Math. Program., Ser. B 108, 251–274 (2006)

Digital Object Identifier (DOI) 10.1007/s10107-006-0710-z

Dimitris Bertsimas · Karthik Natarajan · Chung-Piaw Teo

Persistence in discrete optimization under data uncertainty

Received: May 17, 2004/Accepted: April 15, 2005
Published online: April 25, 2006 –

Abstract. An important question in discrete optimization under uncertainty is to understand the *persistency* of a decision variable, i.e., the probability that it is part of an optimal solution. For instance, in project management, when the task activity times are random, the challenge is to determine a set of critical activities that will potentially lie on the longest path. In the spanning tree and shortest path network problems, when the arc lengths are random, the challenge is to pre-process the network and determine a smaller set of arcs that will most probably be a part of the optimal solution under different realizations of the arc lengths. Building on a characterization of moment cones for single variate problems, and its associated semidefinite constraint representation, we develop a limited marginal moment model to compute the persistency of a decision variable. Under this model, we show that finding the persistency is tractable for zero-one optimization problems with a polynomial sized representation of the convex hull of the feasible region. Through extensive experiments, we show that the persistency computed under the limited marginal moment model is often close to the simulated persistency value under various distributions that satisfy the prescribed marginal moments and are generated independently.

Key words. Persistence · Discrete optimization · Semidefinite programming

1. Introduction

In recent years, there has been a flurry of activity devoted to studying discrete optimization problems under data uncertainty (cf. [11], [6], [1]). Consider the discrete optimization problem:

$$Z_{\max}(\tilde{\boldsymbol{c}}) = \max\left\{\tilde{\boldsymbol{c}}'\boldsymbol{x} : \boldsymbol{x} \in \mathcal{X} \subseteq \{0,1\}^n\right\}, \tag{1}$$

where $\mathcal{X}$ denotes the set of feasible solutions. Suppose the objective coefficients $\tilde{\boldsymbol{c}}$ in $Z_{\max}(\tilde{\boldsymbol{c}})$ are randomly generated. For ease of exposition, we will assume that the set of $\tilde{\boldsymbol{c}}$ such that $Z_{\max}(\tilde{\boldsymbol{c}})$ has multiple optimal solutions has a support with measure zero. Hence for a given objective, the discrete optimization problem has a unique solution. Towards the goal of obtaining insight into the structure of optimal solutions, we would like to

D. Bertsimas: Boeing Professor of Operations Research, Sloan School of Management and Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
e-mail: dbertsim@mit.edu

K. Natarajan: Department of Mathematics, National University of Singapore, Singapore 117543.
e-mail: matkbn@nus.edu.sg

C.-P. Teo: Department of Decision Sciences, Business School, National University of Singapore, Singapore 117591.
e-mail: bizteocp@nus.edu.sg

Mathematics Subject Classification (2000): 90C27, 90C22, 90C15

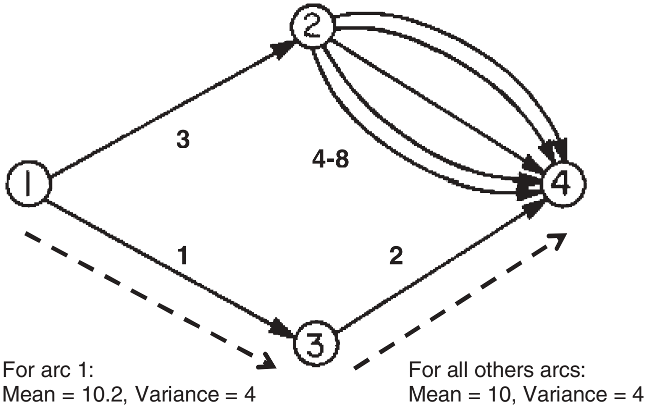

Fig. 1. Deterministic critical path for small sized project from Van Slyke [19]

find the probability that $x_i = 1$ in the optimal solution to $Z_{\max}(\tilde{\boldsymbol{c}})$, which we define as follows.

Definition 1. *The* **persistency** *of a variable x_i is defined to be the probability that $x_i = 1$ in the optimal solution to $Z_{\max}(\tilde{\boldsymbol{c}})$.*

In this paper, we use semi-definite and second-order cone programming to propose an approach to calculate the *persistency*[1] of decision variables in discrete optimization problems under probabilistic information on the objective coefficients. Convex programming techniques have been well developed in the framework of *moment problems* to compute bounds on expected functions of random variables. Problems that have been tackled under this method include computing bounds on prices of options [3], probabilities that a random vector lies in a semi-algebraic set [2, 12], expected order statistics [5] and expected optimal objective value of combinatorial optimization problems [4]. While these approaches focus on computing tight bounds, there has not been much research on obtaining insights into the structure of the solutions under uncertainty. We concretize this problem with an application from project management.

To illustrate the importance of persistence, we use a small project from Van Slyke [19] with eight activities (denoted by arcs) distributed over five paths that need to be completed. The data for the non-negative activity durations include the mean and variance information. Consider the project network and the activity data in Figure 1. The longest path in this graph measures the time to complete the project. The classical CPM/PERT method [15] uses the *expected value* of the activity durations to compute the *critical* (longest) path. Using this approach would identify activities 1 and 2 as critical with an expected duration of 20.2 for the longest path.

Unfortunately, the deterministic approach does not identify the right set of critical activities under data variation. To see why, we simulate the project performance under

[1] In an earlier work, Adams, Lassiter and Sherali studied the question, to what extent the solution to an LP relaxation can be used to fix the value of 0-1 discrete optimization problems. They termed this the persistency problem of 0-1 programming model. The motivation of their work, however is different from ours. See Adams, W.P., Lassiter, J.B., and Sherali, H.D., Persistency in 0-1 Polynomial Programming, Mathematics of Operations Research., Vol. 23, No. 2, 359–389, (1998).

Table 1. Project statistics for the project from Van Slyke [19]

	CPM	Simulation
Expected duration	20.20	22.98
Criticality (1,2)	1.000	0.285
Criticality (3)	0.000	0.715
Criticality (4-8)	0.000	0.143

normally distributed and independent activity durations with the given means and variances. Table 1 indicates the probability that an activity lies on the longest path (termed as criticality index in [19]).

Clearly, the classical method under-estimates the expected duration of the longest path. Furthermore it fails to identify activity 3 as critical. The importance of this activity under data perturbation is evident from the simulation. It is likely to lie on the longest path due to the larger number of parallel paths in the upper part of the graph. Using a deterministic approach would imply that the project manager focuses on the wrong path 71.5 percent of the time!

In the context of project management, persistency as defined in Definition 1 reduces to the notion of criticality indices. It specifies the probability that an activity (decision variable) will lie on the longest path (equal to 1). Computing and identifying this persistence information is useful from a practical perspective. It helps the project manager identify and monitor activities that will have the largest potential to contribute to delays in the completion of the project. In other problems, say for instance the spanning tree and route guidance problems, persistency information can be used to pre-process the network and remove arcs that with high probability are not used in an optimal solution. This allows for problems to be resolved on much smaller networks.

Given a distribution for $\tilde{\boldsymbol{c}}$, the problem of finding the persistency of the variables is in general NP-hard. For example, given that each objective coefficient $\tilde{c}_i$ takes two possible values and the objective coefficients are independently distributed, we need to solve 2^n discrete optimization problems to find the persistency. Another complicating factor that arises in applications is often the incomplete knowledge of distributions (cf. [4]).

In this paper, we formulate a parsimonious model to compute the persistency, by specifying only the range and marginal moments of each $\tilde{c}_i$ in the objective function. The complete distributional information on $\tilde{\boldsymbol{c}}$ and the dependence structure of the $\tilde{c}_i$ are not known. We solve the following model:

$$\sup_{\theta \in \Theta} E_\theta \left(Z_{\max}(\tilde{\boldsymbol{c}}) \right),$$

where Θ represents the class of distributions with the prescribed range and marginal moments for each $\tilde{c}_i$. We show that the above model is tractable for discrete optimization problems where a polynomial sized representation of the convex hull of the feasible region is known. Particularly, by solving a convex program, we show that the primal solutions can be interpreted as persistence of the variables under the distribution that realizes the above supremum either exactly or asymptotically.

Note that a striking feature of our model is the omission of cross moment information among the different random objective coefficients. We do not incorporate constraints to capture conditions such as independently distributed objective coefficients or specifically correlated coefficients. In the latter case, it is conceivable that the estimates provided by our limited marginal moment model may not be precise enough for practical use. Fortunately, extensive experiments seem to indicate that the estimates provided by our model are generally close to the estimates obtained from simulations, provided the random objective coefficients are generated independently. In fact, when the coefficients are independently and normally distributed, our limited marginal moment model using range and the first two moment information already yields good estimates for the persistency values in most cases.

Structure and contributions of the paper:

1. In Sect. 2, we review the duality theory of moment cones and non-negative polynomials and their associated convex cone representations. Particularly, we focus on the characterization of moment cones for single variate problems and their link with positive semi-definite and second-order cone constraints. This forms the basis of our solution methodology for solving the persistency problem.
2. In Sect. 3, we propose a model using limited marginal moment information on the random objective coefficients, to compute the persistency of decision variables. Under this model, we solve the persistency problem with a convex optimization approach and show that the formulation is tight in the limit under the prescribed moments. Furthermore, for a large class of discrete optimization problems, this formulation is shown to be solvable in polynomial time.
3. In Sect. 4, we review generalizations and extensions of the model. In particular, we consider the situation where the moments of the objective coefficients are not explicitly given. Instead, each coefficient is expressed as a (random) solution to another 0-1 stochastic optimization problem with known moment constraints. Interestingly, the approach outlined in this paper applies to this more general problem.
4. In Sect. 5 and Sect. 6, we study the persistency issue in project management and spanning tree problems. Experimental results indicate the potential of the marginal moment model in computing the persistency in discrete optimization problems.

