TMA521/MMA511
Large Scale Optimization
Lecture 7
Cutting plane methods and column generation

Ann-Brith Strömberg

31 Januari 2018

A standard LP problem and its Lagrangian dual

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

- We assume for now that the polyhedron

$$X := \left\{ \boldsymbol{x} \in \mathbb{R}^n_+ \,\middle|\, \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b} \right\}$$

  is bounded

- Let $\left\{\boldsymbol{x}^1, \boldsymbol{x}^2, \ldots, \boldsymbol{x}^K\right\}$ be the set of extreme points in X
- It follows that $X = \operatorname{conv}\left\{\boldsymbol{x}^1, \boldsymbol{x}^2, \ldots, \boldsymbol{x}^K\right\}$

The Lagrangian dual

- The Lagrangian dual of the LP with respect to relaxing the constraints $\boldsymbol{Dx} \leq \boldsymbol{d}$ is

$$v_{\mathrm{L}} := \max_{\boldsymbol{\mu} \geq \mathbf{0}} \quad q(\boldsymbol{\mu}),$$

- Due to convexity of LP, it holds that $v_{\mathrm{L}} = v_{\mathrm{LP}}$
- The Lagrangian dual function is defined as:

$$q(\boldsymbol{\mu}) := \min_{\mathbf{x} \in X} \left\{ \boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{\mu}^\top (\boldsymbol{Dx} - \boldsymbol{d}) \right\} \tag{1a}$$

$$= \min_{i \in \{1,\ldots,K\}} \left\{ \boldsymbol{c}^\top \boldsymbol{x}^i + \boldsymbol{\mu}^\top (\boldsymbol{Dx}^i - \boldsymbol{d}) \right\} \tag{1b}$$

- Solution set to the subproblem at $\boldsymbol{\mu}$:

$$X(\boldsymbol{\mu}) := \operatorname*{argmin}_{\mathbf{x} \in X} \left\{ \boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{\mu}^\top (\boldsymbol{Dx} - \boldsymbol{d}) \right\}$$

$$= \operatorname{conv} \left\{ \operatorname*{argmin}_{\mathbf{x}^i : i=1,\ldots,K} \left\{ \boldsymbol{c}^\top \boldsymbol{x}^i + \boldsymbol{\mu}^\top (\boldsymbol{Dx}^i - \boldsymbol{d}) \right\} \right\}$$

An equivalent formulation of the Lagrangian dual

- Due to (1), for all $\boldsymbol{\mu} \geq \mathbf{0}$ the function q fulfills the inequalities

$$q(\boldsymbol{\mu}) \leq \boldsymbol{c}^\top \boldsymbol{x}^i + \boldsymbol{\mu}^\top (\boldsymbol{D}\boldsymbol{x}^i - \boldsymbol{d}), \qquad i = 1, \dots, K.$$

- The Lagrangian dual can be equivalently formulated as

$$\begin{aligned} v_{\mathrm{L}} := \quad & \max_{\boldsymbol{\mu}, z} \quad z, \\ & \text{subject to} \quad z \leq \boldsymbol{c}^\top \boldsymbol{x}^i + \boldsymbol{\mu}^\top (\boldsymbol{D}\boldsymbol{x}^i - \boldsymbol{d}), \qquad i = 1, \dots, K, \\ & \qquad\qquad\quad\ \boldsymbol{\mu} \geq \mathbf{0}. \end{aligned}$$

- If, at an optimal dual solution $\boldsymbol{\mu}^*$, the solution set $X(\boldsymbol{\mu}^*)$ is a singleton, i.e., $X(\boldsymbol{\mu}^*) = \{\boldsymbol{x}^*\}$, then $\boldsymbol{x}^*$ is optimal (and unique) — thanks to strong duality
- This typically does not happen, unless an optimal solution $\boldsymbol{x}^*$ happens to be an extreme point of X
- But $\boldsymbol{x}^*$ can always be expressed as a convex combination of extreme points of X (cf. the previous lecture)

