

Mixed 0-1 Linear Programs Under Objective Uncertainty: A Completely Positive Representation

Karthik Natarajan

Department of Management Sciences, City University of Hong Kong, Hong Kong, knataraj@cityu.edu.hk

Chung Piaw Teo, Zhichao Zheng

Department of Decision Sciences, NUS Business School, National University of Singapore, Singapore 117591
{bizteocp@nus.edu.sg, zhichao@nus.edu.sg}

In this paper, we analyze mixed 0-1 linear programs under objective uncertainty. The mean vector and the second-moment matrix of the nonnegative objective coefficients are assumed to be known, but the exact form of the distribution is unknown. Our main result shows that computing a tight upper bound on the expected value of a mixed 0-1 linear program in maximization form with random objective is a completely positive program. This naturally leads to semidefinite programming relaxations that are solvable in polynomial time but provide weaker bounds. The result can be extended to deal with uncertainty in the moments and more complicated objective functions. Examples from order statistics and project networks highlight the applications of the model. Our belief is that the model will open an interesting direction for future research in discrete and linear optimization under uncertainty.

Subject classifications: mixed 0-1 linear program; moments; completely positive program.

Area of review: Optimization.

History: Received August 2009; revision received March 2010; accepted June 2010.

1. Introduction

One of the fundamental problems in mixed 0-1 linear programs under uncertainty is to compute the expected optimal objective value. Consider the random optimization problem,

$$\begin{aligned} Z(\tilde{\mathbf{c}}) = \max \quad & \tilde{\mathbf{c}}^T \mathbf{x} \\ \text{s.t.} \quad & \mathbf{a}_i^T \mathbf{x} = b_i \quad \forall i = 1, \dots, m \\ & \mathbf{x} \geq \mathbf{0} \\ & x_j \in \{0, 1\} \quad \forall j \in \mathcal{B} \subseteq \{1, \dots, n\}, \end{aligned} \quad (1)$$

where $\mathbf{x} \in \mathbb{R}_+^n$ is the decision vector and $\tilde{\mathbf{c}}$ is the random objective coefficient vector. The subset $\mathcal{B} \subseteq \{1, \dots, n\}$ indexes the 0-1 decision variables and $\{1, \dots, n\} \setminus \mathcal{B}$ indexes the continuous decision variables. Problem (1) includes the class of 0-1 integer programs and the class of linear programs as special cases. Given distributional information on $\tilde{\mathbf{c}}$, the object of interest is the expected optimal value $\mathbf{E}[Z(\tilde{\mathbf{c}})]$.

This problem has been extensively studied in *network reliability* applications. Network reliability deals with the design and analysis of networks that are subject to random variation in the components. Such network applications arise in, for example, telecommunication, transportation, and power systems. The random weights $\tilde{\mathbf{c}}$ on the edges in the network represent random lengths, capacities, or durations. For designated source node s and destination node t ,

popular reliability measures include the shortest $s-t$ path length, the longest $s-t$ path length in a directed acyclic graph, and the maximum $s-t$ flow. The goal is to compute properties of the network reliability measure, such as the average value or the probability distribution of $Z(\tilde{\mathbf{c}})$. For an excellent review on applications and algorithms for the network reliability analysis problem, the reader is referred to Ball et al. (1995) and the references therein. Under the assumption of independence among the random weights, Hagstrom (1988) showed that computing the expected value of the longest path in a directed acyclic graph is $\#\mathcal{P}$ -complete, when the arc lengths are restricted to taking two possible values each. The expected longest path is not computable in time polynomial in the size of the input unless $\mathcal{P} = \mathcal{NP}$. The $\#\mathcal{P}$ -hardness results for other network reliability measures are discussed in Valiant (1979) and Provan and Ball (1983). Methods developed include identification of efficient algorithms for special cases, enumerative methods, bounding methods, and Monte Carlo methods. For the shortest-path problem with exponentially distributed arc lengths, Kulkarni (1986) developed a Markov-chain based method to compute the expected shortest path. The running time of this algorithm is non-polynomial in the size of the network. Assuming independence and each arc length \tilde{c}_{ij} is exponentially distributed

with mean μ_{ij} , Lyons et al. (1999) developed a lower bound using a convex quadratic optimization problem,