2. Review: Moments, Polynomials and Convex Optimization

Let $\mathbb{P}_{2k}(\Omega)$ denote the cone of univariate non-negative polynomials over the support set Ω:

$$\mathbb{P}_{2k}(\Omega) := \left\{ \boldsymbol{z} \in \Re^{2k+1} : z_0 + z_1 t + \cdots + z_{2k} t^{2k} \geq 0 \text{ for all } t \in \Omega \right\}.$$

Let $\mathbb{M}_{2k}(\Omega)$ denote the conic hull of all vectors of the form $(1, t_j, \ldots, t_j^{2k})$ for $t_j \in \Omega$:

$$\mathbb{M}_{2k}(\Omega) := \left\{ \boldsymbol{y} \in \Re^{2k+1} : \boldsymbol{y} = \sum_j \alpha_j (1, t_j, \ldots, t_j^{2k}) \text{ for all } t_j \in \Omega \text{ with } \alpha_j \geq 0 \right\}.$$

It follows from the definitions above that $\boldsymbol{y} \in \mathbb{M}_{2k}(\Omega)$ (the *moment cone*) if and only if:

$$\mathbb{M}_{2k}(\Omega) = \Big\{ \boldsymbol{y} \in \Re^{2k+1} : \boldsymbol{y} = y_0(1, E[\tilde{c}], \dots, E[\tilde{c}^{2k}]) \\ \text{for some r.v. } \tilde{c} \text{ with support } \Omega \text{ and } y_0 \geq 0 \Big\}.$$

The dual of this moment cone is:

$$\mathbb{M}_{2k}(\Omega)^* = \Big\{ \boldsymbol{z} \in \Re^{2k+1} : \boldsymbol{z}'\boldsymbol{y} \geq 0 \text{ for all } \boldsymbol{y} \in \mathbb{M}_{2k}(\Omega) \Big\}.$$

It is easy to see that $\mathbb{M}_{2k}(\Omega)^* = \mathbb{P}_{2k}(\Omega)$, and hence it follows that:

$$\overline{\mathbb{M}_{2k}(\Omega)} = \mathbb{P}^*_{2k}(\Omega),$$

i.e. the closure of the moment cone is precisely the dual cone of the set of non-negative polynomials on Ω.

For univariate random variables, the cone of moments and non-negative polynomials can be equivalently represented with convex constraints. The moment conditions are related to positive semidefinite conditions on Hankel matrices. We use the notation from [10], [21] to define Hankel matrices. Let $\boldsymbol{M}_k(t)$ denote the rank one matrix:

$$\boldsymbol{M}_k(t) = \begin{pmatrix} 1 & t & \dots & t^k \\ t & t^2 & \dots & t^{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ t^{k+1} & t^{k+2} & \dots & t^{2k} \end{pmatrix}.$$

Let $\boldsymbol{M}_k(t)|_{\boldsymbol{y}}$ denote the basic Hankel matrix obtained by replacing monomial t^i by y_i:

$$\boldsymbol{M}_k(t)|_{\boldsymbol{y}} = \begin{pmatrix} y_0 & y_1 & \dots & y_k \\ y_1 & y_2 & \dots & y_{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_{k+1} & y_{k+2} & \dots & y_{2k} \end{pmatrix}.$$

Proposition 1. *The closure of the moment cone for univariate random variables can be equivalently represented as positive semidefinite conditions on Hankel matrices of the form*:

$$\overline{\mathbb{M}_{2k}(\Re)} = \Big\{ \boldsymbol{y} \in \Re^{2k+1} : \boldsymbol{M}_k(t)|_{\boldsymbol{y}} \succeq \boldsymbol{0} \Big\}$$
$$\overline{\mathbb{M}_{2k}(\Re^+)} = \Big\{ \boldsymbol{y} \in \Re^{2k+1} : \boldsymbol{M}_k(t)|_{\boldsymbol{y}} \succeq \boldsymbol{0},\ t\boldsymbol{M}_{k-1}(t)|_{\boldsymbol{y}} \succeq \boldsymbol{0} \Big\}$$
$$\overline{\mathbb{M}_{2k+1}(\Re^+)} = \Big\{ \boldsymbol{y} \in \Re^{2(k+1)} : \boldsymbol{M}_k(t)|_{\boldsymbol{y}} \succeq \boldsymbol{0},\ t\boldsymbol{M}_k(t)|_{\boldsymbol{y}} \succeq \boldsymbol{0} \Big\}$$
$$\overline{\mathbb{M}_{2k}([0,1])} = \Big\{ \boldsymbol{y} \in \Re^{2k+1} : \boldsymbol{M}_k(t)|_{\boldsymbol{y}} \succeq \boldsymbol{0},\ t\boldsymbol{M}_{k-1}(t)|_{\boldsymbol{y}} - t^2\boldsymbol{M}_{k-1}(t)|_{\boldsymbol{y}} \succeq \boldsymbol{0} \Big\}$$
$$\overline{\mathbb{M}_{2k+1}([0,1])} = \Big\{ \boldsymbol{y} \in \Re^{2(k+1)} : t\boldsymbol{M}_k(t)|_{\boldsymbol{y}} \succeq \boldsymbol{0},\ \boldsymbol{M}_k(t)|_{\boldsymbol{y}} - t\boldsymbol{M}_k(t)|_{\boldsymbol{y}} \succeq \boldsymbol{0} \Big\}.$$

These results follow from well-known representations for the (truncated) Hamburger, Stieltjes and Hausdorff moments problem [12]. It should be noted, that the notion of the closure is introduced here since only a sequence of measures might exist that achieves the moments asymptotically. An example of such a moment vector is $\boldsymbol{y} = (1, 0, 0, 0, 1) \in \overline{\mathbb{M}_4(\Re)}$ with $\boldsymbol{M}_2(t)|_{\boldsymbol{y}} \succeq \mathbf{0}$ but $\boldsymbol{M}_2(t)|_{\boldsymbol{y}} \not\succ \mathbf{0}$. In this case, only the limit of a sequence of measures can be found that achieves the moments (cf. Example 2.37 on Pg. 66 in [7]). Moment representations of other intervals can be obtained from simple transformations of these Hankel matrices. In fact, under only first and second moment information, the moment cone can be characterized with second order cone constraints.

Proposition 2. *The closure of the moment cone for univariate random variables given first and second moments can be equivalently represented as second order cone constraints of the form*:

$$\begin{aligned}
\overline{\mathbb{M}_2(\Re)} &= \left\{ \boldsymbol{y} \in \Re^3 : (y_0 + y_2) \geq \sqrt{(y_0 - y_2)^2 + 4y_1^2} \right\} \\
\overline{\mathbb{M}_2(\Re^+)} &= \left\{ \boldsymbol{y} \in \Re^3 : (y_0 + y_2) \geq \sqrt{(y_0 - y_2)^2 + 4y_1^2},\ y_1 \geq 0 \right\} \\
\overline{\mathbb{M}_2([0,1])} &= \left\{ \boldsymbol{y} \in \Re^3 : (y_0 + y_2) \geq \sqrt{(y_0 - y_2)^2 + 4y_1^2},\ y_1 \geq y_2 \right\}.
\end{aligned}$$

Proof. This result follows from the equivalence of a 2×2 positive semidefinite matrix and a second order cone constraint. Note that:

$$\begin{aligned}
\begin{pmatrix} y_0 & y_1 \\ y_1 & y_2 \end{pmatrix} \succeq \mathbf{0} &\Longleftrightarrow y_0 \geq 0, \quad y_2 \geq 0, \quad y_0 y_2 \geq y_1^2 \\
&\Longleftrightarrow (y_0 + y_2) \geq \sqrt{(y_0 - y_2)^2 + 4y_1^2}.
\end{aligned}$$

Combining with Proposition 1, we obtain the desired second order cone representation. □

3. Marginal Moment Model: Formulation and Analysis

In Problem (1), we assume that we are given the first j_i moments for each objective coefficient $\tilde{c}_i$. Any feasible marginal distribution θ_i satisfies $E_{\theta_i}(\tilde{c}_i^j) = m_{ij}, j = 0, \ldots, j_i$ with the support of the distribution in Ω_i. We denote the vector of marginal moments as $\boldsymbol{m_i} = (m_{i0}, m_{i1}, \ldots, m_{ij_i})$ where $m_{i0} := 1$. Let Θ denote the set of multivariate distributions θ on $\tilde{\boldsymbol{c}}$ such that the marginal distributions satisfy the moment requirements for each $\tilde{c}_i$. We refer to this model as the *Marginal Moments Model* (MMM) [4]. We define:

$$Z^*_{\max} = \sup_{\theta \in \Theta} E_\theta \Big(Z_{\max}(\tilde{\boldsymbol{c}}) \Big), \tag{2}$$

and assume that the convex hull $CH(\mathcal{X})$ is characterized by the set of constraints $\boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}$.

