Available online at www.sciencedirect.com
ScienceDirect
Operations Research Letters 36 (2008) 150–156
Operations Research Letters
www.elsevier.com/locate/orl

Stochastic 0–1 linear programming under limited distributional information

Michael R. Wagner

Department of Management, California State University East Bay, Hayward, CA 94542, USA

Received 12 April 2007; accepted 24 July 2007
Available online 8 September 2007

Abstract

We consider the problem $\min_{\mathbf{x}\in\{0,1\}^n}\{\mathbf{c}'\mathbf{x} : \mathbf{a}_j'\mathbf{x}\leqslant b_j,\ j=1,\ldots,m\}$, where the $\mathbf{a}_j$ are random vectors with unknown distributions. The only information we are given regarding the random vectors $\mathbf{a}_j$ are their moments, up to order k. We give a robust formulation, as a function of k, for the 0–1 integer linear program under this limited distributional information.

Keywords: Stochastic integer programming; Second order cone programming; Semidefinite programming.

1. Introduction

Let $\mathbf{x}=(x_1,x_2,\ldots,x_n)\in\{0,1\}^n$ be a vector of binary variables. A generic 0–1 integer linear program is of the following form: $\min_{\mathbf{x}\in\{0,1\}^n}\{\mathbf{c}'\mathbf{x} : \mathbf{a}_j'\mathbf{x}\leqslant b_j,\ j=1,\ldots,m\}$. We consider a version of this generic problem where the vectors $\mathbf{a}_j,\forall j$ are random. Without loss of generality, the values $b_j,\forall j$ and the vector $\mathbf{c}$ are deterministic and given. A main contribution of our paper is that we consider a limited characterization of the random vectors: We are only given moment information for the random vectors $\mathbf{a}_j,\forall j$ and we do not know their actual distributions. In order to characterize the moment information succinctly, we introduce the following notation, which originally appeared in Bertsimas and Popescu [4]: $J_k=\{\boldsymbol{\kappa}=(k_1,\ldots,k_n) : k_1+\cdots+k_n\leqslant k,\ k_i\in\mathbb{Z}^+,\ i=1,\ldots,n\}$. We are given the following kth order moment knowledge for the random vectors $\mathbf{a}_j,\forall j$: $\mathbb{E}[\mathbf{a}_j^{\boldsymbol{\kappa}}]\triangleq\mathbb{E}[\prod_{i=1}^n a_{ij}^{k_i}],\forall\boldsymbol{\kappa}\in J_k,\ j=1,\ldots,m$, where a_{ij} is the ith element of vector $\mathbf{a}_j$. Let $\sigma_j^k=\{\mathbb{E}[\mathbf{a}_j^{\boldsymbol{\kappa}}] : \boldsymbol{\kappa}\in J_k\}$ denote the collection of all moments up to order k for the random vector $\mathbf{a}_j$. If a distribution $F_{\mathbf{a}_j}(\cdot)$ for $\mathbf{a}_j$ gives moments in agreement with σ_j^k, we denote this as $F_{\mathbf{a}_j}\sim\sigma_j^k$ and say that $F_{\mathbf{a}_j}$ is a *valid* distribution. We assume the vectors $\mathbf{a}_j$ are independent of each other, but we make no assumptions whatsoever about interdependencies between random variables in a given random vector $\mathbf{a}_j$.

Instead of requiring the constraints $\mathbf{a}_j'\mathbf{x}\leqslant b_j,\forall j$ to hold under any realization of the random vectors $\mathbf{a}_j$ under any valid distribution (which results in a *semi-infinite* deterministic problem), we introduce a probabilistic bound on constraint *violations*. We are given values $p_j\in(0,1),\forall j$ such that the probability that $\mathbf{a}_j'\mathbf{x}\leqslant b_j$ is violated is at most p_j. We require that this probabilistic bound holds for any valid distribution for $\mathbf{a}_j$. Therefore, we replace $\mathbf{a}_j'\mathbf{x}\leqslant b_j$ with $\max_{F_{\mathbf{a}_j}\sim\sigma_j^k}\mathbb{P}[\mathbf{a}_j'\mathbf{x}\geqslant b_j]\leqslant p_j$. Our new robust formulation for the 0–1 integer linear program under limited distributional information is the following mathematical program, denoted by $\mathscr{MP}$: $\min_{\mathbf{x}\in\{0,1\}^n}\{\mathbf{c}'\mathbf{x} : \max_{F_{\mathbf{a}_j}\sim\sigma_j^k}\mathbb{P}[\mathbf{a}_j'\mathbf{x}\geqslant b_j]\leqslant p_j,\ j=1,\ldots,m\}$. Let $Z_{\mathscr{MP}}$ denote the optimal value of $\mathscr{MP}$. This model of probabilistically bounded constraint violations is a flexible, tunable approach that can be feasibly applied in practice.

We now argue that defining the deterministic parameters $b_j,\forall j$ and $\mathbf{c}$ a priori is without loss of generality. Note that a deterministic linear objective allows us to also solve $\mathscr{MP}$ when the objective function is instead $\min_{\mathbf{x}}\max_{F_{\mathbf{c}}\sim\sigma_c^k}\mathbb{P}[\mathbf{c}'\mathbf{x}\geqslant v]$, where $v\in\mathbb{R}$ is a given value and F_c and σ_c^k are defined appropriately for a random cost vector $\mathbf{c}$. Furthermore, if any of the b_j parameters are random, we can introduce a new variable x_{n+1} and incorporate b_j into the vector $\mathbf{a}_j$ and the accompanying model of uncertainty.

E-mail address: michael.wagner@csueastbay.edu.

0167-6377/$ - see front matter
doi:10.1016/j.orl.2007.07.003

1.1. Literature review

Stochastic integer programming problems with known distributions have generated interest in the optimization research community for many years and the literature is vast. We reference the interested reader to the website [17], which lists over 250 related journal articles. Additionally, an insightful modern treatment of deterministic (and robust) integer programming can be found in Bertsimas and Weismantel [5].

A popular approach to stochastic integer programming is to consider a *recourse model*. Generically speaking, these models require initial as well as secondary (recourse) decisions. The initial decisions (values for an integer vector) are made when the problem data are uncertain (usually characterized by distributions) and the recourse decision is made once the problem data have been realized. Examples of stochastic integer programming with recourse can be found in Carøe and Schultz [6], Schultz and Tiedemann [14] and in the references therein.

