R
A
P
P
O
R
T

D
E

R
E
C
H
E
R
C
H
E

L R I

IS THE POLYTOPE ASSOCIATED WITH A TWO STAGE STOCHASTIC PROBLEM TDI ?

(Version révisée du rapport n° 1535/2010)

LETOURNEL M / LISSER A / SCHULZ R

Unité Mixte de Recherche 8623
CNRS-Université Paris Sud –LRI

11/2011

Rapport de Recherche N° 1546

CNRS – Université de Paris Sud
Centre d'Orsay
LABORATOIRE DE RECHERCHE EN INFORMATIQUE
Bâtiment 650
91405 ORSAY Cedex (France)

Is the polytope associated with a two stage stochastic problem TDI?

Marc Letournel Abdel Lisser Rüdiger Schulz

April 1, 2011

Abstract

The maximum weight forest problem (MWFP) in a graph is solved by the famous greedy algorithm due to Edmonds (1971) where every edge has a known weight. In particular, the system of constraints on the set of edges is TDI (totally dual integral), since the set of independent edges, i.e., of acyclic subsets of edges, is a matroïd. We extend this approach to the case of two-stage maximum weight forest problem. The set of edges is composed of first stage edges with known weights and second stage ones where the weights are known a priorily in terms of discrete random variables. As the probability distribution is discrete, we transform the stochastic problem into a deterministic equivalent problem. In this article, we prove TDIness for the two stage maximum weight forest problem in the case of only two scenarios. We provide a counter example to prove that the problem is not anymore TDI for more than two scenarios.

1 Introduction, notation and generalities

1.1 Introduction

We consider a graph $G = (V, E)$ where E is a set of edges of cardinality $|E|$, and a cost function c defined on edges. The edges are indexed by $i \in [1, |E|]$ and for any subset F of edges, we call $c(F)$ the sum $\sum_{j \in F} c_j x_j$ where $x_j = 1$ if $j \in F$ and $x_j = 0$ otherwise.
A subset F is said independant if there is no cycle in F. The maximization problem of $c(F)$ for F independent is well known to be connected to matroïds and is efficiently solved in the case of a fixed cost value for every edge (7). When every edge has a fixed value, we say that the problem is deterministic and since independent sets are a matroïd, the greedy algorithm is efficient to provide the maximization problem of $c(F)$ polynomialy. The matroïd structure of independent sets involves the fact that the rank function $r(F)$ (the maximum cardinality of any independent subset included in F) is submodular. In this case, the polytope associated to independence constraint is Totally Dual Integral. This is a very nice property which allows to apply algebraic methods to solve maximization. The problem that we introduce in this paper is to understand what happens when cost function is not deterministic but follows a discrete stochastic distribution π. In our problem, the edges are split into two subsets $E = X \cup Y$. In stochastic programming, the first subset X has a deterministic cost function, this set is called first stage, we set $card(X) = n$, whereas in second subset called second stage, $card(Y) = q$. We notice $N = n + q$ the cardinal of the whole set of edges in G.
Cost function depends on $K \geq 2$ scenarios and cost values are given by a probability distribution $\pi = (\pi_1, \ldots, \pi_K)$. The formulation of this problem is:

$$z_{IP} = \begin{cases} max \sum_{j \in X} c_j x_j + \sum_{k=1}^{k=K} \pi_k \sum_{j \in Y} c_j(k) z_j^k \\ \sum_{j \in S \cap X} x_j + \sum_{j \in S \cap Y} z_j^k \leq r(S), \quad k \in \{1, \ldots, K\}, \quad \forall S \subseteq E \\ (x, z^k) \in \{0,1\}^n \times \{0,1\}^q, k \in \{1, \ldots, K\} \end{cases} \quad (1)$$

We outline that any first stage edge is associated with a single binary variable x_j, while a second stage edge is associated with K binary variables z_j^k.
The most powerful case is when the deterministic problem is associated with a matrix constraint which is Totally Unimodular (TU). Introducing multi stages and dubbing sub matrices of second stage edges is studied by (6). In our case, initial properties are weaker and are formulated in terms of TDI system. Operations that preserve TDI properties are presented in (1) and we will outline connection with our formulation and matroïds intersection, the reader will refer to (4) for further details. Several works have investigated the question of approximation of two stage stochastic versions of classical combinatorial problems. The question of hardness and finding approximate algorithm is studied in (8) for general class of problems and in (5) a general method called boosted sampling turns an approximation algorithm for a deterministic problem into an approximation algorithm for the equivalent two stage stochastic version. Concerning more specificaly the spanning tree problem, (2) study the minimum stochastic spanning tree problem and give approximations algorithm and results of log n-hardness approximation under assumptions of bounded inflation. Furthermore, (Escoffier et al.) study more specificaly the two stage maximum weight spanning tree problem and prove the APX-completeness. They propose an approximation algorithm with performance guarantee or $\dfrac{K}{2K-1}$ where K is the number of scenarios at the second stage. In section 2, we reproduce the basic results of the determinist case. The reader will refer to (7). In the greedy algorithm, we outline the importance of the closure of a subset and the dynamical point of view of the building scheme of the dual solution. We analyze the solution of the greedy algorithm in terms of a continuous solution of the weights.In 3, we present first general results on TDIness for the two stage stochastic problem. In section 4, we deal with the case of two scenarios only, this case is particular since we prove that the system of constraints is TDI exactly as in the deterministic case. In section 5, we show by an example to the contrary that the TDI properties are not preserved in case of strictly more than two scenarios.

2 Deterministic case and threshold for the status of a given edge in a graph

Definitions for matroïds, independent sets and greedy algorithm are given in (7). We outline in this paper a dynamical point of view for the construction of the greedy solution by watching carefully how the closure of a subset of chosen edges is increasing.