Let $x_i(\tilde{\boldsymbol{c}})$ denote the value of the variable x_i in the optimal solution to Problem (1) obtained under $\tilde{\boldsymbol{c}}$. When $\tilde{\boldsymbol{c}}$ is random, $x_i(\tilde{\boldsymbol{c}})$ is a random variable. In our problem, $x_i(\tilde{\boldsymbol{c}}) \in \{0, 1\}$. The objective function can then be expressed as:

$$\begin{aligned}
E_\theta\Big(Z_{\max}(\tilde{\boldsymbol{c}})\Big) &= E_\theta\Big(\sum_{i=1}^{n} \tilde{c}_i x_i(\tilde{\boldsymbol{c}})\Big) \\
&= \sum_{i=1}^{n}\Big(E_\theta\Big(\tilde{c}_i x_i(\tilde{\boldsymbol{c}})\Big| x_i(\tilde{\boldsymbol{c}})=1\Big) P_\theta(x_i(\tilde{\boldsymbol{c}})=1 \\
&\qquad + E_\theta\Big(\tilde{c}_i x_i(\tilde{\boldsymbol{c}})\Big| x_i(\tilde{\boldsymbol{c}})=0\Big) P_\theta(x_i(\tilde{\boldsymbol{c}})=0)\Big) \\
&= \sum_{i=1}^{n}\Big(E_\theta\Big(\tilde{c}_i\Big| x_i(\tilde{\boldsymbol{c}})=1\Big) P_\theta(x_i(\tilde{\boldsymbol{c}})=1)\Big).
\end{aligned}$$

We define

$$w_{ij}(k) = E_\theta\Big(\tilde{c}_i^j \Big| x_i(\tilde{\boldsymbol{c}})=k\Big) P_\theta(x_i(\tilde{\boldsymbol{c}})=k).$$

and obtain

$$E_\theta\Big(Z_{\max}(\tilde{\boldsymbol{c}})\Big) = \sum_{i=1}^{n} w_{i1}(1). \tag{3}$$

Since Problem (1) is a 0-1 optimization problem, we have

$$m_{ij} = E_\theta\Big(\tilde{c}_i^j\Big) = \sum_{k=0}^{1} E_\theta\Big(\tilde{c}_i^j \Big| x_i(\tilde{\boldsymbol{c}})=k\Big) P_\theta(x_i(\tilde{\boldsymbol{c}})=k) = \sum_{k=0}^{1} w_{ij}(k). \tag{4}$$

Furthermore, $E_\theta(x_i(\tilde{\boldsymbol{c}})) = P_\theta(x_i(\tilde{\boldsymbol{c}})=1) = w_{i0}(1)$. Since the vector $(x_1(\tilde{\boldsymbol{c}}), \dots, x_n(\tilde{\boldsymbol{c}})) \in CH(\mathcal{X})$ for all realizations of $\tilde{\boldsymbol{c}}$, taking expectations, we have:

$$\Big(w_{10}(1), w_{20}(1), \dots, w_{n0}(1)\Big) \in CH(\mathcal{X}). \tag{5}$$

This brings us to the following result.

Theorem 1. *$Z^*_{\max}$ is computed by solving*:

$$\begin{aligned}
Z^*_{\max} = \ & \sup \sum_{i=1}^{n} w_{i1}(1) && \text{(6a)}\\
\text{s.t.}\ & \boldsymbol{w_i}(1) + \boldsymbol{w_i}(0) = \boldsymbol{m_i}, \quad i=1,\dots,n && \text{(6b)}\\
& \boldsymbol{A}(w_{10}(1), w_{20}(1), \dots, w_{n0}(1)) \le \boldsymbol{b}, && \text{(6c)}\\
& \boldsymbol{w_i}(k) \in \overline{\mathbb{M}_{j_i}(\Omega_i)}, \quad i=1,\dots,n, \quad k=0,1. && \text{(6d)}
\end{aligned}$$

Note that to compute $Z^*_{\max}$, we have introduced the decision vector:

$$\boldsymbol{w_i}(k) := (w_{i0}(k), w_{i1}(k), \ldots, w_{ij_i}(k))$$

in place of variable x_i in Problem (2). Then (6a), (6b) and (6c) follow from (3), (4) and (5) respectively. (6d) follows from the requirement that the (conditional) moments $\boldsymbol{w_i}(k)$ must lie in the closure of the moment cone $\mathbb{M}_{j_i}(\Omega_i)$ implying that they are valid moments or limit of a sequence of valid moments. It is thus clear that the above formulation constitutes a valid relaxation for $Z^*_{\max}$. To see that the bound obtained is tight, we provide an approach to construct extremal distributions that achieves the bound in Formulation (6). Particularly under the marginal moment model, we identify the persistence of variables in the extremal distributions that exactly or asymptotically achieve $Z^*_{\max}$. This yields a proof to Theorem 1.

Lemma 1. *Let an optimal solution to Formulation* (6) *be denoted by* $\boldsymbol{w}^*_i(1), \boldsymbol{w}^*_i(0)$ *for all* i. *Then, there exists an extremal distribution* θ^* *that exactly or asymptotically achieves the bound* $Z^*_{\max}$ *and satisfies the marginal moment requirements. Under* θ^*, $\boldsymbol{w}^*_i(1)$ *and* $\boldsymbol{w}^*_i(0)$ *are proportional to the moments of the distribution of* $\tilde{c}_i$ *conditional on whether* $x_i = 1$ *or* 0 *in the optimal solution. Furthermore,* $w^*_{i0}(1) = P_{\theta^*}(x_i(\tilde{\boldsymbol{c}}) = 1)$ *is the persistence of variable* x_i *in the optimal solution.*

Proof. We construct a (limiting sequence of) distribution(s) that attains the tight upper bound $Z^*_{\max}$ in the following manner. Let $p \in \{1, \ldots, P\}$ denote the set of extreme point solutions to Problem (1). We let $x_i[p]$ denote the value of the x_i variable at the p^{th} extreme point. In our problem, $x_i[p] \in \{0, 1\}$. Eq. (6c) implies that $(w^*_{10}(1), \ldots, w^*_{n0}(1))$ lies in the convex hull of the set of 0-1 feasible solutions. Expressing it as a convex combination of the extreme points implies that there exist a set of numbers λ^*_p such that:

$$\text{(i)}\quad \lambda^*_p \geq 0 \quad \text{for all } p = 1, \ldots, P$$

$$\text{(ii)}\quad \sum_{p=1}^{P} \lambda^*_p = 1$$

$$\text{(iii)}\quad w^*_{i0}(1) = \sum_{p: x_i[p]=1} \lambda^*_p \quad \text{for all } i = 1, \ldots, n.$$

Clearly, the numbers λ^*_p are not necessarily unique. The moment condition (6d) implies that there exists a (limiting) sequence of measures with moments $\boldsymbol{w}^*_i(1)$ and $\boldsymbol{w}^*_i(0)$ with support contained in Ω_i. We now generate the multivariate distribution θ^* as follows:

(a) Choose a feasible solution $p \in \{1, \ldots, P\}$ to the nominal problem with probability λ^*_p
(b) Generate $\tilde{c}_i \sim \boldsymbol{w}^*_i(1)/w^*_{i0}(1)$ for i s.t. $x_i[p] = 1$ and $\tilde{c}_i \sim \boldsymbol{w}^*_i(0)/w^*_{i0}(0)$ for i s.t. $x_i[p] = 0$.

Here, $\tilde{c}_i \sim \boldsymbol{w}^*_i(1)/w^*_{i0}(1)$ means that we choose a distribution with moments $\boldsymbol{w_i}^*(1)/w^*_{i0}(1)$. Note that if $w^*_{i0}(1) = 0$, then $\lambda^*_p = 0$ for all p s.t. $x_i[p] = 1$, i.e. the feasible solutions with $x_i = 1$ are not chosen in the extremal distribution. Under this distribution, the marginal moments for $\tilde{c}_i$ are computed as follows.

$$\begin{aligned}
&\text{Moment vector for } \tilde{c}_i \\
&= \sum_{p:x_i[p]=1} \lambda_p^* \left(\frac{\boldsymbol{w}_i^*(1)}{w_{i0}^*(1)} \right) + \sum_{p:x_i[p]=0} \lambda_p^* \left(\frac{\boldsymbol{w}_i^*(0)}{w_{i0}^*(0)} \right) \\
&= w_{i0}^*(1) \left(\frac{\boldsymbol{w}_i^*(1)}{w_{i0}^*(1)} \right) + (1 - w_{i0}^*(1)) \left(\frac{\boldsymbol{w}_i^*(0)}{w_{i0}^*(0)} \right) \quad \text{[From (ii) \& (iii)]} \\
&= \boldsymbol{w}_i^*(1) + \boldsymbol{w}_i^*(0) \\
&= \boldsymbol{m}_i.
\end{aligned}$$

Furthermore, under $\tilde{\boldsymbol{c}}$, if we simply pick the p^{th} solution with probability λ_p^*, instead of solving for $Z_{\max}^*(\tilde{\boldsymbol{c}})$, we have

$$\begin{aligned}
E_{\theta^*}[Z_{\max}(\tilde{\boldsymbol{c}})] &\geq \sum_{p=1}^{P} \lambda_p^* \left(\sum_{i:x_i[p]=1} \frac{w_{i1}^*(1)}{w_{i0}^*(1)} \right) \\
&= \sum_{i=1}^{n} \frac{w_{i1}^*(1)}{w_{i0}^*(1)} \left(\sum_{p:x_i[p]=1} \lambda_p^* \right) \\
&= \sum_{i=1}^{n} w_{i1}^*(1) \quad \text{[From (iii)]}.
\end{aligned}$$

Since θ^* generates either exactly or asymptotically an expected optimal objective value that is greater than or equal to the optimal solution from Formulation (6) and satisfies the marginal moment requirements, it attains $Z_{\max}^*$. Clearly under θ^*, $\boldsymbol{w}_i^*(1)$ and $\boldsymbol{w}_i^*(0)$ are proportional to the moments of $\tilde{c}_i$ conditional on whether $x_i = 1$ or 0 in the optimal solution to $Z_{\max}(\tilde{\boldsymbol{c}})$. Furthermore, $w_{i0}^*(1)$ is the persistence of variable x_i in the optimal solution. □

Remarks.

(a) For zero-one optimization problems with a known polynomial sized representation of the convex hull of the feasible region, Formulation (6) is solvable in polynomial time. This implies that for the longest path problem on a directed acylic graph, linear assignment problem, network flow problems like shortest path and spanning tree problems, the persistency problem as defined is solvable in polynomial time.

(b) Formulation (6) can be viewed as the primal moment's version to the polynomial optimization formulation to compute $Z_{\max}^*$ developed in Bertsimas, Natarajan and Teo [4]:

$$
\begin{aligned}
Z^*_{\max} = \min \ & \left(\boldsymbol{p}'\boldsymbol{b} + \sum_{i=1}^{n} \boldsymbol{y_i}'\boldsymbol{m_i} \right) \\
\text{s.t.} \ & \boldsymbol{y_i} + (\boldsymbol{0}, -1, d_i) \in \mathbb{P}_{j_i}(\Omega_i), \ i = 1, \ldots, n \\
& \boldsymbol{y_i} \in \mathbb{P}_{j_i}(\Omega_i), \qquad\qquad\quad i = 1, \ldots, n \\
& \boldsymbol{p}'\boldsymbol{A} = \boldsymbol{d}' \\
& \boldsymbol{p} \geq \boldsymbol{0}.
\end{aligned} \tag{7}
$$

The advantage in the primal formulation is that we obtain important insights into the persistence of decision variables of discrete optimization problems.