Another approach to stochastic integer programming, more related to our paper, is *chance-constrained* programming. In this model, a distribution of the problem data is given and the probability that a constraint holds must be at least some given minimal value. Examples of chance-constrained integer programs can be found in Beraldi and Ruszczynski [2] (in the context of a probabilistic set-covering problem), Dentcheva, et al. [7] and in the references therein.

However, approaches to stochastic integer programming under limited distributional information have been rather scarce. The few existing examples share the common approach of investigating the optimization of an uncertain cost vector over a fixed feasible region. Meilijson and Nadas [10] consider a project management problem where the marginal distributions for each element of the cost vector are known, but the exact form of the joint multivariate distribution of the cost vector is not known; the authors show that an expected cost can be calculated by solving a convex optimization problem. Weiss [18] studies a similar approach in the context of the maximum flow and shortest path problems. Most related to our paper is the article by Bertsimas et al. [3], which considers generic integer programs with a random cost vector characterized by only its moments. For integer programs that are nominally (no uncertainty in the problem data) solvable in polynomial-time, they develop an approach based on semidefinite programming that calculates a tight bound on the expected optimal cost, in polynomial-time.

Central to our paper is the calculation of optimal probabilistic bounds and we rely on the article by Bertsimas and Popescu [4] for useful results. These authors calculate, utilizing semidefinite programming, optimal bounds for $\mathbb{P}[\mathbf{y} \in S]$, where $\mathbf{y}$ is a random vector characterized only by its moments and S is a set taking many structural forms. A key to their analysis is the connection between nonnegative polynomials and semidefinite programming; these connections are also utilized fruitfully in our paper. Another good reference for the relations between polynomials and semidefinite programming (denoted as linear matrix inequalities) is the article by Lasserre [9].

1.2. Contributions

If we only have first order moment information, we show that $\mathscr{MP}$ is equivalent to a deterministic integer linear program. If we have second order moment knowledge at our disposal, we show that $\mathscr{MP}$ is a second order cone program (SOCP) with integer constraints. We propose a branch-and-bound algorithm that solves $\mathscr{MP}$ exactly in this case. We also provide computational studies that compare the performance of our algorithm with a branch-and-bound algorithm for a 0–1 integer linear program where the problem data are known exactly; i.e., we characterize the additional cost of not having complete knowledge of the problem data, a practically useful result. If we have moment knowledge up to order $k \geqslant 3$, we show that, for any $\varepsilon > 0$, there exists a semidefinite program whose value is within ε of the optimal value of $\mathscr{MP}$. We also show that solving a sequence of semidefinite programs allows us to asymptotically find the value of $\mathscr{MP}$.

Next we position our paper by comparing our results with those in the literature. There has been a large amount of research investigating integer programs under uncertainty in the data (see [17]). The majority of these approaches, however, assume that the distributions of the problem data are known. One of our main contributions is that we consider uncertain problem parameters with *unknown distributions*; the only information we are given are the moments. The one article that we are aware of that takes a similar approach to ours is Bertsimas et al. [3]; their article considers the optimization of a random cost vector, with known moments and unknown distribution, over a fixed feasible region. Our approach is similar, except that we also consider uncertainty in the constraint vectors—in other words, we study the optimization of a random cost vector over a *random feasible region*. In the situation where we have lower order moment information, we focus on practically efficient (e.g., branch-and-bound) rather than theoretically efficient (i.e., polynomial-time) algorithms that are the focus of [3]. Therefore, our approach is valid for a larger class of integer programs, since the theoretically efficient approach in [3] is only appropriate for integer programs whose nominal versions are solvable in polynomial-time. For problems where there is higher order moment information available for the data, our approach is more theoretical and structural in spirit and is similar to the general approach in [3].

Outline: In Sections 2.1 and 2.2 we study $\mathscr{MP}$ when we have first and second order moment information, respectively. In Section 3 we study $\mathscr{MP}$ when we have $k \geqslant 3$ order moment information and in Section 4 we numerically solve $\mathscr{MP}$ under second order moment knowledge and compare our solution with that of the problem under perfect knowledge of the constraint vectors.

2. Lower order moment knowledge

In this section, we study $\mathscr{MP}$ when we only have lower order moment information. In particular, we consider $k \in \{1, 2\}$ and we show that in both cases we are able to solve $\mathscr{MP}$ exactly.

2.1. First order moment knowledge: $k=1$

In this section, we require that $\mathbf{a}_j \geqslant 0$ and $b_j \geqslant 0, \forall j$. We are given the first order moments of the constraint vectors; i.e., we are given $\mathbb{E}[\mathbf{a}_j]$, $j=1,\ldots,m$.

Theorem 2.1. *If $\mathbf{a}_j \geqslant 0$, $b_j \geqslant 0, \forall j$ and we are given $\mathbb{E}[\mathbf{a}_j], \forall j$, then an equivalent formulation for $\mathscr{MP}$ is the following deterministic integer linear program, denoted by $\mathscr{IP}$: $\min_{\mathbf{x}\in\{0,1\}^n}\{\mathbf{c}'\mathbf{x} : \mathbb{E}[\mathbf{a}_j]'\mathbf{x} \leqslant p_j b_j, j=1,\ldots,m\}$.*

Proof. It has been shown (e.g., see [4], Section 5) that the standard Markov Inequality is tight; i.e., $\max_{F_{\mathbf{a}_j}\sim\sigma_j^1} \mathbb{P}[\mathbf{a}_j'\mathbf{x} \geqslant b_j] = \min\left(1, \mathbb{E}[\mathbf{a}_j]'\mathbf{x}/b_j\right)$. Since $p_j < 1, \forall j$, applying this equation to our definition of $\mathscr{MP}$, we obtain $\mathscr{IP}$. □

2.2. Second order moment knowledge: $k=2$

In this section we investigate the case where we have first and second order moment information for the constraint vectors. Specifically, we are given $\mathbb{E}[\mathbf{a}_j^{\boldsymbol{\kappa}}], \forall \boldsymbol{\kappa} \in J_2$, $j=1,\ldots,m$. Given this moment information, it is straightforward to calculate the covariance matrix $\boldsymbol{\Gamma}^j$ for each $\mathbf{a}_j$: $\boldsymbol{\Gamma}_{ik}^j = \mathbb{E}\left[\left(a_{ij}-\mathbb{E}[a_{ij}]\right)\left(a_{kj}-\mathbb{E}[a_{kj}]\right)\right], \forall i,k=1,\ldots,n$, where $\boldsymbol{\Gamma}_{ik}^j$ is the value in the ith row and kth column of $\boldsymbol{\Gamma}^j$ and a_{ij} is the ith element of vector $\mathbf{a}_j$. The expected values and variances for the scalar random variables $\mathbf{a}_j'\mathbf{x}, \forall j$, as functions of $\mathbf{x}$, are $\mathbb{E}[\mathbf{a}_j]'\mathbf{x}$ and $\mathbf{x}'\boldsymbol{\Gamma}^j\mathbf{x}$, respectively.

