Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Computers & Operations Research 32 (2005) 1429–1450

computers & operations research

www.elsevier.com/locate/dsw

A survey on benders decomposition applied to fixed-charge network design problems

Alysson M. Costa*

Canada Research Chair in Distribution Management, HEC Montréal, 3000 Chemin de la Côte-Sainte-Catherine, Montréal, Canada H3T 2A7

Abstract

Network design problems concern the selection of arcs in a graph in order to satisfy, at minimum cost, some flow requirements, usually expressed in the form of origin–destination pair demands. Benders decomposition methods, based on the idea of partition and delayed constraint generation, have been successfully applied to many of these problems. This article presents a review of these applications.

Keywords: Benders decomposition; Network design; Fixed charge

1. Introduction

Network design problems are central to a large number of contexts including transportation, telecommunications and power systems. The idea is to establish a network of links (roads, optical fibers, electric lines, etc.) that enables the flow of commodities (people, data packets, electricity, etc.) in order to satisfy some demand characteristics. We are particularly interested in *fixed-charge network design problems*, where, in order to use a link, one must pay a fixed cost representing, for example, the cost of constructing a road, or installing an electric line, etc.

A large number of practical applications may be represented by fixed-charge network design models. One important area is the *service network design problem* which arise, for example, in airline and trucking companies. The idea is to maximize the profit by setting routes and schedules,

* Tel.: +1-514-343-6111x8719; fax: +1-514-343-7121.
E-mail address: alysson@crt.umontreal.ca (A.M. Costa).

0305-0548/$ - see front matter
doi:10.1016/j.cor.2003.11.012

given some resource constraints. For example, airline companies must determine the covered routes and the frequency of the flights considering aircraft and crew availability [1]. Similarly, express package delivery companies (e.g., UPS, Fedex) must establish routes, assign aircraft to them and decide about the flow of packages [2].

Various applications can also be found in telecommunications. Examples are the design of a local access network with one or two technologies [3,4], the project of a terminal layout [5,6], and the interconnection of existing networks [7].

In power systems, fixed-charge network design is used to plan the energy transmission from the generation plants to the consumer centers [8,9] and the energy distribution inside these centers [10,11]. In the latter case, network models can also be used in an operational context to obtain the configuration that minimizes daily loss costs [12].

In all these problems, a proper design can yield better operation levels and cost reductions. The total amount of these reductions is obviously related to each specific problem. However, the economical importance of most of the cited problems and the key role played by the network design in the systems operation suggest that the savings can be significant. For example, Standard and Poor's [13] estimates the revenues of package delivery industry as 52 billions US$ only in the United States.

Probably because of the economical importance of the associated problems, several solution methodologies for network design problems are available. These range from pure heuristic methods to optimal implicit enumeration. Amongst the most successful solution approaches we find Benders decomposition [14]. The basic idea behind this method is to decompose the problem into two simpler parts: the first part, called *master problem*, solves a relaxed version of the problem and obtain values for a subset of the variables. The second part, called *auxiliary problem* (or *subproblem*), obtains the values for the remaining variables while keeping the first ones fixed, and uses these to generate cuts for the master problem. The master and auxiliary problems are solved iteratively until no more cuts can be generated. The conjunction of the variables found in the last master and subproblem iteration is the solution to the original formulation.

The structure of fixed-charge network design problems presents a natural decomposition scheme for the Benders approach: the variables representing the opening of the links are solved in the master problem while the ones representing the actual flow of commodities are kept in the subproblem. Therefore, at each iteration the master solution gives a tentative network for which the subproblem finds the optimal flow of commodities.

Due to the need to solve the master and the auxiliary problems several times, the decomposition approach is only reasonable if these problems can be solved efficiently. This is the case for network design problems, where most of the times, it is much easier to solve the decomposed problems than the original one. Moreover, especially for the subproblem, it is sometimes possible to proceed with further decomposition (by commodity, by period, etc.), resulting in even more efficient solution methods. All this makes Benders decomposition a competitive methodology for this family of problems.

In this article we review the use of Benders decomposition in fixed-charge network design. In Section 2, we present the Benders decomposition approach in further detail. Then, in Section 3 we present the different fixed-charge network problems and a survey of the application of Benders decomposition algorithms to these models. Section 4 closes the paper with a summary and conclusions.

2. Benders decomposition

Benders decomposition [14] is a classical solution approach for combinatorial optimization problems, based on the ideas of partition and delayed constraint generation. Examples of successful applications of this methodology to mixed-integer problems are abundant. We cite, for example, the seminal paper of Geoffrion and Graves [15] on multicommodity distribution network design and the extensions presented by Cordeau et al. [16] on the same problem. Other applications include the locomotive and car assignment problem [17–19], the large-scale water resource management problem [20], and the two stage stochastic linear problem [21].

The method partitions the model to be solved into two simpler problems, named master and auxiliary problems. The master problem is a relaxed version of the original problem, containing only a subset of the original variables and the associated constraints. The auxiliary problem is the original problem with the variables obtained in the master problem fixed.

In the case of the fixed-charge network design problem, the master deals with the integer variables that define the network while the subproblem works with the continuous variables representing the actual flow of commodities for the tentative network obtained in the master problem. Consider, for example, the general formulation presented below, which encompasses most fixed-charge network design models proposed in the literature:

General fixed-charge network design formulation:

$$\text{Minimize} \quad cx + dy \tag{1}$$

$$\text{subject to} \quad Ax + By \geqslant b, \tag{2}$$

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

$$x \geqslant 0, \quad y \geqslant 0 \quad \text{and integer.} \tag{4}$$

Vectors x and y are the continuous and integer variables, respectively, while c and d are the row vectors of the associated costs. Matrices A, B and D and vectors b and e have the appropriate dimensions. This problem can be expressed as

$$\min_{\bar{y}\in Y}\{d\bar{y} + \min_{x\geqslant 0}\{cx : Ax \geqslant b - B\bar{y}\}\}, \tag{5}$$

where $Y = \{y \mid Dy \geqslant e,\ y \geqslant 0 \text{ and integer}\}$. The inner minimization is a linear program. Associating dual variables u to constraints $Ax \geqslant b - B\bar{y}$, we can write the dual version of this problem as

$$\max_{u\geqslant 0}\{u(b - B\bar{y}) : uA \leqslant c\}. \tag{6}$$

This is the Benders decomposition subproblem. Using duality theory, the primal and dual formulations can be interchanged. Therefore, (5) can be rewritten as

$$\min_{\bar{y}\in Y}\{d\bar{y} + \max_{u\geqslant 0}\{u(b - B\bar{y}) : uA \leqslant c\}\}. \tag{7}$$

Note that the feasible space of the subproblem (inner maximization) is independent of the choice made for variables y. Let $F = \{u \mid u \geqslant 0;\ uA \leqslant c\}$ represent this feasible space. We assume that F is not empty for it would correspond to a primal problem either infeasible or unbounded. F is therefore composed of extreme points u^p (for $p = 1,\ldots,P$) and extreme rays r^q (for $q = 1,\ldots,Q$).

