Formulations for the Weight-Constrained Minimum Spanning Tree Problem

Cristina Requejo*, Agostinho Agra†, Adelaide Cerveira** and Eulália Santos‡

*Department of Mathematics and CIDMA, University of Aveiro, Portugal, crequejo@ua.pt
†Department of Mathematics and CIDMA, University of Aveiro, Portugal, aagra@ua.pt
**Department of Mathematics and CIO, University of Trás-os-Montes and Alto Douro, Vila Real, Portugal, adelaide.cerveira@sapo.pt
‡ESTG, Polytechnic Institute of Leiria and CIDMA, University of Aveiro, Portugal, eulalia.santos@sapo.pt

Abstract. We consider the Weight-constrained Minimum Spanning Tree problem (WMST). The WMST aims at finding a minimum spanning tree such that the overall tree weight does not exceed a specified limit on a graph with costs and weights associated with each edge. We present and compare, from the computational point of view, several formulations for the WMST. From preliminary computational results we propose a model that combines a formulation similar to the well known Miller-Tucker-Zemlin formulation with the cut-set inequalities.

Keywords: minimum spanning tree, weight-constraint, extended formulation, knapsack constraint
PACS: 02.10.Ox

INTRODUCTION

We consider an undirected complete graph $G = (V,E)$, with node set $V = \{0,1,\ldots,n\}$ and edge set $E = \{\{i,j\},\ i,j \in V, i \neq j\}$. Associated with each edge $e = \{i,j\} \in E$ we consider nonnegative integer costs c_e and nonnegative integer weights w_e. The Weight Minimum Spanning Tree problem (WMST) is to find a spanning tree $T = (V_T,E_T)$ in G ($V_T \subseteq V$ and $E_T \subseteq E$) of minimum cost $C(T) = \sum_{e\in E_T} c_e$ and with total weight $W(T) = \sum_{e\in E_T} c_e$ not exceeding a given limit H. This is a NP-hard problem [1].

The WMST appears in several real applications. In Henn [3] are mentioned two examples: the design of physical systems subjected to limited budgets and the minimum cost reliability constrained spanning tree with applications to communication network problems.

The WMST is known under several different names. It was first mentioned, under another name, in Aggarwal, Aneja and Nair [1]. In this paper the authors propose an exact algorithm to solve the problem that uses a Lagrangean relaxation to approximate a solution combined with a branch and bound strategy. This solution approach can also be found in the work of Shogan [10]. Another exact approach to solve the problem is given in Hong, Chung and Park [4] where the authors use the matrix-tree theorem. The paper of Ravi and Goemans [9] describes an approximation scheme. Hassin and Levin [6] improve the results in [9] and use an algorithm based on a two-variable extension of the matrix-tree theorem. Henn [3] presents theoretical properties of the convex hull of the set of feasible solutions of the problem and theoretical properties of a Lagrangean relaxation. The author proposes an approximation method that uses a decomposition of the problem. Furthermore, a branch and bound scheme is also presented.

Our purpose is to compare, from the computational point of view, different formulations for the WMST. We introduce three extended formulations that are compared with the classical cut-set formulation. Preliminary results show that very interesting results can be obtained when an extended formulation, similar to the well-known Miller-Tucker-Zemlin (MTZ) reformulation [2], is combined with separation over the cut-set inequalities.

First we review the two classical formulations on the space of the original variables for the WMST. Secondly we present three extended formulations, that is, formulations that use additional variables. Two of them are adapted from formulations for the Minimum Spanning Tree problem: the multicommodity flow formulation and a formulation based on the MTZ inequalities. We also introduce a new formulation based on an extended formulation for the knapsack constraint (the constraint on the total edge weight). Finally we summarize the results and present the main conclusion.