2.1 Polytopes and problems associated

In the deterministic case, we have $n = N = |E|$. We introduce the notations:

$$\pi(r) = \{x \in \{0,1\}^n : \sum_{j \in S} x_j \leq r(S) \quad \forall S \subseteq E\} \tag{2}$$

$$z_{IP} = max\{c^t x : x \in \pi(r)\} \tag{3}$$

The polytope associated to the matroïd:

$$P(r) = \{x \in (\mathbb{R}^+)^n : \sum_{j \in S} x_j \leq r(S) \quad \forall S \subseteq E\} \tag{4}$$

and the linear program associated

$$z_{LP} = max\{c^t x : x \in P(r)\} \tag{5}$$

The dual problem associated with z_{IP} is :

$$z_{LD} = min \sum_{S \subseteq E} r(S) y_S \tag{6}$$

$$st \begin{cases} \sum_{S : j \in S} y_S \geq c_j \quad \forall j \in E \\ y_S \geq 0 \quad \forall S \subset E \end{cases} \tag{7}$$

The problem of finding a maximum-weight independant set in the deterministic case is formulated by (**??**);
It is possible to show that $z_{IP} = z_{LD}$ by exhibiting the optimal solution for z_{LD} which is indeed the greedy solution for z_{IP}. The conclusion is that $\pi(r)$ is TDI and the duality gap is equal to 0.

2.2 The greedy algorithm

This algorithm is due to Edmonds (Nemhauser and Wolsey (7)).
We call $\mathscr{F}$ the set of independant subsets in E. Rank the elements of E so that $c_1 \geq c_2 \geq \ldots \geq c_n$
Let begin by $J_0 = \emptyset$, $t = 1$.
Iteration t : If $c_t \leq 0$ then stop and $S^G = J_{t-1}$.
If $c_t > 0$ and $J_{t-1} \cup \{t\} \in \mathscr{F}$, then set $J_t = J_{t-1} \cup \{t\}$.
If $c_t > 0$ and $S_{t-1} \cup \{t\} \notin \mathscr{F}$ then set $J_t = J_{t-1}$.
If $t = N$ stop and $S^G = J_t$
Set t to $t + 1$
The greedy solution is $\{x_i / i \in S^G\}$.

We call p the number of chosen edges and $S^G = J_p$.

We now present the main result connecting the greedy solution to the dual problem, and we emphasize on the growing closure mechanism. We say that any edge chosen during the greedy algorithm is *greedy* while any non chosen edge is *covered* .

Definition 2.1. *The closure or span of a set A is $sp(A) = \{i \in E : r(A \cup \{i\}) = r(A)\}$.*

We briefly describe the greedy algorithm in terms of closure sets. We refer to an edge $x_i \in E$ either directly as x_i or only via its index i. We assume that indexation of edges is equal to their rank. For the edges of the greedy solution for G, let $\{i_1, \ldots, i_t\}$ be the set of edges indices in decreasing order of edge weigths. Denote $J_\tau = \{i_1, \ldots, i_\tau\}, 1 \leq \tau \leq t$ and $\kappa_\tau = sp(J_\tau)$. The first step of the greedy process is the choice of the heaviest weight and we set $J_1 = \{x_1\}$. We call $i_1 = 1$ the first chosen edge and we notice that $\kappa_1 = sp(J_1)$.

We now choose the second edge i_2 with the highest rank among all remaining edges such that $i_2 \notin \kappa_1$. Up to this point of construction, obviously $i_2 = 2$ since there is no cycle with only two edges. We build now $J_{\tau+1}$:

The iterative mechanism consist in the choice of $i_{\tau+1}$ with the highest rank among all remaining edges such as $i_{\tau+1} \notin \kappa_\tau$ and $c_{i_{\tau+1}} \geq 0$. It is important to notice that $sp(J_\tau) = sp(\kappa_\tau)$.

Definition 2.2. *For a given cost $c \in \mathbb{R}^N$ on G and $b \in \mathbb{R}$, we note $U(b) = \{x_i / c_i \geq b\}$*

Lemma 2.1. *For any graph G of cardinal N, consider specifically any edge x with cost $c \in \mathbb{R}$ considered as a variable, and assume that every other edges $x_1, \ldots, x_{N-1}$ have a given cost $(c_1, \ldots, c_{N-1})$, there exists a threshold $b \in \mathbb{R}^+$ such that if $c > b$ then x belongs to the greedy solution applied to G while x is covered if $c < b$. The threshold b is a function of $(c_1, \ldots, c_{N-1})$. When $c = b$, the status of x can be greedy or covered according to the choice of x or another edge of same weight during the greedy algorithm.*

Proof. We still assume that indexation of edges $x_1, \ldots, x_{N-1}$ is equal to their rank according to a decreasing weight. We dynamically construct greedy solution on $\tilde{G} = G \setminus \{x\}$ without taking in account the specific edge x. We get a finite sequence of sets $\tilde{J}_1, \ldots, \tilde{J}_p$ and of corresponding closures $\tilde{\kappa}_1, \ldots, \tilde{\kappa}_p$ of $\tilde{G}$.
Remark that $\tilde{\kappa}_1, \ldots, \tilde{\kappa}_p$ is a strictly growing sequence for inclusion of subsets such that $G \setminus \{x\} = \tilde{\kappa}_p$.
We reintroduce x as a new edge in the graph $\tilde{G}$. In the case where $x \notin \tilde{\kappa}_p$, that means that as soon as $c > 0$ then x would be chosen during the greedy algorithm applied properly on whole graph G, so that $b = 0$.
In the case where $x \in \tilde{\kappa}_p$ that means that there exists a particular step τ such that $x \in sp(\tilde{J}_{\tau+1})$ and $x \notin sp(\tilde{J}_\tau)$. The threshold b is equal to $c_{i_{\tau+1}}$. In the case where $c = c_{i_{\tau+1}}$ occurs, then x and $x_{i_{\tau+1}}$ have the same cost(and perhaps several other edges) and x or $x_{i_{\tau+1}}$ can be indifferently chosen during greedy algorithm but not both together. □

Theorem 2.3. $b = b(c_1, \ldots, c_{N-1})$ *is a continuous function of* $(c_1, \ldots, c_{N-1})$