Theorem 2.2. *If we are given $\mathbb{E}[\mathbf{a}_j^{\boldsymbol{\kappa}}], \forall \boldsymbol{\kappa} \in J_2$, $j=1,\ldots,m$, then an equivalent formulation for $\mathscr{MP}$ is the following, denoted by $\mathscr{CP}$: $\min_{\mathbf{x}\in\{0,1\}^n}\{\mathbf{c}'\mathbf{x} : \sqrt{\mathbf{x}'\boldsymbol{\Gamma}^j\mathbf{x}} \leqslant \sqrt{\frac{p_j}{1-p_j}}\left(b_j - \mathbb{E}[\mathbf{a}_j]'\mathbf{x}\right), j=1,\ldots,m\}$.*

Proof. Consider the following variant of the Chebyshev Inequality: If y is a random variable with mean μ and variance θ, then $\mathbb{P}[y \geqslant (1+\delta)\mu] \leqslant \frac{\theta}{\theta+\mu^2\delta^2}$, where δ is a given constant. It has been shown (e.g., see [4]) that this bound is tight in the sense that there exists a distribution with mean μ and variance θ where $\mathbb{P}[y \geqslant (1+\delta)\mu] = \theta/(\theta+\mu^2\delta^2)$. Consequentially, setting $\delta = b_j/\left(\mathbb{E}[\mathbf{a}_j]'\mathbf{x}\right)-1$, we can write $\max_{F_{\mathbf{a}_j}\sim\sigma_j^2} \mathbb{P}[\mathbf{a}_j'\mathbf{x} \geqslant b_j] = \frac{\mathbf{x}'\boldsymbol{\Gamma}^j\mathbf{x}}{\mathbf{x}'\boldsymbol{\Gamma}^j\mathbf{x}+(\mathbb{E}[\mathbf{a}_j]'\mathbf{x})^2\delta^2}$. Rearranging terms, we obtain $\mathscr{CP}$. □

Note that if we relax the integer constraint, the resulting problem is a SOCP since the covariance matrix $\boldsymbol{\Gamma}^j$ is positive semidefinite. It is well known that SOCPs can be solved in polynomial-time using interior point methods (e.g., see Renegar [12]). Therefore, this observation leads us to propose a simple branch-and-bound algorithm, with a SOCP solver as a subroutine, for solving $\mathscr{CP}$ exactly.

2.2.1. A branch-and-bound algorithm for $\mathscr{CP}$

In this section we assume that $\mathscr{CP}$ is feasible. Let $Z_{\mathscr{CP}}$ denote the value of $\mathscr{CP}$ and $K=\{0,1\}^n$ denote the set of possible integer solutions for $\mathscr{CP}$. If $X=\{\mathbf{x} : x_i \in \{0,1\}, i \in I, \mathbf{x} \in Y\}$, let $R(X)=\{\mathbf{x} : 0 \leqslant x_i \leqslant 1, i \in I, \mathbf{x} \in Y\}$ denote the set that relaxes the binary constraints of X. Define $\mathscr{CP}(X)$ as $\min_{\mathbf{x}\in X}\{\mathbf{c}'\mathbf{x} : \sqrt{\mathbf{x}'\boldsymbol{\Gamma}^j\mathbf{x}} \leqslant \sqrt{\frac{p_j}{1-p_j}}(b_j - \mathbb{E}[\mathbf{a}_j]'\mathbf{x}), j=1,\ldots,m\}$ and let $Z_{\mathscr{CP}(X)}$ denote its value. Note that $Z_{\mathscr{CP}(\{0,1\}^n)} = Z_{\mathscr{CP}}$ and $\mathscr{CP}(R(X))$ is a SOCP for any set $X \subseteq \{0,1\}^n$. The branch-and-bound algorithm BB for solving $\mathscr{CP}$ is as follows:

BB:

- Initialization:
  1. Set the initial queue $Q=\{K\}$.
  2. Find an initial upper bound U for $Z_{\mathscr{CP}}$; e.g., $U=\sum_{i:\mathbb{E}[c_i]\geqslant 0}\mathbb{E}[c_i]$.
- Loop:
  1. If Q is empty, stop. The current value in $\mathbf{x}^*$ is optimal.
  2. Choose the first element X of the queue Q and remove it. If $Z_{\mathscr{CP}(R(X))} \geqslant U$ or if $\mathscr{CP}(R(X))$ is infeasible, go to step 1.
  3. If the solution to $Z_{\mathscr{CP}(R(X))}$ is integer, set $\mathbf{x}^*$ to be this solution and set $U = Z_{\mathscr{CP}(R(X))}$. Go to step 1.
  4. Find the smallest index i such that x_i is not integer. Insert $X_1 = X \cap \{x_i = 0\}$ and $X_2 = X \cap \{x_i = 1\}$ into the queue Q. Go to step 1.

3. Higher order moment knowledge

We now consider the general case where we have up to kth order moment information for the random vectors $\mathbf{a}_j$: $\mathbb{E}[\mathbf{a}_j^{\boldsymbol{\kappa}}], \forall \boldsymbol{\kappa} \in J_k$, $j=1,\ldots,m$. Before we begin, we need some technical preliminaries.