CP1281, *ICNAAM, Numerical Analysis and Applied Mathematics, International Conference 2010, Vol. III*
edited by T. E. Simos, G. Psihoyios, and Ch. Tsitouras

NATURAL FORMULATIONS

In this section we briefly review the two classical formulations on the space of the original variables for the WMST, one based on a subset of constraints preventing circuits (circuit elimination constraints) and the other based on a subset of constraints (cut-set inequalities) ensuring connectivity. It is well known (see Magnanti and Wolsey [5]) that oriented formulations (based on the underlying directed graph) leads, in general, to tighter formulations (formulations whose lower bounds provided by the linear relaxations are closer to the optimum values). Thus, henceforward we consider the corresponding directed graph, with root node 0, where each edge $e = \{0, j\} \in E$ is replaced with arc $(0, j)$ and each edge $e = \{i, j\} \in E, i \neq 0$, is replaced with two arcs, arc (i, j) and arc (j,i), yielding arc set $A = \{(i, j),\ i \in V \setminus \{0\}, j \in V, i \neq j\}$. These arcs inherit the cost and weight of the ancestor edge.

Considering the binary variables x_{ij} (for all $(i, j) \in A$) indicating whether arc (i, j) is in the MST solution we have the following formulation [5]:

$$
\begin{aligned}
min \quad & \sum_{(i,j)\in A} c_{ij}x_{ij} \\
s.t. \quad & \sum_{i\in V} x_{ij} = 1 \qquad\qquad j \in V \setminus \{0\} & (1)\\
& \text{Connectivity Constraints} & (2)\\
& \sum_{(i,j)\in A} w_{ij}x_{ij} \leq H & (3)\\
& x_{ij} \in \{0,1\} \qquad\qquad (i, j) \in A. & (4)
\end{aligned}
$$

The cardinality constraints (1) ensure there is an arc incident at each node in $V \setminus \{0\}$. The knapsack constraint (3) ensures that the total tree edges weight does not exceed H. For subsets $S,T \subseteq V$, we define $A(S,T) = \{(i, j) \in A\ :\ i \in S, j \in T\}$, when $S = T$ let $A(S) := A(S,S)$, and $S^c = V \setminus S$. In order to ensure connectivity (constraints (2)) the usual approaches consists either in the inclusion of the circuit elimination inequalities

$$\sum_{(i,j)\in A(S)} x_{ij} \leq |S| - 1, S \subset V, S \neq \emptyset \tag{5}$$

or in the inclusion of the cut-set inequalities

$$\sum_{(i,j)\in A(S,S^c)} x_{ij} \geq 1, S \subset V, S \neq \emptyset, 0 \in S. \tag{6}$$

Since the linear relaxation of both models provide the same bound we use the formulation with the cut-set inequalities, denoted by Cut-Set (CS) formulation. As the number of cut-set inequalities increases exponentially with the size of the model these inequalities are introduced in the model as cuts using separation.

EXTENDED FORMULATIONS

It is well known that in order to ensure connectivity/prevent circuits, instead of using one of the families of inequalities (5) and (6) with an exponential number of inequalities, one can use compact extended formulations. In this section we propose three extended formulations. The well-known Multicommodity flow formulation, a formulation based on the Miller-Tucker-Zemlin inequalities and a formulation based on an extended formulation for the knapsack constraint.

The multicommodity flow formulation

To obtain a first extended formulation, in addition to the natural binary variables x_{ij}, we use oriented flow variables f^k_{ij} (for all $(i, j) \in A$ and $k \in V \setminus \{0,i\}$) which specify whether arc (i, j) is used in the path from the root to node k. The

Multicommodity Flow (MF) formulation is as follows:

$$\begin{aligned}
min \quad & \sum_{(i,j)\in A} c_{ij}x_{ij} \\
s.t. \quad & (1),(4) \\
& \sum_{i\in V\setminus\{k\}} f_{ij}^k - \sum_{i\in V\setminus\{0\}} f_{ji}^k = \begin{cases} -1 & j=0 \\ 0 & j\neq 0,k \\ 1 & j=k \end{cases} \qquad j\in V,\ k\in V\setminus\{0\} && (7) \\
& f_{ij}^k \leq x_{ij} \qquad (i,j)\in A, k\in V\setminus\{0,i\} && (8) \\
& \sum_{(i,j)\in A} w_{ij}x_{ij} \leq H && (9) \\
& f_{ij}^k \in \{0,1\} \qquad (i,j)\in A, k\in V\setminus\{0,i\} && (10)
\end{aligned}$$

The flow conservation constraints (7) establishes that the solution must have a path between node 0 and node k (for all $k\in V\setminus\{0\}$). The connecting constraints (8) ensure it is possible to send flow for each node k throughout the arc (i,j) only if the arc is in the solution. Together with the flow conservation constraints they ensure the connectivity of the solution.

