Benders Decomposition

Víctor Blanco
Universidad de Granada

Desafíos de la Matemática Combinatoria
FQM-5849

MINLP School: Theory, algorithms and applications

Sevilla, June 2016

Outline

Benders Decomposition: The Original

$$\min \quad c^t x + f(y)$$
$$\text{s.t. } Ax + F(y) \geq b,$$
$$x \in \mathbb{R}^n_+$$
$$y \in Y \subset \mathbb{R}^m$$

Numerische Mathematik 4, 238—252 (1962)

Partitioning procedures for solving mixed-variables programming problems*

By

J. F. Benders**

I. Introduction

In this paper two slightly different procedures are presented for solving mixed-variables programming problems of the type

$$\max\{c^T x + f(y) \mid A x + F(y) \leqq b,\ x \in R_p,\ y \in S\}, \tag{1.1}$$

where $x \in R_p$ (the p-dimensional Euclidean space), $y \in R_q$, and S is an arbitrary subset of R_q. Furthermore, A is an (m, p) matrix, $f(y)$ is a scalar function and $F(y)$ an m-component vector function both defined on S, and b and c are fixed vectors in R_m and R_p, respectively.

An example is the mixed-integer programming problem in which certain variables may assume any value on a given interval, whereas others are restricted to integral values only. In this case S is a set of vectors in R_q with integral-valued components. Various methods for solving this problem have been proposed by Beale [1], Gomory [9] and Land and Doig [11]. The use of integer variables, in particular for incorporating in the programming problem a choice from a set of alternative discrete decisions, has been discussed by Dantzig [4].

Other examples are those in which certain variables occur in a linear and others in a non-linear fashion in the formulation of the problem (see e.g. Griffith and Stewart [7]). In such cases $f(y)$ or some of the components of $F(y)$ are non-linear functions defined on a suitable subset S of R_q.

Obviously, after an arbitrary partitioning of the variables into two mutually exclusive subsets, any linear programming problem can be considered as being of type (1.1). This may be advantageous if the structure of the problem indicates a natural partitioning of the variables. This happens, for instance, if the problem is actually a combination of a general linear programming and a transportation problem. Or, if the matrix shows a block structure, the blocks being linked only by some columns, to which also many other block structures can easily be reduced. A method of solution for linear programming problems efficiently utilizing such block structures, has been designed by Dantzig and Wolfe [5].

The basic idea behind the procedures to be described in this report is a partitioning of the given problem (1.1) into two sub problems; a programming

---

* Paper presented to the 8th International Meeting of the Institute of Management Sciences, Brussels, August 23—26, 1961.

** Koninklijke/Shell-Laboratorium, Amsterdam (Shell Internationale Research Maatschappij N.V.).

Benders Decomposition: The Original

For fixed $\hat{y} \in S$:

$v(\hat{y})$ is

$$\begin{aligned} \min\ & c^t x + f(\hat{y}) \\ \text{s.t. } & Ax \geq b - F(\hat{y}), \\ & x \in \mathbb{R}^n_+ \end{aligned} \equiv \begin{aligned} f(\hat{y}) + v(\hat{y}) := \min\ & c^t x \\ \text{s.t. } & Ax + F(\hat{y}) \geq b, \\ & x \in \mathbb{R}^n_+ \end{aligned}$$

a linear programming problem, with dual:

$$\begin{aligned} \max\ & (b - F(\hat{y}))^t u \\ \text{s.t. } & A^t u \leq c, \\ & u \geq 0 \end{aligned}$$