Proof. We consider the same mechanism as in lemma 2.1 to get a first threshold $b = b(c_1, \ldots, c_{N-1})$ when removing x from G. We begin to notice that during the greedy algorithm, at every step, when choosing the τ^{th} edge of the greedy solution, $J_\tau \subseteq U(c_{i_\tau}) \subseteq \kappa_\tau$.
We fix $\epsilon > 0$ and for $c = (c_1, \ldots, c_{N-1}) \in \mathbb{R}^{N-1}$, we consider a small perturbation $c' = (c'_1, \ldots, c'_{N-1})$ such that $|c_i - c'_i| < \epsilon \quad \forall i \in [1, \ldots, N-1]$. We note $U'(b) = \{x_i / c'_i \geq b\}$, the set of edges whose slightly modified weights are greater than b.
Obviously, if x is not covered by any independent subset in $U'(b)$, that means that b will decrease. Conversely, if $U'(b)$ contains some new independent subsets that covers x, it will possibly enforce b to increase. we begin to explain how the variation of the threshold is lower bounded: $U(b)$ contains a subset J_τ such that $x \in sp(J_\tau)$ and $U(b) \subset U'(b - 2\epsilon)$. Then $U'(b - 2\epsilon)$ contains a subset $J'_{\tau'}$, step of the greedy algorithm applied to $\tilde{G}$ with c' cost and such that $x \in sp(J'_{\tau'})$. This proves that $b(c') \geq b - 2\epsilon$.
For the upper bound of the variation, we see that since $U(b + \epsilon)$ does not contain any independent set covering x with cost c, and $U'(b + 2\epsilon) \subseteq U(b + \epsilon)$ then $U'(b + 2\epsilon)$ does not contain any independent set covering x in G with c' cost. This proves that $b(c') \leq b(c) + 2\epsilon$.
We conclude that $|b(c') - b(c)| \leq 2\epsilon$, and b is a continuous function of c. □

We now present the connection between dual formulation and the greedy solution in the deterministic case and we will use the same approach in the stochastic case.

2.3 Dynamical construction of the dual solution

This section shortly presents results given in (7) in order to re use the principle of proof in the stochastic case.

Theorem 2.4. *The dual problem z_{LD} and the greedy solution of z_{IP} have the same objective value. $P(r)$ is a integral polytop, and the system defining $P(r)$ is TDI.*

Proof. We assume that indexation of edges is equal to their rank according to a decreasing weight. We dynamically construct both greedy solution and dual solution equals at every step.
According to notations used in 2.2, the optimal solution to z_{LD} is :

$y_{\kappa_t} = c_{j_t} - c_{j_{t+1}} \quad for\ t = 1, \ldots, p-1$
$y_{\kappa_p} = c_{j_p}$
$y_S = 0 \quad otherwise$

First, we need to check that for every $j \in N$: $\sum_{S/j \in S} y_S \geq c_j$.
it is obviously the case for every $j_i \in S^G$ since $j_i \in J_t$ for $i \leq t$.
For any j which has not been chosen during the greedy algorithm, j belongs to some $sp(J_t) \setminus sp(J_{t-1})$
So $\sum_{S/j \in S} y_S = c_{j_t} \geq c_j$.

Secondly, we need to compute the sum : $\sum_{S \subseteq E} r(S) y_S = \sum_{t=1}^{p-1} t(c_{j_t} - c_{j_{t+1}}) + p c_{j_p}$

this leads by splitting in two sums and re-indexing to $\sum_{S \subseteq E} r(S) y_S = \sum_{t=1}^{p} c_{j_t}$.
This shows that z_{IP} and z_{LD} have the same objective value. More precisely, the greedy solution to z_{IP} and z_{LD} have the same objective value. We conclude that the set of inequalities defining $P(r)$ is TDI, and $P(r)$ is an integral polytop. □

In the next section we introduce two stage stochastic problem, where the cost of several edges can have stochastic values. We consider the case of discrete probabilities, where costs belong to a finite set of values. The values of each edges are sampled simultaneously so that there exists a finite number of scenarios.

3 Two stage Stochastic Problem

In this section, we consider the two stage stochastic problem with $K \geq 2$ and we introduce relaxed formulation of (1) and dual formulation. We explain the mechanism of the split of cost for first stage variables and produce first results.

3.1 Linear relaxation and dual formulation associated with the two stage case

The linear program associated to (1) is:

$$z_{LP} = \begin{cases} max \sum_{j \in X} c_j x_j + \sum_{k=1}^{k=K} \pi_k \sum_{j \in Y} c_{jk} z_{jk} : \\ \sum_{j \in S \cap X} x_j + \sum_{j \in S \cap Y} z_{jk} \leq r(S), \quad k \in \{1, \ldots, K\}, \quad \forall S \subseteq E \\ (x, z_k) \in [0,1]^n \times [0,1]^q, k \in \{1, \ldots, K\} \end{cases} \tag{8}$$

and its LP-dual is:

$$z_{LD} = \begin{cases} min \sum_{k=1}^{K} \sum_{S \subseteq E} r(S) y_{S,k} : \\ \sum_{k=1}^{K} \sum_{S \subseteq E: i \in X \cap S} y_{S,k} \geq c_i, \quad i \in X \\ \sum_{S \subseteq E: j \in Y \cap S} y_{S,k} \geq \pi_k c_{jk}, \quad j \in Y, k \in \{1, \dots, K\} \\ y_{S,k} \geq 0, \quad k \in \{1, \dots, K\}, \quad S \subseteq E \end{cases}$$

(9)

3.2 Formal split of the cost of a first stage edge

Since the cost of second stage edges change with scenarios, while first stage edges costs remain the same, when one specific scenario occurs, we apply a greedy algorithm not with the whole cost of first stage edge, but only with a fractional part as described below:
For $E = X \cup Y$ first and second stage edges, and for (c, c_k) weights vectors with $c \in \mathbb{R}^n$, $c_k \in \mathbb{R}^q, k = 1, \dots, K$. Consider a split of the form

$$c_i = \sum_{k=1}^{K} c_i^k, \; i \in X$$

with

$$c_i^k \geq 0, \; i \in X$$

or equivalent vector formulation:

$$c = \sum_{k=1}^{K} c^k$$

with

$$c^k \geq 0$$

For $k \in \{1, \dots, K\}$ we consider the sets $\{i_1^k, \dots, i_{t_k}^k\}$ of indices in the order they are picked by the greedy algorithm applied for each (c^k, c_k) cost vector.
By κ_τ^k we denote the spans of the following subsets of the edge sets in the greedy sequence

$$\{i_1^k, \dots, i_{t_k}^k\}, \; \tau = 1, \dots, t_k, \; k = 1, \dots, K$$

remark

$$r(\kappa_\tau^k) = \tau, \; \tau = 1, \dots, t_k, \; k = 1, \dots, K$$

We introduce a specific condition on first stage edges in every scenario:

Definition 3.1. *For the individual weight vectors (c^k, c_k), $k = 1, \dots, K$, if each first stage edge is either always or never picked simultaneously in every scenario by the greedy algorithm, we say that the status of first stage edges is uniform.*

Lemma 3.1. *Assume there exists a split* $c = \sum_{k=1}^{K} c^k$ *with* $c^k \geq 0$ *fulfilling the condition of definition* (3.1)*, then the system*

$$\left\{ \begin{array}{l} (x, z_1, \ldots, z_K) \in [0,1]^n \times [0,1]^q \times \ldots \times [0,1]^q : \\ \sum_{j \in S \cap X} x_j + \sum_{j \in S \cap Y} z_{jk} \leq r(S), \quad k \in \{1, \ldots, K\}, \quad \forall S \subseteq E \end{array} \right. \quad (10)$$

is totally dual integral.

