TMA521/MMA511
Large Scale Optimization
Lecture 15
Benders' decomposition

Ann-Brith Strömberg

21 February 2018

Benders' decomposition for mixed-integer optimization problems (Lasdon)

- Model:

$$\begin{aligned} \text{minimum} \ & \mathbf{c}^\top \mathbf{x} + f(\mathbf{y}), \\ \text{subject to} \ & \mathbf{A}\mathbf{x} + \mathbf{F}(\mathbf{y}) \geq \mathbf{b}, \\ & \mathbf{x} \geq \mathbf{0}^n, \quad \mathbf{y} \in S. \end{aligned}$$

- The variables **y** are assumed to be "complicating" because:
  - the set S may be complicated, like $S \subseteq \{0, 1\}^p$
  - f and/or **F** may be nonlinear
  - the vector $\mathbf{F}(\mathbf{y})$ may cover every row
- The problem is assumed to be *linear* in **x**, possibly separable (whence **A** is block-diagonal); "easy"

Example

- Block-diagonal structure in $\mathbf{x}$ — Variables $\mathbf{y}$ in "every" row
- Continuous variables $\mathbf{x}$ — Binary constraints on $\mathbf{y}$
- Linear in $\mathbf{x}$ — Nonlinear in $\mathbf{y}$

$$
\begin{array}{rlllll}
\min & \mathbf{c}_1^\top \mathbf{x}_1 + \cdots + \mathbf{c}_n^\top \mathbf{x}_n & + & f(\mathbf{y}) & & \\
\text{s.t.} & \mathbf{A}_1 \mathbf{x}_1 & + & \mathbf{F}_1(\mathbf{y}) & \geq & \mathbf{b}_1 \\
& \quad \ddots & & \vdots & & \vdots \\
& \qquad \mathbf{A}_n \mathbf{x}_n & + & \mathbf{F}_n(\mathbf{y}) & \geq & \mathbf{b}_n \\
& \mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n & & & \geq & \mathbf{0} \\
& & & \mathbf{y} & \in & \{0,1\}^p
\end{array}
$$

Typical application: Multi-stage stochastic programming (optimization under uncertainty)

- Some parameters (constants, e.g., $\mathbf{c}$, $\mathbf{A}$, $\mathbf{b}$) are uncertain
- Choose $\mathbf{y}$ (e.g., investment) such that an *expected* cost over time is minimized
- Uncertain data is represented by future *scenarios* ($\ell \in \mathcal{L}$)
- Variables $\mathbf{x}_\ell$ represent future activities
- $\mathbf{y}$ must be chosen before the outcome of the uncertain parameters is known
- Choose $\mathbf{y}$ such that the expected value over scenarios $\ell \in \mathcal{L}$ of the future optimization w.r.t. $\mathbf{x}_\ell$ ($\Rightarrow \mathbf{x}_\ell(\mathbf{y})$) is minimized

A two-stage stochastic program

$$
\begin{array}{lrcrcll}
\min & \sum_{\ell \in \mathcal{L}} p^{\ell} \cdot \mathbf{c}_{\ell}^{\top} \mathbf{x}_{\ell} & + & \mathbf{d}^{\top} \mathbf{y} & & & \\
\text{s.t.} & \mathbf{A}_{\ell} \mathbf{x}_{\ell} & + & \mathbf{T}_{\ell} \mathbf{y} & = & \mathbf{b}_{\ell}, & \ell \in \mathcal{L} \\
& \mathbf{x}_{\ell} & & & \geq & \mathbf{0}, & \ell \in \mathcal{L} \\
& & & \mathbf{y} & \in & Y &
\end{array}
$$

- Solution idea: Temporarily fix $\mathbf{y}$, solve the remaining problem over $\mathbf{x}$ parameterized over $\mathbf{y} \Rightarrow$ solution $\mathbf{x}(\mathbf{y})$
  Utilize the problem structure to improve the guess of an optimal value of $\mathbf{y}$. Repeat
- Similar to minimizing a function η over two vectors, $\mathbf{v}$ and $\mathbf{w}$:

$$
\inf_{\mathbf{v},\mathbf{w}} \eta(\mathbf{v},\mathbf{w}) = \inf_{\mathbf{v}} \xi(\mathbf{v}), \text{ where } \xi(\mathbf{v}) = \inf_{\mathbf{w}} \eta(\mathbf{v},\mathbf{w}), \ \mathbf{v} \in \mathbb{R}^m
$$

