_{k \in K} \alpha_k = 1, \quad \alpha_k, \beta_j \ge 0. \end{aligned}$$

- The result follows since

$$\operatorname{conv}(X) = \left\{ \sum_{k \in K} \alpha_k v^k + \sum_{j \in J} \beta_j r^j \;\middle|\; \sum_{k \in K} \alpha_k = 1, \alpha_k, \beta_j \ge 0 \right\}$$

Linear Relaxation

- We have characterized the optimal value of the Lagrangian dual as solution to a linear programming problem.
- It is natural to compare the optimal cost Z_{IP} and the optimal cost Z_{LP} of the linear relaxation

$$
\begin{array}{ll}
\max & c^\top x, \\
\text{s.t.} & Ax \leq b, \; Dx \leq d.
\end{array}
$$

- In general, the following ordering holds among Z_{LP}, Z_{IP}, and Z_D:

$$Z_{LP} \geq Z_D \geq Z_{IP}.$$

Linear Relaxation

- We have $Z_{IP} = Z_D$ for all cost vector c, if and only if

$$\text{conv}\left(X \cap \{x \mid Ax \leq b\}\right) = \text{conv}(X) \cap \{x \mid Ax \leq b\}.$$

- We have $Z_{LP} = Z_D$ for all cost vector c, if

$$\text{conv}(X) = \{x \mid Dx \leq d\}.$$

Solution of the Lagrangian Dual

- We outline a method for finding the optimal Lagrangian multipliers λ^*, that solve the Lagrangian dual problem

$$\min \; Z(\lambda), \quad \textsf{s.t.} \;\; \lambda \geq 0.$$

- To keep the presentation simple, we assume that X is finite and $X = \{x^1, \cdots, x^m\}$.
- Given a particular value of λ, we assume that we can calculate $Z(\lambda)$, which we have defined as follows:

$$Z(\lambda) = \max_{i=1,\cdots,m} (c^\top x^i + \lambda^\top (b - Ax^i)).$$

Subgradient

- Let $\mathrm{f}_i = b - Ax^i$ and $h_i = c^\top x^i$. Then,

$$Z(\lambda) = \max_{i=1,\cdots,m} (h_i + \mathrm{f}_i^\top \lambda).$$

- Let $E(\lambda) = \{i \mid Z(\lambda) = h_i + \mathrm{f}_i^\top \lambda\}$.
- For every $i \in E(\lambda^*)$, f_i is a subgradient of the function $Z(\cdot)$ at λ^*.
- $\partial Z(\lambda^*) = \mathrm{conv}(\{\mathrm{f}_i, i \in E(\lambda^*)\})$, i.e., a vector s is a subgradient of the function $Z(\cdot)$ at λ^* if and only if s is a convex combination of the vectors $\mathrm{f}_i, i \in E(\lambda^*)$.

Subgradient Optimization Algorithm

The following algorithm generalizes the steepest ascent algorithm to maximize a nondifferentiable concave function $Z(\cdot)$.

1. Choose a starting point λ^1; let $t = 1$.
2. Given λ^t, choose a subgradient s^t of the function $Z(\cdot)$ at λ^t.
3. If $s^t = 0$, then λ^t is optimal and the algorithm terminates. Else, continue.
4. Let $\lambda_j^{t+1} = \max\{\lambda_j^t - \theta_t s_j^t, 0\}$, where θ_t is a positive stepwise parameter. Increment t and go to Step 2.

- Typically, only the extreme subgradients f_i are used.
- The stopping criterion $0 \in \partial Z(\lambda^t)$ is rarely met. Typically, the algorithm is stopped after a fixed number of iterations.

Stepsize

- It can be proved that $Z(\lambda^t)$ converges for any stepsize sequence θ_t such that

$$\sum_{t=1}^{\infty} \theta_t = \infty, \quad \text{and} \quad \lim_{t \to \infty} \theta_t = 0.$$

- An example of the stepsize sequence is $\theta_t = 1/t$, which leads to slow convergence in practical. Another example is

$$\theta_t = \theta_0 \alpha^t, \quad t = 1, 2, \cdots,$$

where α is a scalar satisfying $0 < \alpha < 1$.