The weighted Miller-Tucker-Zemlin formulation

Here, besides the binary variables x_{ij} to define the topology of the solution we also consider variables p_i $(i=0,\ldots,n)$ which specify the weighted-position of node i in the tree, i.e. the sum of the weights of the arcs in the path between the root and node i. The Weighted Miller-Tucker-Zemlin (WMTZ) formulation is as follows:

$$\begin{aligned}
min \quad & \sum_{(i,j)\in A} c_{ij}x_{ij} \\
s.t. \quad & (1),(4) \\
& w_{ij}x_{ij} + p_i \leq p_j + H(1-x_{ij}) \qquad (i,j)\in A && (11) \\
& \sum_{(i,j)\in A} w_{ij}x_{ij} \leq H && (12) \\
& 0 \leq p_i \leq H \qquad i\in V && (13)
\end{aligned}$$

Constraints (11) are based on the well-known subtour elimination constraints given in Miller et al. [7] for the Traveling Salesman Problem. Following Gouveia [2], constraints (11) can be lifted into the following constraints.

$$w_{ji}x_{ji} + w_{ij}x_{ij} + p_i \leq p_j + H(1-x_{ij}) \qquad (i,j)\in A \qquad (14)$$

$$(H-w_{ji})x_{ji} + w_{ij}x_{ij} + p_i \leq p_j + H(1-x_{ij}) \qquad (i,j)\in A \qquad (15)$$

$$\sum_{k\in V\setminus\{i,j\}} w_{kj}x_{kj} + w_{ij}x_{ij} + p_i \leq p_j + H(1-x_{ij}) \qquad (i,j)\in A \qquad (16)$$

$$\sum_{k\in V\setminus\{i,j\}} (w_{kj}x_{kj} + w_{ik}x_{ik}) + w_{ji}x_{ji} + w_{ij}x_{ij} + p_i \leq p_j + H(1-x_{ij}) \qquad (i,j)\in A \qquad (17)$$

Preliminary computational results indicate that among all the lifted inequalities better computational results are obtained when inequalities (14) are incorporated in the formulation. Thus, henceforward we consider the WMTZ formulation with inequalities (11) replaced by (14).

A weight-extended formulation

We consider the natural binary variables x_{ij} and use oriented weight-flow variables $z_{ijk}^{h_1,h_2}$ (for all $(i,j)\in A$, $k\in V\setminus\{0,i\}$ and $h_1\leq h_2$, $h_1,h_2=0,\ldots,H$) which specify whether arc (i,j) is used in the path from the root to node k and the weight of the path from the root to node i is h_1 and the weight of the path to node j is h_2. Variables

$z_{ijk}^{h_1,h_2}$ for all $h_2 < h_1$ and $h_2 > H$ are set to zero. The Weight-Extended (WE) formulation is as follows:

$$\begin{aligned}
&\min \quad \sum_{(i,j)\in A} c_{ij}x_{ij} \\
&s.t. \quad (1),(4) \\
&\sum_{i\in V\setminus\{k\}} \sum_{h=0|h-w_{ij}\geq 0}^{\bar{h}} z_{ijk}^{h-w_{ij},h} - \sum_{i\in V\setminus\{0\}} \sum_{h=0|h+w_{ji}\leq H}^{\bar{h}} z_{jik}^{h,h+w_{ji}} = \begin{cases} -1 & j=0,\ \bar{h}=0 \\ 0 & j\neq 0,k,\ \bar{h}=H \\ 1 & j=k,\ \bar{h}=H \end{cases} \\
&\qquad\qquad j\in V,\ k\in V\setminus\{0\} \qquad (18) \\
&\sum_{h=0}^{H} z_{ijk}^{h,h+w_{ij}} \leq x_{ij} \qquad (i,j)\in A, k\in V\setminus\{0,i\} \qquad (19) \\
&z_{ijk}^{h_1,h_2} \in \{0,1\} \qquad (i,j)\in A, k\in V\setminus\{0,i\}, \qquad (20)
\end{aligned}$$

