European Journal of Operational Research 226 (2013) 185–202

Contents lists available at SciVerse ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Discrete Optimization

An improved Benders decomposition algorithm for the tree of hubs location problem

Elisangela Martins de Sá, Ricardo Saraiva de Camargo *, Gilberto de Miranda

Department of Industrial Engineering, Federal University of Minas Gerais, Brazil

ARTICLE INFO

Article history:
Received 22 September 2011
Accepted 26 October 2012
Available online 17 November 2012

Keywords:
Tree of hubs location problem
Hub-and-spoke networks
Benders decomposition method
Benders cuts selection scheme

ABSTRACT

The tree of hubs location problem is a particularly hard variant of the so called hub location problems. When solving this problem by a Benders decomposition approach, it is necessary to deal with both optimality and feasibility cuts. While modern implementations of the Benders decomposition method rely on Pareto-optimal optimality cuts or on rendering feasibility cuts based on minimal infeasible subsystems, a new cut selection scheme is devised here under the guiding principle of extracting useful information even when facing infeasible subproblems. The proposed algorithm outperforms two other modern variants of the method and it is capable of optimally solving instances five times larger than the ones previously reported on the literature.

1. Introduction

In many-to-many distribution systems, in which several origin–destination pairs of nodes exchange flows, hub-and-spoke networks have a great application appeal. In such networks, rather than directly connect each pair, hub facilities are used to consolidate, route and distribute the traffic in order to take advantage of economies of scale on inter-hub connections. Flows from the same origin but addressed to different destinations are bundled at the hubs to other traffic that has different origins but the same destination. The consolidation of flows at the hubs allows the exploitation of scale economies due to the use of more efficient and higher volume carriers on inter-hub connections (O'Kelly, 1998).

Usually, in hub-and-spoke networks, it is assumed that there is a inter-hub connection between every hub pair; that no two non-hub nodes can be directly linked; that an origin–destination flow is routed through one or at most two hubs. Moreover, different assumptions may be contemplated including: Single (O'Kelly, 1987; Klincewicz, 1991; Skorin-Kapov et al., 1996; Aykin, 1995; Ernst and Krishnamoorthy, 1998b; Ebery, 2001) or multiple allocation (Campbell, 1994; Skorin-Kapov et al., 1996; Ernst and Krishnamoorthy, 1998a; Mayer and Wagner, 2002; Hamacher et al., 2004; Marín et al., 2006) of the non-hub nodes to the installed hubs, the number of hubs to be located may or may not be known beforehand, direct service between non-hub nodes may be enabled (Aykin, 1994, 1995), capacity constraints on the amount of traffic an installed hub can handle (Campbell, 1994; Aykin, 1994, 1995; Ernst and Krishnamoorthy, 1999; Ebery et al., 2000; Labbé et al., 2005; Costa et al., 2008; Contreras et al., 2009a), consideration of congestion effects at the installed hubs (Elhedhli and Hu, 2005; Camargo et al., 2009a; Elhedhli and Wu, 2010) and flow dependent economies of scale on inter-hub connections (O'Kelly and Bryan, 1998; O'Kelly, 1998; Horner and OKelly, 2001; Klincewicz, 2002; Racunica and Wynter, 2005; Kimms, 2006; Camargo et al., 2009b) among other variants. A general review of different problems is presented on the exhaustive surveys of Campbell et al. (2002) and of Alumur and Kara (2008b).

Recently, more flexible network policies of hub interconnections have been proposed (Nickel et al., 2000; Labbé et al., 2004; Campbell et al., 2005a,b; Contreras et al., 2009b, 2010; Alumur et al., 2009; Calik et al., 2009) in order to broaden the applicability of hub-and-spoke systems to other areas. The idea is to disregard the assumption that every pair of hub has to be directly linked and to adapt the design of the network to the characteristics of the application being addressed.

For instance, depending on the criterion used for establishing the inter-hub connections, many related applications can then be seen as special cases of hub-and-spoke systems: (i) Multi-modal urban transportation (Bruno et al., 1998; Nickel et al., 2000; cGelareh, 2008; Chen et al., 2008; Marín and Jaramillo, 2009), (ii) telecommunication networks (Hu, 1974; Kim and Tcha, 1992; Lee et al., 1994, 1996; Klincewicz, 1998), (iii) tree-shaped facilities location problems (Kim et al., 1996; Puerto and Tamir, 2005), (iv) ring-star network designs (Labbé et al., 2004; Laporte and Marín, 2007), and (v) incomplete hub networks (Campbell et al., 2005a,b;

* Corresponding author. Address: Department of Industrial Engineering, Federal University of Minas Gerais, Av. Antonio Carlos, 6627, 31270-901 Belo Horizonte, MG, Brazil. Tel.: +55 31 3409 4881; fax: +55 31 3409 4883.

E-mail addresses: elisangeladep@ufmg.br (E.M. de Sá), rcamargo@dep.ufmg.br (R.S. de Camargo), miranda@dep.ufmg.br (G. de Miranda).

0377-2217/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ejor.2012.10.051

Yoon and Current, 2008; Alumur and Kara, 2008a; Calik et al., 2009; Alumur et al., 2009).

In this work, the Tree of Hubs Location Problem (THLP), introduced by Contreras et al. (2009b, 2010), is addressed. The THLP is a single allocation hub location problem variant where hubs are located on a network and connected by means of a non-directed tree. Each non-hub node is single assigned to an installed hub. There are a per unit transportation cost associated with each link of the network and a fixed charge to install a hub in a node. The objective is to minimize the total cost that includes the transportation cost and can account or not for the installation costs of the hubs.