- A more sophisticated and popular rule is to let

$$\theta_t = \frac{Z(\lambda^t) - \hat{Z}_D}{\|s^t\|^2} \alpha$$

where α is a scalar satisfying $0 < \alpha < 1$ and $\hat{Z}_D$ is an estimate of the optimal value Z_D.

Outline

1. Lagrangian Relaxation
2. Dantzig-Wolfe decomposition
3. Bender's Decomposition

Mixed Integer Program

- Let us consider a mixed integer program (MIP)

$$\begin{aligned} z_I \quad = \quad \max \quad & c^T x, \\ \text{s.t.} \quad & Ax \le b, Dx \le d, \\ & x \in \mathbb{Z}_+^d \times R_+^p. \end{aligned} \tag{9}$$

- Let X be defined as

$$X = \left\{ x \in \mathbb{Z}_+^d \times R_+^p : \; Dx \le d \right\}.$$

We assume that X is nonempty and D, d have rational entries.

Lagrangian dual

- Let m be the number of rows of A, and take $\lambda \in \mathbb{R}^m_+$. The Lagrangian relaxation with respect to λ as follows.

$$\begin{aligned} z_{LR}(\lambda) \quad = \quad \max \quad & c^\top x + \lambda^\top (b - Ax), \\ \text{s.t.} \quad & Dx \leq d, \\ & x \in \mathbb{Z}^q_+ \times R^p_+. \end{aligned} \tag{10}$$

- Moreover, recall that the Lagrangian dual is defined as

$$z_{LD} \quad = \quad \min\{Z_{LR}(\lambda) : \; \lambda \geq 0\}. \tag{11}$$

- (10) and (11) are related according to the following characterization of z_{LD}.

$$z_{LD} = \max\{c^\top x : \; Ax \leq b, x \in \text{conv}(X)\}.$$

Decomposition of $\mathrm{conv}(X)$

- $\mathrm{conv}(X)$ can be expressed as

$$\mathrm{conv}(X) = \mathrm{conv}\left\{v^1, \ldots, v^n\right\} + \mathrm{cone}\left\{r^1, \ldots, r^\ell\right\},$$

  where $v^1, \ldots, v^n$ are the extreme points of $\mathrm{conv}(X)$ and $r^1, \ldots, r^\ell$ are the extreme rays of $\mathrm{conv}(X)$.
- Any point x in $\mathrm{conv}(X)$ can be written as

$$x = \sum_{k \in [n]} \alpha_k v^k + \sum_{h \in [\ell]} \beta_h r^h$$

  for some $\alpha \in \mathbb{R}^k_+$ and $\beta \in \mathbb{R}^\ell_+$ such that $\sum_{k \in [n]} \alpha_k = 1$.

Dantzig-Wolfe Relaxation

Based on the decomposition of $\mathrm{conv}(X)$, it follows that

$$
\begin{aligned}
z_{\mathrm{LD}} = \max \quad & \sum_{k\in[n]} \left(c^\top v^k\right)\alpha_k + \sum_{h\in[\ell]} \left(c^\top r^h\right)\beta_k, \\
\text{s.t.} \quad & \sum_{k\in[n]} \left(Av^k\right)\alpha_k + \sum_{h\in[\ell]} \left(Ar^h\right)\beta_k \le b, \\
& \sum_{k\in[n]} \alpha_k = 1, \alpha \in \mathbb{R}^k_+, \beta \in \mathbb{R}^\ell_+.
\end{aligned}
\tag{12}
$$

We refer to (12) as the Dantzig-Wolfe relaxation.

