Cutting Plane, Column generation and Dantzig–Wolfe decomposition

23 September 2008

A standard LP problem and its Lagrangian dual

$$
\begin{aligned}
v_{LP} = \text{minimum} \quad & \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}, \\
\text{subject to} \quad & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}, \\
& \boldsymbol{D}\boldsymbol{x} \leq \boldsymbol{d}, \\
& \boldsymbol{x} \in \mathbb{R}^n_+.
\end{aligned}
$$

- Let $X := \{\, \boldsymbol{x} \in \mathbb{R}^n_+ \mid \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b} \,\}$.
- We suppose for now that X is bounded.
- Further, let $P_X := \{\boldsymbol{x}^1, \boldsymbol{x}^2, \dots, \boldsymbol{x}^K\}$ be the set of extreme points in the polyhedron X.

- Its Lagrangian dual with respect to Lagrangian relaxing the constraints $\boldsymbol{D}\boldsymbol{x} \leq \boldsymbol{d}$ is to find

$$
\begin{aligned}
v_{LP} = v_L := \text{maximum} \quad & q(\boldsymbol{\mu}), \\
\text{subject to} \quad & \boldsymbol{\mu} \geq \boldsymbol{0},
\end{aligned}
$$

where

$$
\begin{aligned}
q(\boldsymbol{\mu}) &:= \underset{\boldsymbol{x} \in X}{\text{minimum}} \left\{ \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} + \boldsymbol{\mu}^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x} - \boldsymbol{d}) \right\} \\
&= \underset{i \in P_X}{\text{minimum}} \left\{ \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}^i + \boldsymbol{\mu}^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}^i - \boldsymbol{d}) \right\}.
\end{aligned}
$$

- Equivalent statement:

$$
q(\boldsymbol{\mu}) \leq \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}^i + \boldsymbol{\mu}^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}^i - \boldsymbol{d}), \qquad i \in P_X, \quad \boldsymbol{\mu} \geq \boldsymbol{0}.
$$

- So,

$$
\begin{aligned}
v_L := \text{maximum} \quad & z, \\
\text{subject to} \quad & z \leq \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}^i + \boldsymbol{\mu}^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}^i - \boldsymbol{d}), \qquad i \in P_X, \\
& \boldsymbol{\mu} \geq \boldsymbol{0}.
\end{aligned}
$$

- We know that if at an optimal dual solution $\boldsymbol{\mu}^*$, the set $X(\boldsymbol{\mu}^*)$ is a singleton, then thanks to strong duality this solution is optimal (and it is unique!). This typically does not happen, unless an optimal solution $\boldsymbol{x}^*$ happens to be an extreme point of X. We know, however, that $\boldsymbol{x}^*$ always can be written as a convex combination of such points. Let's see how it can be generated.

A cutting plane method for the Lagrangian dual problem

- Suppose only a subset of P_X is known, and consider the following restriction of the Lagrangian dual problem:

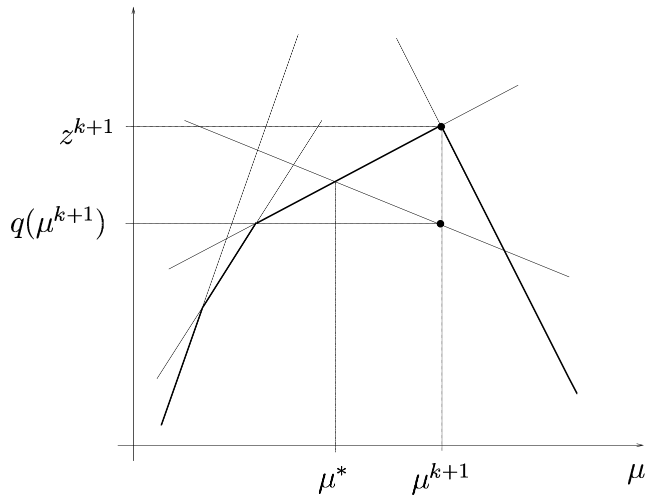
$$\begin{aligned} z^{k+1} := \max\ & z, && \text{(1a)}\\ \text{s.t. } & z \leq \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}^i + \boldsymbol{\mu}^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}^i - \boldsymbol{d}), \quad i = 1, \dots, k, && \text{(1b)}\\ & \boldsymbol{\mu} \geq \boldsymbol{0}. && \text{(1c)} \end{aligned}$$

- How do we determine if we have found the optimal solution? And what IS the optimal solution when we find it?