Proof. Let $(c, c_1, \ldots, c_K)$ be an arbitrary weight vector with $c \in \mathbb{R}^n$, $c_k \in \mathbb{R}^q$, $k = 1, \ldots, K$. Consider the two stage stochastic maximum problem:

$$\left\{ \begin{array}{l} max \sum_{j \in X} c_j x_j + \sum_{k=1}^{k=K} \sum_{j \in Y} c_{jk} z_{jk} \\ \sum_{j \in S \cap X} x_j + \sum_{j \in S \cap Y} z_{jk} \leq r(S), \quad k \in \{1, \ldots, K\}, \quad \forall S \subseteq E \\ (x, z_k) \in \{0,1\}^n \times \{0,1\}^q, k \in \{1, \ldots, K\} \end{array} \right. \quad (11)$$

Consider a split for c fulfilling condition (3.1) and for $k = 1, \ldots, K$, put $x_i^k = 1, z_{ik} = 1$ if the greedy algorithm picks $i \in X$ respectively $i \in Y$, under the vector (c^k, c_k), and $x_i^k = 0, z_{ik} = 0$ otherwise. We obtain :

$$\sum_{k=1}^{k=K} \sum_{j \in X} c_j^k x_j^k + \sum_{k=1}^{k=K} \sum_{j \in Y} c_{jk} z_{jk} = \sum_{j \in X} (\sum_{k=1}^{k=K} c_j^k) x_j + \sum_{k=1}^{k=K} \sum_{j \in Y} c_{jk} z_{jk} = \sum_{j \in X} c_j x_j + \sum_{k=1}^{k=K} \sum_{j \in Y} c_{jk} z_{jk} \quad (12)$$

Feasibility of the scenario specific solutions

$$\sum_{j \in S \cap X} x_j^k + \sum_{j \in S \cap Y} z_{jk} \leq r(S), \quad k \in \{1, \ldots, K\}, \quad \forall S \subseteq E$$

implies feasibility of the two-stage model

$$\sum_{j \in S \cap X} x_j + \sum_{j \in S \cap Y} z_{jk} \leq r(S), \quad k \in \{1, \ldots, K\}, \quad \forall S \subseteq E$$

Now turn to the dual of the LP relaxation aiming at the construction of an optimal solution whose objective value coincides with (12).

$$\left\{ \begin{array}{l} min \sum_{k=1}^{K} \sum_{S \subseteq E} r(S) y_{S,k} : \\ \sum_{k=1}^{K} \sum_{S \subseteq E : i \in X \cap S} y_{S,k} \geq c_i, \quad i \in X \\ \sum_{S \subseteq E : j \in Y \cap S} y_{S,k} \geq c_{jk}, \quad j \in Y, k \in \{1, \ldots, K\} \\ y_{S,k} \geq 0, \quad k \in \{1, \ldots, K\}, \quad S \subseteq E \end{array} \right.$$

(13)

For each $k = 1, \ldots, K$ let, according to the choice $\{i_1^k, \ldots, i_{t_k}^k\}$, $\tau = 1, \ldots, t_k$, $k = 1, \ldots, K$ and in descending order, $\hat{c}_{i_1^k} \geq \ldots \geq \hat{c}_{i_{t_k}^k} \geq 0$ be the weights of the edge picked by the greedy algorithm run on the instance with edge weights (c^k, c_k). According to the deterministic proof of the greedy solution, for the edge sets $S = \kappa^k_\tau,\ \tau = 1, \ldots, t_k,\ k = 1, \ldots, K$ we put

$$y_{S,k} = \hat{c}_{i^k_\tau} - \hat{c}_{i^k_{\tau+1}},\ \tau = 1, \ldots, t_k - 1 \text{ and } y_{S,k} = \hat{c}_{i^k_{t_k}},\ \tau = t_k$$

For the remaining $S \subseteq E$ and $k = 1, \ldots, K$ we put $y_{S,k} = 0$. To check feasibility, fix some $i \in X$, then for each $k = 1, \ldots, K$, there exists a unique index τ^* with $i \in \kappa^k_{\tau^*+1} \setminus \kappa^k_{\tau^*}$. It holds:

$$\sum_{S \subseteq E: i \in X \cap S} y_{S,k} = \sum_{\tau^*}^{t_k} y_{\kappa^k_\tau, k} = \sum_{\tau^*}^{t_k} (\hat{c}_{i^k_\tau} - \hat{c}_{i^k_{\tau+1}}) = \hat{c}_{i^k_{\tau^*}} \geq c^k_i$$

Summing up over k yelds

$$\sum_{k=1}^{K} \sum_{S \subseteq E: i \in X \cap S} y_{S,k} \geq \sum_{k=1}^{K} c^k_i = c_i$$

For $i \in Y$ and $k \in \{1, \ldots, K\}$ again there exists a unique index τ^* with $i \in \kappa^k_{\tau^*+1} \setminus \kappa^k_{\tau^*}$, and we have:

$$\sum_{S \subseteq E: i \in Y \cap S} y_{S,k} = \sum_{\tau^*}^{t_k} y_{\kappa^k_\tau, k} = \sum_{\tau^*}^{t_k} (\hat{c}_{i^k_\tau} - \hat{c}_{i^k_{\tau+1}}) = \hat{c}_{i^k_{\tau^*}} \geq c_{ik}$$