- In effect, we substitute the variable $\mathbf{w}$ by always minimizing over it, and work with the remaining problem in $\mathbf{v}$

Benders' decomposition

- Construct an approximation of this problem over $\mathbf{v}$ by utilizing LP duality
- If the problem over $\mathbf{y}$ is also linear
  - Cutting plane methods
- Benders' decomposition is more general:
  - Solves problems with positive duality gaps!
- Benders' decomposition does *not* rely on the existence of optimal Lagrange multipliers and strong duality

Benders' sub- and master problems

- The basic model revisited:

$$\begin{aligned} \text{minimum} \;\; & \mathbf{c}^\top \mathbf{x} + f(\mathbf{y}), \\ \text{subject to} \;\; & \mathbf{A}\mathbf{x} + \mathbf{F}(\mathbf{y}) \geq \mathbf{b}, \\ & \mathbf{x} \geq \mathbf{0}^n, \quad \mathbf{y} \in S. \end{aligned}$$

- Which values of $\mathbf{y}$ are feasible?
- Choose $\mathbf{y} \in S$ such that the remaining problem in $\mathbf{x}$ is feasible
- I.e., choose $\mathbf{y}$ from the set

$$R := \left\{ \mathbf{y} \in S \,\middle|\, \exists \mathbf{x} \geq \mathbf{0}^n \text{ with } \mathbf{A}\mathbf{x} \geq \mathbf{b} - \mathbf{F}(\mathbf{y}) \right\}$$

Benders' sub- and master problems, cont.

- Apply Farkas' Lemma to this system, or rather to the equivalent system (with $\mathbf{y}$ fixed and slack variables $\mathbf{s}$):

$$\begin{aligned} \mathbf{A}\mathbf{x} - \mathbf{s} &= \mathbf{b} - \mathbf{F}(\mathbf{y}) \\ \mathbf{x} \geq \mathbf{0}^n, &\quad \mathbf{s} \geq \mathbf{0}^m \end{aligned}$$

- From Farkas' Lemma: $\mathbf{y} \in R$ if and only if

$$\mathbf{A}^\top \mathbf{u} \leq \mathbf{0}^n,\ \mathbf{u} \geq \mathbf{0}^m \quad \Longrightarrow \quad \big[\mathbf{b} - \mathbf{F}(\mathbf{y})\big]^\top \mathbf{u} \leq 0$$

- This means that $\mathbf{y} \in R$ if and only if $[\mathbf{b} - \mathbf{F}(\mathbf{y})]^\top \mathbf{u}_i^r \leq 0$ holds for every extreme direction $\mathbf{u}_i^r$, $i = 1, \ldots, n_r$ of the polyhedral cone $C = \{\, \mathbf{u} \in \mathbb{R}_+^m \mid \mathbf{A}^\top \mathbf{u} \leq \mathbf{0}^n \,\}$

  (using the representation theorem for a polyhedral cone)

Benders' subproblem

- Given $\mathbf{y} \in R$, the optimal value in *Benders' subproblem* is

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimum}} \;\; & \mathbf{c}^\top \mathbf{x}, \\ \text{subject to} \;\; & \mathbf{A}\mathbf{x} \geq \mathbf{b} - \mathbf{F}(\mathbf{y}), \\ & \mathbf{x} \geq \mathbf{0}^n. \end{aligned}$$