Dantzig-Wolfe Reformulation

- Moreover, we have
$$z_I = \max\left\{c^\top x : Ax \le b, x \in \operatorname{conv}(X), x_j \in \mathbb{Z},\ \forall j \in [q]\right\}.$$

- Therefore, we deduce
$$\begin{aligned} z_I = \max & \sum_{k\in[n]} \left(c^\top v^k\right)\alpha_k + \sum_{h\in[\ell]} \left(c^\top r^h\right)\beta_k, \\ \text{s.t. } & \sum_{k\in[n]} \left(Av^k\right)\alpha_k + \sum_{h\in[\ell]} \left(Ar^h\right)\beta_k \le b, \\ & \sum_{k\in[n]} \alpha_k = 1, \\ & \alpha \in \mathbb{R}^k_+, \beta \in \mathbb{R}^\ell_+, \\ & \sum_{k\in[n]} \alpha_k v^k_j + \sum_{h\in[\ell]} \beta_h r^h_j \in \mathbb{Z}, \quad j \in [q]. \end{aligned} \tag{13}$$

- Here, (13) is referred to as the Dantzig-Wolfe reformulation.

Pure Binary Programs

- Let us consider a pure binary integer program as follows.

$$\begin{aligned} z_I = \max \quad & c^\top x, \\ \text{s.t.} \quad & Ax \le b, Dx \le d, \\ & x \in \{0,1\}^p. \end{aligned}$$

- We define X as

$$X = \{x \in \{0,1\}^p : Dx \le d\}.$$

- Since X is bounded and finite, $X = \{v^1, \ldots, v^n\}$
- Any point x in X can be expressed as

$$x = \sum_{k \in [n]} \alpha_k v^k, \quad \sum_{k \in [n]} \alpha_k = 1, \quad \alpha \in \{0,1\}^n.$$

Pure Binary Programs

- Then we obtain the Dantzig-Wolfe reformulation.

$$\begin{aligned} z_I = \max \quad & \sum_{k\in[n]} \left(c^\top v^k\right) \alpha_k, \\ \text{s.t.} \quad & \sum_{k\in[n]} \left(Av^k\right) \alpha_k \leq b, \\ & \sum_{k\in[n]} \alpha_k = 1, \quad \alpha \in \{0,1\}^n. \end{aligned}$$

- The Dantzig-Wolfe relaxation

$$\begin{aligned} \max \; & \sum_{k\in[n]} \left(c^\top v^k\right) \alpha_k, \\ \text{s.t.} \; & \sum_{k\in[n]} \left(Av^k\right) \alpha_k \leq b, \\ & \sum_{k\in[n]} \alpha_k = 1, \quad \alpha \geq 0. \end{aligned}$$

Block Diagonal Structure

- We consider the following optimization model with block diagonal structure.

$$
\begin{array}{lllll}
\max & c^{1\top}x^1 + & c^{2\top}x^2 + \cdots & + c^{p\top}x^p, & \\
\text{s.t.} & D^1x^1 & & & \leq d^1, \\
 & & D^2x^2 & & \leq d^2, \\
 & & & \ddots & \\
 & & & D^px^p & \leq d^p, \\
 & A^1x^1 + & A^2x^2 + \cdots & + A^px^p & \leq b, \\
 & x^j \in \{0,1\}^{q_j}, & j \in [p]. & &
\end{array}
$$

- For $j \in [p]$, let X_j be defined as $X_j = \left\{x^j \in \{0,1\}^{q_j} : D^jx^j \leq d^j\right\}$.
- X_j is bounded and finite. Any point x^j in X_j can be written as

$$
x^j = \sum_{v \in X_j} \alpha^j_v v, \quad \sum_{v \in X_j} \alpha^j_v = 1, \quad \alpha^j \in \{0,1\}^{|X_j|}.
$$