- Let $(\boldsymbol{\mu}^{k+1}, z^{k+1})$ be the solution to the above problem. If $z^{k+1} \leq \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}^i + (\boldsymbol{\mu}^{k+1})^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}^i - \boldsymbol{d})$ holds for all $i \in P_X$, then $\boldsymbol{\mu}^{k+1}$ is optimal in the dual! Why?
- How to check optimality: find the most violated dual constraint! That is, solve the subproblem to find

$$\begin{aligned} q(\boldsymbol{\mu}^{k+1}) &:= \underset{\boldsymbol{x} \in X}{\operatorname{minimum}} \left\{ \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} + (\boldsymbol{\mu}^{k+1})^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x} - \boldsymbol{d}) \right\} \qquad (2)\\ &= \underset{i \in P_X}{\operatorname{minimum}} \left\{ \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}^i + (\boldsymbol{\mu}^{k+1})^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}^i - \boldsymbol{d}) \right\}. \end{aligned}$$

- If $z^{k+1} \leq q(\boldsymbol{\mu}^{k+1})$ then $\boldsymbol{\mu}^{k+1}$ is optimal in the dual; otherwise, we have identified a constraint of the form $z \leq \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}^i + \boldsymbol{\mu}^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}^i - \boldsymbol{d})$, where $i \in P_X$, which is violated at $(\boldsymbol{\mu}^{k+1}, z^{k+1})$. Add this inequality and re-solve the LP problem!
- We refer to this algorithm as a *cutting plane* algorithm, for the reason that it is based on adding constraints to the dual problem in order to improve the solution, in the process cutting off the previous point.
- Consider the figure on the next slide. The thick lines correspond to the subset of k inequalities known at iteration k.



- Obviously, $z^{k+1} \geq q(\mu^{k+1})$ must hold, because of the possible lack of constraints. In this case, $z^{k+1} > q(\mu^{k+1})$ holds, so in the next step when we evaluate $q(\mu^{k+1})$ we can identify and add the last lacking inequality; the resulting maximization will then yield the optimal solution μ^* shown in the picture.
- How do we generate a primal optimal solution from this scheme? Let us look at the dual of the problem (1) in this cutting plane algorithm.

Duality relationships and the Dantzig–Wolfe algorithm

- We rewrite the problem (1) as follows:

$$\begin{aligned} &\underset{(z,\boldsymbol{\mu})}{\text{maximize}} \; z, \\ &\text{subject to } z - \boldsymbol{\mu}^{\mathrm{T}}(\boldsymbol{D}\boldsymbol{x}^i - \boldsymbol{d}) \leq \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}^i, \quad i = 1, \ldots, k, \\ &\qquad\qquad \boldsymbol{\mu} \geq \boldsymbol{0}. \end{aligned}$$

- With LP dual variables $\lambda_i \geq 0$ for the linear constraints, we obtain the LP dual to find

$$\begin{aligned} v^{k+1} = \text{minimum} \; & \sum_{i=1}^{k} (\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}^i)\lambda_i, \\ \text{subject to} \; & \sum_{i=1}^{k} \lambda_i = 1, \\ & -\sum_{i=1}^{k} (\boldsymbol{D}\boldsymbol{x}^i - \boldsymbol{d})\lambda_i \geq \boldsymbol{0}, \\ & \lambda_i \geq 0, \qquad i = 1, \ldots, k, \end{aligned}$$

that is,

$$\begin{aligned} v^{k+1} = \text{minimum} \; & \boldsymbol{c}^{\mathrm{T}} \left(\sum_{i=1}^{k} \lambda_i \boldsymbol{x}^i \right), \qquad (3) \\ \text{subject to} \; & \sum_{i=1}^{k} \lambda_i = 1, \\ & \lambda_i \geq 0, \qquad i = 1, \ldots, k, \\ & \boldsymbol{D} \left(\sum_{i=1}^{k} \lambda_i \boldsymbol{x}^i \right) \leq \boldsymbol{d}. \end{aligned}$$

- We maximize $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}$ subject to $\boldsymbol{x}$ lying in the convex hull of the extreme points $\boldsymbol{x}^i$ found so far *and* fulfilling the constraints that are Lagrangian relaxed.

- The problem (3) is known as the *restricted master problem* (RMP) in the Dantzig–Wolfe algorithm.
- In this algorithm, we have at hand a subset $\{1, \dots, k\}$ of extreme points of X (and a dual vector $\boldsymbol{\mu}^k$), and find a feasible solution to the original LP problem by solving the restricted master problem (3). We then generate an optimal dual solution $\boldsymbol{\mu}^{k+1}$ to this restricted problem problem, corresponding to the constraints $\boldsymbol{Dx} \leq \boldsymbol{d}$. If and only if the vector $\boldsymbol{x}^i$ generated in the next subproblem (2) was already included, we have found the optimal solution to the problem.