The solution of the subproblem can be either bounded or unbounded. In the first case, the solution is one of the extreme points u^p $(p=1,\dots,P)$. In the latter situation, there is a direction r^q for which $r^q(b-B\bar{y})>0$. The unbounded situation results in an unfeasible primal problem and must be avoided. We must therefore eliminate the values of $\bar{y}$ that would yield an unbounded inner dual problem. This is done by explicitly considering the restrictions

$$r^q(b-B\bar{y})\leqslant 0, \quad q=1,\dots,Q. \tag{8}$$

With this restrictions in the external formulation, the maximum value of the inner problem is the value of one of the extreme points of F. Problem (7) becomes

$$\min_{\bar{y}\in Y} \quad \{d\bar{y}+\max\{u^p(b-B\bar{y}):\ p=1,\dots,P\}\}$$

$$\text{s.t} \quad r^q(b-B\bar{y})\leqslant 0, \quad q=1,\dots,Q, \tag{9}$$

or, with the use of an auxiliary continuous variable z:

$$\text{Minimize} \quad dy+z \tag{10}$$

$$\text{subject to} \quad z\geqslant u^p(b-B\bar{y}), \quad p=1,\dots,P, \tag{11}$$

$$r^q(b-B\bar{y})\leqslant 0, \quad q=1,\dots,Q, \tag{12}$$

$$y\in Y, \quad z\geqslant 0. \tag{13}$$

Formulation (10)–(13) is called the Benders reformulation and its drawback is that the number of extreme points and extreme rays is usually extremely large. To overcome this limitation, Benders proposed to delay the generation of constraints (11) and (12). Initially, only constraints (13) are considered, yielding the first master problem:

$$\text{Minimize} \quad dy+z \tag{14}$$

$$\text{subject to} \quad y\in Y, \quad z\geqslant 0. \tag{15}$$

This problem is a relaxed version of (10)–(13) and therefore the objective value $dy+z$ is a lower bound to the original problem. Once this problem is solved, a tentative configuration of variables y is used in subproblem (6). This subproblem is solved and the result is either unbounded, in which case a constraint of type (12) is inserted in the master, or it is an extreme point. In the latter situation, the conjunction of the solution of the master with the primal solution of the subproblem provides a complete solution (upper bound) to the original problem, while the dual solution of the subproblem is used to generate a constraint of type (11), which is also inserted in the master.

The master and the auxiliary problems are solved iteratively, until the upper and lower bounds are sufficiently close. Because of this iterative constraint generation scheme, Benders decomposition can be regarded as sort of a “dual version” of Dantzig–Wolfe column-generation: the application of Dantzig–Wolfe column generation to the dual of the original problem.

Several extensions have been proposed for the Benders decomposition method. One of the most important ones is presented by Geoffrion [22] who suggests a “generalized Benders decomposition” approach. Geoffrion makes use of non-linear duality theory and extends the Benders method to the case where the subproblem is a convex optimization problem. This development enables the application of Benders decomposition to a whole new set of problems, particularly those where the joint problem is generally non-convex but can be made convex by fixing one set of variables. Values for these variables are therefore obtained in the master problem and fixed in the subproblem which then becomes a convex optimization problem for which many optimization methods are available.

Magnanti and Wong [23] have studied the influence of cuts in a Benders decomposition algorithm. They showed that the use of stronger cuts, when available, may have a great impact in the convergence of the algorithm by reducing the number of iterations. The idea is simple. Many times the solution of subproblem (6) is not unique. Indeed, for network design problems this is usually true, since the subproblems are variations of *shortest route*, *transshipment* and other network optimization problems, known for their degeneracy. Therefore, one can try to judiciously choose the dual variables in order to obtain better cuts. To determine what is a better cut the authors use the notion of dominance: for the presented reformulation of problem (1)–(4), a cut generated from the extreme point u^1 dominates a cut generated from the extreme point u^2 if

$$u^1(b - B\bar{y}) \geqslant u^2(b - B\bar{y}) \tag{16}$$

for all $\bar{y} \in Y$ with strict inequality for at least one point. A cut that is dominated by no other cut is Pareto-optimal. Let Y^{LP} be the polyhedron defined by the linear relaxation of Y, and $ri(Y^{\mathrm{LP}})$ be the relative interior of Y^{LP}. The following problem yields a Pareto-optimal cut for the case of the general problem (1)–(4):

$$\text{Maximize} \quad u(b - By^{\mathrm{o}}) \tag{17}$$

$$\text{subject to} \quad uA \leqslant c, \tag{18}$$

$$v(\bar{y}) = u(b - B\bar{y}), \tag{19}$$

$$u \geqslant 0, \tag{20}$$

where $y^{\mathrm{o}} \in ri(Y^{LP})$ and $v(\bar{y})$ is the optimal value of subproblem (6) when variables y are fixed in $\bar{y}$. The objective function maximizes the strength of the cut for y^{o} while the constraints define the feasible space as the optimal solutions of the original subproblem.

The authors also analyze the impact of different formulations on the performance of a Benders decomposition algorithm. The conclusion is that two different formulations for a problem can yield very different performances. In fact, tighter formulations yield stronger Benders cuts, to the extent that for the Convex Hull formulation, only one cut is needed. However, adding constraints to a formulation not only strengthens the Benders cuts but also complicates the solution of the linear subproblems, yielding a trade-off between the quality of the Benders cuts and the computational effort needed to solve the subproblems.

In a later work, Magnanti et al. [24] have extended the study to the context of the *uncapacitated network design problem*. They showed the generality of Benders decomposition by proving that some known cuts [25–28] are actually Benders cuts for particular choices of the integer variables (the tentative network configuration). Moreover, they showed that in this particular case, Pareto-optimal cuts can be generated at the price of solving *k minimum cost flow problems* (one for each commodity). The authors also advocate the use of Benders decomposition in conjunction with other approaches, based on their own experience with the use of a dual-ascent method, as will be seen in the next section.

3. Literature review

Network design problems concern the selection of arcs in a graph in order to satisfy, at minimum cost, some flow requirements usually expressed in the form of origin–destination pair demands (for surveys of these problems see Magnanti and Wong, [28], Minoux, [29] and Balakrishnan et al. [30]). The *fixed-charge network design problem* (FNDP) has the particularity that each arc has an associated fixed-cost which must be paid if the arc is part of the solution.

In this section we present a review of the Benders decomposition algorithms proposed for the FNDP. We organize our presentation according to the different problems. First, we characterize the models as uncapacitated or capacitated, and we then enumerate the most common problems within each category.

The idea is to present different versions of fixed-charge network design problems, giving for each one, a mathematical formulation and general comments on the applications of Benders decomposition methods for that problem found in the literature.

Concerning the mathematical developments, we note that there are two main types of formulations that are used in the literature: node-arc and arc-path models. For ease of presentation, we concentrate on node-arc models and we develop the arc-path model only for the last and more general case of Section 3.2.3.

In all developments we denote by $G(N,A,K)$ an undirected graph where N is the set of nodes, A is the set of edges, and K is the set of commodities to be transported (people or goods, data packets, electric power, etc.). Variables x_{ij}^k represent the actual flow of commodity k going from node i to node j, and variables y_{ij} are integer variables associated with the utilization of the link (uncapacitated case) or to its capacity (capacitated case).

Note that, although the graph is undirected, it makes sense to define both variables x_{ij}^k and x_{ji}^k, because the flow itself is directed. Thus, we consider that an edge in A is doubly represented by the arcs (i,j) and (j,i). However, whenever we refer to the integer variables (or their coefficients) we assume that only the edge (i,j) with $i<j$ is considered.

3.1. Uncapacitated network design problems (UNDP)

We first consider *uncapacitated network design problems*. In these problems, there is no limit on the flow that can circulate through the selected links. Many real-life applications may be well represented by such models [4,5,31].

3.1.1. The origin–destination pairs UNDP

We start with the *origin–destination pairs uncapacitated network design problem*. In this case, to each commodity $k=1,\dots,|K|$ is associated a demand d_k, an origin node $O(k)$ and a destination node $D(k)$. In some cases, the fixed costs fully represent the real cost of the network, i.e., there is no volume cost associated with the flow of a commodity on one link (see, e.g., [32], and references therein). However, in order to appropriately model some problems, one must sometimes include volume-based costs, as in the work of Randazzo et al. [4].