4. Extensions

We now show that the conditional moments approach can be extended to computing persistency for more general two-step discrete optimization problems. The class of two-step discrete optimization problems that we study is described next.

Inner Step: In the inner step, we consider R different discrete optimization problems:

$$
Z_{\max}(\tilde{\boldsymbol{c}}^{\boldsymbol{r}}) = \max \left\{ \tilde{\boldsymbol{c}}^{\boldsymbol{r}\prime}\boldsymbol{x}^{\boldsymbol{r}} : \boldsymbol{x}^{\boldsymbol{r}} \in \mathcal{X}^r \subseteq \{0, 1\}^{n_r} \right\} \quad \text{for } r = 1, \ldots, R, \tag{8}
$$

where $\mathcal{X}^r$ is the set of feasible solutions for the r^{th} problem and $\tilde{\boldsymbol{c}}^{\boldsymbol{r}} := (\tilde{c}^r_1, \ldots, \tilde{c}^r_{n_r})$ are the random objective coefficients for the r^{th} problem, with given moment constraints.

Outer Step: In the outer step, we consider a single discrete optimization problem that links the R inner discrete optimization problems as follows:

$$
Z_{\max}(\tilde{\boldsymbol{c}}^{\boldsymbol{1}}, \ldots, \tilde{\boldsymbol{c}}^{\boldsymbol{R}}) = \max \left\{ \sum_{r=1}^{R} Z_{\max}(\tilde{\boldsymbol{c}}^{\boldsymbol{r}}) y_r : (y_1, \ldots, y_R) \in \mathcal{Y} \subseteq \{0, 1\}^R \right\}, \tag{9}
$$

where the decision vector $\boldsymbol{y} := (y_1, \ldots, y_R)$ lies in the 0-1 feasible region $\mathcal{Y}$. We would like to estimate the persistency of the variable y_r (i.e., the probability that the coefficient $Z_{\max}(\tilde{\boldsymbol{c}}^{\boldsymbol{r}})$ will be a part of the optimal solution) and the persistency of the variable x^r_i (i.e., the probability that x^r_i will take a value of 1 in the optimal solution), given that $y_r = 1$ in the optimal solution.

Such two-step optimization problems can be used to study:

(a) **Multiobjective optimization problems:** Consider the case where the feasible region for the R inner optimization problems are the same, i.e., $\mathcal{X}^1 = \cdots = \mathcal{X}^R = \mathcal{X}$ and $n_1 = \cdots = n_R = n$. Let $\tilde{\boldsymbol{c}}^{\boldsymbol{1}}, \ldots, \tilde{\boldsymbol{c}}^{\boldsymbol{R}}$ denote R different objective vectors for the discrete optimization problem. Then, $Z_{\max}(\tilde{\boldsymbol{c}}^{\boldsymbol{1}}, \ldots, \tilde{\boldsymbol{c}}^{\boldsymbol{R}})$ represents a possible formulation to find a good solution to this multiobjective problem. For example, when

$\mathcal{Y}=\left\{\boldsymbol{y} \in \{0,1\}^R : \sum_{i=1}^{R} y_i = 1\right\}$, $Z_{\max}(\tilde{\boldsymbol{c}}^{\boldsymbol{1}}, \ldots, \tilde{\boldsymbol{c}}^{\boldsymbol{R}})$ reduces to $\max\left\{Z_{\max}(\tilde{\boldsymbol{c}}^{\boldsymbol{r}}) : r=1,\ldots,R\right\}$, which is equivalent to

$$\max\left\{\max\left(\tilde{\boldsymbol{c}}^{\boldsymbol{r}\prime}\boldsymbol{x}^{\boldsymbol{r}} : \boldsymbol{x}^{\boldsymbol{r}} \in \mathcal{X}\right) : r=1,\ldots,R\right\}$$
$$= \max\left\{\max\left(\tilde{\boldsymbol{c}}^{\boldsymbol{1}\prime}\boldsymbol{x}, \ldots, \tilde{\boldsymbol{c}}^{\boldsymbol{R}\prime}\boldsymbol{x}\right) : \boldsymbol{x} \in \mathcal{X}\right\}.$$

Under randomly generated coefficients $\tilde{\boldsymbol{c}}^{\boldsymbol{r}}$, we focus on computing the persistency in this multiobjective optimization problem. In the outer step, persistency identifies the importance of the r^{th} objective among the R different objectives. In the inner step, persistency identifies the probability that the variable x_i^r takes a value of 1 given that the r^{th} objective is persistent among the R different objectives. The variable $\sum_{r=1}^{R} x_i^r y_r$ can be used to identify the persistency that the variable x_i takes a value of 1 in the optimization problem under data uncertainty:

$$\max\left\{\max\left(\tilde{\boldsymbol{c}}^{\boldsymbol{1}\prime}\boldsymbol{x}, \ldots, \tilde{\boldsymbol{c}}^{\boldsymbol{R}\prime}\boldsymbol{x}\right) : \boldsymbol{x} \in \mathcal{X}\right\}.$$

(b) **Portfolio of optimization problems:** Consider the case where one solves R different discrete optimization problems with the r^{th} problem denoted as $Z_{\max}(\tilde{\boldsymbol{c}}^{\boldsymbol{r}})$. If the solution of each problem requires the utilization of certain capacitated pool of resources, then not all the discrete optimization problems can be solved at the same time. In this case, $Z_{\max}(\tilde{\boldsymbol{c}}^{\boldsymbol{1}}, \ldots, \tilde{\boldsymbol{c}}^{\boldsymbol{R}})$ represents a possible formulation to identify the important problems in this portfolio of optimization problems with respect to the feasible region $\mathcal{Y}$. Under uncertainty in the objective coefficients, persistency in the outer step identifies the importance of the r^{th} optimization problem and persistency in the inner step identifies the probability that the variable x_i^r takes a value of 1 given that the r^{th} problem is included among the optimization problems to be solved.

We now show how marginal moment information can be used to study two-step discrete optimization problems of the type defined under data uncertainty. As before, we assume that for each objective coefficient $\tilde{c}_i^r$, we know a limited set of marginal moments $\boldsymbol{m}_{\boldsymbol{i}}^{\boldsymbol{r}}$. We let Θ denote the feasible set of multivariate distributions θ for the combined objective vector $(\tilde{\boldsymbol{c}}^{\boldsymbol{1}}, \ldots, \tilde{\boldsymbol{c}}^{\boldsymbol{R}})$ that satisfy the marginal moment requirements. We are interested in solving the following model:

$$Z_{\max}^{*} = \sup_{\theta \in \Theta} E_{\theta}\left(Z_{\max}(\tilde{\boldsymbol{c}}^{\boldsymbol{1}}, \ldots, \tilde{\boldsymbol{c}}^{\boldsymbol{R}})\right). \tag{10}$$

We assume that the convex hull of the feasible region for the r^{th} problem $CH(\mathcal{X}^r)$, is characterized by the set of constraints $\boldsymbol{A}^{\boldsymbol{r}}\boldsymbol{x}^{\boldsymbol{r}} \leq \boldsymbol{b}^{\boldsymbol{r}}$. Similarly, we assume that the convex hull of the feasible region of the variables in the outer step problem $CH(\mathcal{Y})$ are characterized by the set of constraints $\boldsymbol{G}\boldsymbol{y} \leq \boldsymbol{g}$. We can then express Formulation (9) as an equivalent non-linear integer optimization problem:

$$Z_{\max}(\tilde{\boldsymbol{c}}^{\mathbf{1}},\ldots,\tilde{\boldsymbol{c}}^{\boldsymbol{R}}) = \max \left(\sum_{r=1}^{R} y_r \tilde{\boldsymbol{c}}^{\boldsymbol{r}\prime} \boldsymbol{x}^{\boldsymbol{r}}\right)$$
$$\text{s.t } \boldsymbol{A}^{\boldsymbol{r}} \boldsymbol{x}^{\boldsymbol{r}} y_r \le \boldsymbol{b}^{\boldsymbol{r}} y_r, \quad r=1,\ldots,R$$
$$\boldsymbol{G}\boldsymbol{y} \le \boldsymbol{g}.$$

We let $x_i^r(\tilde{\boldsymbol{c}})$ denote the value of the variable x_i^r in the optimal solution obtained under $\tilde{\boldsymbol{c}}$. In addition, we let $y_r(\tilde{\boldsymbol{c}})$ denote the corresponding optimal 0-1 solution for y_r. The variables y_r are used to model the outer stage problem and incorporating it in the convex marginal moments model is the key contribution in this Section. We next state the main result for two-step discrete optimization problems of the type defined under marginal moment constraints and then sketch the proof.