3.1. Nonnegative Polynomials and Semidefinite Programming

We begin this technical section with the notion of a sum-of-squares (SOS) polynomial. A function $f(\mathbf{x})$ is a SOS polynomial if there exist polynomial functions $h_i(\mathbf{x}), i=1,\ldots,N$ (for some finite N) such that $f(\mathbf{x}) = \sum_{i=1}^N h_i^2(\mathbf{x})$.

A function that is a SOS polynomial is clearly nonnegative. For univariate functions, a SOS representation of a function is equivalent to the nonnegativity of the function. However, for multivariate functions, this is not the case: There exist nonnegative functions that do not have a SOS representation (e.g., see Reznick [13]).

We next present several useful results that link nonnegative polynomials with positive semidefinite matrices. We first present results for univariate polynomials and then multivariate polynomials.

Lemma 3.1 (*Bertsimas and Popescu [4]*). *The univariate polynomial $g(x)=\sum_{r=0}^{2k} y_r x^r$ satisfies $g(x) \geqslant 0$ for all $x \in \mathbb{R}$ if and only if there exists a matrix $\mathbf{X}=[x_{ij}]_{i,j=0,\ldots,k}$, such that $y_r = \sum_{i,j:i+j=r} x_{ij}$, $r=0,\ldots,2k$, $\mathbf{X} \succeq 0$.*

Lemma 3.2 (*Bertsimas and Popescu [4]*). *The univariate polynomial $g(x)=\sum_{r=0}^{k} y_r x^r$ satisfies $g(x) \geqslant 0$ for all $x \in [a,\infty)$*

if and only if there exists a matrix $\mathbf{X}=[x_{ij}]_{i,j=0,\ldots,k}$, *such that* $0=\sum_{i,j:i+j=2l-1} x_{ij}$, $l=1,\ldots,k$, $\sum_{r=l}^{k} y_r \binom{r}{l} a^{r-l} = \sum_{i,j:i+j=2l} x_{ij}$, $l=0,\ldots,k$, and $\mathbf{X}\succeq 0$.

For the next lemma, we need additional notation: let $\mathbf{x}_{(d)}$ be a vector of all monomials with degree less than or equal to d of the vector $\mathbf{x}=(x_1,\ldots,x_n)$. For convenience, we use $\mathbf{x}_{(d)}=(1,x_1,\ldots,x_n,x_1x_2,\ldots,x_1x_n,\ldots,x_n^2,\ldots,x_n^d)$. The dimension of $\mathbf{x}_{(d)}$ is $\sum_{i=0}^{d}\binom{n}{i}=\binom{n+d}{d}$.

Lemma 3.3 (*Bertsimas and Popescu [4]*). *The multivariate polynomial* $f(\mathbf{x})$ *of degree* $2d$ *has a SOS decomposition if and only if there exists a positive semidefinite matrix* $\mathbf{Q}$ *for which* $f(x)=\mathbf{x}_{(d)}'\mathbf{Q}\mathbf{x}_{(d)}$.

Finally, we present a useful theorem due to Putinar [11].

Theorem 3.4 (*Putinar [11]*). *Suppose that the set* $K=\{\mathbf{x}\in\mathbb{R}^n \mid g_i(\mathbf{x})\geqslant 0,\ i\in I\}$ *is compact and there exists a polynomial* $h(\mathbf{x})$ *of the form* $h(\mathbf{x})=h_0(\mathbf{x})+\sum_{i\in I} h_i(\mathbf{x})g_i(\mathbf{x})$, *such that* $\{\mathbf{x}\in\mathbb{R}^n \mid h(\mathbf{x})\geqslant 0\}$ *is compact and* $h_i(\mathbf{x})$, $i\in I\cup\{0\}$ *are polynomials that have a SOS representation. Then, for any polynomial* $g(\mathbf{x})$ *that is strictly positive for all* $\mathbf{x}\in K$, *there exist polynomials* $s_i(\mathbf{x})$, $i\in I\cup\{0\}$ *that are SOS polynomials such that* $g(\mathbf{x})=s_0(\mathbf{x})+\sum_{i\in I} s_i(\mathbf{x})g_i(\mathbf{x})$.

Note that this theorem does not provide any information regarding the dimensions of the polynomials $s_i(\mathbf{x})$. However, it has been noted by numerous authors (e.g., [4,9]) that the conditions of Theorem 3.4 are not that restrictive. For example, the conditions hold if (1) the variables are all binary or (2) if there is a single constraint function g_i such that $\{g_i(\mathbf{x})\geqslant 0\}$ is compact or (3) if there is a constraint $g(\mathbf{x})=a^2-\|\mathbf{x}\|^2\geqslant 0$ for a sufficiently large value of a.

We conclude this technical section with a useful characterization of positive semidefinite matrices, which is traditionally known as Sylvester's Criterion.

Definition 3.5. If $\mathbf{X}$ is an n-dimensional symmetric matrix, then for any $S\subseteq\{1,\ldots,n\}$, $S\neq\emptyset$, $\mathbf{X}_S$ is the submatrix of $\mathbf{X}$ that corresponds to elements in the rows and columns indexed by S. $\mathbf{X}_S$ is commonly known as a principal minor of the matrix $\mathbf{X}$.

Lemma 3.6 (*Swamy [16]*). $\mathbf{X}$ *is positive semidefinite if and only if the determinants of all its principal minors are nonnegative*: $\mathbf{X}\succeq 0 \iff \det(\mathbf{X}_S)\geqslant 0$, $\forall S\subseteq\{1,\ldots,n\}$, $S\neq\emptyset$.

Remark 3.7. Note that Sylvester's Criterion is markedly simpler in the characterization of positive definite matrices, since only the *leading* principal minors need to be considered: $X\succ 0 \iff \det(\mathbf{X}_S)>0$, $S=\{1,\ldots,k\}$, $k=1,\ldots,n$.

3.2. An Asymptotically Exact Approximation for $\mathscr{MP}$

For simplicity, we first consider a single constraint and later generalize to the full problem. Consider the simplified problem $\mathscr{S}$ and let $Z_{\mathscr{S}}$ denote its value: $\min_{\mathbf{x}\in\{0,1\}^n}\{\mathbf{c}'\mathbf{x} : \max_{F_{\mathbf{a}}\sim\sigma^k}\mathbb{P}[\mathbf{a}'\mathbf{x}\geqslant b]\leqslant p\}$.

