Lecture 11: Benders decomposition

Michael Patriksson

28 September 2009

A standard LP problem and its Lagrangian dual

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

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

Benders decomposition for mixed-integer optimization problems—Lasdon (1970)

- Model:

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

- The variables $\mathbf{y}$ are "difficult" because:
  - the set S may be complicated, like $S \subseteq \{0,1\}^p$
  - f and/or $\mathbf{F}$ may be nonlinear
  - the vector $\mathbf{F}(\mathbf{y})$ may cover every row, while the problem in $\mathbf{x}$ for fixed $\mathbf{y}$ may separate

- The problem is *linear*, possibly separable in $\mathbf{x}$; "easy"

Example: Block-angular structure in x, binary constraints on y, linear in x, nonlinear in y

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

Applications

- **Typical application:** Multi-stage stochastic programming (optimization under uncertainty)
  - Some parameters (constants) are uncertain
  - Choose $\mathbf{y}$ (e.g., investment) such that an *expected* cost over time is minimized
  - Uncertainty in data is represented by future *scenarios* (ℓ)
  - Variables $\mathbf{x}_\ell$ represent future activities
  - $\mathbf{y}$ must be chosen before the outcome of the uncertain parameters is known
  - Choose $\mathbf{y}$ s.t. the expected value over scenarios ℓ of the future optimization over $\mathbf{x}_\ell$ ($\Rightarrow \mathbf{x}_\ell(\mathbf{y})$) is the best

A two-stage stochastic program, I

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

A two-stage stochastic program, I

- **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

- **Benders decomposition:** construct an approximation of this problem over $\mathbf{v}$ by utilizing LP duality

- If the problem over $\mathbf{y}$ is also linear

$\Rightarrow$ cutting plane methods from above

- 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

The Benders sub- and master problems, I

- The model revisited:

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

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

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

The Benders sub- and master problems, II

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

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

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

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

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

The Benders sub- and master problems, III

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

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \ & \mathbf{c}^{\mathrm{T}}\mathbf{x}, \\ \text{subject to} \ & \mathbf{A}\mathbf{x} \geq \mathbf{b} - \mathbf{F}(\mathbf{y}), \\ & \mathbf{x} \geq \mathbf{0}^n \end{aligned}$$

- By LP duality, this equals the problem to

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

provided that the first problem has a finite solution

The Benders sub- and master problems, IV

- We prefer the dual formulation, since its constraints do not depend on $\mathbf{y}$
- Moreover, the *extreme directions* of its feasible set are given by the vectors $\mathbf{u}_i^r$, $i = 1, \ldots, n_r$, discussed above
- Let $\mathbf{u}_i^p$, $i = 1, \ldots, n_p$, denote the *extreme points* of this set
- This completes the subproblem
- Let's now study the *restricted master* problem (RMP) of Benders' algorithm

The Benders sub- and master problems, V

- The original model:

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

- This is equivalent to

$$\begin{aligned} & \min_{\mathbf{y} \in S} \; \left\{ f(\mathbf{y}) + \min_{\mathbf{x}} \left\{ \mathbf{c}^{\mathrm{T}}\mathbf{x} \mid \mathbf{A}\mathbf{x} \geq \mathbf{b} - \mathbf{F}(\mathbf{y}); \mathbf{x} \geq \mathbf{0}^n \right\} \right\} \\ = & \min_{\mathbf{y} \in R} \; \left\{ f(\mathbf{y}) + \max_{\mathbf{u}} \left\{ [\mathbf{b} - \mathbf{F}(\mathbf{y})]^{\mathrm{T}}\mathbf{u} \mid \mathbf{A}^{\mathrm{T}}\mathbf{u} \leq \mathbf{c};\ \mathbf{u} \geq \mathbf{0}^m \right\} \right\} \\ = & \min_{\mathbf{y} \in R} \; \left\{ f(\mathbf{y}) + \max_{i=1,\dots n_p} \left\{ [\mathbf{b} - \mathbf{F}(\mathbf{y})]^{\mathrm{T}}\mathbf{u}_i^p \right\} \right\} \end{aligned}$$

The Benders sub- and master problems, VI

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

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

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

The Benders sub- and master problems, VII

- Suppose that not the whole sets of constraints in the latter problem is known
- This means that not all extreme points and directions for the dual problem are known
- Replace "$i = 1, \dots, n_p$" with "$i \in I_1$" and "$i = 1, \dots, n_r$" with "$i \in I_2$" where $I_1 \subset \{1, \dots, n_p\}$ and $I_2 \subset \{1, \dots, n_r\}$
- Since not all constraints are included, we get a lower bound on the optimal value of the original problem

The Benders sub- and master problems, VIII

- Suppose that $(z^0, \mathbf{y}^0)$ is a finite optimal solution to this problem

- To check if this is indeed an optimal solution to the original problem: check for the most violated constraint, which we
  - either satisfy $\Rightarrow$ $\mathbf{y}^0$ is optimal
  - or not $\Rightarrow$ include this new constraint, extending either the set I_1 or I_2, and possibly improving the lower bound

The Benders sub- and master problems, IX

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

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

  $\Rightarrow$ a new extreme point or 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

The Benders sub- and master problems, X

- If this problem has an unbounded solution, then it is unbounded along an extreme direction: $[\mathbf{b} - \mathbf{F}(\mathbf{y}^0)]^{\mathrm{T}} \mathbf{u}_i^r > 0$

$\Rightarrow$ Add the constr. $0 \geq [\mathbf{b} - \mathbf{F}(\mathbf{y})]^{\mathrm{T}} \mathbf{u}_i^r$ to RMP (enlarge I_2)

- Suppose instead that the optimal solution is finite:

$\Rightarrow$ Let $\mathbf{u}_i^p$ be an optimal extreme point
If $z^0 < f(\mathbf{y}^0) + [\mathbf{b} - \mathbf{F}(\mathbf{y}^0)]^{\mathrm{T}} \mathbf{u}_i^p$, add the constraint
$z \geq f(\mathbf{y}) + [\mathbf{b} - \mathbf{F}(\mathbf{y})]^{\mathrm{T}} \mathbf{u}_i^p$ to RMP (enlarge I_1)

- If $z^0 \geq f(\mathbf{y}^0) + [\mathbf{b} - \mathbf{F}(\mathbf{y}^0)]^{\mathrm{T}} \mathbf{u}_i^p$ then equality must hold ($>$ cannot happen—why?)

$\Rightarrow$ We then have an optimal solution to the original problem and terminate

Convergence, I

- Suppose that 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

- 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)

Convergence, II

- Note the resemblance to the Dantzig–Wolfe algorithm! In fact, if f and $\mathbf{F}$ both are linear, then the methods coincide, in the sense that (the duals of) their subproblems and restricted master problems are identical!
- Modern implementations of the Dantzig–Wolfe and Benders algorithms are inexact, that is, at least their RMP:s are not solved exactly
- Moreover, their RMP:s are often restricted such that there is an additional "box constraint" added. This constraint forces the solution to the next RMP to be relatively close to the previous one

Convergence, III

- The effect is that of a stabilization; otherwise, there is a risk that the sequence of solutions to the RMP:s "jump about," and convergence becomes slow as the optimal solution is approached
- This was observed quite early on with the Dantzig–Wolfe algorithm, which even can 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