Theorem 2. $Z^*_{\max}$ *is computed by solving*:

$$Z^*_{\max} = \sup \sum_{r=1}^{R}\sum_{i=1}^{n_r} w_{i1}^r(1) \tag{11a}$$
$$\text{s.t. } \boldsymbol{w}_{\boldsymbol{i}}^{\boldsymbol{r}}(1) + \boldsymbol{w}_{\boldsymbol{i}}^{\boldsymbol{r}}(0) = \boldsymbol{m}_{\boldsymbol{i}}^{\boldsymbol{r}}, \quad i=1,\ldots,n_r,\ r=1,\ldots,R \tag{11b}$$
$$\boldsymbol{A}^{\boldsymbol{r}}(w_{10}^r(1),\ldots,w_{n_r0}^r(1)) \le \boldsymbol{b}^{\boldsymbol{r}} y_r, \quad r=1,\ldots,R \tag{11c}$$
$$\boldsymbol{G}(y_1,\ldots,y_R) \le \boldsymbol{g}, \tag{11d}$$
$$\boldsymbol{w}_{\boldsymbol{i}}^{\boldsymbol{r}}(k) \in \overline{\mathbb{M}_{j_i^r}(\Omega_i^r)}, \qquad i=1,\ldots,n_r,\ r=1,\ldots,R,\ k=0,1. \tag{11e}$$

For the extremal distribution θ^* *that exactly or asymptotically achieves* $Z^*_{\max}$, $\boldsymbol{w}_{\boldsymbol{i}}^{\boldsymbol{r}*}(1)$ *and* $\boldsymbol{w}_{\boldsymbol{i}}^{\boldsymbol{r}*}(0)$ *are proportional to the moments of the distribution of* $\tilde{c}_i^r$ *conditional on whether* $x_i^r y_r = 1$ *or* 0 *in the optimal solution. Furthermore,* $w_{i0}^{r*}(1) = P_{\theta^*}(x_i^r(\tilde{\boldsymbol{c}}) y_r(\tilde{\boldsymbol{c}}) = 1)$ *is the persistence of variable* $x_i^r y_r$ *in the optimal solution and* $y_r^* = P_{\theta^*}(y_r(\tilde{\boldsymbol{c}}) = 1)$ *is the persistence of variable* y_r *in the optimal solution.*

Proof. To prove the result, we introduce two sets of decision variables:

$$w_{ij}^r(k) = E_\theta\left((\tilde{c}_i^r)^j \,\middle|\, x_i^r(\tilde{\boldsymbol{c}}) y_r(\tilde{\boldsymbol{c}}) = k\right) P_\theta(x_i^r(\tilde{\boldsymbol{c}}) y_r(\tilde{\boldsymbol{c}}) = k),$$

and with a slight abuse of notation:

$$y_r = P_\theta(y_r(\tilde{\boldsymbol{c}}) = 1).$$

Clearly for the two-step discrete optimization problem, $x_i^r(\tilde{\boldsymbol{c}})y_r(\tilde{\boldsymbol{c}}) \in \{0, 1\}$. Conditioning based on this, we express the objective function in $Z_{\max}(\tilde{\boldsymbol{c}}^{\boldsymbol{1}}, \dots, \tilde{\boldsymbol{c}}^{\boldsymbol{R}})$ as:

$$\begin{aligned} E_\theta\Big(Z_{\max}(\tilde{\boldsymbol{c}}^{\boldsymbol{1}}, \dots, \tilde{\boldsymbol{c}}^{\boldsymbol{R}})\Big) &= E_\theta\Big(\sum_{r=1}^{R}\sum_{i=1}^{n_r} \tilde{c}_i^r x_i^r(\tilde{\boldsymbol{c}})y_r(\tilde{\boldsymbol{c}})\Big) \\ &= \sum_{r=1}^{R}\sum_{i=1}^{n_r}\Big(E_\theta\Big(\tilde{c}_i^r x_i^r(\tilde{\boldsymbol{c}})y_r(\tilde{\boldsymbol{c}})\Big| x_i^r(\tilde{\boldsymbol{c}})y_r(\tilde{\boldsymbol{c}}) = 1\Big) \\ &\qquad \times P_\theta(x_i^r(\tilde{\boldsymbol{c}})y_r(\tilde{\boldsymbol{c}}) = 1) \\ &\qquad + E_\theta\Big(\tilde{c}_i^r x_i^r(\tilde{\boldsymbol{c}})y_r(\tilde{\boldsymbol{c}})\Big| x_i^r(\tilde{\boldsymbol{c}})y_r(\tilde{\boldsymbol{c}}) = 0\Big) \\ &\qquad \times P_\theta(x_i^r(\tilde{\boldsymbol{c}})y_r(\tilde{\boldsymbol{c}}) = 0)\Big) \\ &= \sum_{r=1}^{R}\sum_{i=1}^{n_r} w_{i1}^r(1). \end{aligned}$$

Furthermore $E_\theta(x_i^r(\tilde{\boldsymbol{c}})y_r(\tilde{\boldsymbol{c}})) = P_\theta(x_i^r(\tilde{\boldsymbol{c}})y_r(\tilde{\boldsymbol{c}}) = 1) = w_{i0}^r(1)$ and $E_\theta(y_r(\tilde{\boldsymbol{c}})) = P_\theta(y_r(\tilde{\boldsymbol{c}}) = 1) = y_r$. Since:

$$\boldsymbol{A^r x^r}(\tilde{\boldsymbol{c}})y_r(\tilde{\boldsymbol{c}}) \leq \boldsymbol{b^r} y_r(\tilde{\boldsymbol{c}}),$$

we can take expectations on both sides to obtain (11c):

$$\boldsymbol{A^r}(w_{10}^r(1), \dots, w_{n_r0}^r(1)) \leq \boldsymbol{b^r} y_r.$$

Similarly, constraints (11b) and (11d)–(11e) follow from the variable definitions and the moment restrictions. Hence, Formulation (11) provides a valid relaxation to compute an upper bound on $Z^*_{\max}$.

We next construct a distribution θ^* that attains the bound $Z^*_{\max}$ either exactly or asymptotically. The construction is based on a two step convex decomposition, first in the variables y_r and then in the variables $\boldsymbol{x^r}$:

(a) Express the optimal $(y_1^*, \dots, y_R^*)$ as a convex combination of the extreme points of the 0-1 feasible region $\mathcal{Y}$ defined by constraints (11d). Particularly, we choose the extreme point $(y_1[q], \dots, y_R[q])$ with probability ψ_q^* for the set of extreme points $q \in \{1, \dots, Q\}$.
(b) Based on the decomposition in (a), for a particular $r \in \{1, \dots, R\}$, we need to consider two possible cases:

Case 1. Suppose $y_r[q] = 1$. Since $(w_{10}^{r*}(1)/y_r^*, \dots, w_{n_r0}^{r*}(1)/y_r^*)$ lies in the convex hull of the set of 0-1 feasible solutions, we can express it as a convex combination of the extreme points of $\mathcal{X}^r$. Particularly, for the set of extreme points $p \in \{1, \dots, P_r\}$, we choose the p^{th} feasible solution with probability $\lambda_p^{r\,*}$. We generate $\tilde{c}_i^r$ as follows:

(a) Choose a feasible solution $p \in \{1, \dots, P_r\}$ to the nominal problem with probability $\lambda_p^{r\,*}$

(b) Generate $\tilde{c}_i^r \sim \boldsymbol{w}_{\boldsymbol{i}}^{\boldsymbol{r}*}(1)/w_{i0}^{r*}(1)$ for $i : x_i[p]=1$ and $\tilde{c}_i \sim \boldsymbol{w}_{\boldsymbol{i}}^{\boldsymbol{r}*}(0)/w_{i0}^{r*}(0)$ for $i : x_i[p]=0$.

Case 2. Suppose $y_r[q]=0$. We generate $\tilde{c}_i^r$ as follows:

(a) Generate $\tilde{c}_i^r \sim \boldsymbol{w}_{\boldsymbol{i}}^{\boldsymbol{r}}(0)/w_{i0}^r(0)$ for all i.

With this two-step construction, the moment vector for $\tilde{c}_i^r$ for each i, r is:

$$
\begin{aligned}
&\sum_{q:y_r[q]=1} \psi_q^* \left\{ \sum_{p:x_i[p]=1} \lambda_p^{r*} \left(\frac{\boldsymbol{w}_{\boldsymbol{i}}^{\boldsymbol{r}*}(1)}{w_{i0}^{r*}(1)} \right) + \sum_{p:x_i[p]=0} \lambda_p^{r*} \left(\frac{\boldsymbol{w}_{\boldsymbol{i}}^{\boldsymbol{r}*}(0)}{w_{i0}^{r*}(0)} \right) \right\} \\
&\quad + \sum_{q:y_r[q]=0} \psi_q^* \left\{ \frac{\boldsymbol{w}_{\boldsymbol{i}}^{\boldsymbol{r}*}(0)}{w_{i0}^{r*}(0)} \right\} \\
&= \sum_{q:y_r[q]=1} \psi_q^* \left\{ \frac{w_{i0}^{r*}(1)}{y_r^*} \left(\frac{\boldsymbol{w}_{\boldsymbol{i}}^{\boldsymbol{r}*}(1)}{w_{i0}^{r*}(1)} \right) + \left(1 - \frac{w_{i0}^{r*}(1)}{y_r^*} \right) \left(\frac{\boldsymbol{w}_{\boldsymbol{i}}^{\boldsymbol{r}*}(0)}{w_{i0}^{r*}(0)} \right) \right\} \\
&\quad + \sum_{q:y_r[q]=0} \psi_q^* \left\{ \frac{\boldsymbol{w}_{\boldsymbol{i}}^{\boldsymbol{r}*}(0)}{w_{i0}^{r*}(0)} \right\} \\
&= \boldsymbol{w}_{\boldsymbol{i}}^{\boldsymbol{r}*}(1) + (y_r^* - w_{i0}^{r*}(1)) \left(\frac{\boldsymbol{w}_{\boldsymbol{i}}^{\boldsymbol{r}*}(0)}{w_{i0}^{r*}(0)} \right) + \left(1 - y_r^* \right) \left(\frac{\boldsymbol{w}_{\boldsymbol{i}}^{\boldsymbol{r}*}(0)}{w_{i0}^{r*}(0)} \right) \\
&= \boldsymbol{w}_{\boldsymbol{i}}^{\boldsymbol{r}*}(1) + \boldsymbol{w}_{\boldsymbol{i}}^{\boldsymbol{r}*}(0) = \boldsymbol{m}_{\boldsymbol{i}}^{\boldsymbol{r}}.
\end{aligned}
$$