$$\mathbf{E}[Z(\tilde{\mathbf{c}})] \geq \min \left\{ \sum_{(i,j) \in E} \mu_{ij} x_{ij}^2 : \mathbf{x} \in \mathcal{X} \right\},$$

where \mathcal{X} denotes the (s, t) -path polytope. For shortest-path problems on complete graphs with n vertices and independent and exponentially distributed arc lengths with means μ , Davis and Prieditis (1993) proved the following exact result,

$$\mathbf{E}[Z_n(\tilde{\mathbf{c}})] = \frac{\mu}{(n-1)} \sum_{k=1}^{n-1} \frac{1}{k}.$$

Similar formulas and asymptotic expressions have been developed for other random optimization problems including the spanning tree (Frieze 1985), assignment (Aldous 2001, Mézard and Parisi 1987, Linusson and Wästlund 2004, Nair et al. 2006), traveling salesman (Wästlund 2010), and Steiner tree problem (Bollobás et al. 2004). In general, when the deterministic problem is itself \mathcal{NP} -hard, computing the expected optimal value is even more challenging. It is then natural to develop polynomial-time computable bounds.

One of the fundamental assumptions underlying most of the network reliability literature is that the probability distributions for the random weights are known. In this paper, we adopt the *distributional robustness* approach where information on only a few moments of the random coefficients are assumed to be known. The bound computed is distributionally robust, i.e., it is valid across the set of distributions satisfying the given moment information. Such a “moment”-based approach has become a popular technique to find bounds in optimization problems (Bertsimas et al. 2010, 2004, 2006; Bertsimas and Popescu 2005; Calafiore and Ghaoui 2006; Delage and Ye 2010; Lasserre 2002; Mishra et al. 2008; Natarajan et al. 2009; Popescu 2007; Vandenbergh et al. 2007). Links with conic programming have made the approach attractive from a theoretical and computational perspective (Bertsimas and Popescu 2005, Nesterov 2000, Todd 2001). In addition, conic programs provide additional insights into the structure of the optimal solution. One such parameter of importance is the *persistence* of a binary variable, which is defined as the probability that the variable takes value 1 in an optimal solution (Bertsimas et al. 2006). The notion of persistence generalizes “criticality index” in project networks and “choice probability” in discrete choice models (Bertsimas et al. 2006, Mishra et al. 2008, Natarajan et al. 2009).

In this paper, we develop moment-based bounds for mixed 0-1 linear programs. Suppose $\tilde{\mathbf{c}}$ is nonnegative with known first-moment vector $\boldsymbol{\mu} = \mathbf{E}[\tilde{\mathbf{c}}]$ and second-moment matrix $\boldsymbol{\Sigma} = \mathbf{E}[\tilde{\mathbf{c}}\tilde{\mathbf{c}}^T]$. The central problem we solve is

$$\sup_{\tilde{\mathbf{c}} \sim (\boldsymbol{\mu}, \boldsymbol{\Sigma})^+} \mathbf{E}[Z(\tilde{\mathbf{c}})],$$

where $\tilde{\mathbf{c}} \sim (\boldsymbol{\mu}, \boldsymbol{\Sigma})^+$ denotes the set of feasible multivariate distributions supported on \mathbb{R}_+^n with first-moment vector $\boldsymbol{\mu}$ and second-moment matrix $\boldsymbol{\Sigma}$. In the situation without the support requirement, the corresponding problem with sup replaced by inf in the model reduces to a simple Jensen bound (Bertsimas et al. 2010). Our results naturally extend to lower bounds on the expected optimal objective value of a mixed 0-1 linear program in minimization form with random objective. For the longest-path problem arising in project networks, the bound corresponds to the worst-case expected project completion time. For the maximum flow problem, the bound corresponds to the worst-case expected flow supported by the network. In the shortest-path context, this is a lower bound along the lines of the Lyons et al. bound (Lyons et al. 1999), but valid over a larger set of distributions.

Structure of the Paper. In §2, we review several existing moment models that are based on semidefinite programs, followed by a discussion on completely positive programs. Detailed descriptions are provided in Appendices I and II. In §3, we develop a completely positive program to compute the bound. The persistency of the variables under an extremal distribution are obtained from the optimal solution to the completely positive program. In §4, we provide some important extensions to our model. In §5, we present applications of our model in order statistics and project management with computational results. We conclude in §6.