Consider the random variable $\mathbf{a}'\mathbf{x}$. Let $M_r(\mathbf{x})=\mathbb{E}[(\mathbf{a}'\mathbf{x})^r]$, $r=0,\ldots,k$, denote the rth moment of $\mathbf{a}'\mathbf{x}$. Note that $M_r(\mathbf{x})$ is a multivariate polynomial of degree r. Our subsequent analysis is based on the approach in [4] to calculate optimal bounds on $\mathbb{P}[\mathbf{y}\in S]$ where $\mathbf{y}$ is a random vector with given moments and S is a set taking many structural forms; the difference in our approach is that we are analyzing a random variable $\mathbf{a}'\mathbf{x}$ that is a function of the decision variable $\mathbf{x}$ of the optimization problem $\mathscr{S}$ (and $\mathscr{MP}$). We begin with an assumption that allows us to use a strong duality result.

Assumption 3.8. For any $\mathbf{x}\in\{0,1\}^n$, the moment vector $(M_0(\mathbf{x}),\ldots,M_r(\mathbf{x}))$ is in the interior of the set of all possible-moment vectors for $\mathbf{a}'\mathbf{x}$.

Lemma 3.9. *An equivalent formulation for* $\max_{F_{\mathbf{a}}\sim\sigma^k}\mathbb{P}[\mathbf{a}'\mathbf{x}\geqslant b]$ *is the dual* $\mathscr{D}$: $\min_{\mathbf{y}}\{\sum_{r=0}^{k} y_r M_r(\mathbf{x}) : \sum_{r=0}^{k} y_r z^r\geqslant 0,\ \forall z\in\mathbb{R},\ \sum_{r=0}^{k} y_r z^r\geqslant 1,\ \forall z\geqslant b\}$.

Proof. The expression $\max_{F_{\mathbf{a}}\sim\sigma^k}\mathbb{P}[\mathbf{a}'\mathbf{x}\geqslant b]$ is equivalent to $\mathscr{P}$: $\max_{\mu}\{\int_{\mathbb{R}} 1(z\geqslant b)\,\mathrm{d}\mu : \int_{\mathbb{R}} z^r\,\mathrm{d}\mu = M_r(\mathbf{x}),\ r=0,\ldots,k,\ \mu\geqslant 0\ a.e.\}$, where μ is a distribution for $z=\mathbf{a}'\mathbf{x}$. The dual of $\mathscr{P}$ is $\mathscr{D}$: $\min_{\mathbf{y}}\{\sum_{r=0}^{k} y_r M_r(\mathbf{x}) : \sum_{r=0}^{k} y_r z^r\geqslant 0,\ \forall z\in\mathbb{R},\ \sum_{r=0}^{k} y_r z^r\geqslant 1,\ \forall z\geqslant b\}$.

Let the optimal values of the primal and dual problems be denoted as $Z_{\mathscr{P}}$ and $Z_{\mathscr{D}}$, respectively. Assumption 3.8 allows strong duality $Z_{\mathscr{P}}=Z_{\mathscr{D}}$ to hold (e.g., see Shapiro [15]) and the proof is complete. □

Note that the value of any feasible solution for $\mathscr{D}$ is an upper bound for $\max_{F_{\mathbf{a}}\sim\sigma^k}\mathbb{P}[\mathbf{a}'\mathbf{x}\geqslant b]$. Therefore, the following is an equivalent description of $\mathscr{S}$: $\min_{\mathbf{x},\mathbf{y}}\{\mathbf{c}'\mathbf{x} : \sum_{r=0}^{k} y_r M_r(\mathbf{x})\leqslant p,\ \sum_{r=0}^{k} y_r z^r\geqslant 0,\ \forall z\in\mathbb{R},\ \sum_{r=0}^{k} y_r z^r\geqslant 1,\ \forall z\geqslant b,\ \mathbf{x}\in\{0,1\}^n\}$. Note that this representation of $\mathscr{S}$ is a semi-infinite mathematical program (i.e., a finite number of variables and an infinite number of constraints). We next show that $\mathscr{S}$ can be characterized with a finite number of constraints, albeit with an increase in the dimension of the problem.

Theorem 3.10. *There exists a finite-dimensional vector* $\tilde{\mathbf{c}}$ *and a finite number of polynomial functions* $g_i(\cdot)$, $i\in I$ *such that* $\mathscr{S}$ *is equivalent to* $\min_{\mathbf{w}}\{\tilde{\mathbf{c}}'\mathbf{w} : g_i(\mathbf{w})\geqslant 0,\ i\in I\}$.

Proof. Consider the representation of $\mathscr{S}$ that was just derived. For simplicity, we assume k is even. Applying Lemma 3.1, we rewrite the second polynomial constraint as a semidefinite constraint: $y_r=\sum_{i,j:i+j=r} p_{ij}$, $r=0,\ldots,k$, $\mathbf{P}\succeq 0$, where $\mathbf{P}=[p_{ij}]_{i,j=1,\ldots,k/2}$ is a symmetric matrix of dimension $k/2$. Lemma 3.6 states that $\mathbf{P}\succeq 0$ is equivalent to $\det(\mathbf{P}_S)\geqslant 0$, $\forall S\subseteq\{1,\ldots,k/2\}$, $S\neq\emptyset$, where $\mathbf{P}_S$ is the submatrix of $\mathbf{P}$ consisting of the rows and columns indexed by S and $\det(\mathbf{P}_S)$ is its determinant. For any S, $\det(\mathbf{P}_S)$ is a polynomial in the coefficients of the matrix $\mathbf{P}$. Therefore, the second polynomial constraint of $\mathscr{S}$ can be represented as a finite set of inequalities:

$$y_r \geqslant \sum_{i,j:i+j=r} p_{ij}, \quad r=0,\dots,k, \tag{1}$$

$$y_r \leqslant \sum_{i,j:i+j=r} p_{ij}, \quad r=0,\dots,k, \tag{2}$$

$$\det(\mathbf{P}_S) \geqslant 0 \quad \forall S \subseteq \{1,\dots,k/2\}, \quad S \neq \emptyset. \tag{3}$$