The extremal distribution θ^* hence satisfies the marginal moment requirements for each $\tilde{c}_i^r$. Furthermore, we have:

$$
\begin{aligned}
E_{\theta^*}[Z_{\max}(\tilde{\boldsymbol{c}}^{\boldsymbol{1}}, \ldots, \tilde{\boldsymbol{c}}^{\boldsymbol{R}})] &\geq \sum_{q=1}^{Q} \psi_q^* \left\{ \sum_{r:y_r[q]=1} \left\{ \sum_{p=1}^{P_r} \lambda_p^{r*} \left(\sum_{i:x_i[p]=1} \frac{w_{i1}^{r*}(1)}{w_{i0}^{r*}(1)} \right) \right\} \right\} \\
&= \sum_{q=1}^{Q} \psi_q^* \left\{ \sum_{r:y_r[q]=1} \left\{ \sum_{i=1}^{n_r} \frac{w_{i1}^{r*}(1)}{w_{i0}^{r*}(1)} \left(\sum_{p:x_i[p]=1} \lambda_p^{r*} \right) \right\} \right\} \\
&= \sum_{q=1}^{Q} \psi_q^* \left\{ \sum_{r:y_r[q]=1} \left\{ \sum_{i=1}^{n_r} \left(\frac{w_{i1}^{r*}(1)}{y_r^*} \right) \right\} \right\} \\
&= \sum_{r=1}^{R} \left\{ \sum_{q:y_r[q]=1} \psi_q^* \left(\sum_{i=1}^{n_r} \frac{w_{i1}^{r*}(1)}{y_r^*} \right) \right\} \\
&= \sum_{r=1}^{R} \sum_{i=1}^{n_r} w_{i1}^{r*}(1),
\end{aligned}
$$

which proves the desired result. □

5. Application in Project Management

We now study an application of solving the persistency problem in project management. The deterministic problem specifies a directed acyclic graph representation of a project where arcs denote activities. The arc lengths denote time to complete individual activities and the longest path from the start node s to end node t measures the time to complete the project. Formally, the problem is defined as:

Given a directed acyclic graph $G(V \cup \{s, t\}, E)$ with (non-negative) arc lengths $\tilde{c}_{ij}$ for arcs (i, j), find the longest path from s to t.

This problem can be formulated as a linear optimization problem:

$$
\begin{aligned}
Z_{\max}(\tilde{\boldsymbol{c}}) = \max \ & \sum_{(i,j)\in E} \tilde{c}_{ij} x_{ij} \\
\text{s.t.} \ & \sum_{j:(i,j)\in E} x_{ij} - \sum_{j:(j,i)\in E} x_{ji} = \begin{cases} 1, & \text{if } i = s, \\ -1, & \text{if } i = t, \\ 0, & \text{if } i \in V, \end{cases} \\
& x_{ij} \geq 0, \ \forall (i, j) \in E,
\end{aligned}
\tag{12}
$$

and is well known to be solvable in polynomial time. Furthermore, there exist even more efficient algorithms to solve this problem [13]. In the stochastic project management problem, the arc lengths $\tilde{c}_{ij}$ are random. Under uncertainty, we need to identify the persistent (critical) activities for the project manager to focus on.

5.1. Computational Tests

The computations to compare our marginal approach with more traditional project management techniques were carried out on a Windows XP platform on a Pentium IV 2.4 GHz machine. The solver for semidefinite, second order cone and linear optimization in SeDuMi version 1.05 [18] was integrated with MATLAB 6.5 to test the method. The network flow formulation (12) was incorporated into Formulation (6) to solve the marginal moment model.

Example 1. Importance of incorporating variability

We first study the project from Van Slyke [19] under our marginal moments model. As observed earlier, the behavior of this small and seemingly simple project is altered drastically under uncertainty.

We use *Monte Carlo Simulations* (MCS) to analyze the project. Four different distributions with 20000 samples for each were used to simulate the project performance. The first three distributions were multivariate normal $N(\boldsymbol{\mu}, \boldsymbol{Q})$ with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{Q}$. The correlation between all activities were set to zero except for activities 1 and 3 which was varied in $\{-1, 0, 1\}$, to capture some possible dependencies. The fourth distribution simulated was a triangular distribution $Tri(\underline{\boldsymbol{c}}, \boldsymbol{m}, \overline{\boldsymbol{c}})$ specified by three parameters – the minimum ($\underline{\boldsymbol{c}}$), mode ($\boldsymbol{m}$) and maximum value ($\overline{\boldsymbol{c}}$). Under independence, we set these three estimates to (7.2, 7.55, 15.85) for activity 1 and (7.14, 7.2, 15.66) for the remaining activities. All the distributions were chosen to match the known means and variances for the activity durations. Lastly, we use our proposed marginal moments

Table 2. Project statistics for small sized project from Van Slyke [19]

Method	PERT	MCS				MMM
Data	$\boldsymbol{\mu}$	Normal $(\boldsymbol{\mu}, \boldsymbol{Q})$			$Tri(\underline{\boldsymbol{c}}, \boldsymbol{m}, \overline{\boldsymbol{c}})$	(μ_i, Q_{ii})
Dependence	-	$\sigma_{13} = -1$	$\sigma_{13} = 0$	$\sigma_{13} = 1$	Independence	Arbitrary
Expected duration	20.2	23.33	22.98	22.56	23.16	26.30
Persistency (1,2)	1.0	0.325	0.285	0.190	0.260	0.345
Persistency (3)	0.0	0.675	0.715	0.810	0.740	0.655
Persistency (4-8)	0.0	0.135	0.143	0.162	0.148	0.131

model (MMM) with range $[0, \infty)$ and known mean and variance information for the activities. Without assuming the exact distribution, the worst case expected completion time is computed by solving Formulation (6). The complete project statistics with the persistency (criticality indices) of the activities are provided in Table 2. The total computational times were under 5 CPU sec for this problem.

From Table 2, it is clear that persistency is sensitive to the correlation structure as should be expected. In fact, as the degree of dependence among activity durations increases, these variations in persistency could potentially become significant. Nonetheless, our marginal moment model identifies arc 3 as the most critical arc, in agreement with the simulation results obtained for different distributions.

Example 2. Comparison with other worst case approaches
The second example is a larger project taken from Kleindorfer [8]. The project consists of forty activities distributed over fifty one paths. The data for activity durations is provided in Table 3.

The moment based approach for this project is compared with other worst case approaches developed by Meilijson and Nadas [16] and Klein Haneveld [9]. Under complete marginal distribution information (MDM), namely $\tilde{c}_i \sim \theta_i$, but with no assumption on the dependence the problem that they solve is:

$$\sup_{\tilde{c}_i \sim \theta_i \forall i} E\Big(Z_{\max}(\tilde{\boldsymbol{c}})\Big) = \min_{\boldsymbol{d}} \left(Z^*_{\max}(\boldsymbol{d}) + \sum_{i=1}^{n} E_{\theta_i}[\tilde{c}_i - d_i]^+ \right).$$

This problem is the dual to a limiting version of our MMM approach [4], provided the (infinite) moment sequences uniquely characterize the probability distribution under study. To solve this model, successive piecewise linearization has been proposed in [17]. Such an approach is cumbersome due to lack of precision demonstrating an additional advantage of our convex optimization approach.

The MDM approach for this project was solved under a triangular and a two atom discrete distribution for the activity durations. Under the triangular distribution, the term $E_{\theta_i}[\tilde{c}_i - d_i]^+$ is nonlinear and of third degree in variable d_i. After testing, we found twenty linear pieces to be sufficient for each activity for obtaining accurate estimates. The two atom distribution was chosen with the probability of each atom set to 0.5. For the MMM approach, the lower and upper bound on the range of the triangular distribution was used in conjunction with the mean and variance. The project statistics are tabulated in Table 4. Activities that are not mentioned in the table have a persistency of 0. For this project, the CPU times was under 0.5 seconds.

Table 3. Activity duration estimates for Kleindorfer project [8]

Data	Triangular			Discrete		Marginal	
Activity	Min	Mode	Max	Atom 1	Atom 2	Mean	Var
1,40	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	9.00	10.00	11.00	9.591	10.409	10.00	0.167
3,18,27	1.00	2.00	4.00	1.709	2.957	2.333	0.389
4	2.00	5.00	6.00	3.483	5.183	4.333	0.722
5	12.00	12.50	13.00	12.295	12.705	12.50	0.042
6,32,35	1.00	1.50	2.00	1.295	1.705	1.50	0.042
7,9,25	1.00	3.00	4.00	2.043	3.291	2.667	0.389
8,22,34	3.00	4.00	5.00	3.591	4.409	4.00	0.167
10	15.00	17.00	18.00	16.043	17.291	16.667	0.389
11	2.00	18.00	24.00	10.024	19.310	14.667	21.556
12,13,20,31	1.00	2.00	3.00	1.591	2.409	2.00	0.167
14,15,21	4.00	5.00	7.00	4.709	5.957	5.333	0.389
16	5.00	11.00	17.00	8.551	13.449	11.00	6.00
17	1.00	5.00	6.00	2.92	5.080	4.00	1.167
19,23	2.00	3.00	4.00	2.591	3.409	3.00	0.167
24	14.00	14.50	15.00	14.295	14.705	14.50	0.042
26	7.00	19.00	31.00	14.101	23.899	19.00	24.00
27	1.00	2.00	4.00	1.709	2.957	2.333	0.389
28	3.00	4.50	5.00	3.742	4.592	4.167	0.181
29	1.00	8.00	15.00	5.142	10.858	8.00	8.167
30	2.00	4.00	5.00	3.043	4.291	3.667	0.389
33	8.00	10.00	12.00	9.183	10.817	10.00	0.667
36	3.00	7.00	11.00	5.367	8.633	7.00	2.667
37	5.00	10.00	21.00	8.658	15.342	12.00	11.167
38	13.00	13.50	14.00	13.295	13.705	13.50	0.042
39	1.00	12.00	19.00	6.963	14.371	10.667	13.722

Table 4. Project statistics for project from Kleindorfer [8]

Method	PERT	MDM		MMM
Distribution	-	Triangular	Discrete	-
Expected duration	53.667	62.047	63.910	64.344
Persistency (1,40)	1.000	1.000	1.000	1.000
Persistency (3)	1.000	0.910	1.000	0.856
Persistency (2,5,24,38)	0.000	0.090	0.000	0.144
Persistency (10,16,29,37)	0.000	0.362	0.500	0.339
Persistency (11,26,36,39)	1.000	0.548	0.500	0.517

Table 4 indicates that the activities (2, 5, 24, 38) are not identified under the discrete distribution. The specification of θ_i will potentially affect the values of the persistency. However, a complete specification of θ_i is itself normally not available in practice. Furthermore, having to determine the breakpoints for the MDM approach is itself a non-trivial task. In contrast, MMM solves a single tractable convex formulation to compute the persistency values displayed in Figure 2.