As remarked by Contreras et al. (2010), the design of the high-speed train network in Spain, which will place a train station within 50 kilometer of every Spanish city with more the 10,000 inhabitants and which it is expected to be finished by 2020, is a concrete example of an application of the THLP. The design of rapid transit systems for urban areas, in which citizens travel between origin–destination pairs of stations, can also be seen as another case of the THLP.

The design of such networks has to be carefully done, since it involves great amounts of resources and has a major impact in the operational costs and in the overall efficiency of the service afterwards. Hence efficient exact methods capable of solving more realistic, large-scale instances in reasonable time are of the utmost importance, despite posing as a difficult challenge.

Benders decomposition method (Benders, 1962) is one of the classical exact methods suitable to address the THLP. Its prior successfulness in solving hub-and-spoke network problems (Camargo et al., 2008; Gelareh and Nickel, 2008; Camargo et al., 2009b; Camargo et al., 2009a; Contreras et al., 2011; Contreras et al., 2011b; Contreras et al., 2011a), as well as other problems (Costa, 2005; Cordeau et al., 2006; Papadakos, 2009; Fortz and Poss, 2009), has rekindled the interest of the research community for it and thus motivated the current work.

Benders decomposition is a method for solving mixed integer programming (MIP) problems that have special structure in the constraint set, i.e. when fixing the complication variables (integer variables), the mathematical program reduces to an ordinary, easy to solve linear problem. The technique relies on projection and problem manipulation, followed by solution strategies of dualization, outer linearization and relaxation (Lasdon, 1972; Minoux and Vajda, 1986).

In general terms, the complicating variables of the original problem are projected out, resulting into an equivalent model with fewer variables, but many more constraints. When attaining optimality, a large number of these constraints will not be binding, suggesting then a strategy of relaxation that ignores all but a few of these constraints.

These constraints are then added on demand by observing two levels of coordination. At a higher level, known as master problem (MP), a relaxed version of the original problem having the set of the integer variables and its associated constraints is responsible for fixing the values of these integer variables and for providing a lower bound (LB) for the problem. At a lower level, known as subproblem (SP), the dual of the original problem with the values of the integer variables temporarily fixed by the MP is responsible for rendering a new cut or a Benders cut to be added to the MP and for generating a upper bound (UB) for the problem. The algorithm iterates, solving the MP and the SP one at a time, until the UB and the LB converge towards an optimal solution, if one exists.

Over the years, since its introduction by Benders (1962), different improvements have been proposed in order to speedup the performance of the Benders decomposition algorithm. Basically, these enhancements can be classified into three categories: (i) Reformulations for tightening the linear programming (LP) relaxation; (ii) algorithmic add-ons; and (iii) Benders cut selection schemes.

Geoffrion and Graves (1974) are the first authors to discuss how a tighter LP relaxation improves the overall efficiency of the algorithm, while McDaniel et al. (1977) show how the use of an initial set of Benders cuts, generated from the LP relaxation of the MP prior to starting the main iterations, also tightens the LP relaxation and betters the performance of the method.

Aiming at MIPs which depend on big-M coefficients to model logical implications, Codato and Fischetti (2006) propose a reformulation method for removing this dependency and, therefore, for improving the LP relaxation of the formulation based on the addition of combinatorial inequalities to the MP instead of the traditional Benders cuts. Their method appears to be more suitable for MIPs whose objective function depends only on the integer variables.

By applying local branching throughout the solution process, Rei et al. (2009) describe a Benders decomposition algorithm that improves both the LB and UB at each iteration.

Algorithmic add-ons are commonly used to ease the computational effort of the technique. Cote and Laughton (1984), and Poojari and Beasley (2009) use heuristics for solving the MP and for generating Benders cuts from the found integer feasible solutions, alleviating then the resolution of MP to optimality.

Working with two-stage stochastic problems, Birge and Louveaux (1988) show how to enrich the MP by disaggregating the Benders cuts into multiple cuts when the dual SPs are decomposable. Fortz and Poss (2009) embed the Benders algorithm into a branch-and-cut framework, generating Benders cuts in the nodes of branch-and-bound tree of the MP. They obtain good results on a multi-layer network design problem.

Deploying a great variety of enhancements, Contreras et al. (2011) include in their algorithm: Generation of multiple Benders cuts based on the work of Birge and Louveaux (1988), reduction tests during the main iterations relying on the ideas of Mitchell (1997), and the use of a heuristic for generating optimality cuts prior to the start of the main iterations of the algorithm and for obtaining a good initial UB. The applied reduction tests are able to eliminate a large number of the integer variables for the uncapacitated multiple allocation hub location problem.

Cut selection schemes have attracted the interest of many researchers since the work of Magnanti and Wong (1981). Magnanti andWong introduce the concepts of Pareto-optimal cuts, after noticing that cuts of different strength can be assembled, when the Benders dual SPs are degenerate. They demonstrate how the strongest possible one or the Pareto-optimal cut can be obtained by solving an additional SP, given that a fixed core point or a point belonging to the relative interior of the MP convex hull is known. Although it requires the resolution of one more SP per Benders cycle, their scheme proves to be very efficient for solving network problems.

The scheme of Magnanti andWong for generating Pareto-optimal cuts has one major drawback: Numerical instability due to a normalization constraint present in the additional SP. Addressing this disadvantage, Papadakos (2008) shows how to disregard this constraint when a new core-point is utilized at each Benders iteration. Papadakos (2008) demonstrates that the convex combination of the current MP solution for the integer variables and the previous used core-point suffices for obtaining a new core-point at each cycle. Nevertheless, finding a starting core-point still remains a challenge for some problems.

Working in the context of multi-commodity capacitated network design problems, Costa et al. (2009) establish the relationships between three classes of inequalities: Benders, metric and cutset. They show that feasibility Benders cuts are always metric, and how to strengthen the Benders cuts in order to attain metric inequalities.