Benders Decomposition: The Original

$$\begin{aligned} \min \quad & c^t x + f(y) \\ \text{s.t.} \quad & Ax + F(y) \geq b, \\ & x \in \mathbb{R}^n \\ & y \in Y \subset \mathbb{R}^m \end{aligned}$$

$$\begin{aligned} \max \quad & (b - F(\hat{y}))^t u \\ \text{s.t.} \quad & A^t u \geq c, \\ & u \geq 0 \end{aligned}$$

$$\begin{aligned} \min_y \max_u \quad & f(y) + (b - F(y))^t u \\ \text{s.t.} \quad & A^t u \leq c, \\ & u \geq 0, \\ & y \in Y \subset \mathbb{R}^m \end{aligned}$$

$$\equiv$$

$$\begin{aligned} & \min_y \theta \\ \text{s.t.} \quad & \theta \geq f(y) + (b - F(y))^t u, \forall u \geq 0, \\ & y \in Y \subset \mathbb{R}^m \end{aligned}$$

Benders Decomposition: The Original

$$
\begin{aligned}
\min \;\; & c^t x + f(y) \\
& \text{s.t. } Ax + F(y) \geq b, \\
& \quad x \in \mathbb{R}^n \\
& \quad y \in Y \subset \mathbb{R}^m
\end{aligned}
\quad \equiv \quad
\begin{aligned}
& \min_y \theta \\
& \text{s.t. } \theta \geq f(y) + (b - F(y))^t u, \forall u \geq 0, \\
& \quad y \in Y \subset \mathbb{R}^m
\end{aligned}
$$

The right problem has infinitely many constraints!!!

Add them sequentially (and as many as needed)...

Benders Decomposition: The Original

$$\min \theta$$
$$\text{s.t. } \theta \geq f(y) + (b - F(y))^t u, \forall u \geq 0, \quad \text{(M)}$$
$$y \in Y \subset \mathbb{R}^m$$

$$\min c^t x$$
$$\text{s.t. } Ax + F(\hat{y}) \geq b, \quad \text{(SP}(y))$$
$$x \in \mathbb{R}^n$$

$k = 0$, $y_0 \in Y$, $UB = \infty$, $LB = -\infty$,

Repeat until $UB - LB < \varepsilon$:

- Set $y = y_k$.
- Solve (SP(y)):
- If (SP(y)) is feasible and with optimal multipliers u_k
  - Update $UB = \min\{UB, v(y_k)\}$.
  - Add the (optimality) cut $\theta \geq (b - F(y))^t u_k$ and solve (M): $LB = \theta_k$.
- If (SP(y)) is infeasible (dual unbounded):
  - Find an extreme direction v_k (such that $(b - F(y))^t v_k > 0$)
  - Add the (feasibility) cut $(b - F(y))^t v_k \leq 0$ and solve (M): $LB = \theta_k$.

Example

$$\begin{aligned}\min\ & 2x_1 + 3x_2 + 2y \\ \text{s.t.}\ & x_1 + 2x_2 + y \geq 3, \\ & 2x_1 - x_2 + 3y \geq 4, \\ & x_1, x_2, y \geq 0.\end{aligned}$$

1. Set an initial feasible y: $y_0 = 0$.
2. Solve SP:

$$\begin{aligned}\min\ & 2x_1 + 3x_2 \\ \text{s.t.}\ & x_1 + 2x_2 + y \geq 3 - y_0, \\ & 2x_1 - x_2 \geq 4 - 3y_0, \\ & x_1, x_2, \geq 0.\end{aligned}$$

$x^* = (2.2, 0.4)$ and $u^* = (1.60, 0.20)$: $UB = f^* = 5.6$

3. Add Optimality Cut to MP:

$$\begin{aligned}\min\ & \theta \\ \text{s.t.}\ & \theta \geq 2y - 2.2y + 5.6 = -0.2y + 5.6 \qquad \theta \geq 0\end{aligned}$$

$y^* = 2.545$, $LB = \theta^* = 5.909$.

Feasibility Cuts

If the subproblem is infeasible (and then, the dual unbounded), the extreme rays to add feasibility cuts in the form:

$$\theta \geq (b - F(y))^t v_k > 0$$

can be found solving:

$$\max (b - F(y))^t v$$
$$\text{s.t.} \quad A^t v \geq 0$$

Generalized Benders Decomposition

$$
\begin{aligned}
\min \quad & f(x, y) \\
\text{s.t.} \quad & G(x, y) \geq 0, \qquad \text{(P)} \\
& x \in X, \\
& y \in Y.
\end{aligned}
$$

Such that:

- Y is the set of complicating variables.
- For fixed y, (P) is easy-to-solve (the problem become convex, or combinatorial...)

JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 10, No. 4, 1972

Generalized Benders Decomposition[1]

A. M. GEOFFRION[2]

Communicated by A. V. Balakrishnan

Abstract. J. F. Benders devised a clever approach for exploiting the structure of mathematical programming problems with *complicating variables* (variables which, when temporarily fixed, render the remaining optimization problem considerably more tractable). For the class of problems specifically considered by Benders, fixing the values of the complicating variables reduces the given problem to an ordinary linear program, parameterized, of course, by the value of the complicating variables vector. The algorithm he proposed for finding the optimal value of this vector employs a cutting-plane approach for building up adequate representations of (i) the extremal value of the linear program as a function of the parameterizing vector and (ii) the set of values of the parameterizing vector for which the linear program is feasible. Linear programming duality theory was employed to derive the natural families of *cuts* characterizing these representations, and the parameterized linear program itself is used to generate what are usually *deepest* cuts for building up the representations.

In this paper, Benders' approach is generalized to a broader class of programs in which the parametrized subproblem need no longer be a linear program. Nonlinear convex duality theory is employed to derive the natural families of cuts corresponding to those in Benders' case. The conditions under which such a generalization is possible and appropriate are examined in detail. An illustrative specialization is made to the variable factor programming problem introduced by R. Wilson, where it offers an especially attractive approach. Preliminary computational experience is given.

[1] Paper received April 10, 1970; in revised form, January 28, 1971. An earlier version was presented at the Nonlinear Programming Symposium at the University of Wisconsin sponsored by the Mathematics Research Center, US Army, May 4–6, 1970. This research was supported by the National Science Foundation under Grant No. GP-8740.

[2] Professor, University of California at Los Angeles, Los Angeles, California.

GBD

$$\begin{aligned} \min \quad & f(x, y) \\ \text{s.t.} \quad & G(x, y) \geq 0, \\ & x \in X, \\ & y \in Y. \end{aligned}$$

is equivalent to

$$\begin{aligned} \min_y \min_x \, & f(x, y) \\ \text{s.t.} \quad & G(x, y) \geq 0, \\ & x \in X, \\ & y \in Y. \end{aligned}$$

$$\begin{aligned} \min_y \, & v(y) \\ \text{s.t.} \quad & y \in V \cap Y. \end{aligned}$$

$V = \{y : \exists x \in X : G(x, y) \geq 0\}$ and $v(y) = \min_{x \in X} \{f(x, y) : G(x, y) \geq 0\}$.

GBD

$$\min_y v(y)$$
$$\text{s.t. } y \in V \cap Y.$$

- The original problem is infeas./unbounded iff the projection is.
- The projection of an optimal solution of the original problem onto the y-space is an optimal solution of projection ("iff").
- Also for ϵ approximations...
- Under some conditions (convexity, ...): $y \in V$ iff $\sup_x \lambda^t G(x, y) \geq 0$, $\forall \lambda \in \{\lambda \in \mathbb{R}^m_+ : \sum_i \lambda_i = 1\}$.

Benders Decomposition

From duality:

$$v(y) = \sup_{u \geq 0} \inf_{x} f(x, y) + u^t G(x, y)$$

forall $y \in V \cap Y$.
Hence, the problem is equivalent to:

$$\min_{y} \sup_{u \geq 0} \inf_{x} f(x, y) + u^t G(x, y)$$
$$\text{s.t.} \quad \sup_{x} \lambda^t G(x, y) \geq 0, \ \forall \lambda \in \{\lambda \in \mathbb{R}^m_+ : \sum_i \lambda_i = 1\}.$$