The flow conservation constraints (18) establishes that the solution must have a path between node 0 and node k (for all $k \in V\setminus\{0\}$). The connecting constraints (19) ensure it is possible to send flow for each node k throughout the arc (i,j) only if the arc is in the solution. Together with the flow conservation constraints they ensure the connectivity of the solution.

COMPUTATIONAL RESULTS AND CONCLUSIONS

In order to compare the proposed formulations and following Pisinger [8] we randomly generate a set of test instances corresponding to complete graphs. We perform all tests on an Intel(R) Core(TM)2 Duo CPU 2.00 GHz processor and 1.99Gb of RAM using the Xpress Release 2009 (Xpress-Optimizer 20.00.05 and Xpress-Mosel 3.0.0) [11]. As expected, given the size of the model, the WE formulation could only be used in instances up to ten nodes. Although the linear relaxations of the MF formulation and the CS formulation provide the same lower bounds, the MF formulation could only be used in solving instances up to 100 nodes. Hence we focus on the comparison of the CS formulation with the WMTZ formulation and with an hybrid procedure that results from the WMTZ formulation strengthened with the cut-set inequalities. Using these two models and the hybrid procedure we solved instances up to 250 nodes. From the preliminary computational results, the hybrid procedure (WMTZ+C), that is, the WMTZ formulation with the addition of cuts obtained from the separation of the cut-set inequalities at the root of the Branch and Bound, outperformed the two other formulations. The next table exemplifies these computational results for an instance with 100 nodes .

$\lvert V\rvert$	opt	time with CS	time with WMTZ	time with WMTZ+C
100	1189	>1000s	>1000s	9s

REFERENCES

1. V. Aggarwal, Y. P. Aneja, K. P. K. Nair, *Minimal spanning tree subject to a side constraint*, Computers and Operations Research **9**, 287–296 (1982).
2. L. Gouveia, *Using the Miller-Tucker-Zemlin constraints to formulate a minimal spanning tree problem with hop constraints*, Computers and Operations Research **22**, no. 9, 959–970 (1995).
3. S. T. Henn, *Weight-constrained minimum spanning tree problem*, Ph.D. thesis, Department of Mathematics, University of Kaiserslautern, Kaiserslautern, Germany (2007).
4. S-P. Hong, S-J. Chung, B. H. Park, *A fully polynomial bicriteria approximation scheme for the constrained spanning tree problem*, Operations Research Letters **32**, 233–239 (2004).
5. T. Magnanti, L. Wolsey, *Optimal trees*, Network Models, Handbooks in Operations Research and Management Science, Vol. 7 (M.O. Ball, T.L. Magnanti, C.L. Monma, and G.L. Nemhauser, eds.), Elsevier Science Publishers, North-Holland (1995).
6. R. Hassin, and A. Levin, *SIAM Journal on Computing* **33**, 261Ű268 (2004).
7. C. Miller, A. Tucker, R. Zemlin, *Integer programming formulations and travelling salesman problems*, Journal of the Association for Computing Machinery **7**, 326–329 (1960).
8. D. Pisinger, *Where are the hard knapsack problems?*, Tech. Rep. 2003/08, DIKU, University of Copenhagen, Denmark (2003).
9. R. Ravi, M. X. Goemans, *The constrained minimum spanning tree problem*, Proc. of the Scandinavian Workshop on Algorithmic Theory, LNCS **1097**, 66–75 (1996).
10. A. Shogan, *Constructing a minimal-cost spanning tree subject to resource constraints and flow requirements*, Networks **13**, 169–190 (1983).
11. FICO Xpress Optimization Suite (www.fico.com/xpress) (2009).