Notations and Definitions. Throughout this paper, we use small letters to denote scalars, bold letters to denote vectors, and capital letters to denote matrices. Random terms are denoted using the tilde notation. The trace of a matrix A , denoted by $\text{tr}(A)$, is sum of the diagonal entries of A . The inner product between two matrices of appropriate dimensions A and B is denoted as $A \cdot B = \text{tr}(A^T B)$. I_n is used to represent the identity matrix of dimension $n \times n$. For any convex cone \mathcal{H} , the dual cone is denoted as \mathcal{H}^* , and the closure of the cone is denoted as $\overline{\mathcal{H}}$. \mathcal{S}_n denotes the cone of $n \times n$ symmetric matrices, and \mathcal{S}_n^+ denotes the cone of $n \times n$ positive semidefinite matrices,

$$\mathcal{S}_n^+ := \{A \in \mathcal{S}_n \mid \forall \mathbf{v} \in \mathbb{R}^n, \mathbf{v}^T A \mathbf{v} \geq 0\}.$$

$A \succeq 0$ indicates that the matrix A is positive semidefinite and $B \succeq A$ indicates $B - A \succeq 0$. Two cones of special interest are the cone of completely positive matrices and the cone of copositive matrices. The cone of $n \times n$ *completely positive matrices* is defined as

$$\mathcal{CP}_n := \{A \in \mathcal{S}_n \mid \exists V \in \mathbb{R}_+^{n \times k}, \text{ such that } A = VV^T\},$$

or equivalently,

$$\mathcal{CP}_n := \left\{ A \in \mathcal{S}_n \mid \exists \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k \in \mathbb{R}_+^n, \right. \\ \left. \text{such that } A = \sum_{i=1}^k \mathbf{v}_i \mathbf{v}_i^T \right\}.$$

The above is called the *rank 1 representation* of the completely positive matrix A . The cone of $n \times n$ *copositive matrices* is defined as

$$\mathcal{C}\mathcal{O}_n := \{A \in \mathcal{S}_n \mid \forall \mathbf{v} \in \mathbb{R}_+^n, \mathbf{v}^T A \mathbf{v} \geq 0\}.$$

$A \succeq_{cp}$ (\succeq_{co}) 0 indicates that the matrix A is completely positive (copositive).

2. Literature Review

2.1. Related Moment Models

Over the last couple of decades, research in semidefinite programming (SDP) has experienced an explosive growth (Todd 2001). Besides the development of theoretically efficient algorithms, the modeling power of SDP has made it a highly attractive tool for optimization problems. The focus in this section is on SDP-based moment models related to our problem of interest. The explicit formulations of these models are provided in Appendix I, which is part of the electronic companion that is part of the online version at <http://or.journal.informs.org/>.

Marginal Moment Model (MMM). Under the MMM (Bertsimas et al. 2004, 2006), information on $\tilde{\mathbf{c}}$ is described only through marginal moments of each \tilde{c}_j . No explicit assumption on independence or the dependence structure of the coefficients is made. Whereas an arbitrary set of marginal moments can be specified in MMM, we restrict our attention to the first two moments. Suppose for each nonnegative coefficient \tilde{c}_j , the mean μ_j , and second moment Σ_{jj} is known. Under the MMM, the bound is computed over all joint distributions with the specified marginal moments, i.e., solving

$$\sup_{\tilde{c}_j \sim (\mu_j, \Sigma_{jj})^+, \forall j=1, \dots, n} \mathbf{E}[Z(\tilde{\mathbf{c}})].$$

For 0-1 integer programs, Bertsimas et al. (2004, 2006) showed that this bound can be computed in polynomial time if the deterministic problem is solvable in polynomial time. Using SDP, they developed a computational approach to compute the bound and the persistency values under an extremal distribution. When the objective coefficients are generated independently, they observed that the qualitative insights in the persistency estimates obtained from MMM are similar to the simulation results. However, it is conceivable that because the dependence structure is not captured, the bounds and persistency estimates need not always be good. In addition, the results are mainly useful for polynomial-time solvable 0-1 integer programs where the linear constraints characterizing the convex hull are explicitly known. Natarajan et al. (2009) extended the MMM to general integer programs and linear programs. Their formulation is based on a characterization of the convex hull of the binary reformulation, which is difficult to do typically.