Using the notation introduced at the beginning of the section, we present below one formulation for this class of problems:

(F_1) *Origin–destination pairs UNDP formulation*:

$$\text{Minimize} \quad \sum_{(i,j)\in A}\left(\sum_{k\in K} c_{ij}^k x_{ij}^k + f_{ij} y_{ij}\right) \tag{21}$$

$$\text{subject to} \quad \sum_{j\mid(i,j)\in A} x_{ij}^k - \sum_{j\mid(j,i)\in A} x_{ji}^k = \begin{cases} d_k, & i=O(k),\ \forall k\in K,\\ 0, & i\notin\{O(k),\ D(k)\},\ \forall k\in K,\\ -d_k, & i=D(k),\ \forall k\in K,\end{cases} \tag{22}$$

$$x_{ij}^k \leqslant d_k y_{ij}, \quad \forall (i,j)\in A,\ \forall k\in K, \tag{23}$$

$$x_{ij}^k \geqslant 0, \quad \forall (i,j)\in A,\ \forall k\in K, \tag{24}$$

$$y_{ij}\in\{0,1\}, \quad \forall (i,j)\in A. \tag{25}$$

Variables y_{ij} are associated with the construction of link $(i,j)\in A : y_{ij}=1$ if (i,j) belongs to the final solution, otherwise $y_{ij}=0$. The objective function (21) is the sum of variable and fixed costs. In this function, c_{ij}^k is the linear cost associated with the flow of one unit of commodity k through link (i,j) and f_{ij} is the fixed cost associated with the selection of link (i,j) in the final solution. Some non-linear versions of this function can be considered by "generalized Benders decomposition" (see, for example, the work of Hoang [33] later on this section or the design of electrical networks in Section 3.2.3). However, for simplicity, we concentrate on linear objective functions.

The set of restrictions (22) ensure that the flow of commodity k leaves the origin node $O(k)$ and arrives at the destination node $D(k)$, and are also responsible for flow conservation. Constraints (23) limit the flow of commodities to selected arcs, constraints (24) define the non-negativity of variables x_{ij}^k, and constraints (25) impose the integrality of variables y_{ij}. Constraints (23) can be substituted by aggregate constraints, in the form

$$\sum_{k\in K} x_{ij}^k \leqslant \left(\sum_{k\in K} d_k\right) y_{ij}, \quad \forall (i,j)\in A. \tag{26}$$

This aggregation can substantially reduce the number of constraints. However, as discussed in the last section, it has been shown that the Benders cuts obtained with disaggregated (tighter) formulations are more efficient than the ones obtained with aggregated ones [24].

Application of Benders decomposition: The basic uncapacitated case was solved via Benders decomposition in an early work of Magnanti et al. [24]. There, the master problem proposes a tentative network by setting the integer variables and the subproblem finds the continuous variables flow distribution. The authors use a set of instances of up to 30 nodes and 130 arcs to prove two points: (i) that the use of auxiliary techniques can improve the performance of the Benders decomposition (in this case, a preprocessing based on a dual-ascent procedure in order to eliminate variables), and (ii) that the intelligent choice of cuts, specifically the Pareto-optimal cuts (see Section 2), can strongly affect the performance of the algorithm.

The work of Magnanti et al. was extended ten years later by Gutierrez et al. [34], who proposed a robust approach able to consider the uncertainty in the transportation costs c_{ij}^k and in the demands d_k. The data uncertainty is described though a set of scenarios, each one with different values for c_{ij}^k. Note that, since the authors deal with the uncapacitated case, the changes in c_{ij}^k may reflect both changes in the cost itself or in the demand (with the appropriate scaling). The solution of the model is done via a multi-master Benders algorithm, where an individual master problem is associated with each possible scenario. Each time a master problem is solved, there is a cross generation of Benders cuts, i.e., the subproblem generates a cut for each of the master problems. The authors use a set of instances based on the one presented by Magnanti et al. [24] with 11 scenarios for each original problem and conclude that the cross-generation of Benders cuts accelerates the convergence of the algorithm.

An interesting version of the uncapacitated network design problem with one facility has been proposed by Hoang [33]. The author develops a formulation of this problem with non-linear costs. The objective is to select among several projects (each corresponding to a set of investments in the network) respecting some budget constraints. In this case, no explicit upper bounds on the arc capacities like (23) exist: they are indirectly accounted for by the convexity in the cost functions. Generalized Benders decomposition is used, together with a Lagrangian heuristic to solve the master, and a convex cost multicommodity algorithm [35] is applied to the solution of the subproblem. Two instances are successfully solved, the largest one having 155 nodes, 376 arcs and 720 commodities.

3.1.2. The single-origin UNDP (local access network design)

An important special case of the problem presented in Section 3.1.1 is where all demand points must be connected to a single origin. This is the situation, for example, in the design of *local access networks* (LAN) [3,36]. When the objective function is well represented only by the variable costs, the LAN Design can be formulated as the easy *single-source transshipment problem*. On the other hand, if there are only fixed-charge costs, it becomes an NP-hard *Steiner problem*, or a *minimum spanning tree problem* (if all the nodes must be reached). In the latter case, an important situation is where the number of links connecting directly or indirectly to the source is limited. This problem is known as the *capacitated minimum spanning tree problem* [5,6] (we classify this problem as uncapacitated since the flow on the arc itself is not subject to any upper bound).

We present a mathematical formulation for the general case where each link has a fixed and variable associated cost. In (F_2) we consider each demand as a different commodity, with different transportation costs c_{ij}^k that are dependent both on the link and on the commodity. Node O is the origin for all the commodities.

(F_2) *Single-origin UNDP formulation*:

$$\text{minimize} \quad \sum_{(i,j)\in A} \left(\sum_{k\in K} c_{ij}^k x_{ij}^k + f_{ij} y_{ij} \right) \tag{27}$$

$$\text{subject to} \quad \sum_{j\,|\,(i,j)\in A} x_{ij}^k - \sum_{j\,|\,(j,i)\in A} x_{ji}^k = \begin{cases} d_k, & i = O, \ \forall k \in K, \\ 0, & i \notin \{O,\ D(k)\}, \ \forall k \in K, \\ -d_k, & i = D(k), \ \forall k \in K, \end{cases} \tag{28}$$

$$x_{ij}^k \leqslant d_k y_{ij}, \quad \forall (i,j) \in A, \ \forall k \in K, \tag{29}$$

$$x_{ij}^k \geqslant 0, \quad \forall (i,j) \in A, \ \forall k \in K, \tag{30}$$

$$y_{ij} \in \{0,1\}, \quad \forall (i,j) \in A. \tag{31}$$

Besides the fact that all commodities are now originating at one single-origin node, (F_2) is identical to (F_1). However, the presence of a single-origin node yields a simplified formulation for the case where the variable costs are independent on the commodity. (F_2') presents this simplified single-origin, single-commodity formulation.

(F_2') *Single-origin, single-commodity UNDP formulation*:

$$\text{Minimize} \quad \sum_{(i,j)\in A} (c_{ij}x_{ij} + f_{ij}y_{ij}) \tag{32}$$

$$\text{subject to} \quad \sum_{j \mid (i,j)\in A} x_{ij} - \sum_{j \mid (j,i)\in A} x_{ji} = \begin{cases} D, & i = O, \\ -d_i, & i \neq O, \end{cases} \tag{33}$$

$$x_{ij} \leqslant D y_{ij}, \quad \forall (i,j) \in A, \tag{34}$$

$$x_{ij} \geqslant 0, \quad \forall (i,j) \in A, \tag{35}$$

$$y_{ij} \in \{0,1\}, \quad \forall (i,j) \in A. \tag{36}$$

The parameter d_i represents the demand at each node and D is the sum of the demands in all nodes excluded the origin. The objective function (32) of formulation (F_2') still minimizes the sum of fixed and variable costs. However, there is no need for a differentiation among commodities. The only requirement is that the correct amount of flow leaves the node origin and reaches each one of the demand nodes, which is ensured by constraints (33). Constraints (34) still forbid the flow in closed links.

As pointed out by Magnanti and Wong [23], selecting the proper formulation is an important factor that affects the computational performance of Benders decomposition. In this case, we note that since constraints (34) of formulation (F_2') make use of the sum of all the demands D, they are an equivalent of the aggregated constraints presented in (26). Therefore, we should expect this formulation to be weaker than formulation (F_2), which uses the disaggregated constraints as expressed in (29).