Block Diagonal Structure

- The Dantzig-Wolfe reformulation is given by

$$\begin{aligned} \max \quad & \sum_{v \in X_1} \left(c^{1\top} v \right) \alpha_v^1 + \sum_{v \in X_2} \left(c^{2\top} v \right) \alpha_v^2 + \cdots + \sum_{v \in X_p} \left(c^{p\top} v \right) \alpha_v^p, \\ \text{s.t.} \quad & \sum_{v \in X_1} \left(A^1 v \right) \alpha_v^1 + \sum_{v \in X_2} \left(A^2 v \right) \alpha_v^2 + \cdots + \sum_{v \in X_p} \left(A^p v \right) \alpha_v^p \le b, \\ & \sum_{v \in X_j} \alpha_v^j = 1, \quad \alpha^j \in \{0,1\}^{|X_j|}, \quad j \in [p]. \end{aligned}$$

- The Dantzig-Wolfe relaxation is given by

$$\begin{aligned} \max \quad & \sum_{v \in X_1} \left(c^{1\top} v \right) \alpha_v^1 + \sum_{v \in X_2} \left(c^{2\top} v \right) \alpha_v^2 + \cdots + \sum_{v \in X_p} \left(c^{p\top} v \right) \alpha_v^p, \\ \text{s.t.} \quad & \sum_{v \in X_1} \left(A^1 v \right) \alpha_v^1 + \sum_{v \in X_2} \left(A^2 v \right) \alpha_v^2 + \cdots + \sum_{v \in X_p} \left(A^p v \right) \alpha_v^p \le b, \\ & \sum_{v \in X_j} \alpha_v^j = 1, \quad \alpha^j \ge 0, \quad j \in [p]. \end{aligned}$$

Column Generation: Master Problem

- The Dantzig-Wolfe relaxation has variables $\alpha_1, \ldots, \alpha_n$ for the extreme points of $\mathrm{conv}(X)$ and variables $\beta_1, \ldots, \beta_\ell$ for the extreme rays of $\mathrm{conv}(X)$.
- n and ℓ are potentially very large. In this case, we may apply the column generation technique.
- The column generation procedure works as follows. We start with $N \subseteq [n]$ and $L \subseteq [\ell]$. Then we have the master problem

$$\begin{aligned} \max \quad & \sum_{k \in N} \left(c^\top v^k\right) \alpha_k + \sum_{h \in L} \left(c^\top r^h\right) \beta_k, \\ \text{s.t.} \quad & \sum_{k \in N} \left(Av^k\right) \alpha_k + \sum_{h \in L} \left(Ar^h\right) \beta_k \leq b, \\ & \sum_{k \in N} \alpha_k = 1, \quad \alpha \in \mathbb{R}^k_+, \beta \in \mathbb{R}^\ell_+. \end{aligned}$$

- 子问题的构造基于对偶理论，其目标是在当前主问题解的对偶向量λ 下，检验是否存在未加入的列（变量）仍可提升目标函数。
- 设当前主问题对偶解为λ，则任意潜在变量$x \in \text{conv}(X)$ 的reduced cost 为：

$$\bar{c}(x) = c^\top x - \lambda^\top Ax.$$

- 若存在$x \in \text{conv}(X)$ 使得$\bar{c}(x) > 0$，则当前主问题并非最优。
- 因此，构造如下子问题以判定最优性并生成新列：

$$\max_{x \in \text{conv}(X)} \left(c^\top + \lambda^\top (b - Ax) \right).$$

- 三种情形：
  - 若最优值> 0：存在可提升目标的极点/极射线；
  - 若最优值$= 0$：当前已最优，无需再加列；
  - 若最优值为$+\infty$：存在无界方向（极射线）可继续改进。