- By LP duality, this is equivalent to

$$\begin{aligned} \underset{\mathbf{u}}{\text{maximum}} \;\; & [\mathbf{b} - \mathbf{F}(\mathbf{y})]^\top \mathbf{u}, \\ \text{subject to} \;\; & \mathbf{A}^\top \mathbf{u} \leq \mathbf{c}, \\ & \mathbf{u} \geq \mathbf{0}^m, \end{aligned}$$

provided that the primal problem has a *finite* solution

Benders' subproblem

- We prefer the dual formulation, since its constraints do not depend on $\mathbf{y}$
- Moreover, the *extreme directions* of the dual feasible set are given by the vectors $\mathbf{u}_i^r$, $i = 1, \ldots, n_r$:
  $C = \{ \mathbf{u} \in \mathbb{R}_+^m \mid \mathbf{A}^\top \mathbf{u} \leq \mathbf{0}^n \}$
- Let $\mathbf{u}_i^p$, $i = 1, \ldots, n_p$, denote the *extreme points* of the set $\{ \mathbf{u} \in \mathbb{R}_+^m \mid \mathbf{A}^\top \mathbf{u} \leq \mathbf{c} \}$

The master problem (MP) of Benders' algorithm

- The original model:

$$\begin{aligned} \text{minimum} \quad & \mathbf{c}^\top \mathbf{x} + f(\mathbf{y}), \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} + \mathbf{F}(\mathbf{y}) \geq \mathbf{b}, \\ & \mathbf{x} \geq \mathbf{0}^n, \quad \mathbf{y} \in S \end{aligned}$$

- This is equivalent to

$$\min_{\mathbf{y}\in S} \left\{ f(\mathbf{y}) + \min_{\mathbf{x}} \left\{ \mathbf{c}^\top \mathbf{x} \,|\, \mathbf{A}\mathbf{x} \geq \mathbf{b} - \mathbf{F}(\mathbf{y}); \mathbf{x} \geq \mathbf{0}^n \right\} \right\}$$

$$= \min_{\mathbf{y}\in R} \left\{ f(\mathbf{y}) + \max_{\mathbf{u}} \left\{ [\mathbf{b} - \mathbf{F}(\mathbf{y})]^\top \mathbf{u} \,|\, \mathbf{A}^\top \mathbf{u} \leq \mathbf{c};\ \mathbf{u} \geq \mathbf{0}^m \right\} \right\}$$

$$= \min_{\mathbf{y}\in R} \left\{ f(\mathbf{y}) + \max_{i=1,\dots n_p} \left\{ [\mathbf{b} - \mathbf{F}(\mathbf{y})]^\top \mathbf{u}_i^p \right\} \right\}$$

The master problem, continued

$$\min_{\mathbf{y}\in R} \left\{ f(\mathbf{y}) + \max_{i=1,\dots n_p} \left\{ [\mathbf{b} - \mathbf{F}(\mathbf{y})]^\top \mathbf{u}_i^p \right\} \right\}$$

$$\begin{aligned} = \min_{\mathbf{y},z} \ & z \\ \text{s.t. } & z \geq f(\mathbf{y}) + [\mathbf{b} - \mathbf{F}(\mathbf{y})]^\top \mathbf{u}_i^p, \quad i = 1,\dots,n_p, \\ & \mathbf{y} \in R, \end{aligned}$$

$$\begin{aligned} = \min_{\mathbf{y},z} \ & z \\ \text{s.t. } & z \geq f(\mathbf{y}) + [\mathbf{b} - \mathbf{F}(\mathbf{y})]^\top \mathbf{u}_i^p, \quad i = 1,\dots,n_p, \\ & 0 \geq [\mathbf{b} - \mathbf{F}(\mathbf{y})]^\top \mathbf{u}_i^r, \qquad\qquad i = 1,\dots,n_r, \\ & \mathbf{y} \in S. \end{aligned}$$

The restricted master problem

- Suppose that only a subset of the constraints in the latter problem is known
- This means that *not all* extreme points and directions for the dual problem are known
- Let $I_1 \subset \{1, \ldots, n_p\}$ and $I_2 \subset \{1, \ldots, n_r\}$
- Replace "$i = 1, \ldots, n_p$" by "$i \in I_1$" and "$i = 1, \ldots, n_r$" by "$i \in I_2$" $\Rightarrow$ *restricted master problem:*

$$
\begin{aligned}
\min_{\mathbf{y}, z} \;\; & z \\
\text{s.t. } & z \geq f(\mathbf{y}) + [\mathbf{b} - \mathbf{F}(\mathbf{y})]^\top \mathbf{u}_i^p, && i \in I_1, \\
& 0 \geq [\mathbf{b} - \mathbf{F}(\mathbf{y})]^\top \mathbf{u}_i^r, && i \in I_2, \\
& \mathbf{y} \in S.
\end{aligned}
$$