We also simulated the project performance under independent triangular distributions. The persistency for the activities is displayed in Figure 3. The computational

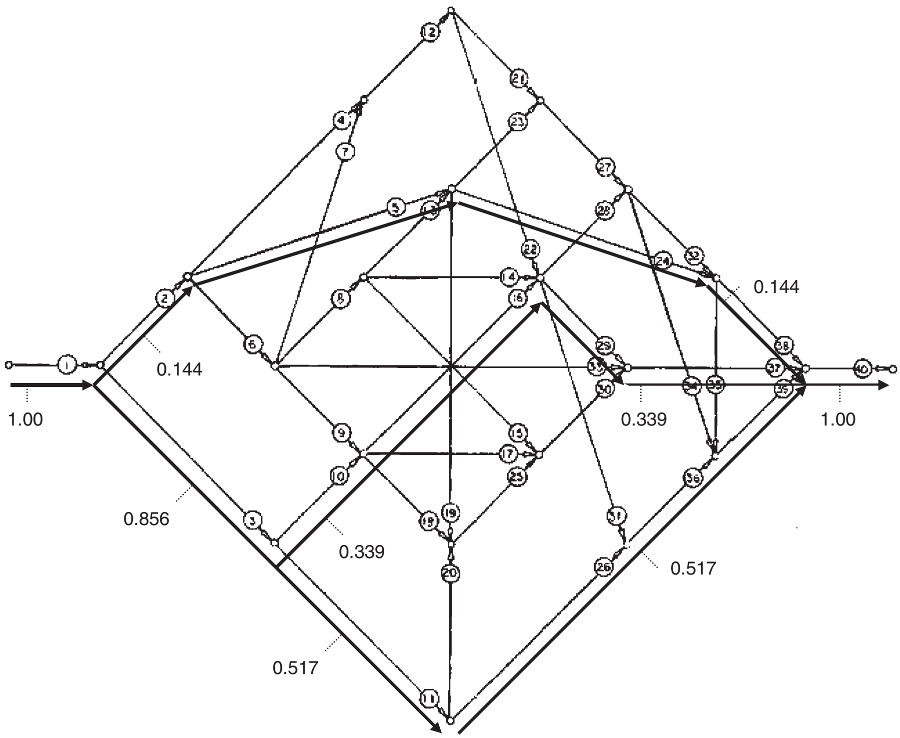

Fig. 2. Persistent activities under MMM for Kleindorfer project

time was 942 CPU sec for solving the linear optimization problem (12) with 20000 samples and the expected completion time was computed to be 56.77. While we obtain more persistent activities under the simulation (20 activities) as compared to the marginal moments approach (15 activities), the probabilities of these extra activities are small (less than 0.08). In this case, under triangular and independently distributed activity durations, the persistency values under simulation is observed to be fairly close in agreement with that of the MMM approach.

6. Application in the Minimum Spanning Tree Problem

In this section, we study the minimum spanning tree (MST) problem under edge length (cost) uncertainty. Formally, the MST problem is defined as:

Given an undirected graph $G(V,E)$ with $|V| = N$ nodes and (non-negative) arc lengths $\tilde{c}_e$ for arcs $e \in E$, find a tree that spans the N nodes with minimum total sum of the cost of edges.

This problem is known to be solvable in polynomial time with the greedy algorithm. In fact, the compact linear representation of the convex hull of spanning trees is explicitly known [14]:

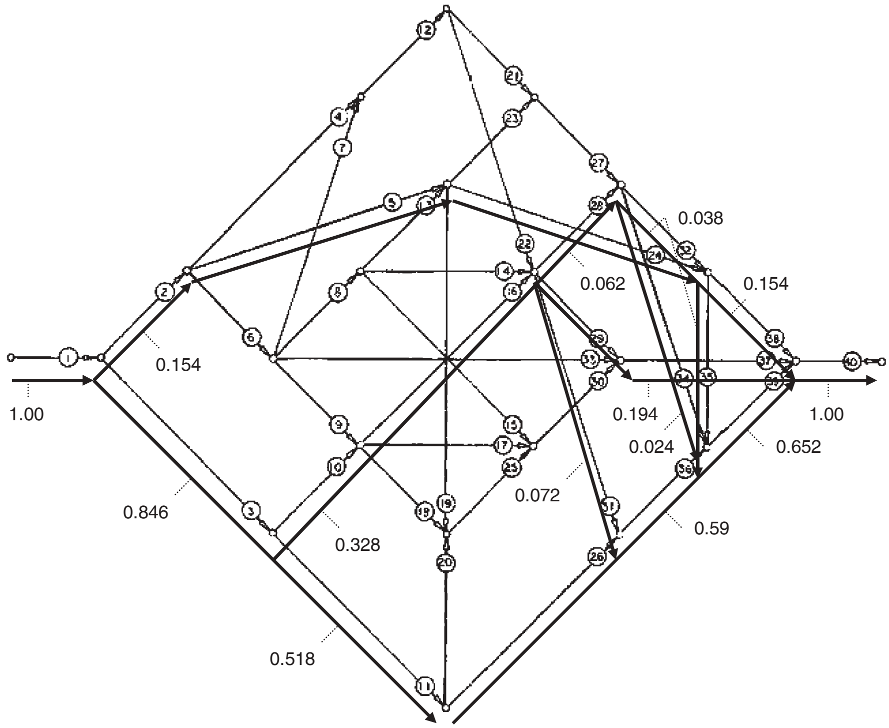

Fig. 3. Persistent activities under triangular distribution and independence

$$
\begin{aligned}
Z_{\min}(\tilde{\boldsymbol{c}}) = \min \sum_{e \in E} \tilde{c}_e (y_{ij} + y_{ji}) \\
\text{s.t.} \quad & \sum_{j:(j,r)\in E'} f^k_{jr} - \sum_{j:(r,j)\in E'} f^k_{rj} = -1, && \forall k \neq r \\
& \sum_{j:(j,v)\in E'} f^k_{jv} - \sum_{j:(v,j)\in E'} f^k_{vj} = 0, && \forall v \neq r, v \neq k, \forall k \\
& \sum_{j:(j,k)\in E'} f^k_{jk} - \sum_{j:(k,j)\in E'} f^k_{kj} = 1, && \forall k \neq r \\
& f^k_{ij} \leq y_{ij}, && \forall (i,j), \forall k \neq r \\
& \sum_{(i,j)\in E'} y_{ij} = N - 1 \\
& y_{ij}, f^k_{ij} \geq 0, && \forall (i,j), \forall k.
\end{aligned}
\tag{13}
$$

This compact representation is obtained from a directed multicommodity flow problem. Here $G(V, E')$ represent the directed version of the original graph obtained by introducing two directed arcs (i, j) and (j, i) for each $e \in E$ that connects nodes i and j. Let $k, k \neq r$ represents a commodity that must be delivered to each node k starting from any predetermined node r. The variables f_{ij}^k denote the directed flow of commodity k on arc (i, j) and y_{ij} denotes the capacity of the flow on arc (i, j) for each commodity k. The sum of the optimal variables $(y_{ij} + y_{ji})$ provide the 0-1 solutions to indicate if an arc connecting i and j lies in the optimal spanning tree. For a complete graph on N nodes, this linear formulation has $O(N^3)$ variables and constraints.

Under interval uncertainty on the edge lengths, Yaman, Karasan and Pinar [20] use this compact representation to solve a mixed integer program to find a robust spanning tree. To reduce the dimension of the problem, they implement a pre-processing routine that identifies edges that will not be a part of the robust spanning tree solution. By solving the mixed integer program on this reduced set of variables, they obtain significant computational savings in solving the robust problem. We now extend this idea to the notion of re-optimization algorithms with a focus on the minimum spanning tree problem.

Consider the setting, where one needs to repeatedly solve instances of the spanning tree problem under perturbations in the objective coefficients. Assume that we obtain K different samples of the objective coefficients denoted as $\boldsymbol{c}^{\mathbf{1}}, \dots, \boldsymbol{c}^{\boldsymbol{K}}$ from a multivariate distribution θ that satisfies some known marginal moment conditions. The generic re-optimization algorithm is specified as:

Algorithm A$(V, E, \boldsymbol{c}^{\mathbf{1}}, \dots, \boldsymbol{c}^{K})$**:**

1. For $k = 1, \dots, K$:
   (a) Solve the MST problem on $G(V, E)$ with objective $\boldsymbol{c}^{\boldsymbol{k}}$.

We now propose a modified re-optimization algorithm that uses the marginal moments model in a preprocessing step. Particularly, consider using the MMM approach to compute the persistence of variables by solving the convex model under limited marginal moment information. We use the compact convex hull representation to solve our marginal moments model. Note that since we are interested in the minimum spanning tree, we compute the persistence of the variables simply by replacing the sup by inf in Formulation (6). We then pick variables with say the L highest persistency values and solve the re-optimization algorithm over this smaller subset of variables. The proposed algorithm is:

Algorithm B $(V, E, \boldsymbol{c}^{\mathbf{1}}, \dots, \boldsymbol{c}^{K}, L)$**:**

1. **Preprocessing**:
   (a) Solve the marginal moments model for $G(V, E)$ under given moment information.
   (b) Identify the subset of variables $E_L \subseteq E$ with L highest persistency values.
2. Call **Algorithm A**$(V, E_L, \boldsymbol{c}^{\mathbf{1}}, \dots, \boldsymbol{c}^{K})$.