1. 解主问题（restricted master problem），得到对偶变量λ；
2. 构造子问题：

$$\max_{x \in \mathrm{conv}(X)} \quad c^\top x + \lambda^\top (b - Ax);$$

3. 若子问题最优值> 0，则：
   - 若最优解为极点v^k，则有

     $$(Av^k - b)^\top \lambda < c^\top v^k,$$

     加入v^k 到主问题；
   - 若子问题无界，存在极射线r^h，则有

     $$(Ar^h)^\top \lambda < c^\top r^h,$$

     加入r^h 到主问题；
4. 若子问题最优值≤ 0，则当前解已为全局最优，算法终止。

注： 子问题等价于在X 上求解线性目标$\left(c - A^\top \lambda\right)^\top x$ 的最优解，是pricing problem 的具体实现。

Outline

Bender's Decomposition

- We use the Lagrangian relaxation framework to deal with complicating constraints.
- In this section, we learn the Bender's reformulation technique that can deal with complicating variables.
- Consider the following mixed-integer program.

$$
\begin{aligned}
z_I = \max \quad & c^\top x + q^\top y, \\
\text{s.t.} \quad & Ax + Gy \leq b, \\
& x \in \mathbb{Z}_+^d, y \in \mathbb{R}_+^p.
\end{aligned}
$$

Bender's Decomposition

- Here, the integer variables x are complicating variables. If we fix the x part, then the optimization problem becomes

$$\begin{aligned} z_{LP}(x) = \max \quad & q^\top y, \\ \text{s.t.} \quad & Gy \leq b - Ax, \\ & y \in \mathbb{R}^p_+. \end{aligned}$$

- Taking the dual of it, we deduce

$$\begin{aligned} \min \quad & u^\top (b - Ax), \\ \text{s.t.} \quad & G^\top u \geq q, \\ & u \geq 0. \end{aligned}$$

- Here, the feasible set of the dual does not depend on x.

Bender's Decomposition

- Let Q denote the feasible set of the dual:

$$Q = \left\{ u : G^\top u \geq q, u \geq 0 \right\}.$$

- Suppose that Q can be expressed as

$$Q = \operatorname{conv}\left\{ v^1, \ldots, v^n \right\} + \operatorname{cone}\left\{ r^1, \ldots, r^\ell \right\}.$$

for some vectors $v^1, \ldots, v^n$ and $r^1, \ldots, r^\ell$.
- We will prove the following theorem.

Theorem (Bender's Decomposition)

The mixed integer program can be reformulated as

$$\begin{aligned} z_I = \max \quad & \eta, \\ \textit{s.t.} \quad & \eta \leq c^\top x + (b - Ax)^\top v^k, \quad k \in [n], \\ & (b - Ax)^\top r^h \geq 0, \quad h \in [\ell], \\ & x \in \mathbb{Z}_+^d, \quad \eta \in \mathbb{R}. \end{aligned}$$

Projection Theorem of Egon Balas

Theorem

Let $P = \left\{(x, y) \in \mathbb{R}^d \times \mathbb{R}^p : Ax + Gy \leq b, y \geq 0\right\}$*. Suppose that* $C = \left\{u : G^\top u \geq 0, u \geq 0\right\}$ *can be expressed as* $C = \operatorname{cone}\left\{r^1, \ldots, r^\ell\right\}$*. Then* $\operatorname{proj}_x(P)$*, the projection of* P *onto the* x*-space, is given by*

$$\operatorname{proj}_x(P) = \left\{x \in \mathbb{R}^d : (b - Ax)^\top r^h \geq 0, h \in [\ell]\right\}.$$

- Let $\bar{x} \in \mathbb{R}^d$. Note that $\bar{x} \notin \operatorname{proj}_x(P)$ holds if and only if there is no $y \in \mathbb{R}^p$ that satisfies $Gy \leq b - A\bar{x}$ and $y \geq 0$.
- By Farkas' Lemma, the system $Gy \leq b - A\bar{x}, y \geq 0$ is infeasible if and only if there exists $u \in C$ such that $u^\top(b - A\bar{x}) < 0$.
- Since $C = \operatorname{cone}\left\{r^1, \ldots, r^\ell\right\}$, such a vector u exists if and only if $(b - A\bar{x})^\top r^h \leq 0$ for some $h \in [\ell]$, in which case, $\bar{x} \notin \left\{x \in \mathbb{R}^d : (b - Ax)^\top r^h \geq 0, h \in [\ell]\right\}$.