Similarly, applying Lemma 3.2, the third constraint in $\mathscr{S}$ is equivalent to

$$0 \geqslant \sum_{i,j:i+j=2l-1} q_{ij}, \quad l=1,\dots,k, \tag{4}$$

$$0 \leqslant \sum_{i,j:i+j=2l-1} q_{ij}, \quad l=1,\dots,k, \tag{5}$$

$$\sum_{r=l}^{k} y_r \binom{r}{l} b^{r-l} \geqslant \sum_{i,j:i+j=2l} q_{ij}, \quad l=1,\dots,k, \tag{6}$$

$$\sum_{r=l}^{k} y_r \binom{r}{l} b^{r-l} \leqslant \sum_{i,j:i+j=2l} q_{ij}, \quad l=1,\dots,k, \tag{7}$$

$$\mathbf{Q} \succeq 0, \tag{8}$$

where $\mathbf{Q} = [q_{ij}]_{i,j=1,\dots,k/2}$ is a symmetric matrix of dimension $k/2$. Again applying Lemma 3.6, Eq. (8) can be written as another finite set of inequalities

$$\det(\mathbf{Q}_S) \geqslant 0 \quad \forall S \subseteq \{1,\dots,k/2\}, \quad S \neq \emptyset, \tag{9}$$

where $\mathbf{Q}_S$ are the submatrices of $\mathbf{Q}$. Finally, notice that

$$\{0,1\}^n = \{\mathbf{x} \in \mathbb{R}^n | x_i^2 - x_i \leqslant 0, x_i - x_i^{mathrm2} \geqslant 0\}, \tag{10}$$

a finite set of inequalities. Combining Eqs. (1)–(7), (9), and (10) and the first polynomial constraint of $\mathscr{S}$, we obtain an equivalent description of the feasible region of $\mathscr{S}$ using a finite set of polynomial inequalities. Letting $\mathbf{p}$ (resp. $\mathbf{q}$) denote the vector of coefficients of matrix $\mathbf{P}$ (resp. $\mathbf{Q}$), defining the aggregate variable $\mathbf{w} = (\mathbf{x}, \mathbf{y}, \mathbf{p}, \mathbf{q})$ and $\tilde{\mathbf{c}} = (\mathbf{c}, 0, 0, 0)$ completes the proof. Note that the dimension of $\mathbf{w}$ is $k^2/2 + k + n + 1$. □

Let $K = \{\mathbf{w} : g_i(\mathbf{w}) \geqslant 0, i \in I\}$. Note that Theorem 3.10 gives the following equivalent description of $\mathscr{S}$:

$$Z_{\mathscr{S}} = \min_{\mathbf{w} \in K} \tilde{\mathbf{c}}'\mathbf{w}. \tag{11}$$

Theorem 3.11. *If K satisfies the conditions of Theorem 3.4, then for any $\varepsilon > 0$, there exists a semidefinite program whose value is greater than or equal to $Z_{\mathscr{S}} - \varepsilon$.*

Proof. Note that Eq. (11) implies

$$\tilde{\mathbf{c}}'\mathbf{w} - Z_{\mathscr{S}} \geqslant 0 \quad \forall \mathbf{w} \in K, \tag{12}$$

with equality at the optimal value of $\mathbf{w}$. Fix $\varepsilon > 0$ and let $z = Z_{\mathscr{S}} - \varepsilon$. Consider the following strict inequality version of Eq. (12):

$$\tilde{\mathbf{c}}'\mathbf{w} - z > 0 \quad \forall \mathbf{w} \in K. \tag{13}$$

Theorem 3.4 states that there exist an integer d, that depends on z and ε, and polynomials $s_i(\mathbf{w}), i \in I \cup \{0\}$ of degree $2d$ that are SOS such that

$$\tilde{\mathbf{c}}'\mathbf{w} - z = s_0(\mathbf{w}) + \sum_{i \in I} s_i(\mathbf{w}) g_i(\mathbf{w}). \tag{14}$$

Let l denote the dimension of $\mathbf{w}$ and let $\beta = (\beta_1,\dots,\beta_l)$ and $\mathbf{w}^\beta = w_1^{\beta_1} \cdots w_l^{\beta_l}$. The highest dimensional polynomial inequality defining K is the first constraint in $\mathscr{S}$, which has dimension k. Therefore, for each $i \in I$, there exists a sequence $\{\gamma_\beta^i\}$ such that $g_i(\mathbf{w}) = \sum_{\beta \in J_k} \gamma_\beta^i \mathbf{w}^\beta, i \in I$, where J_k is redefined as $J_k = \{(k_1,\dots,k_l) : k_1 + \cdots + k_l \leqslant k, k_i \in \mathbb{Z}^+, i = 1,\dots,l\}$. Similarly, there exist sequences $\{s_\beta^i\}, i \in I \cup \{0\}$ such that $s_i(\mathbf{w}) = \sum_{\beta \in J_{2d}} s_\beta^i \mathbf{w}^\beta, i \in I \cup \{0\}$. Eq. (14) can be written as

$$\begin{aligned}\tilde{\mathbf{c}}'\mathbf{w} - z &= \sum_{\beta \in J_{2d}} s_\beta^0 \mathbf{w}^\beta + \sum_{i \in I} \left(\sum_{\beta \in J_{2d}} s_\beta^i \mathbf{w}^\beta \right) \left(\sum_{\alpha \in J_k} \gamma_\alpha^i \mathbf{w}^\alpha \right) \\ &= \sum_{\beta \in J_{2d}} s_\beta^0 \mathbf{w}^\beta + \sum_{i \in I} \sum_{\substack{\beta \in J_{2d} \\ \alpha \in J_k}} s_\beta^i \gamma_\alpha^i \mathbf{w}^{\beta+\alpha}.\end{aligned} \tag{15}$$

Recall that the first n elements of $\tilde{\mathbf{c}}$ correspond to $\mathbf{c}$ (and $\mathbf{x}$) and the remaining coefficients are zero. We equate the coefficients of both sides of Eq. (15) to obtain the following constraints.

$$c_i = s_\beta^0 + \sum_{i \in I} (s_\beta^i \gamma_0^i + s_0^i \gamma_\beta^i),$$

$$\beta_i = 1, \quad \beta_j = 0, \quad j \neq i, \quad i = 1,\dots,n;$$

$$0 = s_\beta^0 + \sum_{i \in I} \sum_{\substack{\delta \in J_{2d} \\ \alpha \in J_k \\ \delta + \alpha = \beta}} s_\delta^i \gamma_\alpha^i,$$

$$\forall \beta \in J_{2d+k} \backslash \{\beta_i = 1, \beta_j = 0, j \neq i, i = 1,\dots,n, \beta = 0\};$$

$$-z = s_0^0 + \sum_{i \in I} s_0^i \gamma_0^i.$$

Next, we express the SOS polynomials $s_i(\cdot)$ as semidefinite constraints, utilizing Lemma 3.3: There exist positive semidefinite matrices $\mathbf{S}^i \succeq 0, i \in I \cup \{0\}$ such that $s_i(\mathbf{x}) = \mathbf{x}'_{(d)} \mathbf{S}^i \mathbf{x}_{(d)}$.