Application of Benders decomposition: Randazzo and Luna [3] present a comparison of a Lagrangean relaxation-based branch-and-bound, a branch-and-cut and a Benders decomposition methodologies for the LAN design problem presented in this section. Prior to the Benders decomposition, the authors use a linear relaxation and a shortest-path algorithm to obtain a feasible solution, which is used to generate initial values for the dual variables. Redundant strengthening constraints are used in the master, to increase the probability of generating a feasible solution. The subproblem is decomposed into k easy network flow problems, one for each commodity. Benders decomposition outperformed the two other methods on three small sets of instances. Perhaps the most important conclusion obtained by the authors is the fact that Benders decomposition, although slower than the branch-and-cut on six of the 30 instances, was the only algorithm able to solve to optimality all instances within the limit time (24 h), leading to the idea that it is a more robust method.

In an earlier article, Gavish [37] has presented a Benders decomposition algorithm for the *capacitated minimal spanning tree problem*. The algorithm is rather standard: the master problem fixes the integer variables yielding two situations (i) the selected edges form a spanning tree: then, either the solution is optimal (if the number of links connecting to the source respects the limit) or the links connected to the source can be used to generate a cut and solve the master problem again, (ii) the graph is not connected, in which case a cut can be generated and the master problem is solved again. The algorithm was tested on networks containing up to 12 nodes. The number of generated cuts was very large, resulting in a poor performance of the algorithm.

3.1.3. The two-technologies single-origin UNDP (two technologies local access network design problem)

A third common situation in uncapacitated network design problems is the case where two technologies, each having their own advantages, are available [4,31,38]. One technology may have a larger fixed-cost but a smaller variable cost and vice versa (as happens with optical fibers and copper in telecommunications networks).

The formulation is an extension of formulation (F_2), with the duplication of variables x_{ij}^k and y_{ij}. Therefore, in (F_3), y_{ijt} is the binary variable associated with the construction of the link (i,j) with technology t. Analogously, x_{ijt}^k is the flow of commodity k through that link.

(F_3) *Two-technologies single-origin UNDP formulation*:

$$\text{Minimize} \quad \sum_{(i,j)\in A} \sum_{t=1}^{2} \left(\sum_{k\in K} c_{ijt}^k x_{ijt}^k + f_{ijt} y_{ijt} \right) \tag{37}$$

$$\text{subject to} \quad \sum_{t=1}^{2} \left(\sum_{j\,|\,(i,j)\in A} x_{ijt}^k - \sum_{j\,|\,(j,i)\in A} x_{jit}^k \right) = \begin{cases} d_k, & i = O, \ \forall k \in K, \\ 0, & i \notin \{O, \ D(k)\}, \ \forall k \in K, \\ -d_k, & i = D(k), \ \forall k \in K, \end{cases} \tag{38}$$

$$x_{ijt}^k \leqslant d_k y_{ijt}, \quad \forall (i,j)\in A, \ \forall k \in K, \ t = 1,2, \tag{39}$$

$$\sum_{t=1}^{2} y_{ijt} \leqslant 1, \quad \forall (i,j)\in A, \tag{40}$$

$$x_{ijt}^k \geqslant 0, \quad \forall (i,j)\in A, \ \forall k \in K, \ t = 1,2, \tag{41}$$

$$y_{ijt} \in \{0,1\}, \quad \forall (i,j)\in A, \ t = 1,2. \tag{42}$$

In (F_3), the goal (37) is to minimize the sum of variable and fixed costs for both technologies. This must be done while respecting the demand requirements (38) and the fact that, for both technologies, there can only be flow in a link if the link is open (39). Each link uses a single technology, as expressed by constraints (40).

We observe that (F_3) is the most complete of the four formulations. In fact, (F_1), (F_2) and (F_2') can all be regarded as special cases of (F_3). Moreover, (F_3) easily adapts to the case where more technologies are available (although the case with two technologies is the most common in practice).

Straightforward extensions of (F_3) include models that require a cost for connecting links with different technologies. These models are useful for the common situation where extra equipment is needed to perform the connection. Other models may be used to limit the number of such connections or restrict the nodes where it may take place.

Additional restrictions may force some 'primary nodes' to be served by the 'primary' technology [36]. Carneiro et al. [11] present an interesting application of this model for the problem of designing power distribution secondary networks. In this case, all the demand must be supplied by the 'secondary' technology. However, it is still advantageous to use the primary technology (in this case, the higher voltages) until very close to the demand points for it has much lower variable costs. The conversion from the first technology to the second requires the installation of additional equipment in the nodes of the network (electric transformers) and is limited to the capacity of these transformers.

Application of Benders decomposition: Randazzo et al. [4] use Benders decomposition to solve this problem. The algorithm is very similar to the one presented by Randazzo and Luna [3], but without the use of linear relaxation. The first feasible solution is obtained by means of a shortest-path algorithm considering only the variable costs. This initial solution is used to generate initial lower and upper bounds. Additional constraints are used in the master in order to guarantee an arborescent structure during its solution. The auxiliary subproblems are still easy network flow problems decomposable by commodity. The tests were conducted on a limited set of instances of up to 41 nodes (all instances could be solved to optimality with a linear relaxation model) for which Benders decomposition was usually faster (especially for the bigger instances) than Cplex 3.0.

3.2. Capacitated network design problems (CNDP)

We now consider the *capacitated network design problems*. As before, flow in a link is only allowed if a fixed cost is paid. However, the amount of this flow is limited to a certain capacity. Examples of applications of CNDP frequently occur in the design of telecommunications networks [39,40]. In this section, we analyze some of the most common models. These models are presented in the next subsections and classified according to the structure of the cost–capacity relationship.

3.2.1. The single-facility CNDP

The first CNDP to be considered is the *single-facility CNDP*. In this case, there is only one type of technology, which can be installed multiple times on each link. Each unit of the facility installed enables a given amount of flow and has a given cost. Examples of this problem can be found in Magnanti and Mirchandani [41] and [42].

In (F_4) we present a mathematical formulation for the case with origin–destination pair demands.

(F_4) *Capacitated single-facility network design problem*:

$$\text{Minimize} \quad \sum_{(i,j)\in A}\left(\sum_{k\in K} c_{ij}^k x_{ij}^k + f_{ij}y_{ij}\right) \tag{43}$$

$$\text{subject to} \quad \sum_{j\,|\,(i,j)\in A} x_{ij}^k - \sum_{j\,|\,(j,i)\in A} x_{ji}^k = \begin{cases} d_k, & i = O(k),\ \forall k\in K,\\ 0, & i \notin \{O(k),\ D(k)\},\ \forall k\in K,\\ -d_k, & i = D(k),\ \forall k\in K,\end{cases} \tag{44}$$

$$\sum_{k \in K} x_{ij}^k \leqslant C y_{ij}, \quad \forall (i,j) \in A, \tag{45}$$

$$x_{ij}^k \geqslant 0, \quad \forall (i,j) \in A, \ \forall k \in K, \tag{46}$$

$$y_{ij} \in \mathbb{Z}, \quad \forall (i,j) \in A. \tag{47}$$

In (F_4), there are two important modifications with respect to (F_1). First, one single constraint of type (45) is generated for each arc (instead of k constraints per arc). These constraints now link the x_{ij}^k variables together, as in constraints (26). Moreover, in the right-hand side of (45), d_k has been changed for C, the amount of capacity given by each unit of the *purchased* facility. The second modification is that constraints (46) indicate that y_{ij} is integer. These two modifications enable the selection of an integer number of unities of the facility to be installed in each individual link, therefore, giving the link the total capacity of $y_{ij} \times C$.

Modifications in order to make (F_4) a *single-origin* model analogous to the one presented in Section 3.1.2 are straightforward and equivalent to those required in the uncapacitated case.