So:

$$\min_{y} \theta$$
$$\text{s.t.} \quad \theta \geq \inf_{x} f(x, y) + u^t G(x, y), \forall u \geq 0,$$
$$\inf_{x} \lambda^t G(x, y) \geq 0, \ \forall \lambda \in \{\lambda \in \mathbb{R}^m_+ : \sum_i \lambda_i = 1\}.$$

Classical GBD

Input : $y_0 \in V \cap Y$, $u_0 \geq 0$ optimal multiplier, $p = 1$, $q = 0$, $UB = v(y_0)$, $LB = -\infty$, $\varepsilon > 0$ (tolerance)

1. Solve

$$\hat{\theta} = \min_{y,\theta} \theta$$

$$s.t. \theta \geq \inf_{x \in X} f(x,y) + u_k G(x,y),\ k = 0, \dots, p, \text{(Benders Optimality Cuts)}$$

$$\inf_{x \in X} \lambda_k^t G(x,y) \geq 0,\ k = 1, \dots, q \text{(Benders Feasibility Cuts).}$$

   with solution $\hat{y}$.
   if $UB \leq \hat{\theta} - \varepsilon$ **then** STOP;

2. Solve $v(\hat{y})$:
   if $v(\hat{y}) < \infty$ **then**
   - **if** $v(\hat{y}) \leq \hat{\theta} - \varepsilon$ **then**
     - STOP
   - **else**
     - Increase $p \mapsto p + 1$ and compute a multiplier u_p. GO TO 1

   else
   - Increase $q \mapsto q + 1$ and determine λ_q such that $\sup_{x \in X} \lambda_q^t G(x,y) < 0$. GO TO 1

Benders Decomposition

- If Y is a finete discrete set, X nonempty and convex and G convex for each fixed $y \in Y$. Then, Benders terminates in a finite number of steps.
- Benders decompostion is useful when for each u, λ, $\sup_{x \in X} f(x, y) + uG(x, y)$ and $\sup_{x \in X} \lambda G(x, y)$ can be explicitly computed with little effort as a function of y.

Example 1: Linearly separable

$f(x, y) = f_1(x) + f_2(x)$, $G(x, y) = G_1(x) + G_2(x)$

- $\inf_{x \in X} f(x, y) + uG(x, y) = v(\hat{y}) + (f_2(y) - f_2(\hat{y})) + u(G_2(y) - G_2(\hat{y}))$.
- $\inf_{x \in X} \lambda G(x, y) = \inf_{x \in X} \{\lambda G_1(x)\} + \lambda G_2(y)$

Optimality Cuts

By Lagrangian duality:

$$\nabla_y v(y) = \nabla_y f(\hat{x}, \hat{y}) + \hat{u}^t \nabla_y G(\hat{x}, \hat{y})$$

Hence, optimality cuts can be written in the form:

$$\theta \geq v(\hat{y}) + \nabla_y v(y)^t (y - \hat{y})$$

Example: Toy Bilinear Problems

$$\begin{aligned} \min \quad & xy \\ \text{s.t.} \quad & l_x \le x \le u_x, \\ & y \in \{0,1\} \end{aligned}$$

For fixed $\hat{y} \in [l_y, u_y]$:

$$\begin{aligned} v(\hat{y}) = \min \quad & \hat{y}x \\ \text{s.t.} \quad & l_x \le x \le u_x, \end{aligned}$$

$$\begin{aligned} \min \quad & \theta \\ \text{s.t.} \quad & \theta \ge v(\hat{y}) + \nabla_y v(y)^t (y - \hat{y}), \\ & y \in \{0,1\} \end{aligned}$$

- $v(y) = l_x y$, so $\nabla v(y) = l_x$.
- If $y = 0$: $\theta \ge 0 + l_x(y - 0) = l_x y$.
- If $y = 1$: $\theta \ge l_x + l_x(y - 1) = l_x y$.

$$\begin{aligned} \min \quad & \theta \\ \text{s.t.} \quad & \theta \ge l_x y, \\ & y \in \{0,1\} \end{aligned}$$

- If $l_x < 0$: $\theta^* = l_x$, $y^* = 1$, $x^* = l_x$.
- If $l_x \ge 0$, $\theta^* = 0$, $y^* = 0$, $x^* \in [l_x, u_x]$.