Writing $\mathbf{S}^i = [s_{\delta,\alpha}^i]_{\delta,\alpha \in J_d}$, the following constraints are clear:

$$s_\beta^i = \sum_{\substack{\delta,\alpha \in J_d \\ \delta + \alpha = \beta}} s_{\delta,\alpha}^i \quad \forall \beta \in J_{2d}, \quad i \in I \cup \{0\}.$$

Therefore, the following semidefinite program finds a SOS representation of Eq. (13) and a value $Z(d)$ that is within ε of the optimal solution $Z_{\mathscr{S}}$.

$$
\begin{aligned}
Z(d) \triangleq \max_{\mathbf{S}^i, z}\ & z \\
\text{s.t.}\quad & c_i = s^0_\beta + \sum_{i \in I} (s^i_\beta \gamma^i_0 + s^i_0 \gamma^i_\beta), \\
& \beta_i = 1,\ \ \beta_j = 0,\ \ j \neq i, i = 1, \ldots, n; \\
& 0 = s^0_\beta + \sum_{i \in I} \sum_{\substack{\delta \in J_{2d} \\ \alpha \in J_k \\ \delta + \alpha = \beta}} s^i_\delta \gamma^i_\alpha, \\
& \forall \beta \in J_{2d+k} \backslash \{\beta_i = 1, \beta_j = 0, \\
& j \neq i, i = 1, \ldots, n, \beta = 0\}; \\
& -z = s^0_0 + \sum_{i \in I} s^i_0 \gamma^i_0; \\
& s^i_\beta = \sum_{\substack{\delta, \alpha \in J_d \\ \delta + \alpha = \beta}} s^i_{\delta, \alpha}, \\
& \forall \beta \in J_{2d},\ \ i \in I \cup \{0\}; \\
& \mathbf{S}^i \succeq 0, \quad i \in I \cup \{0\}. \qquad \square
\end{aligned}
$$

Note that Theorem 3.11 only proves the existence of a semidefinite program for finding an ε-approximate solution to $\mathscr{S}$. However, for any fixed value of d, the semidefinite program gives a lower bound for $Z_{\mathscr{S}}$. As d increases, the quality of the lower bound increases as well. Asymptotically, we obtain the solution value $Z_{\mathscr{S}}$. This observation is formalized in the following corollary.

Corollary 3.12. *The semidefinite programs with values* $Z(d)$ *asymptotically approach* $Z_{\mathscr{S}}$; *i.e.*, $\lim_{d \to \infty} Z(d) \uparrow Z_{\mathscr{S}}$.

The analysis in this section for the single constraint in $\mathscr{S}$ can simply be repeated for each constraint of $\mathscr{MP}$. Therefore, letting $K_{\mathscr{MP}}$ denote the feasibility set for $\mathscr{MP}$ that corresponds to K in this section, we have the following results.

Theorem 3.13. *If* $K_{\mathscr{MP}}$ *satisfies the conditions of Theorem* 3.4, *then for any* $\varepsilon > 0$, *there exists a semidefinite program whose value is greater than or equal to* $Z_{\mathscr{MP}} - \varepsilon$.

Corollary 3.14. *The semidefinite programs* (*appropriated generalized for* $\mathscr{MP}$) *with values* $Z(d)$ *asymptotically approach* $Z_{\mathscr{MP}}$; *i.e.*, $\lim_{d \to \infty} Z(d) \uparrow Z_{\mathscr{MP}}$.

4. Computational study

In this section we implement the branch-and-bound algorithm `BB`, given in Section 2.2.1, for solving $\mathscr{CP}$ ($\mathscr{MP}$ under second order moment knowledge) exactly. We compare its performance with a standard branch-and-bound algorithm for solving the underlying 0–1 integer linear program with known data. In particular, we examine how well our robust formulation approximates the actual underlying integer linear program.

We implement `BB` in MATLAB [1]. More specifically, we modify the MATLAB `bintprog` (a branch-and-bound algorithm for solving 0–1 linear programs) procedure to utilize the Disciplined Convex Programming [8] package to solve the SOCP relaxations. We compare our algorithm's performance with that of `bintprog` on the underlying 0–1 integer linear program where the problem data are known exactly. Our implementation of `BB` utilizes the default options for `bintprog`: The branching strategy is to choose the variable with the maximum infeasibility (i.e., the variable with value closest to 0.5) and the node search strategy is to choose the node with the lowest bound on the objective function.

4.1. Experimental design

Consider fixed values of m and n. For each constraint $j = 1, \ldots, m$ of $\mathscr{CP}$, we randomly generate the covariance matrix $\boldsymbol{\Gamma}^j$ of the constraint vector $\mathbf{a}_j$ in the following manner: $\boldsymbol{\Gamma}^j = \boldsymbol{\Lambda}'_j \boldsymbol{\Lambda}_j$, where $\boldsymbol{\Lambda}_j$ is an $n \times n$ matrix whose elements are i.i.d. realizations of the standard normal distribution; this design results in a positive semidefinite matrix $\boldsymbol{\Gamma}^j$, which is necessary for a covariance matrix. Likewise, the elements of the mean vector $\mathbb{E}[\mathbf{a}_j]$ are i.i.d. realizations of the standard normal distribution. The scalar value b_j is an i.i.d. realization of a normal distribution with a mean of 50 and a variance of 1; we utilize this specific normal distribution to decrease the frequency of infeasible integer programs. We fix $p_j = p, \forall j$ and we consider $p \in \{0.01, 0.05\}$. Finally, the elements of the cost vector $\mathbf{c}$ are also i.i.d. realizations of the standard normal distribution.