Cross-Moment Model (CMM). Under the CMM, information on $\tilde{\mathbf{c}}$ is described through both the marginal and the cross moments. Suppose the mean vector and second-moment matrix of the objective coefficients is known. Mishra et al. (2008) computed the upper bound on the expected maximum of n random variables where the support is in \mathbb{R}^n ,

$$\sup_{\tilde{\mathbf{c}} \sim (\boldsymbol{\mu}, \boldsymbol{\Sigma})} \mathbf{E}[\max(\tilde{c}_1, \tilde{c}_2, \dots, \tilde{c}_n)].$$

The SDP formulation developed therein is based on an extreme point enumeration technique. Bertsimas et al. (2010) showed that generalizing this model to general linear programs leads to \mathcal{NP} -hard problems. In a related vein, Lasserre (2010) developed a hierarchy of semidefinite relaxations that uses higher-order moment information to solve parametric polynomial optimization problems.

Generalized Chebyshev Bounds. In a related vein, Vandenberghe et al. (2007) used SDP to bound the probability that a random vector lies within a set defined by several strict quadratic inequalities. For a given set $C \subseteq \mathbb{R}^n$ defined by

$$C = \{\mathbf{c} \in \mathbb{R}^n \mid \mathbf{c}^T A_i \mathbf{c} + 2\mathbf{b}_i^T \mathbf{c} + d_i < 0, \forall i = 1, \dots, m\},$$

they computed the tight lower bound on the probability that $\tilde{\mathbf{c}}$ lies in the set C ,

$$\inf_{\tilde{\mathbf{c}} \sim (\boldsymbol{\mu}, \boldsymbol{\Sigma})} \mathbf{P}(\tilde{\mathbf{c}} \in C).$$

We refer to this as the VBC approach. For linear programs with random objective, the VBC approach can be used to bound the probability that a particular basis is optimal. This follows from the optimality conditions for linear programming, which is a set of linear inequalities in $\tilde{\mathbf{c}}$. For other multivariate generalizations of Chebyshev's inequality, the reader is referred to Bertsimas and Popescu (2005), Isii (1959), Lasserre (2002), Vandenberghe et al. (2007), Zuluaga and Pena (2005).

2.2. Completely Positive Programs and \mathcal{NP} -Hard Problems

One of the shortcomings of the existing SDP-based moment models is the lack of bounds for general mixed 0-1 linear programs under cross-moment information. Our goal is to develop a parsimonious model that can cover this important class of problems while capturing first- and second-moment conditions. The approach is based on recent results that show that several \mathcal{NP} -hard optimization problems can be expressed as the linear programs over the convex cone of the copositive matrices. This is called a *copositive program* (COP) (Bomze et al. 2000, Burer 2009, de Klerk and Pasechnik 2002). Each COP is associated with a dual problem over the convex cone of completely positive matrices. Such a program is called a *completely positive program*

(CPP) (Berman and Shaked-Monderer 2003). A review on COP and CPP is provided in Appendix II.

Burer (2009) recently showed that the nonconvex quadratic programs with a mixture of binary and continuous variables can be expressed as CPPs. Under an easily enforceable condition on the feasible region, he proved the equivalence of the following two formulations,

$$\begin{aligned} \max \quad & \mathbf{x}^T Q \mathbf{x} + 2\mathbf{c}^T \mathbf{x} \\ \text{s.t.} \quad & \mathbf{a}_i^T \mathbf{x} = b_i \quad \forall i = 1, \dots, m \\ & \mathbf{x} \geq \mathbf{0} \\ & x_j \in \{0, 1\} \quad \forall j \in \mathcal{B} \\ \max \quad & Q \cdot X + 2\mathbf{c}^T \mathbf{x} \\ \text{s.t.} \quad & \mathbf{a}_i^T \mathbf{x} = b_i \quad \forall i = 1, \dots, m \\ & \mathbf{a}_i^T X \mathbf{a}_i = b_i^2 \quad \forall i = 1, \dots, m \\ = \quad & X_{jj} = x_j \quad \forall j \in \mathcal{B} \\ & \begin{pmatrix} 1 & \mathbf{x}^T \\ \mathbf{x} & X \end{pmatrix} \succeq_{cp} \mathbf{0}. \end{aligned}$$