2-Stage LP with Recourse

$$\min \; c^t x + E_\xi[Q(x,\xi)]$$
$$\text{s.t. } Ax = b,$$
$$x \in \{0,1\}^n$$

where $Q(x,\xi) = \min\{q(\xi)^t y : W\, y \geq h(\xi) - Tx,\, y \geq 0\}$.

- Given a first stage decision, x, the realization of the r.v. ξ is observed.
- In the second stage, ξ is known and y must be taken to satisfy $Wy \geq \xi - Tx$ and $y \geq 0$.
- y is assumed to cause a penalty of $q(\xi)$.

If ξ has a discrete distribution with finite support $\{\xi_1, \ldots, \xi_s\}$, with $\mathbb{P}[\xi = \xi_i] = p_i$:

$$\min \; c^t x + \sum_{i=1}^{s} p_i Q(x, \xi_i)$$
$$\text{s.t. } Ax = b,$$
$$x \in \{0,1\}^n$$

where $Q(x,\xi) = \min\{q_i^t y : W\, y \geq h(\xi_i) - Tx,\, y \geq 0\}$.

2-Stage Binary Programming with Recourse

$$\min \; c^t x + \sum_{i=1}^{s} p_i Q(x, \xi_i)$$
$$\text{s.t. } Ax = b,$$
$$x \in \{0,1\}^n$$

where $Q(x, \xi_i) = \min\{q_i^t y : W\, yv \geq h(\xi_i) - T_i x,\; y \geq 0\}$.

$$\min \; c^t x + \theta$$
$$\text{s.t. } Ax = b,$$
$$x \in \{0,1\}^n,$$
$$\theta \geq Q(x)$$

where $Q(x) = \sum_{i=1}^{s} p_i Q(x, \xi_i)$.

2-Stage LP with Recourse: Optimality Cuts

Fix a solution $x = \hat{x}$. If $Q(x, \xi_i) = \min\{q_i^t y : W\, yv \geq h(\xi_i) - Tx, y \geq 0\}$ is feasible, its dual is:

$$\begin{aligned} \max \;\; & u_i^t (h(\xi_i) - T_i \hat{x}) \\ \text{s.t.} \;\; & u_i^t W \leq q_i^t \\ & u \geq 0. \end{aligned}$$

So, optimality cuts are in the form:

$$\theta \geq \sum_{i=1}^{s} p_i \hat{u}_i^t (h(\xi_i) - T_i x)$$

2-Stage LP with Recourse: Feasibility Cuts

One way to find extreme directions of the dual problem is solving the following LP:

$$\max \lambda^t (h(\xi_i) - T_i \hat{x})$$
$$\text{s.t. } \lambda^t W \leq 0,$$
$$\sum \lambda_i \leq 1, \lambda_i \geq 0.$$

with such a solution, the feasibility cut for those realizations (ξ_i) with positive obj. val of the problem above is:

$$\lambda^t (h(\xi_i) - T_i \hat{x}) \leq 0 \Rightarrow \lambda^t T_i \hat{x} \geq \lambda^t h(\xi)$$

Uncapacitated Facility Location

- A set of potential customers J.
- A set of potential facility locations I.
- Allocation costs between customers and facilities: c_{ij}, $i \in I$, $j \in J$.
- Opening costs of facilities f_i, $\forall i \in I$.

$$\min \sum_{i \in I} f_i y_i + \sum_i \sum_j c_{ij} x_{ij}$$

$$\text{s.t.} \quad \sum_{i \in I} x_{ij} = 1, \forall j \in J,$$

$$x_{ij} \leq y_i, \forall i \in I, j \in J,$$

$$x_{ij} \geq 0, \forall i \in I, j \in J,$$

$$y_i \in \{0, 1\}, \forall i \in I,$$

Redesigning Benders Decomposition for Large Scale Facility Location

Matteo Fischetti[1], Ivana Ljubić[2], Markus Sinnl[3]