A cutting plane method for the Lagrangian dual problem

- Suppose that only a subset, $\left\{\boldsymbol{x}^1, \boldsymbol{x}^2, \ldots, \boldsymbol{x}^k\right\}$ of the extreme points $\left\{\boldsymbol{x}^1, \boldsymbol{x}^2, \ldots, \boldsymbol{x}^K\right\}$ is known (hence, $k \leq K$), and consider the following relaxation of the Lagrangian dual problem:

$$(\boldsymbol{\mu}^k, z^k) \in \operatorname*{argmax}_{\mu, z} z, \tag{2a}$$

$$\text{subject to } z \leq \boldsymbol{c}^\top \boldsymbol{x}^i + \boldsymbol{\mu}^\top (\boldsymbol{D}\boldsymbol{x}^i - \boldsymbol{d}), \; i = 1, \ldots, k, \tag{2b}$$

$$\boldsymbol{\mu} \geq \boldsymbol{0} \tag{2c}$$

- For $k = 1, \ldots, K$ it holds that $z^k \geq v_{\mathrm{L}}$
- How do we determine whether an optimal solution to the Lagrangian dual is found?
- If it holds that

$$z^k \leq \boldsymbol{c}^\top \boldsymbol{x}^i + (\boldsymbol{\mu}^k)^\top (\boldsymbol{D}\boldsymbol{x}^i - \boldsymbol{d}), \quad i = 1, \ldots, K,$$

then $\boldsymbol{\mu}^k$ is optimal in the dual. *Why?*

Check for optimality—generate new inequality

- How check for optimality? Find the most violated constraint
- Solve the subproblem

$$
\begin{aligned}
q(\boldsymbol{\mu}^k) &:= \min_{\mathbf{x} \in X} \left\{ \boldsymbol{c}^\top \boldsymbol{x} + (\boldsymbol{\mu}^k)^\top (\boldsymbol{D}\boldsymbol{x} - \boldsymbol{d}) \right\} \\
&= \min_{i \in \{1,\dots,K\}} \left\{ \boldsymbol{c}^\top \boldsymbol{x}^i + (\boldsymbol{\mu}^k)^\top (\boldsymbol{D}\boldsymbol{x}^i - \boldsymbol{d}) \right\}
\end{aligned} \tag{3}
$$

- If $z^k \leq q(\boldsymbol{\mu}^k)$ then $\boldsymbol{\mu}^k$ is optimal in the dual
- Otherwise, we have identified a constraint of the form

$$
z \leq \boldsymbol{c}^\top \boldsymbol{x}^{k+1} + \boldsymbol{\mu}^\top (\boldsymbol{D}\boldsymbol{x}^{k+1} - \boldsymbol{d}), \tag{4}
$$

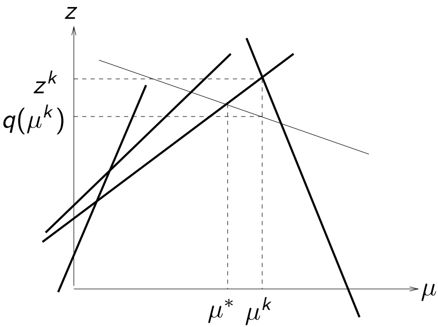
  which is violated at $(\boldsymbol{\mu}^k, z^k)$
  - The latter means that the following (strict) inequality holds:

$$
z^k > \boldsymbol{c}^\top \boldsymbol{x}^{k+1} + (\boldsymbol{\mu}^k)^\top (\boldsymbol{D}\boldsymbol{x}^{k+1} - \boldsymbol{d})
$$

- Add the inequality (4) to (2b), update $k := k + 1$, and resolve the LP (2)

Cutting plane algorithm

- We call this a *cutting plane* algorithm
- It is based on the addition of constraints to the dual problem, in order to improve the solution, in the process of *cutting off* the previous (optimal) point
- The **thick** lines correspond to the subset of k inequalities known at iteration k



Cutting plane algorithm

- Obviously, $z^k \geq q(\mu^k)$ must hold, because of the possible lack of constraints
- For the case in the figure, $z^k > q(\mu^k)$ holds $\Longrightarrow$ in the next iteration $q(\mu^k)$ is evaluated and the last lacking inequality (the thin solid line) will be identified
- The resulting maximization will then yield the optimal solution μ^*, as illustrated in the figure
- How do we generate an optimal primal solution from this scheme?
    - Let us look at the LP dual of the Lagrangian dual problem (2) in this cutting plane algorithm

Duality relations and the Dantzig–Wolfe algorithm

- We rewrite the relaxed Lagrangian dual problem (2) as

$$
\begin{aligned}
z^k := \quad & \max_{(z,\boldsymbol{\mu})} \quad z, \\
& \text{subject to} \quad z - \boldsymbol{\mu}^\top(\boldsymbol{D}\boldsymbol{x}^i - \boldsymbol{d}) \leq \boldsymbol{c}^\top \boldsymbol{x}^i, \quad i = 1, \dots, k, \\
& \qquad\qquad\qquad \boldsymbol{\mu} \geq \boldsymbol{0}
\end{aligned}
$$

- Defining the LP dual variables $\lambda_i \geq 0$, $i = 1, \dots, k$, we obtain the LP dual:

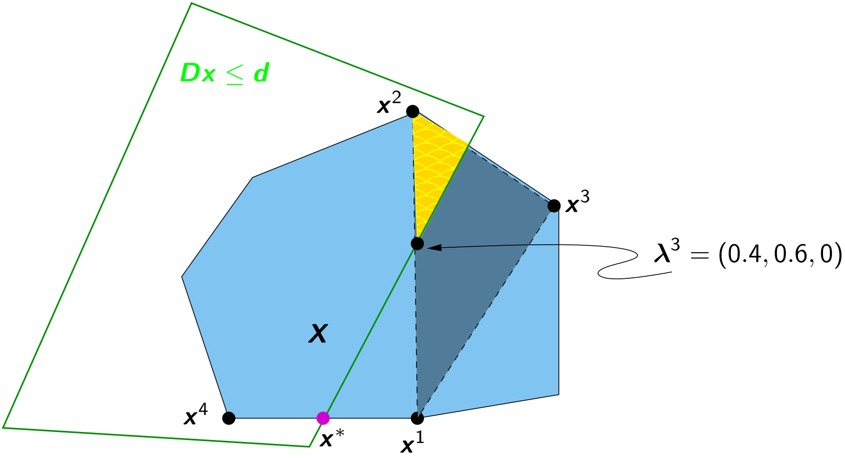
$$
\begin{aligned}
z^k = \quad & \min \ \textstyle\sum_{i=1}^k (\boldsymbol{c}^\top \boldsymbol{x}^i)\lambda_i, \\
& \text{subject to} \quad \textstyle\sum_{i=1}^k \lambda_i = 1, \\
& \qquad -\textstyle\sum_{i=1}^k (\boldsymbol{D}\boldsymbol{x}^i - \boldsymbol{d})\lambda_i \geq \boldsymbol{0}, \\
& \qquad \lambda_i \geq 0, \qquad i = 1, \dots, k
\end{aligned}
$$

The linear programming dual rewritten

$$
\begin{aligned}
z^k = \quad & \min \quad \mathbf{c}^\top \left(\sum_{i=1}^{k} \lambda_i \mathbf{x}^i \right), & (5) \\
& \text{subject to} \quad \sum_{i=1}^{k} \lambda_i = 1, & \\
& \qquad\qquad \lambda_i \geq 0, \qquad i = 1, \dots, k, & \\
& \mathbf{D} \left(\sum_{i=1}^{k} \lambda_i \mathbf{x}^i \right) \leq \mathbf{d} &
\end{aligned}
$$

- Minimize $\mathbf{c}^\top \mathbf{x}$ when $\mathbf{x}$ lies in the convex hull of the extreme points $\mathbf{x}^i$ found so far and fulfills the constraints that are Lagrangian relaxed

An illustration in the $\boldsymbol{x}$-space



The Dantzig-Wolfe algorithm

- The problem (5) is known as the restricted master problem (RMP) in the Dantzig–Wolfe algorithm (to be developed next)
- In this algorithm, we have at hand a subset $\{\boldsymbol{x}^1, \ldots, \boldsymbol{x}^k\}$ of the extreme points of X (and a dual vector $\boldsymbol{\mu}^{k-1}$)
- Find a feasible solution to the original LP by solving the RMP
- Then generate an optimal dual solution $\boldsymbol{\mu}^k$ to this RMP, where the dual variables $\boldsymbol{\mu}$ correspond to the primal constraints $\boldsymbol{D}\boldsymbol{x} \leq \boldsymbol{d}$
- The vector $\boldsymbol{x}^{k+1}$ generated in the next subproblem (3) was already included in the RMP $\iff$ An optimal solution to the original LP is found and verified

Three algorithms which are "dual" to each other

- "Cutting plane" applied to the Lagrangian dual of an LP

$$\Longleftrightarrow$$

- "Dantzig–Wolfe decomposition" applied to the original LP

$$\Longleftrightarrow$$

- Benders decomposition applied to the dual of the original LP

Column generation

- Consider an LP with *very* many variables:
  $c_j, x_j \in \mathbb{R}$, $\boldsymbol{a}_j, \boldsymbol{b} \in \mathbb{R}^m$, $n \gg m$

$$
\begin{aligned}
\text{minimize} \quad & z = \sum_{j=1}^{n} c_j x_j && \text{(6a)}\\
\text{subject to} \quad & \sum_{j=1}^{n} \boldsymbol{a}_j x_j = \boldsymbol{b} && \text{(6b)}\\
& x_j \geq 0, \qquad j = 1, \dots, n && \text{(6c)}
\end{aligned}
$$