Expressing the problem as a COP or CPP does not resolve the difficulty of the problem, because capturing these cones is generally suspected to be difficult (Bomze et al. 2000, Burer 2009, Dür and Still 2007). For instance, the problem of testing if a given matrix is copositive is known to be in the co- \mathcal{NP} -complete class (Murty and Kabadi 1987). However, such a transformation shifts the difficulty into the respective convex cones, so that whenever something more is known about copositive or completely positive matrices, it can be applied uniformly to many \mathcal{NP} -hard problems (Burer 2009). One such result is a well-known hierarchy of linear and semidefinite representable cones that approximate the copositive and completely positive cone (Bomze et al. 2000, Parrilo 2000, de Klerk and Pasechnik 2002). For the numerical experiments, we restrict our attention to the following simple relaxation of CPP,

$$A \succeq_{cp} \mathbf{0} \implies A \succeq \mathbf{0}, \quad A \geq \mathbf{0}.$$

We exploit the power of COP and CPP to develop the general result.

3. The Cross-Moment Model for Mixed 0-1 Linear Programs

3.1. Problem Notations and Assumptions

Denote the linear portion of the feasible region in (1) as

$$\mathcal{L} := \{\mathbf{x} \geq \mathbf{0} \mid \mathbf{a}_i^T \mathbf{x} = b_i, \quad \forall i = 1, \dots, m\},$$

and the entire feasible region as $\mathcal{L} \cap \{0, 1\}^{\mathcal{B}}$. The problem of interest is

$$(P) \quad Z_p = \sup_{\tilde{\mathbf{c}} \sim (\boldsymbol{\mu}, \Sigma)^+} \mathbf{E} \left[\max_{\mathbf{x} \in \mathcal{L} \cap \{0, 1\}^{\mathcal{B}}} \tilde{\mathbf{c}}^T \mathbf{x} \right].$$

The key assumptions under which the problem is analyzed are discussed next.

Assumptions. (A1) The set of distributions of $\tilde{\mathbf{c}}$ is defined by the nonnegative support \mathbb{R}_+^n , finite mean vector $\boldsymbol{\mu}$, and finite second-moment matrix Σ . This set is assumed to be nonempty.

$$(A2) \quad \mathbf{x} \in \mathcal{L} \implies x_j \leq 1, \quad \forall j \in \mathcal{B}.$$

(A3) The feasible region $\mathcal{L} \cap \{0, 1\}^{\mathcal{B}}$ is nonempty and bounded.

The nonnegativity of $\tilde{\mathbf{c}}$ in Assumption (A1) is guaranteed when the objective denotes price, time, length, or demand. Checking for the existence of multivariate distributions with nonnegative support satisfying a given mean and second-moment matrix is, however, a difficult problem (Bertsimas and Popescu 2005, Kemperman and Skibinsky 1993, Murty and Kabadi 1987). For the case when the first two moments are calculated from empirical distributions or from common multivariate distributions, (A1) is verifiable by construction. In general, to characterize the feasibility of the first and second moments in a support $\Omega \subseteq \mathbb{R}^n$, the moment cone is defined as

$$\mathcal{M}_2(\Omega) = \{\lambda(1, \boldsymbol{\mu}, \Sigma) \mid \lambda \geq 0, \boldsymbol{\mu} = \mathbf{E}[\tilde{\mathbf{c}}], \Sigma = \mathbf{E}[\tilde{\mathbf{c}}\tilde{\mathbf{c}}^T]\}$$

for some random vector $\tilde{\mathbf{c}}$ with support Ω .