[1] Department of Information Engineering, University of Padua, Italy, matteo.fischetti@unipd.it
[2] ESSEC Business School of Paris, Cergy-Pontoise, France, ivana.ljubic@essec.edu
[3] Department of Statistics and Operations Research, University of Vienna, Austria, markus.sinnl@univie.ac.at

February 11, 2016

Abstract

The Uncapacitated Facility Location (UFL) problem is one of the most famous and most studied problems in the Operations Research literature. Given a set of potential facility locations, and a set of customers, the goal is to find a subset of facility locations to open, and to allocate each customer to open facilities, so that the facility opening plus customer allocation costs are minimized. In our setting, for each customer the allocation cost is assumed to be a linear or *separable convex quadratic function*.

Motivated by recent UFL applications in business analytics, we revise approaches that work on a projected decision space and hence are intrinsically more scalable for large scale input data. Our working hypothesis is that many of the exact (decomposition) approaches that have been proposed decades ago and discarded soon after, need to be redesigned to draw the advantage of the new hardware and software technologies. To this end, we "thin out" the classical models from the literature, and use (generalized) Benders cuts to replace a huge number of allocation variables by a small number of continuous variables that model the customer allocation cost directly. Our results show that Benders decomposition allows for a significant boost in the performance of a Mixed-Integer Programming solver. We report the optimal solution of a large set of previously unsolved benchmark instances widely used in the available literature. In particular, dramatic speedups are achieved for UFL's with separable quadratic allocation costs—which turn out to be much easier than their linear counterpart when our approach is used.

Introduction

levance and importance of mathematical modeling and optimization tools have been widely accept professionals working in the field of business analytics. Predictive and prescriptive data analytics vadays impossible without efficient optimization tools capable of dealing with large amount of da ese recent synergies between operations research and business analytics impose new challenges for the n neration of exact algorithms. Despite the huge success of general purpose solvers in the last decade, findi imal solutions for Mixed-Integer Programming (MIP) models involving millions of variables still rema of reach for most of the important combinatorial optimization problems. This article studies linear a vex quadratic variant of one of the most famous and most studied problems in the Operations Resea rature: the Uncapacitated Facility Location (UFL) problem. UFL with linear costs and its cardinali strained variant known as the p-median problem play a prominent role in the area of clustering a ssification, where they are used for unsupervised learning; for further references regarding the interp operations research and data mining, see e.g., Meisel and Mattfeld [40], Olafsson et al. [43]. UFL w adratic allocation costs, on the other side, appears as an important subproblem in the design of ene tribution networks.

Fischetti, Ljubic, Sinnl. *Management Science, 2016.*

Uncapacitated Facility Location

For each fixed $\hat{y}$, the projected problem is:

$$\min \sum_i \sum_j c_{ij} x_{ij}$$

$$s.t. \sum_i x_{ij} = 1, \forall j \in J,$$

$$x_{ij} \leq \hat{y}_i, \forall i \in I, j \in J$$

$$x_{ij} \geq 0.$$

and separable for $j \in J$:

$$\min \sum_i c_{ij} x_{ij}$$

$$s.t. \sum_i x_{ij} = 1, \forall i \in I,$$

$$x_{ij} \leq \hat{y}_i, i \in I$$

$$x_{ij} \geq 0, i \in I.$$

Uncapacitated Facility Location

Optimality Cuts:

$$v(y_i) = \min \sum_i c_{ij} x_{ij}$$

$$s.t. \sum_i x_{ij} = 1, \forall i \in I,$$

$$x_{ij} \leq \hat{y}_i, i \in I$$

$$x_{ij} \geq 0, i \in I.$$

The Lagrangean function is: $\sum_i c_{ij} \hat{x}_{ij} + \hat{u}_0 (1 - \sum_i x_{ij}) + \sum_i \hat{u}_i (\hat{x}_i - y_i)$, so the Benders cut is:

$$\theta \geq v(\hat{y}) - \sum_i \hat{u}_i (y_i - \hat{y}_i)$$

(Actually, $\hat{x}$ and $\hat{u}$ can be explicitly constructed from $\hat{y}$, Fischeti, Ljubic & Sinnl, 2015)