Non negativity of the dual solution is immediate. If $i \notin \kappa^k_{t_k}$, then its edge weight is non-positive, and the dual constraint involving i is fulfilled. For the dual objective, it holds

$$\sum_{k=1}^{K} \sum_{S \subseteq E} r(S) y_{S,k} \tag{14}$$

$$= \sum_{k=1}^{K} \sum_{\tau=1}^{t_k} r(\kappa^k_\tau) y_{\kappa^k_\tau, k} = \sum_{k=1}^{K} \sum_{\tau=1}^{t_k - 1} \tau (\hat{c}_{i^k_\tau} - \hat{c}_{i^k_{\tau+1}}) + t_k \hat{c}_{i^k_{t_k}} = \sum_{k=1}^{K} \sum_{\tau=1}^{t_k} \hat{c}_{i^k_\tau} \tag{15}$$

$$= \sum_{k=1}^{K} (\sum_{i \in X} c^k_i x^k_i + \sum_{i \in Y} c_{ik} z_{ik}) = \sum_{k=1}^{K} (\sum_{i \in X} c^k_i x_i + \sum_{i \in Y} c_{ik} z_{ik}) \tag{16}$$

$$= \sum_{i \in X} (\sum_{k=1}^{K} c^k_i) x_i + \sum_{k=1}^{K} \sum_{i \in Y} c_{ik} z_{ik} \tag{17}$$

$$= \sum_{i \in X} c_i x_i + \sum_{k=1}^{K} \sum_{i \in Y} c_{ik} z_{ik} \tag{18}$$

which coincides with (12). Hence the system in question is totally dual integral.

□

4 Two stage problem with only two scenarios

The case $K = 2$ is very different from the case $K \geq 3$. We prove in this section that in the case $K = 2$, the optimal value of the problem is integer. The main idea is to prove the existence of a correct split of first stage edges costs according to (3.1) by induction on the number of first stage edges. Yet, the case of only one first stage edge gives a basic settlement for any value of K, and the case of only two first stage edges is a central point in the proof. When there exists a split $c = \sum_{k=1}^{K} c^k$ with $c^k \geq 0$ fulfilling the condition of definition (3.1), we speak about a first stage edge i as *'covered'* if the split turns to $x_i^k = 0$ $k = 1, \ldots, K$ when applying greedy algorithms, and as *'chosen'* if the split turns to $x_i^k = 1$ $k = 1, \ldots, K$.

4.1 The case of only one edge in the first level

We consider the case where only one edge x_1 belongs to the first stage.

Theorem 4.1. *With only one edge in the first stage and $K = 2$, the primal problem z_{IP} (1) and dual problem z_{LP} (3.1) have the same integer value. This entails that the system is TDI.*

Proof. It suffices to prove that it is possible to correctly split the cost c_1 into two parts in order to get the same status in both scenarios.
According to lemma 2.1, there exist two thresholds $b_1^1 = b_1^1(c_{11}, \ldots, c_{q1})$ and $b_1^2 = b_1^2(c_{12}, \ldots, c_{q2})$ that determine the status of x_1 in each scenario.
In the case where $b_1^1 + b_1^2 \leq c_1$ then it is possible to split c_1 with respect to $b_1^1 \leq c_1^1$ and $b_1^2 \leq c_1^2$.
In the case where $b_1^1 + b_1^2 \geq c_1$ then it is possible to split c_1 with respect to $b_1^1 \geq c_1^1$ and $b_1^2 \geq c_1^2$.
From the point of view of a single scenario, these different cases can be summarized into one single criteria: for $c_1^1 \in [min(b_1^1, c_1 - b_1^2), max(b_1^1, c_1 - b_1^2)]$, the status of x_1 into both scenarios during greedy algorithm is the same. □

The case of only one edge in the first stage with any number of scenarios can easily be answered in the same manner:

Theorem 4.2. *In the case of any number of scenarios $K \geq 2$, with only one edge in the first stage, the primal problem z_{IP} (1) and dual problem z_{LP} (3.1) have the same integer value. This entails that the system is TDI.*

Proof. It suffices to prove that it is possible to correctly split the cost c_1 into K parts in order to get the same status in every scenario.
According to lemma 2.1, there exist K thresholds $b_1^k = b_1^k(c_{1k}, \ldots, c_{qk})$, $k = 1, \ldots, K$. In the case where $\sum_{k=1}^{K} b_1^k \leq c_1$ then it is possible to split c_1 with respect to $b_1^k \leq c_1^k$, $\forall k$
In the case where $\sum_{k=1}^{K} b_1^k \leq c_1$ then it is possible to split c_1 with respect to $b_1^k \geq c_1^k$, $\forall k$

□

4.2 The case of two edges in the first stage

We consider the case where two edges x_1 and x_2 belong to the first stage with respectively costs c_1 and c_2. We proceed in the same way than with one single edge by spliting fixed costs into two parts: $c_1 = c_1^1 + c_1^2$ and $c_2 = c_2^1 + c_2^2$.

Theorem 4.3. *With only two edges in first stage and $K = 2$, the primal problem z_{IP} (1) and dual problem z_{LP} (3.1) have the same integer value. This entails that the system is TDI.*

Proof. In the proof of theorem 4.1, we have seen that an accurate split of the cost of one edge is given by a compact interval $[min(b_1^1, c_1 - b_1^2), max(b_1^1, c_1 - b_1^2)]$, where b_1^1 and b_1^2 are two continuous functions of the other costs of all edges -independently of their stage.
Since there are only two scenarios, any split of the cost $c_2 = c_2^1 + c_2^2$ can be interpreted as the variation of a single parameter c_2^1. That means that b_1^1 and b_1^2 can be seen as two functions depending of a simple variable c_2^1 and several fixed parameters $c_2, c_{11}, \ldots, c_{q1}$ and respectively $c_2, c_{12}, \ldots, c_{q2}$:

$$b_1^1 = b_1^1(c_2^1, c_{11}, \ldots, c_{q1})$$

and

$$b_1^2 = b_1^2(c_2 - c_2^1, c_{12}, \ldots, c_{q2})$$

Considering the standalone variation of the value $c_2^1 \in [0, c_2]$, we get two continuous functions $f_1(c_2^1) = min(b_1^1, c_1 - b_1^2)$ and $g_1(c_2^1) = max(b_1^1, c_1 - b_1^2)$, defining the border line of a never empty area of the two dimensional space for variables $(c_1^1, c_2^1) \in [0, c_1] \times [0, c_2]$ where x_1 has a common status in scenario $\mathscr{S}_1$ and $\mathscr{S}_2$ ie $x_1^1 = x_1^2$. See figure 1