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

Basic feasible solutions

- $B := \{m \text{ elements from the set } \{1, \dots, n\}\}$ 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 $\boldsymbol{x}_B \geq \mathbf{0}^m$
- A better basic feasible solution can be found by computing the reduced costs: $\bar{c}_j := c_j - \boldsymbol{c}_B^\top \boldsymbol{B}^{-1}\boldsymbol{a}_j$ for $j \notin B$
  - Let $\bar{c}_s := \underset{j \notin B}{\text{minimum}} \{\bar{c}_j\}$
  - If $\bar{c}_s < 0 \implies$ a better solution will be received if x_s enters the basis
  - If $\bar{c}_s \geq 0 \implies \boldsymbol{x}_B$ is an optimal basic solution

Generating columns

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

- The incoming column is then chosen by solving a subproblem

$$\bar{c}(\boldsymbol{a}') := \underset{\mathbf{a} \in S}{\text{minimum}} \left\{ c - \boldsymbol{c}_B^\top \boldsymbol{B}^{-1} \boldsymbol{a} \right\}$$

- $\boldsymbol{a}'$ is a column having the least reduced cost w.r.t. the basis B

- If $\bar{c}(\boldsymbol{a}') < 0$, let the column $\begin{bmatrix} c(\boldsymbol{a}') \\ \boldsymbol{a}' \end{bmatrix}$ enter the LP (6)

Example: The cutting stock problem

- **Supply:** rolls of e.g. paper of length L
- **Demand:** b_i roll pieces of length $\ell_i < L$, $i = 1, \dots, m$
- **Objective:** minimize the number of rolls needed for producing the demanded pieces

First formulation

$$x_k = \begin{cases} 1 & \text{if roll } k \text{ is used} \\ 0 & \text{otherwise} \end{cases} \qquad y_{ik} = \begin{cases} \text{\# of pieces of length } \ell_i \\ \text{that are cut from roll } k \end{cases}$$

$$\begin{aligned} \text{minimize} \quad & \sum_{k=1}^{K} x_k, \\ \text{subject to} \quad & \sum_{i=1}^{m} \ell_i y_{ik} \leq L x_k, \quad k = 1, \dots, K \\ & \sum_{k=1}^{K} y_{ik} = b_i, \qquad i = 1, \dots, m \\ & x_k \text{ binary}, \quad k = 1, \dots, K \\ & y_{ik} \in \mathbb{Z}_+^{m \cdot K}, \; i = 1, \dots, m, \; k = 1, \dots, K \end{aligned}$$

First formulation

- The value of the continuous relaxation is $\frac{\sum_{i=1}^{m} \ell_i b_i}{L}$, which can be very bad if, e.g., $\ell_i = \lfloor L/2 + 1 \rfloor$ and L is large
- A large duality gap $\Rightarrow$ potentially bad performance of IP solvers
- Also, there are a lot of symmetries (i.e., equivalently good, but differently denoted, solutions) in the "first formulation"
- Symmetries are extremely difficult to handle for integer programming solvers

Second formulation

- **Cut pattern** number 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
- **Variables:** x_j = number of times that pattern j is used

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

- **Bad news:** n = total number of feasible cut patterns—a HUGE integer
- **Good news:** the value of the continuous relaxation is often very close to that of the optimal solution

$\Rightarrow$ Relax integrality constraints, solve an LP instead of an ILP

Starting solution

Natural: m unit columns (yields lots of waste) $\Longrightarrow$

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

New columns generated by the subproblem

Generate better patterns using the dual variable values $\pi_i \implies$ a new column $\overline{\boldsymbol{a}}_k$ computed from:

$$1 - \max_{a_{ik}} \left\{ \sum_{i=1}^{m} \pi_i a_{ik} \right\} \qquad \left[\Leftrightarrow \text{minimize } \left(c_k - \underbrace{\boldsymbol{c}_B^\top \boldsymbol{B}^{-1}}_{\boldsymbol{\pi}} \boldsymbol{a}_k \right) \right]$$

$$\text{subject to } \sum_{i=1}^{m} \ell_i a_{ik} \leq L,$$

$$a_{ik} \geq 0, \text{ integer}, \qquad i = 1, \ldots, m$$

- Solution to this integer knapsack problem: new column $\overline{\boldsymbol{a}}_k$
- If $1 - \boldsymbol{\pi}^\top \overline{\boldsymbol{a}}_k < 0$, then the column $\overline{\boldsymbol{a}}_k$ will improve the formulation
- If $1 - \boldsymbol{\pi}^\top \overline{\boldsymbol{a}}_k \geq 0$, then the columns already generated are sufficient to determine the optimum (of the LP problem)