- Three algorithms which are "dual" to each other:

$$\begin{gathered}\text{Cutting plane applied to the Lagrangian dual}\\ \Longleftrightarrow\\ \text{Dantzig–Wolfe applied to the original LP}\\ \Longleftrightarrow\\ \text{Benders decomposition applied to the dual LP.}\end{gathered}$$

Column generation

An LP with very many variables $c_j, x_j \in \mathbb{R}$, $\boldsymbol{a}_j, \boldsymbol{b} \in \mathbb{R}^m$, $m \ll n$

$$\begin{aligned}\text{minimize} \quad & z = \sum_{j=1}^{n} c_j x_j\\ \text{subject to} \quad & \sum_{j=1}^{n} \boldsymbol{a}_j x_j = \boldsymbol{b}\\ & x_j \geq 0, \qquad j = 1, \dots, n\end{aligned}$$

The matrix $(\boldsymbol{a}_1, \dots, \boldsymbol{a}_n)$ is too large to handle. Assume that m is relatively small $\Longrightarrow$ the basic matrix is not too large $(m \times m)$

Basic feasible solutions

$B = \{m$ elements from the set $\{1, \dots, n\}\}$ is a basis if the corresponding matrix $\boldsymbol{B} = (\boldsymbol{a}_j)_{j \in B}$ has an inverse, $\boldsymbol{B}^{-1}$

A basic solution is given by $\boldsymbol{x}_B = \boldsymbol{B}^{-1}\boldsymbol{b}$ and $x_j = 0$, $j \notin B$. It is feasible if $\boldsymbol{x}_B \geq \boldsymbol{0}^m$

A better basic feasible solution can be found by computing reduced costs: $\bar{c}_j = c_j - \boldsymbol{c}_B^{\mathrm{T}} \boldsymbol{B}^{-1} \boldsymbol{a}_j$ for $j \notin B$

Let $\bar{c}_s = \underset{j \notin B}{\text{minimum}}\ \bar{c}_j$

If $\bar{c}_s < 0 \Longrightarrow$ a better solution is received if x_s enters the basis

If $\bar{c}_s \geq 0 \Longrightarrow \boldsymbol{x}_B$ is an optimal basic solution

Suppose the columns $\boldsymbol{a}_j$ are defined by a set $S = \{\boldsymbol{a}_j \mid j = 1, \dots, n\}$ being, e.g., solutions to a system of equations (extreme points, integer points, ...)

The incoming column is then chosen by solving a "subproblem":

$$\bar{c}(\boldsymbol{a}') = \underset{\boldsymbol{a} \in S}{\text{minimum}} \left\{ c - \boldsymbol{c}_B^{\mathrm{T}} \boldsymbol{B}^{-1} \boldsymbol{a} \right\}$$

$\boldsymbol{a}'$ is a column having the least reduced cost wrt basis B

If $\bar{c}(\boldsymbol{a}') < 0$ let the column $\begin{pmatrix} c(\boldsymbol{a}') \\ \boldsymbol{a}' \end{pmatrix}$ enter problem

Example: The Cutting Stock Problem

Supply: rolls of e.g. paper of length L

Demand: b_i pieces of length $\ell_i < L$, $i = 1, \dots, m$

Objective: minimize the number of rolls needed to satisfy the demanded of the pieces

First formulation:
Let

$$x_k = \begin{cases} 1 & \text{roll } k \text{ is used} \\ 0 & \text{otherwise} \end{cases} \quad y_{ik} = \begin{cases} 1 & \text{piece } i \text{ is cut from roll } k \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} \text{minimize} \quad & \sum_{k=1}^{M} x_k \\ \text{s.t.} \quad & \sum_{i=1}^{m} \ell_i y_{ik} \leq L x_k, \qquad k = 1, \dots, M \\ & \sum_{k=1}^{M} y_{ij} = b_i, \qquad i = 1, \dots, m \\ & x_k, y_{ik} \geq 0, \text{ binary,} \end{aligned}$$

The value of the LP-relaxation is $\frac{\sum \ell_i b_i}{L}$ which can be very bad if $\ell_i = \lfloor L/2 + 1 \rfloor$ for large L (large duality gap, potentially bad performance of ILP-solvers).