The parameters $\mathbf{c}$, $\boldsymbol{\Gamma}^j$, $\mathbb{E}[\mathbf{a}_j]$, b_j and p_j, $j = 1, \ldots, m$ are the problem data for $\mathscr{CP}$. We compare the performance of $\mathscr{CP}$ with that of a 0–1 integer linear program with known data. To create this data, we let the constraint vector $\mathbf{a}_j$ be a realization of a multivariate normal distribution with covariance matrix $\boldsymbol{\Gamma}^j$ and mean vector $\mathbb{E}[\mathbf{a}_j]$. The parameters $\mathbf{c}$, $\mathbf{a}_j$ and b_j, $j = 1, \ldots, m$ are used to solve the deterministic 0–1 integer linear program that underlies $\mathscr{CP}$.

We consider the following combinations of (m, n): (5,2), (10,5) and (25,10). For each combination of (m, n) we repeat the above procedure 50 times and report the average objective values of `BB` and `bintprog`. We also report the average runtimes of `BB` and `bintprog`.

4.2. Results

In this section, we present the results of our computational study.

Our computational results indicate that the objective value output by our robust formulation is a rather tight approximation for the underlying 0–1 integer linear program, for the smaller problem instances and at a higher computational cost. Therefore, if extra computational power is available, requiring exact knowledge of the constraint vectors is not absolutely necessary

Table 1
Comparison of BB with bintprog for $p = 0.01$

(m, n)	BB objective	BB runtime (s)	bintprog Objective	bintprog Runtime (s)
(5, 2)	−0.9491	0.3344	−1.0526	0.0191
(10, 5)	−0.8882	1.9325	−1.9307	0.0213
(25, 10)	0.0000	14.3731	−4.2250	0.0288

All values reported are the average of 50 trials and are accurate to four decimal places.

Table 2
Comparison of BB with bintprog for $p = 0.05$

(m, n)	BB objective	BB runtime (s)	bintprog Objective	bintprog Runtime (s)
(5, 2)	−0.9727	0.2916	−0.9790	0.0156
(10, 5)	−1.9432	1.2941	−2.0725	0.0194
(25, 10)	−1.5077	25.2953	−3.9960	0.0269

All values reported are the average of 50 trials and are accurate to four decimal places.

for smaller instances. Our results suggest that, practically, statistical estimates of the constraint vectors can provide a good approximation of the underlying 0–1 integer linear program. This observation is important since, in practice, increasing computational power is generally less costly than obtaining more accurate information about the constraint vectors. Our study also suggests that the approximating power of our robust formulation deteriorates as the problem size increases. We also note that the approximating performance of BB improves as p, the probability that a constraint is violated, is increased. Finally, note that in Table 1, for $m = 25$ and $n = 10$, it seems that the experimental design resulted in feasible regions where the zero vector was the only feasible solution. In Table 2, we see that by relaxing the constraints (i.e., increasing p), we obtain non-trivial solutions for $m = 25$ and $n = 10$.

Acknowledgements

We would like to thank Joy Bhadury for discussions that generated ideas that developed into this research project. We would also like to thank Ioana Popescu for reading an early draft and providing constructive feedback. We also thank the anonymous referee for thoughtful comments, which improved the quality and clarity of our paper. Finally, this research was made possible, in part, by financial support from the Office of Research and Sponsored Programs at California State University East Bay.

References

[1] ⟨http://www.mathworks.com/⟩.
[2] P. Beraldi, A. Ruszczynski, The probabilistic set covering problem, Oper. Res. 50 (2002) 956–967.
[3] D. Bertsimas, K. Natarajan, C. Teo, Probabilistic combinatorial optimization: moments, semidefinite programming, and asymptotic bounds, SIAM J. Optim. 15 (1) (2004) 185–209.
[4] D. Bertsimas, I. Popescu, Optimal inequalities in probability theory: a convex optimization approach, SIAM J. Optim. 15 (3) (2005) 780–804.
[5] D. Bertsimas, R. Weismantel, Optimization over Integers, Dynamic Ideas, 2005.
[6] C. Carøe, R. Schultz, Dual decomposition in stochastic integer programming, Oper. Res. Lett. 24 (1999) 37–45.
[7] D. Dentcheva, A. Prekopa, A. Ruszczynski, Bounds for probabilistic integer programming problems, Discrete Appl. Math. 124 (2002) 55–65.
[8] M. Grant, S. Boyd, Y. Ye, Cvx: Matlab software for disciplined convex programming, ⟨http://www.stanford.edu/ ∼boyd/cvx/⟩.
[9] J. Lasserre, Global optimization with polynomials and the problem of moments, SIAM J. Optim. 11 (3) (2001) 796–817.
[10] I. Meilijson, A. Nadas, Convex majorization with an application to the length of critical path, J Appl. Probab. 16 (1979) 671–677.
[11] M. Putinar, Positive polynomials on compact semi-algebraic sets, Indiana Univ. Math. J. 42 (1993) 969–984.
[12] J. Renegar, A Mathematical View of Interior-Point Methods in Convex Optimization, MPS-SIAM Series on Optimization, 2001.
[13] B. Reznick, Some concrete aspects of hilbert's 17th problem, in: Real Algebraic Geometry and Ordered Structures, Contemp. Math., vol. 253, AMS, Providence, RI, 2000, pp. 251–272.
[14] R. Schultz, S. Tiedemann, Conditional value-at-risk in stochastic programs with mixed-integer recourse, Math. Programming 105 (2006) 365–386.
[15] A. Shapiro, On duality theory of conic linear programs, in: Semi-Infinite Programming: Recent Advances, Kluwer Academic Publishers, Dordrecht, 2001, pp. 135–165.
[16] K. Swamy, On sylvester's criterion for positive-semidefinite matrices, IEEE Trans. Autom. Control 18 (1973) 306.
[17] M.H. van der Vlerk, Stochastic integer programming bibliography, World Wide Web, ⟨http://mally.eco.rug.nl/biblio/stoprog.html⟩, 1996–2003.
[18] G. Weiss, Stochastic bounds on distributions of optimal value functions with applications to pert, network flows and reliability, Ope. Res. 34 (1986) 595–605.