From the theory of moments (Karlin and Studden 1966, Kemperman and Skibinsky 1993), it is well known that the dual of this moment cone is given as

$$\mathcal{M}_2(\Omega)^* = \{(w_0, \mathbf{w}, \mathbf{W}) \mid w_0 + \mathbf{w}^T \mathbf{c} + \mathbf{c}^T \mathbf{W} \mathbf{c} \geq 0 \text{ for all } \mathbf{c} \in \Omega\}.$$

Then, the dual of the dual of the moment cone is simply the closure of the moment cone, i.e.,

$$\overline{\mathcal{M}_2(\Omega)} = (\mathcal{M}_2(\Omega)^*)^*.$$

For $\Omega = \mathbb{R}_+^n$, the dual of the moment cone is equivalently the cone of copositive matrices and the closure of the moment cone is thus the cone of completely positive matrices. Testing for (A1) is a difficult problem because

$$(1, \boldsymbol{\mu}, \Sigma) \in \overline{\mathcal{M}_2(\mathbb{R}_+^n)} \iff \begin{pmatrix} 1 & \boldsymbol{\mu}^T \\ \boldsymbol{\mu} & \Sigma \end{pmatrix} \succeq_{cp} \mathbf{0}.$$

Assumption (A2) is easy to enforce and is based on Burer’s paper (Burer 2009). If $\mathcal{B} = \emptyset$, then the assumption is vacuous. For problems such as the longest-path problem on a directed acyclic graph, (A2) is implied from the network flow constraints. When $\mathcal{B} \neq \emptyset$ and the assumption is not implied in the constraints, one can add the constraints $x_j + s_j = 1$ and $s_j \geq 0$.

Assumption (A3) ensures that $\mathbf{E}[Z(\tilde{\mathbf{c}})]$ is finite, and hence the supremum is finite.

3.2. Formulation

Denote $x_j(\mathbf{c})$ to be the value of the variable x_j in an optimal solution to Problem (1) obtained under the specific \mathbf{c} . When $\tilde{\mathbf{c}}$ is random, $\mathbf{x}(\tilde{\mathbf{c}})$ is also random. For continuous distributions, the support of $\tilde{\mathbf{c}}$ over which Problem (1) has multiple optimal solutions has measure zero. For discrete distributions with possibly multiple optimal solutions in a support of strictly positive measure, we define $\mathbf{x}(\mathbf{c})$ to be an optimal solution randomly selected from the set of optimal solutions at \mathbf{c} . Next, we define

$$\begin{aligned} \mathbf{p} &:= \mathbf{E}[\mathbf{x}(\tilde{\mathbf{c}})], \\ Y &:= \mathbf{E}[\mathbf{x}(\tilde{\mathbf{c}})\tilde{\mathbf{c}}^T], \\ X &:= \mathbf{E}[\mathbf{x}(\tilde{\mathbf{c}})\mathbf{x}(\tilde{\mathbf{c}})^T]. \end{aligned}$$

Note that the matrix X is symmetric, but Y is not. Then

$$\begin{aligned} \mathbf{E}[Z(\tilde{\mathbf{c}})] &= \mathbf{E}\left[\sum_{j=1}^n \tilde{c}_j x_j(\tilde{\mathbf{c}})\right] \\ &= \sum_{j=1}^n Y_{jj} \\ &= \mathbf{I}_n \cdot Y. \end{aligned}$$

Define the vector $\mathbf{y}(\tilde{\mathbf{c}})$ as:

$$\mathbf{y}(\tilde{\mathbf{c}}) = \begin{pmatrix} 1 \\ \tilde{\mathbf{c}} \\ \mathbf{x}(\tilde{\mathbf{c}}) \end{pmatrix}.$$

Then,

$$\begin{aligned} \mathbf{E}[\mathbf{y}(\tilde{\mathbf{c}})\mathbf{y}(\tilde{\mathbf{c}})^T] &= \begin{pmatrix} 1 & \mathbf{E}[\tilde{\mathbf{c}}^T] & \mathbf{E}[\mathbf{x}(\tilde{\mathbf{c}})^T] \\ \mathbf{E}[\tilde{\mathbf{c}}] & \mathbf{E}[\tilde{\mathbf{c}}\tilde{\mathbf{c}}^T] & \mathbf{E}[\tilde{\mathbf{c}}\mathbf{x}(\tilde{\mathbf{c}})^T] \\ \mathbf{E}[\mathbf{x}(\tilde{\mathbf{c}})] & \mathbf{E}[\mathbf{x}(\tilde{\mathbf{c}})\tilde{\mathbf{c}}^T] & \mathbf{E}[\mathbf{x}(\tilde{\mathbf{c}})\mathbf{x}(\tilde{\mathbf{c}})^T] \end{pmatrix} \\ &= \begin{pmatrix} 1 & \boldsymbol{\mu}^T & \mathbf{p}^T \\ \boldsymbol{\mu} & \Sigma & Y^T \\ \mathbf{p} & Y & X \end{pmatrix}. \end{aligned}$$