Application of Benders decomposition: Sridhar and Park [32] work with a version of this problem that does not have variable costs. Besides the arc capacity constraints, their model also includes node capacity constraints, in the form

$$\sum_{j \,|\, (i,j) \in A} \sum_{k \in K} x_{ij}^k + \sum_{k \,|\, D(k)=i} d_k \leqslant \kappa_i \quad \forall i, \tag{48}$$

indicating that the node capacity κ_i must be greater or equal to all the flow passing through, originating or ending at node i. The capacity κ_i of each node is a parameter determined by the desired network performance. The model is solved using an embedded Benders enumeration algorithm, initiated with a lower bound given by a cutting plane algorithm based on two polyhedral cuts and a feasible solution given by a heuristic algorithm [43]. The method is applied to instances of up to 20 nodes and several demand levels, for which the Benders-and-cut approach outperformed a conventional branch-and-bound.

Ouorou et al. [44] deal with a similar single-facility capacitated problem. In this case, however, the demands are *elastic*, i.e., one of the goals of the optimization problem is to establish the optimal demand levels (this can be done considering that the demands are dependent on the asked prices, as is often the case in telecommunications systems). Another important difference is the fact that the authors deal with a *capacity expansion* problem, i.e., there is already sufficient capacity installed and the question is whether it may be advantageous to expand the capacity in order to serve consumers who are willing to pay higher tariffs in exchange for better service levels. A generalized Benders decomposition scheme is used, yielding a mixed-integer master problem that deals with the capacity expansions and a convex network flow auxiliary subproblem that solves the demand and flow levels. The algorithm chosen for solving the subproblems is an adaptation of the proximal decomposition method [45] to the case of elastic demands which alternates between shortest path calculations and primal-dual updates efficiently distributed among arcs and commodities. The authors solve instances based on real-life data with up to 19 nodes, 68 arcs and 342 commodities and the results suggest that Benders decomposition is an effective strategy.

3.2.2. The two-facility CNDP

The *two-facility CNDP* is the analogous capacitated model for the uncapacitated situation presented in Section 3.1.3. The extension here is that besides choosing the technology for each link, we must also choose the capacity to be installed. This capacity must come in multiples of the unitary facility capacity for the chosen technology. Therefore, given two technologies with two different unitary fixed-costs f_{ij1} and f_{ij2} and two different unitary capacities C_1 and C_2, formulation (F_5) models the problem of finding the cheapest network that supplies the demand requirements in the form of origin–destination pair demands.

(F_5) *Two-facility CNDP formulation*:

$$\text{Minimize} \quad \sum_{(i,j)\in A} \sum_{t=1}^{2} \left(\sum_{k\in K} c_{ijt}^k x_{ijt}^k + f_{ijt} y_{ijt} \right) \tag{49}$$

$$\text{subject to} \quad \sum_{t=1}^{2} \left(\sum_{j\,|\,(i,j)\in A} x_{ijt}^k - \sum_{j\,|\,(j,i)\in A} x_{jit}^k \right) = \begin{cases} d_k, & i = O(k),\ \forall k\in K, \\ 0, & i \notin \{O(k),\ D(k)\},\ \forall k\in K, \\ -d_k, & i = D(k),\ \forall k\in K, \end{cases} \tag{50}$$

$$\sum_{k\in K} x_{ijt}^k \leqslant C_t y_{ijt}, \quad \forall (i,j)\in A,\ t=1,2, \tag{51}$$

$$y_{ijt} \leqslant U_t z_{ijt}, \quad \forall (i,j)\in A,\ t=1,2, \tag{52}$$

$$\sum_{t=1}^{2} z_{ijt} \leqslant 1, \quad \forall (i,j)\in A, \tag{53}$$

$$x_{ijt}^k \geqslant 0, \quad \forall (i,j)\in A,\ \forall k\in K,\ t=1,2, \tag{54}$$

$$y_{ijt} \in \mathbb{Z}, \quad \forall (i,j)\in A,\ t=1,2, \tag{55}$$

$$z_{ijt} \in \{0,1\}, \quad \forall (i,j)\in A,\ t=1,2. \tag{56}$$

Additional variables z_{ijt} are necessary to limit to one the number of technologies in a link. Variable z_{ijt} equals 1 if the link (i,j) has technology t installed and equals 0 otherwise. The coefficient U_t is the maximum number of units of capacity t that can be installed on a link.

The objective function (49) minimizes the sum of variable and fixed costs for both technologies. The demand requirements are supplied by one of the two technologies as stated by (50). Constraints (51) enforce the capacity limits and constraints (52) in conjunction with (53) ensure that only one type of technology may be installed in each individual link. The definition of x_{ijt}^k as positive, y_{ijt} as integer and z_{ijt} as binary variables is effected by constraints (54), (55) and (56), respectively.

Magnanti et al. [46] report a very similar case in the design of private communication networks. In this case, both technologies may coexist in a link, having their capacities added up. Therefore, there is no need for the additional z_{ijt} variables nor for constraints (52) and (53). The technology index t may be also dropped from variables x_{ij}^k. Constraints (51) are therefore replaced by

$$\sum_{k\in K} x_{ij}^k \leqslant C_1 y_{ij1} + C_2 y_{ij2}, \quad \forall (i,j)\in A. \tag{57}$$

As in the last subsection, the conversion of (F_5) into a *single-origin* model is straightforward.

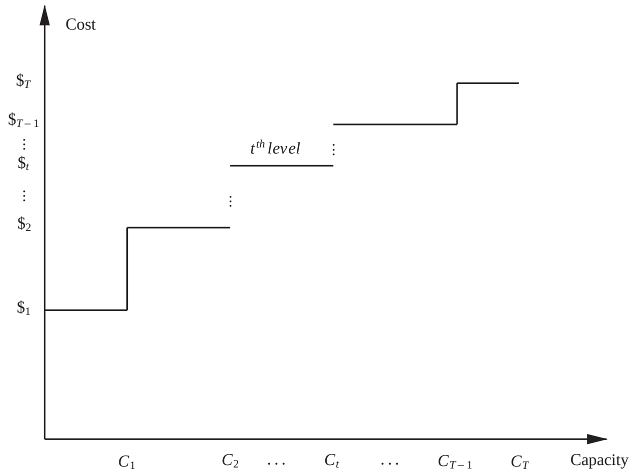

Fig. 1. Step-increasing $cost \times capacity$ function.

We are not aware of any authors that have used Benders decomposition to solve the problem presented in this section.

3.2.3. The step-increasing cost CNDP

The last and more general case is the *step-increasing cost CNDP* [47–49]. Here, the cost of *purchasing* capacity for a link is given by a step-increasing cost-capacity function. Formulation (F_6) models this last case. This formulation creates additional binary variables to cope with the cost–capacity function discontinuities. In (F_6), C_t is the allowed capacity of a link if that link is in the tth level of the cost function (refer to Fig. 1) and T is the total number of these levels.

(F_6) *The step-increasing cost network design problem* (*node-arc formulation*):

$$\text{Minimize} \quad \sum_{(i,j)\in A}\left(\sum_{k\in K} c_{ij}^k x_{ij}^k + \sum_{t=1}^{T} f_{ijt} y_{ijt}\right) \tag{58}$$

$$\text{subject to} \quad \sum_{j\,|\,(i,j)\in A} x_{ij}^k - \sum_{j\,|\,(j,i)\in A} x_{ji}^k = \begin{cases} d_k, & i = O(k), \quad \forall k \in K, \\ 0, & i \notin \{O(k),\ D(k)\}, \quad \forall k \in K, \\ -d_k, & i = D(k), \quad \forall k \in K, \end{cases} \tag{59}$$

$$\sum_{k\in K} x_{ij}^k \leqslant \sum_t C_t y_{ijt}, \quad \forall (i,j) \in A, \tag{60}$$

$$\sum_{t=1}^{T} y_{ijt} \leqslant 1, \quad \forall (i,j) \in A, \tag{61}$$

$$x_{ij}^k \geqslant 0, \quad \forall (i,j) \in A,\ \forall k \in K, \tag{62}$$

$$y_{ijt} \in \{0,1\},\ \forall (i,j) \in A,\ t = 1,\ldots,T. \tag{63}$$

Again, the objective function is to minimize the sum of variable and fixed costs, as expressed by (58). Constraints (59) guarantee that the demand requirements will be satisfied while constraints (60) limits the flow on each link according to the purchased capacity. Constraints (61) state that only one variable y_{ijt} may be selected, for these variables indicate the purchased capacity level, as shown in Fig. 1.