Second formulation:

Cut pattern j contains a_{ij} pieces of length ℓ_i

Feasible pattern if $\sum_{i=1}^{m} \ell_i a_{ij} \leq L$, where $a_{ij} \geq 0$, integer

Integer variables: x_j = number of times pattern j is used

Bad news: n = total number of feasible cut pattern — very large integer

$$\begin{aligned} \text{minimize} \quad & \sum_{j=1}^{n} x_j \\ \text{subject to} \quad & \sum_{j=1}^{n} a_{ij} x_j = b_i, \qquad i = 1, \ldots, m \\ & x_j \geq 0, \ \text{integer}, \qquad j = 1, \ldots, n \end{aligned}$$

Good news: the value of the LP-relaxation is often very close to the value of the optimal solution[a]. We may relax the integrality constaints and solve an LP instead of an ILP!

[a] Marcotte 1985: The cutting stock problem and integer rounding, Mathematical Programming 33

Start solution

Trivial: m unit columns (gives lots of waste):

$$\begin{aligned} \text{minimize} \quad & \sum_{j=1}^{m} x_j \\ \text{subject to} \quad & x_j = b_j, \quad j = 1, \ldots, m \\ & x_j \geq 0, \quad j = 1, \ldots, m \end{aligned}$$

New columns

Generate better patterns using the dual variables π:

$$\begin{aligned} 1 - \underset{a_{ij}}{\text{maximum}} \quad & \sum_{i=1}^{m} \pi_i a_{ij} \qquad \left[\text{minimize} \, (c_j - \underbrace{\boldsymbol{c}_B^{\mathrm{T}} \boldsymbol{B}^{-1}}_{\pi} \boldsymbol{a}_j) \right] \\ \text{subject to} \quad & \sum_{i=1}^{m} \ell_i a_{ij} \leq L, \\ & a_{ij} \geq 0, \ \text{integer}, \qquad i = 1, \ldots, m \end{aligned}$$

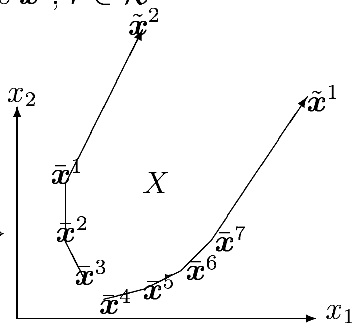
Solution to this knapsack problem: New column $\boldsymbol{a}_j$

Formulation of LP on column generation form—Dantzig–Wolfe decomposition

Let $X = \{\boldsymbol{x} \in \mathbb{R}^n_+ \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\}$ (or $\boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}$) be a polyhedron with the *extreme points* $\bar{\boldsymbol{x}}^p$, $p \in \mathcal{P}$ and the *extreme recession directions* $\tilde{\boldsymbol{x}}^r$, $r \in \mathcal{R}$

$\mathcal{P} = \{1, 2, \ldots, 7\}$

$\mathcal{R} = \{1, 2\}$



$$\boldsymbol{x} \in X \iff \left(\begin{array}{rcl} \boldsymbol{x} & = & \sum_{p \in \mathcal{P}} \lambda_p \bar{\boldsymbol{x}}^p + \sum_{r \in \mathcal{R}} \mu_r \tilde{\boldsymbol{x}}^r \\ & & \sum_{p \in \mathcal{P}} \lambda_p = 1 \\ & & \lambda_p \geq 0, \quad p \in \mathcal{P} \\ & & \mu_r \geq 0, \quad r \in \mathcal{R} \end{array} \right)$$

$\boldsymbol{x} \in X$ is a convex combination of the extreme points plus a conical combination of the extreme directions

This *inner representation* of the set X can be used to reformulate a linear optimization problem according to the *Dantzig-Wolfe decomposition principle*, which is then solved by column generation.

An LP and its complete master problem

$$\begin{array}{llrl} \text{[LP1]} & z^* = \text{minimum} & \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} & \\ & \text{subject to} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} & (\text{“simple” constraints}) \\ & & \boldsymbol{D}\boldsymbol{x} = \boldsymbol{d} & (\text{complicating constraints}) \\ & & \boldsymbol{x} \geq \boldsymbol{0} & \end{array}$$