Convex Uncapacitated Facility Location

$$
\begin{aligned}
\min \sum_{i \in I} f_i y_i + \sum_i \sum_j c_{ij} x_{ij}^2 \\
\text{s.t.} \quad & \sum_{i \in I} x_{ij} = 1, \forall j \in J, \\
& x_{ij} \leq y_i, \forall i \in I, j \in J, \\
& x_{ij} \geq 0, \forall i \in I, j \in J, \\
& y_i \in \{0, 1\}, \forall i \in I,
\end{aligned}
$$

$$
\begin{aligned}
\min \sum_{i \in I} f_i y_i + \sum_i \sum_j c_{ij} z_{ij} \\
\text{s.t.} \quad & \sum_{i \in I} x_{ij} = 1, \forall j \in J, \\
& x_{ij} \leq y_i, \forall i \in I, j \in J, \\
& x_{ij}^2 \leq z_{ij} y_i, \forall i \in I, j \in J, \\
& x_{ij} \geq 0, \forall i \in I, j \in J, \\
& y_i \in \{0, 1\}, \forall i \in I,
\end{aligned}
$$

Explicit Benders Cuts: $\theta \geq v(\hat{y}) - \sum_{i \in I} (u_i^* + v_i^* z_i^*)(y_i - y_i^*)$

Modern Benders Decomposition

The success in recent implementations of GBD comes from:

- Commercial Solvers as Gurobi, CPLEX, Xpress allow the control of callbacks.
- Benders cuts can be incorporated into a branch-and-cut scheme, as Lazy Constraints.
- Stabilization methods that allow directing the search (Kelley, 1960) or Level stabilizations:

$$\min_{y,\theta} \ \theta + \frac{1}{2t}\|y - y_k\|^2$$

$$\min_{y,\theta} \ \theta \qquad \|y - y_k\|^2 \leq R.$$

$$\min_{y,\theta} \frac{1}{2}\|y - y_k\|^2 \qquad \theta \leq L.$$

- Combinatorial Benders Cuts.

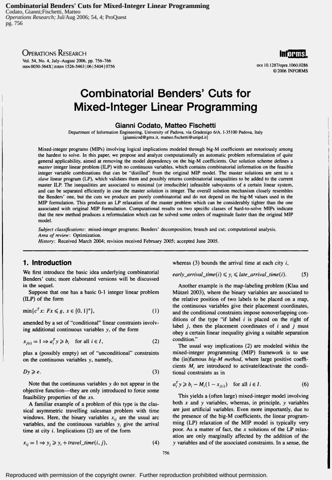
CBC

$\min\{c^t x \text{ or } d^t y : x \in P_X, y \in P_Y, (x, y) \in P_{XY}, x \in \mathbb{Z}_+^{n_1} \times \{0,1\}^{n_2}, y \geq 0\}$.

- When solving SP: UB^*.
- Next time we solve SP add $obj_{SP} \leq UB - \varepsilon$.
- If feasible: Update UB.
- Otherwise add cuts in the form:

$$\sum_{i \in C} x_{ji} \leq |C| - 1$$

where C is a inclusion-minimal set such that the SP is not feasible (computable via IIS). Useful in Map Labeling , Statistical Classification, ...