It is usual to express the node-arc formulation (F_6) as an equivalent *arc-path* formulation. In this case, let $\mathbb{P}(k)$ be the set of possible paths between nodes $O(k)$ and $D(k)$. The coefficient $a_{kp(i,j)}$ equals 1 if the commodity k flowing on path p uses arc (i,j), and $a_{kp(i,j)} = 0$, otherwise. Also let n_{kp} be the flow of commodity k on path p. Formulation (F_6') presents an arc-path formulation for this last case.

(F_6') *The step-increasing cost network design problem* (*arc-path formulation*):

$$\text{Minimize} \quad \sum_{(i,j)\in A} \left(\sum_{k\in K} c_{ij}^k x_{ij}^k + \sum_{t=1}^{T} f_{ijt} y_{ijt} \right) \tag{64}$$

$$\text{subject to} \quad x_{ij}^k = \sum_p a_{kp(i,j)} n_{kp}, \quad \forall (i,j)\in A, \ \forall k \in K, \tag{65}$$

$$\sum_p n_{pk} = d_k, \quad \forall k \in K, \tag{66}$$

$$\sum_{k\in K} x_{ij}^k \leqslant \sum_t C_t y_{ijt}, \quad \forall (i,j)\in A, \tag{67}$$

$$\sum_{t=1}^{T} y_{ijt} \leqslant 1, \quad \forall (i,j)\in A, \tag{68}$$

$$n_{kp} \geqslant 0, \quad \forall k \in K, \ \forall p \in \mathbb{P}, \tag{69}$$

$$x_{ij}^k \geqslant 0, \quad \forall (i,j)\in A, \ \forall k \in K, \tag{70}$$

$$y_{ijt} \in \{0,1\}, \quad \forall (i,j)\in A, \ t = 1,\dots,T. \tag{71}$$

Here, x_{ij}^k still represents the flow on each arc. Constraints (65) define the flow of a commodity on an arc as the sum of the flows of that commodity on all paths that use that arc. Equalities (66) are the demand constraints and inequalities (67) are the capacity constraints. Constraints (68) still state that only one variable y_{ijt} may be selected.

Arc-path formulations can be used to generate a natural reduced version of the network design problem by simply considering a limited number of paths for each origin–destination pair demand. We have exemplified for (F_6) but all the former formulations also have an equivalent arc-path model.

Application of Benders decomposition: Gabrel et al. [48] deal with this problem by using a set of metric inequalities to define the feasible region. The idea is to define a vector $\lambda = (\lambda_1, \dots, \lambda_a, \dots, \lambda_{|A|}) \in \mathbb{R}_+^{|A|}$. Then, $\Theta(\lambda)$ denotes the quantity $\sum_{k=1}^{K} d_k \times l_k(\lambda)$, where $l_k(\lambda)$ is the length of a shortest chain joining $O(k)$ and $D(k)$, when each arc in A is given the length $\lambda_a \geqslant 0$. Therefore, let $y = (y_1, \dots, y_a, \dots, y_{|A|})$ be the selected capacities on each link. This solution is feasible if and only if [50] for all $\lambda \in \mathbb{R}_+^{|A|}$, we have

$$\sum_{a=1}^{|A|} \lambda_a y_a \geqslant \Theta(\lambda). \tag{72}$$

The authors then use these inequalities within a constraint generation approach where the master problem consists of maximizing the objective function (58) while the subproblem generates the violated metric inequalities (72). They also extend the approach with the generation of multiple violated constraints by the use of bipartition inequalities (i.e., metric inequalities corresponding to bipartitions of the node set $|N|$).

Mahey et al. [49] have modified the problem of Gabrel et al. [48] to introduce the issue of quality of service. Their version of the problem neglects the variable costs and uses an average delay function that can be regarded as a measure of the level of service or of the reliability of the network. The authors introduce variables $n_{(i,j)c}$, the flow of commodities in a link of capacity C. Therefore, the average delay on arc (i,j) given by Kleinrock's law (see e.g., Bertsekas and Gallager, [51]) is proportional to the non-linear expression $(n_{(i,j)c})/(C - n_{(i,j)c})$ which is introduced in the objective function. The term is convex for a fixed capacity but not jointly convex in $(C, n_{(i,j)c})$. Therefore, the authors use a generalized Benders decomposition approach that obtains a solution for the variables y_{ijt} in the master problem. The capacities are fixed in the auxiliary problem that reduces to a convex cost multicommodity flow problem, which is solved via a proximal decomposition method that works as a separable augmented Lagrangean algorithm [45]. The auxiliary problem feeds the master problem with feasibility and optimality cuts and the algorithm iterates until no more cuts can be generated and the solution is proved to be optimal. The authors have also tested the effect of additional cuts (e.g., connexity cuts and spanning tree cuts) and concluded that they could strongly reduce the number of necessary iterations until convergence. The authors have tested instances derived from real private communications networks and containing up to 25 nodes, 62 arcs and 500 commodities.

An important class of problems that can be modelled as step-increasing cost network design problems is the planning of electric power transmission and distribution networks. In the power systems transmission (distribution) design problem, the goal is to establish a minimum cost set of transmission (distribution) lines in order to enable the flow of one commodity—electrical power—from the generation plants (substations) to the consumer centers.

Power systems transmission and distribution network design problems are generally non-convex. However, once the integer variables are fixed, the remaining model becomes a convex transportation problem. Generalized Benders decomposition proposed by Geoffrion [22] is therefore a natural solution approach, as argued in Section 2. Benchakroun et al. [52] use decomposition to solve the *multi-period power system distribution network design problem*. It is basically the *step-increasing cost network design problem* with the addition of several specific constraints related to the power flow characteristics, node capacities, radiality and voltage drop limits. A model with linear constraints is proposed and solved via a generalized Benders decomposition. Gascon et al. [10] have improved the method by developing a Lagrangean relaxation based heuristic procedure to solve the master problem. Tests were conducted on examples of up to three periods, three sources (substations), ten load locations and 31 admissible arcs which proved the quality of the method.

For the *transmission network design problem*, the application of the method goes back to Pereira et al. [53]. The authors propose two models for the problem (one simplified transportation model and a more sophisticated one including power flow equations) and solve both of them via Benders decomposition. When solving the second model the cuts generated for the simplified one are kept. The idea of keeping the cuts obtained for a simplified model in the resolution of a more complex formulation was extended by Romero and Monticelli [8,54] who proposed a three-phase hierarchical heuristic approach that gradually increases the complexity of the model. For each level of complexity, a solution is obtained via Benders decomposition and the cuts are propagated in the next phase. Validations tests were conducted on a real-life network of 45 nodes. A similar algorithm with only two phases is proposed by Oliveira et al. [55] and used to solve a real network of 79 bus (nodes) and 155 circuits (arcs).

Tsamasphyrou et al. [56] introduced stochasticity in this problem by considering several availability (on the lines and generation units) and demand scenarios. The authors carried tests on an IEEE network of 24 nodes and 40 lines, considering cases of up to 2000 scenarios. Benders decomposition was considered for cases ranging from the situation where a single cut was generated at each iteration to the situation where, at each iteration, a cut was generated *for each* scenario. The authors concluded that the best approach lied somewhere between these two extremes, and divided the scenarios into a number (algorithm parameter) of sets, being one cut generated for each set of scenarios. This strategy divided the computation times by a factor ranging from 1.8 (1000 scenarios) to 3.8 (2000 scenarios).