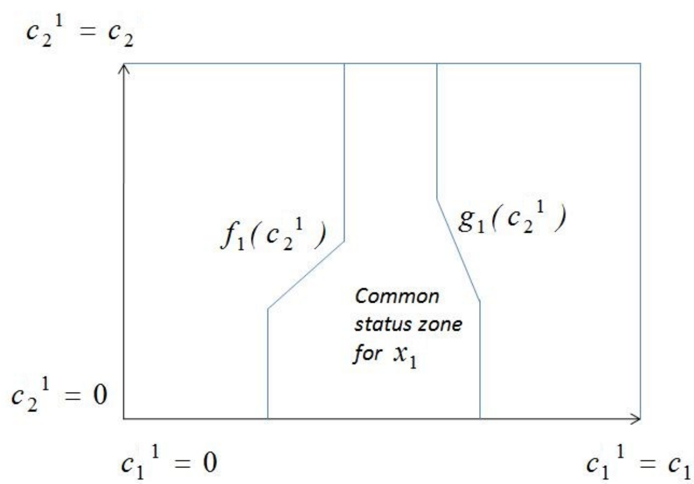

Figure 1: common status for x_1 between two continuous parametric curves

Functions f_1 and g_1 define two continuous parametric curves reaching respectively the points $(f_1(0), 0)$ to $(f_1(c_2), c_2)$ and $(g_1(0), 0)$ to $(g_1(c_2), c_2)$ in the space $(c_1^1, c_2^1) \in [0, c_1] \times [0, c_2]$.
The same analysis with the second first stage edge leads to introduce two similar thresholds:

$$b_2^1 = b_2^1(c_1^1, c_{11}, \ldots, c_{q1})$$

and

$$b_2^2 = b_2^2(c_1 - c_1^1, c_{12}, \ldots, c_{q2})$$

A split of c_2 between these two thresholds gives the same status of x_2 in both scenarios.
We then consider in the same manner two continuous functions $f_2(c_1^1) = min(b_2^1, c_2 - b_2^2)$ and $g_2(c_1^1) = max(b_2^1, c_2 - b_2^2)$ defining the border line of a non empty area of the same two dimensional space for variables $(c_1^1, c_2^1) \in [0, c_1] \times [0, c_2]$ where x_2 has a common status in scenario $\mathscr{S}_1$ and $\mathscr{S}_2$.

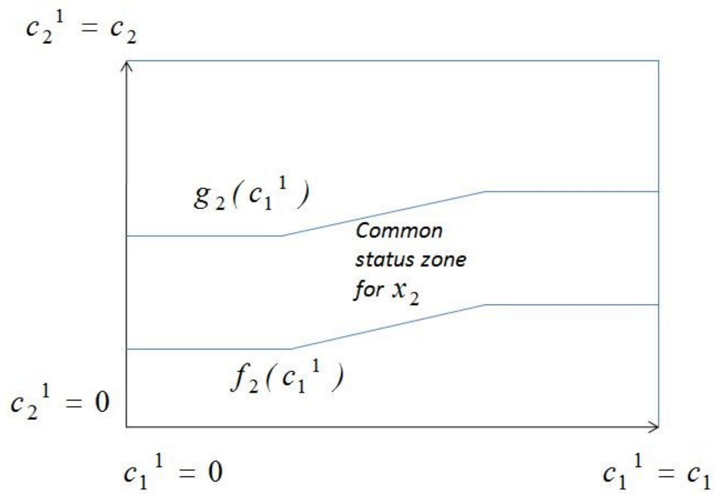

Figure 2: common status for x_2 between two continuous parametric curves

Functions f_2 and g_2 define two continuous parametric curves reaching respectively the points $(0, f_2(0))$ to $(c_1, f_2(c_1))$ and $(0, g_2(0))$ to $(c_1, g_2(c_1), 0)$ in the space $[0, c_1] \times [0, c_2]$. See figure 2
According to the theorem of intermediate values for continuous functions, there exist crossing values for curves (f_1, f_2), (g_1, f_1), (f_1, g_2) and (g_1, g_2) which define a zone with continuous parametric curves for border line and where x_1 and x_2 have simultaneously the same status in $\mathscr{S}_1$ and $\mathscr{S}_2$. See figure 3

□

4.3 The case of any number of first stage edges with only two scenarios

Theorem 4.4. *With only two scenarios, the primal problem z_{IP} and dual problem z_{LP} have the same integer value and the system is TDI.*

Proof. We prove this theorem by induction based on the number of first stage edges. We claim the following assumption:
$H(n) \Leftrightarrow$ "with n edges in the first stage, there exist a correct split of the costs of every edges of first stage in terms of $c_1 = c_1^1 + c_1^2, \ldots, c_n = c_n^1 + c_n^2$ such that these edges get the same respective uniform status in scenario $\mathscr{S}_1$ and $\mathscr{S}_2$ with respect to condition (3.1). When focusing on the part of these splits concerning the first scenario, the correct split is given by $(c_1^1, \ldots, c_n^1) \in \Omega_{(1,\ldots,n)}$ where $\Omega_{(1,\ldots,n)} \subset [0, c_1] \times \ldots \times [0, c_n]$ is a regular compact zone whose border is given by regular (continuous) parametric hypersurfaces.
These hypersurfaces are specific thresholds of the kind

$$f_i(c_1^1, \ldots, c_{i-1}^1, c_{i+1}^1, \ldots, c_n^1, c_{11}, \ldots, c_{1q}, c_{11}, \ldots, c_{q1}, c_{12}, \ldots, c_{q2}) = min(b_i^1, c_i - b_i^2),\ i \in \{1, \ldots, n\}$$

and

$$g_i(c_1^1, \ldots, c_{i-1}^1, c_{i+1}^1, \ldots, c_n^1, c_{11}, \ldots, c_{q1}, c_{12}, \ldots, c_{q2}) = max(b_i^1, c_i - b_i^2),\ i \in \{1, \ldots, n\}$$

or intersections of such thresholds."
The case of $n = 2$ has been proved in section 4.2.