Proof of Bender's Decomposition

- Let $P = \{(x, y) \in \mathbb{R}^d \times \mathbb{R}^p : Ax + Gy \leq b, y \geq 0\}$. Note that

$$\begin{aligned} z_I = \max \quad & c^\top x + z_{LP}(x), \\ \text{s.t.} \quad & x \in \mathbb{Z}^d_+. \end{aligned}$$

- Here, $z_{LP}(x) > -\infty$ if and only if there exists some $y \geq 0$ such that $Gy \leq b - Ax$, which is equivalent to $x \in \mathrm{proj}_x(P)$.
- Therefore, it follows that

$$\begin{aligned} z_I = \max \quad & c^\top x + z_{LP}(x), \\ \text{s.t.} \quad & x \in \mathrm{proj}_x(P) \cap \mathbb{Z}^d_+. \end{aligned}$$

Proof of Bender's Decomposition

- Recall that $Q = \{u : G^\top u \geq q, u \geq 0\}$ and

$$Q = \operatorname{conv}\{v^1, \ldots, v^n\} + \operatorname{cone}\left\{r^1, \ldots, r^\ell\right\}.$$

- Then $C = \{u : G^\top u \geq 0, u \geq 0\}$ is the recession cone of Q, so we have $C = \operatorname{cone}\{r^1, \ldots, r^\ell\}$.
- Then it follows from projection theorem of Egon Balas that $\operatorname{proj}_x(P) = \{x \in \mathbb{R}^d : (b - Ax)^\top r^h \geq 0, h \in [\ell]\}$.
- Therefore, we deduce that

$$\begin{aligned} z_I = \max \quad & c^\top x + z_{LP}(x), \\ \text{s.t.} \quad & (b - Ax)^\top r^h \geq 0, \quad h \in [\ell], \\ & x \in \mathbb{Z}_+^d. \end{aligned}$$

Proof of Bender's Decomposition

- Moreover, note that for any $x \in \text{proj}_x(P), z_{LP}(x) > -\infty$, so strong duality implies that

$$\begin{aligned} z_{LP}(x) = \min \quad & u^\top (b - Ax), \\ \text{s.t.} \quad & G^\top u \geq q, \\ & u \geq 0. \end{aligned}$$

- If $z_{LP}(x)$ is finite, then it means that Q is non-empty and

$$z_{LP}(x) = \min_{k \in [n]} \left\{ (b - Ax)^\top v^k \right\}.$$

- If $z_{LP}(x) = +\infty$, then Q is empty, so the above equation also holds. Hence,

$$\begin{aligned} z_I = \max \quad & c^\top x + \min_{k \in [n]} \left\{ (b - Ax)^\top v^k \right\}, \\ \text{s.t.} \quad & (b - Ax)^\top r^h \geq 0, \quad h \in [\ell], \\ & x \in \mathbb{Z}^d_+. \end{aligned}$$

Proof of Bender's Decomposition

- We may move the term $\min_{k\in[n]}\left\{(b-Ax)^\top v^k\right\}$ in the objective to constraints, after which we deduce that

$$\begin{aligned} z_I = \max \quad & \eta, \\ \text{s.t.} \quad & \eta \le c^\top x + \min_{k\in[n]}\left\{(b-Ax)^\top v^k\right\}, \\ & (b-Ax)^\top r^h \ge 0, \quad h \in [\ell], \\ & x \in \mathbb{Z}^d_+, \quad \eta \in \mathbb{R}. \end{aligned}$$

which is equivalent to Bender's reformulation as required.

Bender's Decomposition Algorithm

- The Bender's reformulation has an enormous number of constraints.
- A natural approach is to work with a small subset of the constraints and add new ones as cutting planes.
- The Bender's decomposition algorithm is the row generation framework for Bender's reformulation.