Binato et al. [9] solve the *transmission network design* with additional constraints to deal with the power transmission stability problem. The authors make use of a disjunctive parameter in order to cope with the non-selected candidate arcs. Benders decomposition had been proposed before for this problem [57] but numerical problems caused by the large values of the disjunctive parameter had limited the validity of the results. Binato et al. [9] have proposed a new scheme where the disjunctive parameter is increased along the Benders iterations. Additional cuts are added at each iteration based on relaxations of the auxiliary problem and on Gomory's lifting procedure. A single real network of 46 nodes and 237 candidate arcs was considered and the authors showed that the proposed modifications helped reduce the computational time significantly.

4. Summary and conclusions

Numerous practical applications can be formulated as network design problems. In these problems, the idea is to obtain a least cost network in order to satisfy some flow constraints, commonly expressed in the form of origin–destination demands. We have presented a review on Benders decomposition methods applied to network design. Formulations for these problems usually contain one set of integer variables associated with the selection of the arcs in the network, and one set of continuous variables associated with commodity flows. This structure offers a natural framework for the decomposition approach which consists of isolating the integer variables in the master problem and the flow variables in the auxiliary subproblem. Moreover, the relative ease of solving the auxiliary subproblem in network design formulations make of Benders decomposition one of the most appropriate approaches. Indeed, in most of the surveyed articles (see Table 1 for a summary), validations tests have indicated that Benders decomposition is an efficient method for solving network design problems, and may outperform traditional techniques such as Branch-and-Bound or Lagrangian relaxation. Efficient solution methodologies have also been obtained by combining Benders decomposition

Table 1
Summary of the main Benders decomposition applications

Reference	Type of problem	Additional constraints	Features of the method	Size of the problems solved
Gavish [37]	Single origin UNDP	Constraints on the capacities of the edges connecting the origin node	Cuts are generated if the master solution is not connected or if the flows obtained by the auxiliary subproblem violate the capacities	$\|N\| \leqslant 12$
Hoang [33]	Origin–destination pairs UNDP	Budgetary constraints, non-linear costs, no explicit upper bounds on arc flow	Generalized Benders decomposition with master problem solved with a Lagrangian relaxation heuristic and subproblem via a convex cost multicommodity algorithm [35]	$\|N\| \leqslant 155$, $\|A\| \leqslant 376$
Magnanti et al. [24]	Origin–destination pairs UNDP	None	Preprocessing with dual-ascent procedure to eliminate variables. Use of Pareto-optimal cuts	$\|N\| \leqslant 30$, $\|A\| \leqslant 130$
Pereira et al. [53]	Step-increasing cost CNDP	Power flow equations, decision also on the generation levels	Standard algorithm with cuts generated for a simplified model kept in the resolution of a more complex formulation	$\|N\| \leqslant 46$, $\|A\| \leqslant 73$
Benchakroun et al. [52], Gascon et al. [10]	Step-increasing cost CNDP	Multi-period situation. Node-capacity, power flow and voltage drop limits constraints	Generalized Benders decomposition with Lagrangean relaxation heuristic to solve the master	$\|N\| \leqslant 13$ (3 sources and 10 loads) $\|A\| \leqslant 31$
Romero and Monticelli [8,54], Oliveira et al. [55]	Step-increasing cost CNDP	Power flow equations, decision also on the generation levels	Hierarchical heuristic approach with Benders decomposition used to solve each phase of a multi-step algorithm	$\|N\| \leqslant 79$, $\|A\| \leqslant 155$
Gutierrez et al. [34]	Origin–destination pairs UNDP	None (Robust approach with multi-scenario objective function)	One master problem for each scenario. Cross generation of Benders cuts	$\|N\| \leqslant 33$, $\|A\| \leqslant 100$ (11 scenarios)
Gabrel et al. [48]	Step-increasing cost CNDP	None	Generation of multiple violated constraints with bipartition inequalities	$\|N\| \leqslant 20$, $\|A\| \leqslant 36$

Reference	Type of problem	Additional constraints	Features of the method	Size of the problems solved
Tsamasphyrou et al. [56]	Step-increasing cost CNDP	Power flow equations, stochasticity in the demands and availability of lines and generation unities	Scenarios divided into a number (algorithm parameter) of sets and one cut generated for each set (instead of the extreme situations of one single cut for all scenarios or one cut for each scenario)	$\|N\| = 24$, $\|A\| = 40$ (up to 2000 scenarios)
Sridhar and Park [32]	Single facility CNDP	Node-capacity constraints	Embedded Benders decomposition initiated with a lower bound given by a cutting plane algorithm and a feasible solution given by a heuristic method	$\|N\| \leqslant 20$ (for various demand levels)
Randazzo and Luna [3]	Single origin UNDP	Redundant strengthening constraints to promote feasible solution in the Benders iterations	Dual variables in Benders decomposition initialized with feasible solution obtained with a linear relaxation and shortest-path algorithm. Auxiliary subproblem decomposed by commodity	$\|N\| \leqslant 100$, $\|A\| \leqslant 200$
Randazzo et al. [4]	Two-technologies single origin UNDP	Redundant strengthening constraints to promote feasible solution in the Benders iterations	Dual variables in Benders decomposition initialized with feasible solution obtained with a shortest-path algorithm. Auxiliary subproblem decomposed by commodity	$\|N\| \leqslant 41$, $\|A\| \leqslant 417$
Binato et al. [9]	Step-increasing cost CNDP	Constraints to ensure the power system stability based on a disjunctive parameter	Iterative increasing of disjunctive parameter solves earlier Benders numerical problem. In each iteration, several cuts are generated based on relaxed versions of the subproblem and on Gomory's lifting procedure	$\|N\| = 66$, $\|A\| = 237$
Mahey et al. [49]	Step-increasing cost CNDP	None (Objective function with non-linear term representing the quality of service.)	Generalized Benders decomposition with additional cuts. Convex optimization subproblem solved with the aid of a proximal decomposition method	$\|N\| \leqslant 25$, $\|A\| \leqslant 62$
Ouorou et al. [44]	Single facility CNDP	Elastic demands; fixed-open links (expansion problem)	Generalized Benders decomposition. Convex optimization subproblem solved with the aid of a proximal decomposition method	$\|N\| \leqslant 19$, $\|A\| \leqslant 68$

with other techniques, as proposed by Magnanti et al. [24]. A rich variety of these successful hybrid approaches is available in the literature.

In spite of this success, Benders decomposition has been mostly ignored for many years, not only for network design problems but for some of the other applications mentioned in Section 2. We believe that this tendency is slowly changing, given the increasing number of researchers using this technique, as shown in this article.

Acknowledgements

The author is grateful to Jean-François Cordeau and Bernard Gendron for their valuable comments which improved both the content as well as the presentation of the paper. The author also acknowledges CAPES-Brazil for the financial support (grant BEX 1097/01-6).