Let $X = \{\, \boldsymbol{x} \geq 0 \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \,\}$ with the extreme points $\bar{\boldsymbol{x}}^p$, $p \in \mathcal{P}$ and the extreme directions $\tilde{\boldsymbol{x}}^r$, $r \in \mathcal{R} \Longrightarrow$

$$\begin{array}{rll} \text{[LP2]}\ z^* = \min & \sum_{p \in \mathcal{P}} \lambda_p (\boldsymbol{c}^{\mathrm{T}} \bar{\boldsymbol{x}}^p) + \sum_{r \in \mathcal{R}} \mu_r (\boldsymbol{c}^{\mathrm{T}} \tilde{\boldsymbol{x}}^r) & \\ \text{s.t.} & \sum_{p \in \mathcal{P}} \lambda_p (\boldsymbol{D} \bar{\boldsymbol{x}}^p) + \sum_{r \in \mathcal{R}} \mu_r (\boldsymbol{D} \tilde{\boldsymbol{x}}^r) = \boldsymbol{d} & \mid \boldsymbol{\pi} \\ & \sum_{p \in \mathcal{P}} \lambda_p = 1 & \mid q \\ & \lambda_p, \mu_r \geq 0, \ \forall p, r & \end{array}$$

Number of constraints in [LP2] equals to “the number of constraints in $\boldsymbol{D}\boldsymbol{x} = \boldsymbol{d}$” + 1

Number of columns very large (# extreme pts./dirs. to X)

The dual of [LP2] is given by (not all extreme pts./dirs. found yet: $\bar{\mathcal{P}} \subseteq \mathcal{P}$; $\bar{\mathcal{R}} \subseteq \mathcal{R}$)

$$\begin{array}{lllll} \text{[DLP2]}\ z^* \leq \max\limits_{(\boldsymbol{\pi}, q)} & \boldsymbol{d}^{\mathrm{T}}\boldsymbol{\pi} + q & & & \\ \text{s.t.} & (\boldsymbol{D}\bar{\boldsymbol{x}}^p)^{\mathrm{T}}\boldsymbol{\pi} + q \leq (\boldsymbol{c}^{\mathrm{T}}\bar{\boldsymbol{x}}^p), & p \in \bar{\mathcal{P}} & \mid \lambda_p \\ & (\boldsymbol{D}\tilde{\boldsymbol{x}}^r)^{\mathrm{T}}\boldsymbol{\pi} \leq (\boldsymbol{c}^{\mathrm{T}}\tilde{\boldsymbol{x}}^r), & r \in \bar{\mathcal{R}} & \mid \mu_r \end{array}$$

with solutions $(\bar{\boldsymbol{\pi}}, \bar{q})$

Reduced cost for the variable λ_p, $p \in \mathcal{P} \setminus \bar{\mathcal{P}}$ is given by

$$(\boldsymbol{c}^{\mathrm{T}}\bar{\boldsymbol{x}}^p) - (\boldsymbol{D}\bar{\boldsymbol{x}}^p)^{\mathrm{T}}\bar{\boldsymbol{\pi}} - \bar{q} = (\boldsymbol{c} - \boldsymbol{D}^{\mathrm{T}}\bar{\boldsymbol{\pi}})^{\mathrm{T}}\bar{\boldsymbol{x}}^p - \bar{q}$$

Reduced cost for the variable μ_r, $r \in \mathcal{R} \setminus \bar{\mathcal{R}}$ is given by

$$(\boldsymbol{c}^{\mathrm{T}}\tilde{\boldsymbol{x}}^r) - (\boldsymbol{D}\tilde{\boldsymbol{x}}^r)^{\mathrm{T}}\bar{\boldsymbol{\pi}} = (\boldsymbol{c} - \boldsymbol{D}^{\mathrm{T}}\bar{\boldsymbol{\pi}})^{\mathrm{T}}\tilde{\boldsymbol{x}}^r$$

Column generation

The least reduced cost is found by solving the subproblem

$$\min_{\boldsymbol{x}\in X} (\boldsymbol{c} - \boldsymbol{D}^{\mathrm{T}}\boldsymbol{\pi})^{\mathrm{T}}\boldsymbol{x} \quad \left(\text{alt:} \quad \min_{\boldsymbol{x}\in X} (\boldsymbol{c} - \boldsymbol{D}^{\mathrm{T}}\bar{\boldsymbol{\pi}})^{\mathrm{T}}\boldsymbol{x} - \bar{q}\right)$$

Gives as solution an extreme point, $\bar{\boldsymbol{x}}^p$, or an extreme direction $\tilde{\boldsymbol{x}}^r$

$\Longrightarrow$ a new column in [LP2]: (if < 0)

Either $\begin{pmatrix} \boldsymbol{c}^{\mathrm{T}}\bar{\boldsymbol{x}}^p \\ \boldsymbol{D}\bar{\boldsymbol{x}}^p \\ 1 \end{pmatrix}$ or $\begin{pmatrix} \boldsymbol{c}^{\mathrm{T}}\tilde{\boldsymbol{x}}^r \\ \boldsymbol{D}\tilde{\boldsymbol{x}}^r \\ 0 \end{pmatrix}$ enters the problem and improves the solution