- Since not all constraints are included, this yields a *lower bound* on the optimal value of the original problem

Restricted master problem, continued

- Suppose that $(z^0, \mathbf{y}^0)$ is a finite optimal solution to the restricted master problem
- To check whether this is an optimal solution to the original problem: check for the most violated constraint, which is
  - either satisfied, $\Rightarrow$ $\mathbf{y}^0$ is optimal
  - or not, $\Rightarrow$ include this new constraint, extending either the set I_1 or I_2, and possibly improving the lower bound.

Find new constraints of the master problem

- The search for a new constraint is done by solving the dual of Benders' subproblem at $\mathbf{y} = \mathbf{y}^0$:

$$\begin{aligned} \underset{\mathbf{u}}{\text{maximum}} \;\; & [\mathbf{b} - \mathbf{F}(\mathbf{y}^0)]^\top \mathbf{u}, \\ \text{subject to} \;\; & \mathbf{A}^\top \mathbf{u} \leq \mathbf{c}, \\ & \mathbf{u} \geq \mathbf{0}^m, \end{aligned}$$

  $\Rightarrow$ the solution is a new dual extreme point or dual extreme direction, due to a "new" objective

- The solution $\mathbf{u}(\mathbf{y}^0)$ to this (dual) problem corresponds to a *feasible* (primal) solution $(\mathbf{x}(\mathbf{y}^0), \mathbf{y}^0)$ to the original problem, and therefore also an *upper bound* on the optimal value, provided that it is finite

Add new constraints to the master problem

- If the subproblem problem has an unbounded solution, then it is unbounded along an extreme direction: $[\mathbf{b} - \mathbf{F}(\mathbf{y}^0)]^\top \mathbf{u}_i^r > 0$
  - Add the constraint $\boxed{0 \geq \left[\mathbf{b} - \mathbf{F}(\mathbf{y})\right]^\top \mathbf{u}_i^r}$ to RMP (enlarge I_2)
- Suppose instead that the optimal solution is finite:
  - Let $\mathbf{u}_i^p$ be an optimal extreme point
  - If $z^0 < f(\mathbf{y}^0) + [\mathbf{b} - \mathbf{F}(\mathbf{y}^0)]^\top \mathbf{u}_i^p$, add the constraint $\boxed{z \geq f(\mathbf{y}) + \left[\mathbf{b} - \mathbf{F}(\mathbf{y})\right]^\top \mathbf{u}_i^p}$ to RMP (enlarge I_1)
- If $z^0 \geq f(\mathbf{y}^0) + [\mathbf{b} - \mathbf{F}(\mathbf{y}^0)]^\top \mathbf{u}_i^p$ then equality must hold ("$>$" cannot happen—why?)
  - We then have an optimal solution to the original problem and terminate

Convergence

- Suppose that the set S is closed and bounded and that f and $\mathbf{F}$ are both continuous on S
- Then, provided that the computations are exact, we terminate in a finite number of iterations with an optimal solution
- The proof is due to the finite number of constraints in the complete master problem, that is, the number of extreme points and directions in any polyhedron
- A numerical example of the use of Benders' decomposition is found in Lasdon (1970, Sections 7.3.3–7.3.5)

Discussion

- Note the resemblance to the Dantzig–Wolfe algorithm! If f and $\mathbf{F}$ are both linear, then they coincide: (the duals of) their subproblems and RMP:s are identical
- Modern implementations of the DW and Benders' algorithms are inexact: at least their RMP:s are not solved exactly
- Their RMP:s are often restricted with an additional "box constraint", which forces the solution to the next RMP to be fairly close to the previous one
- The effect is stability; otherwise, the sequence of solutions to the RMP:s may "jump about" and convergence becomes slow
- This was observed quite early for the DW algorithm, which can even be enriched with non-linear "penalty" terms in the RMP to further stabilize convergence
- In any case, convergence holds also under these modifications, except perhaps for the finiteness