The key advantage in using Algorithm B to re-optimize is that the dimension of the optimization problem can often significantly reduced. However this approach is interesting, only if there is a small loss in accuracy in going from the original distribution θ to the marginal moment model.

We now consider a computational experiment where under independently and normally generated cost coefficients, the results from the re-optimization algorithm B are promising with a small loss in accuracy but significant savings in computational times.

6.1. Computational Tests

For testing, we consider the minimum spanning tree problem on two complete graphs with $N = 15$ and 30 vertices respectively. The graphs were generated by choosing N points randomly on the square [0, 10] × [0, 10]. For each arc connecting nodes i and j, the data for the arc lengths are:

- Mean μ_{ij} set to the Euclidean distances between points i and j
- Standard deviation σ_{ij} chosen randomly from $[0, \mu_{ij}/3]$
- Range $[\mu_{ij} - 3\sigma_{ij}, \mu_{ij} + 3\sigma_{ij}]$.

The parameters were chosen such that the probability of a cost coefficient lying outside the range for a normal distribution are negligible (less than 0.003). Under this information, the preprocessing step in Algorithm B identifies 66 arcs with non-zero persistence values (plotted in Figure 4) for the 15 node, 105 arc graph and 191 arcs for the 30 node, 435 arc graphs respectively.

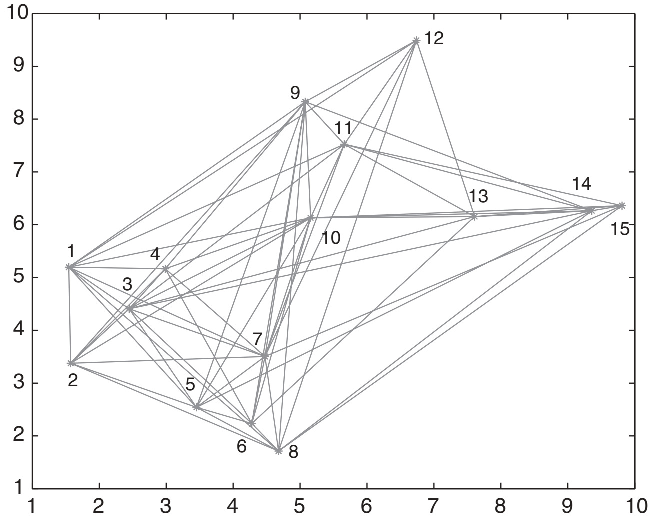

Fig. 4. Persistent arcs under MMM for 15 node graph

To test the re-optimization techniques, $K = 20000$ samples of the cost coefficients were generated independently from normal distributions with the given means and variances. The deterministic MST problems were solved with Prim's greedy algorithm with a binary heap implementation with running time complexity of $O(|E| \log V)$. For the 15 node graph, the results are plotted in Figures 5 and 6. The results for the two graphs are summarized in Table 5. For this example, under normal and independently distributed arc lengths, as the number of arcs L chosen in the preprocessing step decreases, the reduction in computational time is observed to be more significant that the increase in expected objective value. These empirical results suggest the potential of using the marginal moments model in re-optimization algorithms.

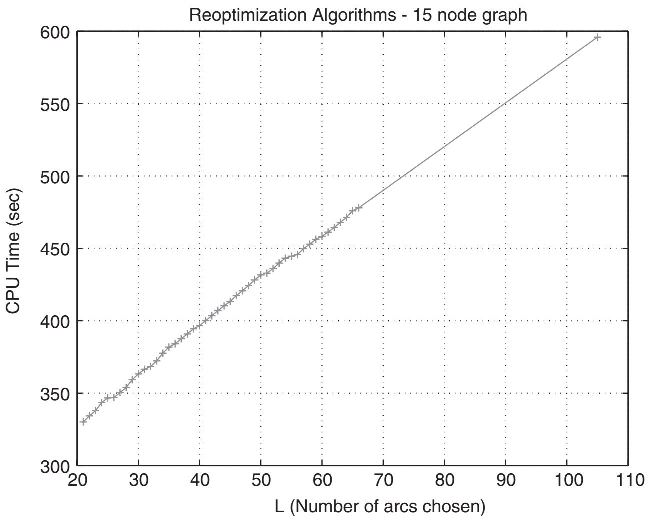

Fig. 5. CPU time for re-optimization algorithms for 15 node graph (L = 105 is Algorithm A)

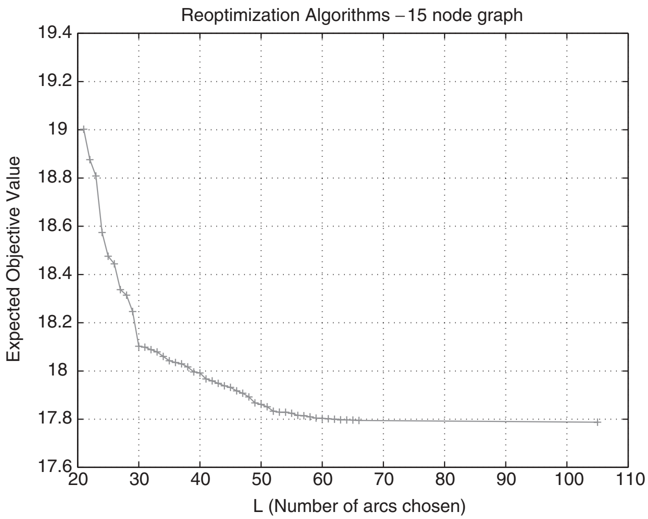

Fig. 6. Objective for re-optimization algorithms for 15 node graph (L = 105 is Algorithm A)

7. Conclusions

In this paper, we have formalized the notion of persistency in discrete optimization problems. While evaluating persistency in general is a difficult problem, we have characterized a marginal moments model under which the persistency of decision variables can be computed using convex optimization techniques. For easily solvable discrete optimization problems, computing this persistence is easy with semi-definite and second order cone optimization techniques. Experimental results and simulation justify the potential of the approach in identifying persistency values.

Table 5. Sample performance of re-optimization algorithms for spanning tree problem

			15 Node Graph		
Algorithm	E_L	CPU sec	Exp. Obj	% Decrease (CPU sec)	% Increase (Exp. Obj)
A	105	595.864	17.788	0.000	0.000
B	66	477.980	17.795	19.784	0.042
B	55	444.582	17.825	25.389	0.208
B	42	403.515	17.959	32.281	0.963
B	21	330.159	19.002	44.592	6.829
			30 Node Graph		
Algorithm	E_L	CPU sec	Exp. Obj	% Decrease (CPU sec)	% Increase (Exp. Obj)
A	435	2142.8	28.652	0.000	0.000
B	191	1522.7	28.823	28.938	0.596
B	146	1376.9	29.258	35.744	2.113
B	86	1174.3	29.888	45.200	4.314
B	37	995.97	32.930	53.521	14.929

Acknowledgements. The authors would like to thank two anonymous referees whose valuable comments helped improve the overall presentation of the paper.

References

1. Atamturk, A.: Strong formulations of robust mixed 0-1 integer programming. To appear in Math. Prog. (2005)
2. Bertsimas, D., Popescu, I.: Optimal inequalities in probability: A convex optimization approach. SIAM J. Optim. **15**, No. 3, 780–804 (2005)
3. Bertsimas, D., Popescu, I.: On the relation between option and stock prices: An optimization approach. Oper. Res. **50**, No. 2, 358–374 (2002)
4. Bertsimas, D., Natarajan K., Chung Piaw Teo.: Tight bounds on expected order statistics. To appear in Probability in Engineering and Informational Sciences (2005)
5. Bertsimas, D., K. Natarajan, Chung Piaw Teo: Tight bounds on expected order statistics. Working Paper (2005)
6. Bertsimas, D., Sim, M.: Robust discrete optimization and network flows. Math. Prog. **98**, No. 1–3, 49–71 (2003)
7. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press (2003)
8. Kleindorfer, G. B.: Bounding distributions for stochastic acyclic networks. Oper. Res. **19**, No. 6, 1586–1601 (1971)
9. Klein Haneveld, W. K.: Robustness against dependence in PERT: An application of duality and distributions with known marginals. Math. Prog. Study **27**, 153–182 (1986)
10. Kojima, M., Kim, S., Waki, H.: A general framework for convex relaxation of polynomial optimization problems over cones. Journal of Operations Research Society of Japan **46**, No. 2, 125–144 (2003)
11. Kouvelis, P., Yu, G.: Robust discrete optimization and its applications. Kluwer Academic Publishers, Norwell, MA (1997)
12. Lasserre, J. B.: Bounds on measures satisfying moment conditions. Ann. Appl. Prob. **12**, 1114–1137 (2002)
13. Lawler, E. L.: Combinatorial Optimization: Networks and Matroids (1976)
14. Magnanti, T. L., Wolsey, L. A.: Optimal Trees. Network Models: Handbooks in Operations Research and Management Science **7**, 503–615 (1995)
15. Malcolm, D. G., Roseboom, J. H., Clark, C. E., Fazar, W.: Application of a technique for research and development program evaluation. Oper. Res. **7**, 646–669 (1959)
16. Meilijson, I., Nadas, A.: Convex majorization with an application to the length of critical path. J. Appl. Prob. **16**, 671–677 (1979)
17. Nadas, A.: Probabilistic PERT. IBM Journal of Research and Development **23**, 339–347 (1979)
18. Sturm, J. F.: SeDuMi v. 1.03, Available from http://fewcal.kub.nl/sturm/software/sedumi.html.
19. Van Slyke, R. M.: Monte Carlo methods and the PERT problem. Oper. Res. **11**, 839–860 (1963)

20. Yaman, H., Karasan, O. E., Pinar, M. C.: The robust spanning tree problem with interval data. Oper. Res. Lett. **29**, 31–40 (1995)
21. Zuluaga, L., Pena, J. F.: A conic programming approach to generalized tchebycheff inequalities. Mathematics of Operations Research **30**, 369–388 (2005)