Combinatorial Benders' Cuts for Mixed-Integer Linear Programming
Codato, Gianni;Fischetti, Matteo
Operations Research; Jul/Aug 2006; 54, 4; ProQuest
pg. 756

OPERATIONS RESEARCH
Vol. 54, No. 4, July–August 2006, pp. 756–766
ISSN 0030-364X | EISSN 1526-5463 | 06 | 5404 | 0756

INFORMS
DOI 10.1287/opre.1060.0286

Combinatorial Benders' Cuts for Mixed-Integer Linear Programming

Gianni Codato, Matteo Fischetti
Department of Information Engineering, University of Padova, via Gradenigo 6/A, I-35100 Padova, Italy
{giannicod@gmx.it, matteo.fischetti@unipd.it}

Mixed-integer programs (MIPs) involving logical implications modeled through big-M coefficients are notoriously among the hardest to solve. In this paper, we propose and analyze computationally an automatic problem reformulation of quite general applicability, aimed at removing the model dependency on the big-M coefficients. Our solution scheme defines a *master* integer linear problem (ILP) with no continuous variables, which contains combinatorial information on the feasible integer variable combinations that can be "distilled" from the original MIP model. The master solutions are sent to a *slave* linear program (LP), which validates them and possibly returns combinatorial inequalities to be added to the current master ILP. The inequalities are associated to minimal (or irreducible) infeasible subsystems of a certain linear system, and can be separated efficiently in case the master solution is integer. The overall solution mechanism closely resembles the Benders' one, but the cuts we produce are purely combinatorial and do not depend on the big-M values used in the MIP formulation. This produces an LP relaxation of the master problem which can be considerably tighter than the one associated with original MIP formulation. Computational results on two specific classes of hard-to-solve MIPs indicate that the new method produces a reformulation which can be solved some orders of magnitude faster than the original MIP model.

Subject classifications: mixed-integer programs; Benders' decomposition; branch and cut; computational analysis.
Area of review: Optimization.
History: Received March 2004; revision received February 2005; accepted June 2005.

1. Introduction

We first introduce the basic idea underlying combinatorial Benders' cuts; more elaborated versions will be discussed in the sequel.

Suppose that one has a basic 0-1 integer linear problem (ILP) of the form

$$\min\{c^T x : Fx \leqslant g, \; x \in \{0,1\}^n\}, \tag{1}$$

amended by a set of "conditional" linear constraints involving additional continuous variables y, of the form

$$x_{j(i)} = 1 \Rightarrow a_i^T y \geqslant b_i \quad \text{for all } i \in I, \tag{2}$$

plus a (possibly empty) set of "unconditional" constraints on the continuous variables y, namely,

$$Dy \geqslant e. \tag{3}$$

Note that the continuous variables y do not appear in the objective function—they are only introduced to force some feasibility properties of the xs.

A familiar example of a problem of this type is the classical asymmetric travelling salesman problem with time windows. Here, the binary variables x_{ij} are the usual arc variables, and the continuous variables y_i give the arrival time at city i. Implications (2) are of the form

$$x_{ij} = 1 \Rightarrow y_j \geqslant y_i + travel\_time(i, j), \tag{4}$$

whereas (3) bounds the arrival time at each city i,

$$early\_arrival\_time(i) \leqslant y_i \leqslant late\_arrival\_time(i). \tag{5}$$

Another example is the map-labeling problem (Klau and Mützel 2003), where the binary variables are associated to the relative position of two labels to be placed on a map, the continuous variables give their placement coordinates, and the conditional constraints impose nonoverlapping conditions of the type "if label i is placed on the right of label j, then the placement coordinates of i and j must obey a certain linear inequality giving a suitable separation condition."

The usual way implications (2) are modeled within the mixed-integer programming (MIP) framework is to use the (in)famous *big-M method*, where large positive coefficients M_i are introduced to activate/deactivate the conditional constraints as in

$$a_i^T y \geqslant b_i - M_i(1 - x_{j(i)}) \quad \text{for all } i \in I. \tag{6}$$

This yields a (often large) mixed-integer model involving both x and y variables, whereas, in principle, y variables are just artificial variables. Even more importantly, due to the presence of the big-M coefficients, the linear programming (LP) relaxation of the MIP model is typically very poor. As a matter of fact, the x solutions of the LP relaxation are only marginally affected by the addition of the y variables and of the associated constraints. In a sense, the