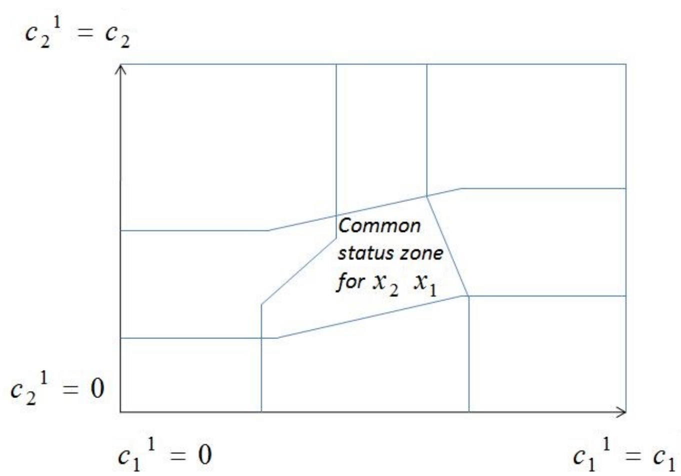

Figure 3: common status for x_1 and x_2

Assume that $H(n)$ is true for some value n, we consider a graph G with $n+1$ first stage edges and q second stage edges. We split the value c_{n+1} into $c_{n+1} = c^1_{n+1} + c^2_{n+1}$. To every value c^1_{n+1}, by application of $H(n)$ it corresponds a regular compact zone $\Omega(c^1_{n+1})$ into which borders are given by specific thresholds of the kind :

$$f_i(c^1_1, \ldots, c^1_{i-1}, c^1_{i+1}, \ldots, c^1_n, c^1_{n+1}, c_{11}, \ldots, c_{q1}, c_{12}, \ldots, c_{q2}) = min(b^1_i, c_i - b^2_i)$$

and

$$g_i(c^1_1, \ldots, c^1_{i-1}, c^1_{i+1}, \ldots, c^1_n, c^1_{n+1}, c_{11}, \ldots, c_{q1}, c_{12}, \ldots, c_{q2}) = max(b^1_i, c_i - b^2_i)$$

or intersections of such thresholds.
Since functions involved in type f or g are min or max of continuous functions of the kind of b^1_i or b^2_i defined in 2.3, the collection $\Omega_{(1,\ldots,n)} = \{\Omega(c^1_{n+1}), c^1_{n+1} \in [0, c_{n+1}]\}$ defines a continuous zone into which all first stage edges $x_1, \ldots, x_n$ have respectively a uniform status in both scenarios according to condition (3.1).
Consider now the proper thresholds for x_{n+1} given by

$$f_{n+1}(c^1_1, \ldots, c^1_n, c_{11}, \ldots, c_{q1}, c_{12}, \ldots, c_{q2}) = min(b^1_{n+1}, c_{n+1} - b^2_{n+1})$$

and

$$g_{n+1}(c^1_1, \ldots, c^1_n, c_{11}, \ldots, c_{q1}, c_{21}, \ldots, c_{q2}) = max(b^1_{n+1}, c_{n+1} - b^2_{n+1})$$

These two functions define respectively two regular (continuous) parametric hypersurfaces in between which x_{n+1} has a common status in scenario $\mathscr{S}_1$ and $\mathscr{S}_2$. If we call Ω_{n+1} the subset of $[0, c_1] \times \ldots \times [0, c_{n+1}]$ between these two hypersurfaces, then the crossing part $\Omega_{(1,\ldots,n)} \cap \Omega_{n+1} = \Omega_{(1,\ldots,n+1)}$ is a non empty zone where all $n+1$ first stage edges have a uniform status in both scenarios. The border line of this intersection is of the same kind of those described in $H(n)$ and that ends proof for $H(n+1)$.For illustration with $n = 3$, see figure (4). $\square$

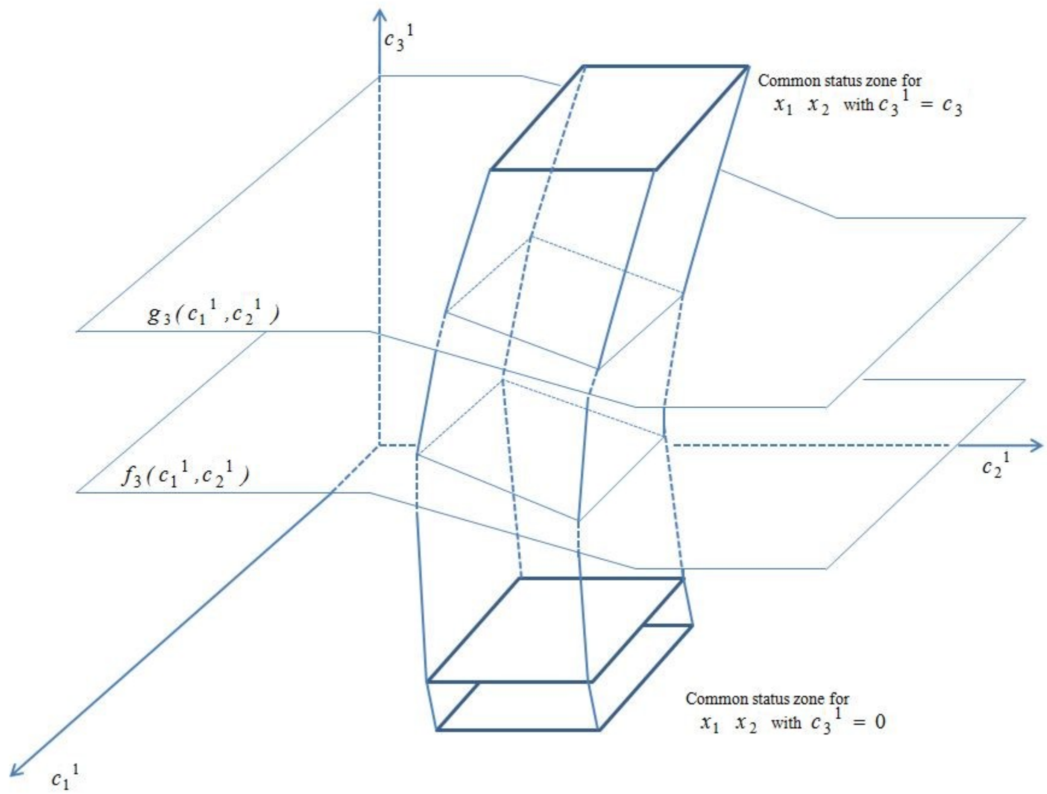

Figure 4: intersection of surfaces f_{n+1} and g_{n+1} and zone $U_{(1,\ldots,n)}$ in the case $n = 2$