References

[1] Lederer PJ, Nambimadom RS. Airline network design. Operations Research 1998;46:785–804.
[2] Armacost AP, Barnhart C, Ware KA. Composite variable formulations for express shipment service network design. Transportation Science 2002;36:1–20.
[3] Randazzo CD, Luna HPL. A comparison of optimal methods for local access uncapacitated network design. Annals of Operations Research 2001;106:263–86.
[4] Randazzo CD, Luna HPL, Mahey P. Benders decomposition for local access network design with two technologies. Discrete Mathematics and Theoretical Computer Science 2001;4:235–46.
[5] Gavish B. Formulations and algorithms for the capacitated minimal directed tree problem. Journal of the Association for Computing Machinery 1983;30:118–32.
[6] Gouveia L. A 2n constraint formulation for the capacitated minimal spanning tree problem. Operations Research 1995;43:130–41.
[7] Myung Y, Lee C, Tcha D. On the generalized minimum spanning tree problem. Networks 1995;26:231–41.
[8] Romero R, Monticelli A. A hierarchical decomposition approach for transmission network expansion planning. IEEE Transactions on Power Systems 1994;9:373–80.
[9] Binato S, Pereira MVF, Granville S. A new Benders Decomposition approach to solve power transmission network design problems. IEEE Transactions on Power Systems 2001;16:235–40.
[10] Gascon V, Benchakroun A, Ferland JA. Electricity distribution planning model: a network design approach for solving the master problem of the Benders decomposition method. INFOR 1993;31:205–20.
[11] Carneiro MS, França P, Silveira PD. Long-range planning of power distribution systems: secondary networks. Computers and Electrical Engineering 1996;22:179–91.
[12] Cavelluci C, Lyra Filho C. Minimization of energy losses in electric power distribution systems by intelligent search strategies. International Transactions in Operational Research 1997;4:23–33.
[13] Standard and Poor's. Commercial transportation industry survey, 2000.
[14] Benders JF. Partitioning procedures for solving mixed variables programming problems. Numerische Mathematik 1962;4:238–52.
[15] Geoffrion AM, Graves GW. Multicommodity distribution system design by Benders decomposition. Management Science 1974;20:822–44.
[16] Cordeau J-F, Pasin F, Solomon MM. An integrated model for logistics network design. Annals of Operations Research, to appear.
[17] Florian M, Guérin G, Bushel G. The engine scheduling problem on a railway network. INFOR 1976;14:121–8.
[18] Cordeau J-F, Soumis F, Desrosiers J. A Benders decomposition approach for the locomotive and car assignment problem. Transportation Science 2000;34:133–49.

[19] Cordeau J-F, Soumis F, Desrosiers J. Simultaneous assignment of locomotives and cars to passenger trains. Operations Research 2001;49:531–48.
[20] Cai X, McKinney DC, Lasdon LS, Watkins Jr. DW. Solving large nonconvex water resources management models using generalized Benders decomposition. Operations Research 2001;49:235–45.
[21] Zhao G. A log-barrier method with Benders decomposition for solving two-stage stochastic linear programs. Mathematical Programming 2001;90:507–36.
[22] Geoffrion AM. Generalized Benders decomposition. Journal of Optimization Theory and Applications 1972;10: 237–60.
[23] Magnanti T, Wong RT. Accelerating Benders decomposition: algorithmic enhancement and model selection criteria. Operations Research 1981;23:464–84.
[24] Magnanti T, Mireault P, Wong RT. Tailoring Benders decomposition for uncapacitated network design. Mathematical Programming Study 1986;26:112–54.
[25] Boyce DE, Farhi A, Weischedel R. Optimal network problem: a branch-and-bound algorithm. Environment and Planning 1973;5:519–33.
[26] Hoang HH. A computational approach to the selection of an optimal network. Management Science 1973;19: 488–98.
[27] Gallo G. Lower planes for the network design problem. Networks 1983;13:411–26.
[28] Magnanti T, Wong RT. Network design and transportation planning: models and algorithms. Transportation Science 1984;18:1–56.
[29] Minoux M. Network synthesis and optimum network design problems: models, solution methods and applications. Networks 1989;19:313–60.
[30] Balakrishnan A, Magnanti TL, Mirchandani P. Network design. In: Dell'Amico M, Maffioli F, Martello S, editors. Annotated bibliographies in combinatorial optimization. New York: Wiley; 1997. p. 311–34.
[31] Balakrishnan A, Magnanti TL, Mirchandani P. Modeling and heuristic worst-case performance analysis of the two-level network design problem. Management Science 1994;40:846–67.
[32] Sridhar V, Park JS. Benders-and-cut algorithm for fixed-charge capacitated network design problem. European Journal of Operations Research 2000;125:622–32.
[33] Hoang HH. Topological optimization of networks: a nonlinear mixed integer model employing generalized Benders decomposition. IEEE Transactions on Automatic Control 1982;27:164–9.
[34] Gutierrez GJ, Kouvelis P, Kurawarwala AA. A robustness approach to uncapacitated network design problems. European Journal of Operations Research 1996;94:362–76.
[35] Nguyen S. An algorithm for the traffic assignment problem. Transportation Science 1974;8:203–16.
[36] Gouveia L. Using generalized capacitated trees for designing the topology of local access networks. Telecommunications Systems 1997;7:315–37.
[37] Gavish B. Topological design of centralized computer networks—formulations and algorithms. Networks 1982;12: 355–77.
[38] Gouveia L, Telhada J. An augmented arborescence formulation for the two-level network design problem. Annals of Operations Research 2001;106:47–61.
[39] Gendron B, Crainic T, Frangioni A. Multicommodity capacitated network design. In: Sanso B, Soriano P, editors. Telecommunications network planning. Dordrecht: Kluwer; 1999. p. 1–19.
[40] Riis M, Andersen KA. Capacitated network design with uncertain demand. INFORMS Journal on Computing 2002;14:247–60.
[41] Magnanti T, Mirchandani P. Shortest-paths, single origin–destination network design and associated polyhedra. Networks 1993;23:103–21.
[42] Barahona F. Network design using cut inequalities. SIAM Journal of Optimization 1996;6:823–37.
[43] Sridhar V, Park JS, Gavish B. LP-based heuristic algorithms for interconnecting token rings via source routing bridges. Journal of Heuristics 2000;6:149–66.
[44] Ouorou A, Luna HPL, Mahey P. Multicommodity network expansion under elastic demands. Optimization and Engineering 2001;2:277–92.
[45] Mahey P, Ourou A, LeBlanc L, Chifflet J. A new proximal decomposition algorithm for routing in telecommunications networks. Networks 1998;31:227–38.

[46] Magnanti T, Mirchandani P, Vachani R. Modeling and solving the two-facility capacitated network loading problem. Operations Research 1995;43:142–57.
[47] Stoer M, Dahl G. A polyhedral approach to multicommodity survivable network design. Numerische Mathematik 1994;68:149–67.
[48] Gabrel V, Knippel A, Minoux M. Exact solution of multicommodity network optimization problems with general step cost functions. Operations Research Letters 1999;25:15–23.
[49] Mahey P, Benchakroun A, Boyer F. Capacity and flow assignment of data networks by generalized Benders decomposition. Journal of Global Optimization 2001;20:173–93.
[50] Onaga K, Kakusho O. On feasibility conditions of multicommodity flows in networks. IEEE Transactions on Circuit Theory 1971;18:425–9.
[51] Bertsekas DP, Gallager RG. Data networks. Englewood Cliffs, NJ: Prentice-Hall; 1987.
[52] Benchakroun A, Ferland JA, Cléroux R. Distribution system planning through a generalized Benders decomposition approach. European Journal of Operations Research 1992;62:149–62.
[53] Pereira MVF, Pinto LMVG, Cunha SHF, Oliveira GC. A decomposition approach to automated generation-transmission expansion planning. IEEE Transactions on Power Apparatus and Systems 1985;104: 3074–81.
[54] Romero R, Monticelli A. A zero-one implicit enumeration method for optimizing investments in transmissions expansion plannings. IEEE Transactions on Power Systems 1994;9:1385–91.
[55] Oliveira GC, Costa APC, Binato S. Large scale transmission network planning using optimization and heuristics techniques. IEEE Transactions on Power Systems 1995;10:1828–34.
[56] Tsamasphyrou P, Renaud A, Carpentier P. Transmission network planning under uncertainty with Benders decomposition. Lecture Notes in Economics and Mathematical Systems 2000;481:457–72.
[57] Granville S, Pereira MVF. Analysis of the linearized power flow model in Benders decomposition. Technical Report SOL 85-04, Stanford University; 1985.