Master Problem

- At iteration t, we have $N_t \subseteq [n]$ and $L_t \subseteq [\ell]$. Then we solve

$$\begin{aligned} z_I^t = \max \quad & \eta, \\ \textbf{s.t.} \quad & \eta \le c^\top x + (b - Ax)^\top v^k, \quad k \in N_t, \\ & (b - Ax)^\top r^h \ge 0, \quad h \in L_t, \\ & x \in \mathbb{Z}_+^d, \eta \in \mathbb{R}. \end{aligned}$$

  This is the **master problem.**

- Assume that we get a solution (x^t, η^t) after solving the master problem at iteration t. Then we attempts to find a violated inequality among

$$\begin{aligned} & \eta \le c^\top x + (b - Ax)^\top v^k, \quad k \in [n] \backslash N_t, \\ & (b - Ax)^\top r^h \ge 0, \quad h \in [\ell] \backslash L_t. \end{aligned}$$

Subproblem

- The question is
  - does there exists $k_t \in [n]$ such that
$$\eta^t > c^\top x^t + (b - Ax^t)^\top v^{k_t}?$$
  - does there exists $h_t \in [\ell]$ such that
$$(b - Ax^t)^\top r^{h_t} < 0?$$
- To answer this, we solve
$$\begin{aligned} z_{LP}(x^t) = \max \quad & q^\top y, \\ \text{s.t.} \quad & Gy \leq b - Ax^t, \\ & y \in \mathbb{R}^p_+. \end{aligned}$$
- This is the **subproblem** for the Bender's decomposition algorithm.

Solving the Subproblem

- If $z_{LP}\left(x^t\right) = +\infty$, then for any $M > 0$, there exists $y \geq 0$ such that $Ax^t + Gy \leq b$ and $c^\top x^t + q^\top y > M$, in which case $z_I = +\infty$.
- If $z_{LP}\left(x^t\right)$ is finite, then

$$z_{LP}\left(x^t\right) = \min_{k\in[n]} \left(b - Ax^t\right)^\top v^k = \left(b - Ax^t\right)^\top v^{k_t}$$

  for some k_t.
- Hence, we deduce that

$$c^\top x^t + z_{LP}\left(x^t\right) = c^\top x^t + \left(b - Ax^t\right)^\top v^{k_t}.$$

- Moreover, if $z_{LP}\left(x^t\right) = -\infty$, then the subproblem is infeasible, in which case, there exists $h_t \in [\ell]$

$$\left(b - Ax^t\right)^\top r^{h_t} < 0.$$

Bender's decomposition algorithm

1. At iteration t, solve the master problem with $N_t \subseteq [n]$ and $L_t \subseteq [\ell]$.
2. If $z_I^t = -\infty$, then the mixed-integer program is infeasible.
3. Let (x^t, η^t) be an optimal solution to the master problem. Solve the subproblem with x^t.
4. If $z_{LP}(x^t) = +\infty$ then the mixed-integer program is unbounded.
5. If $z_{LP}(x^t) = -\infty$ then there exists $h_t \in [\ell]$ such that $(b - Ax)^\top r^{h_t} < 0$.
   Add constraint $(b - Ax)^\top r^{h_t} \geq 0$ and update $L_{t+1} = L_t \cup \{h_t\}$.
6. If $z_{LP}(x^t)$ is finite. Let y^t be an optimal solution and $k_t \in \operatorname{argmin}_{k \in [n]} \{(b - Ax^t)^\top > v^k\}$.
   If $c^\top x^t + q^\top y^t \geq \eta^t$, then we conclude that (x^t, y^t) is an optimal solution.
   If $c^\top x^t + q^\top y^t < \eta^t$, then we add constraint $\eta \leq c^\top x + (b - Ax)^\top v^{k_t}$ and update $N_{t+1} = N_t \cup \{k_t\}$.