5 Multiple scenarios-The case of more than two scenarios

In this section, we study the case where there exist more than two scenarios in the second stage. We will change our point of view by exhibiting an example to the contrary where a fractional solution for (x, z) still compatible with all requirements leads to a higher value than for all integer vectors (x, z). We present a graph where there exist 3 first stage edges not directly connected, and 6 second stage edges.
For all first stage edges, the cost values are $c_1 = c_2 = c_3 = 5$.
In scenario $\mathscr{S}_1$, the cost function for second stage edges is $c_{11} = c_{21} = 6$ and $c_{31} = c_{41} = c_{51} = c_{61} = 0$.
In scenario $\mathscr{S}_2$, the cost function for second stage edges is $c_{32} = c_{42} = 6$ and $c_{12} = c_{22} = c_{52} = c_{62} = 0$.
In scenario $\mathscr{S}_3$, the cost function for second stage edges is $c_{53} = c_{63} = 6$ and $c_{13} = c_{23} = c_{33} = c_{43} = 0$.
There is no integer solution (x, z) where it is possible to take all second stage edges with positive strictly cost and strictly more than only one first stage edge, otherwise there would be a cycle (see figure 5). We can afford that best integer value is less or equal than $6 * 6 + 5 = 41$. Now we propose to take $x_1 = x_2 = x_3 = \dfrac{1}{2}$ and in second stage only edges with strictly positive cost: $z_{11} = z_{21} = 1$; $z_{32} = z_{42} = 1$; $z_{53} = z_{63} = 1$. This fractional solution gives a positive value of $6 * 6 + 3 * 5 * \dfrac{1}{2} = 43,5$. This clearly shows that this

system is not TDI. All requirements are satisfied:

$$\left\{ \begin{array}{l} \displaystyle\sum_{j \in S \cap X} x_j + \sum_{j \in S \cap Y} z_{jk} \leq r(S), \quad k \in \{1, \ldots, 3\}, \quad \forall S \subseteq E \\ (x, z_k) \in [0,1]^n \times [0,1]^q, k \in \{1, \ldots, 3\} \end{array} \right. \tag{19}$$

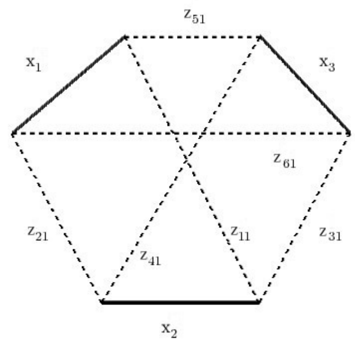

Figure 5: Complete graph for one scenario

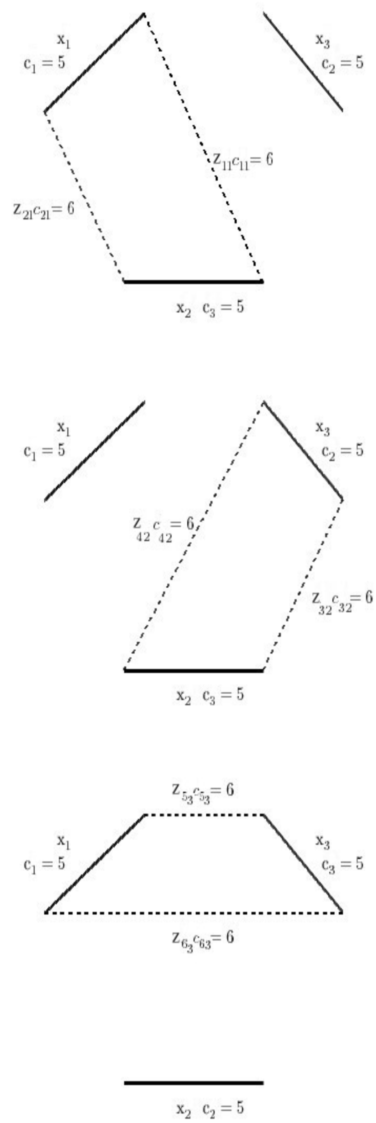

Figure 6: three scenarios with cost function

6 Conclusion

In this paper, we solved TDIness problem for two-stage maximum weight forest problem for different instances with at least two scenarios.

We use an analytic approach with a variable cost split between several scenarios for general graphs. The main difference between instances with two and strictly more than two scenarios dwells in the fact that in some cases, for a specific scenario, the status of a first stage edge can not be chosen by balancing the cost on the other scenarios in order to find an equilibrium between more than two situations. This problem is described with the generic counter example of section 5. In further research, we will focus on approximation methods for the maximum weight stochastic tree based on variations of weights.

References

[1] Cook, W. (1983). operations that preserve total dual integrality. *Elsevier Science Publishers B.V*, 2:31–35.

[2] Dhamdhere, K., Ravi, R., and Singh, M. (2005). On two-stage stochastic minimum spanning trees. *IPCO 2005, LNCS 3509*, pages 321–334.

[Escoffier et al.] Escoffier, B., L.GourvÃÍs, Monnot, J., and Spanjaard, O. Two-stage stochastic matching and spanning tree problems : polynomial instances and approximation. *preprint Elsevier.*

[4] Franck, A. (1981). A weighted matroï£¡d intersection. *Journal of Algorithms*, 2:328–336.

[5] Gupta, A., Pal, M., Ravi, R., and Sinha, A. (2004). Boosted sampling: approximation algorithms for stochastic optimization. *Proc. of the 36th Annual ACM Symposium on Theory of Computong (STOC'04)*, pages 417–426.

[6] Kong, N., Schaeffer, A., and Shabbir, A. (2007). Totally unimodular programs. 1.

[7] Nemhauser, G. and Wolsey, L. (1999). *Integer And Combinatorial Optimization*. Wiley-Interscience series in Discrete Mathematics and Optimization.

[8] Ravi, R. and Sinha, A. (2006). Hedging uncertainty: Approximation algorithms for stochastic optimization problems. *Math. Program.*, 108:97–114.