Because $\tilde{\mathbf{c}} \geq \mathbf{0}$ and $\mathbf{x}(\tilde{\mathbf{c}}) \geq \mathbf{0}$, $\mathbf{y}(\tilde{\mathbf{c}})$ is a nonnegative vector. Hence, $\mathbf{y}(\tilde{\mathbf{c}})\mathbf{y}(\tilde{\mathbf{c}})^T$ is a completely positive matrix. Because the set of all completely positive matrices is convex, by taking the expectation over all the possibilities of $\tilde{\mathbf{c}}$, $\mathbf{E}[\mathbf{y}(\tilde{\mathbf{c}})\mathbf{y}(\tilde{\mathbf{c}})^T]$ is a completely positive matrix.

Because $\mathbf{a}_i^T \mathbf{x}(\tilde{\mathbf{c}}) = b_i$ for all realizations of $\tilde{\mathbf{c}}$, by taking the expectations, we get

$$\mathbf{a}_i^T \mathbf{p} = b_i \quad \forall i = 1, \dots, m.$$

Using a lifting technique, we obtain

$$\begin{aligned} b_i^2 &= \mathbf{a}_i^T \mathbf{x}(\tilde{\mathbf{c}})(\mathbf{a}_i^T \mathbf{x}(\tilde{\mathbf{c}})) \\ &= \mathbf{a}_i^T \mathbf{x}(\tilde{\mathbf{c}})(\mathbf{a}_i^T \mathbf{x}(\tilde{\mathbf{c}}))^T \\ &= \mathbf{a}_i^T (\mathbf{x}(\tilde{\mathbf{c}})\mathbf{x}(\tilde{\mathbf{c}})^T) \mathbf{a}_i. \end{aligned}$$

Taking expectations again,

$$\mathbf{a}_i^T X \mathbf{a}_i = b_i^2 \quad \forall i = 1, \dots, m.$$

In addition, $\forall j \in \mathcal{B}$, $x_j(\tilde{\mathbf{c}}) = x_j(\tilde{\mathbf{c}})^2$, and hence

$$\begin{aligned} X_{jj} &= \mathbf{E}[x_j(\tilde{\mathbf{c}})^2] \\ &= \mathbf{E}[x_j(\tilde{\mathbf{c}})] \\ &= p_j. \end{aligned}$$

By considering \mathbf{p} , Y , and X as the decision variables, we construct a completely positive program relaxation to (P) as follows,

$$\begin{aligned} \text{(C)} \quad Z_C &= \max \quad \mathbf{I}_n \cdot Y \\ \text{s.t.} \quad \mathbf{a}_i^T \mathbf{p} &= b_i \quad \forall i = 1, \dots, m \\ \mathbf{a}_i^T X \mathbf{a}_i &= b_i^2 \quad \forall i = 1, \dots, m \\ X_{jj} &= p_j \quad \forall j \in \mathcal{B} \subseteq \{1, \dots, n\} \\ \begin{pmatrix} 1 & \boldsymbol{\mu}^T & \mathbf{p}^T \\ \boldsymbol{\mu} & \Sigma & Y^T \\ \mathbf{p} & Y & X \end{pmatrix} &\succeq_{cp} \mathbf{0}. \end{aligned}$$

Note that from Assumption (A3), the variables \mathbf{p} and X are bounded. Moreover, each Y_{ij} is bounded by the positive semidefiniteness of the 2×2 matrix,

$$\begin{pmatrix} \Sigma_{ii} & Y_{ij} \\ Y_{ij} & X_{jj} \end{pmatrix}.$$

Hence, we can use “max” instead of “sup” in (C).

Because the model is based on a completely positive program, we refer to it as the *Completely Positive Cross-Moment Model* (CPCMM).

From the construction of the model, it is clear that $Z_P \leq Z_C$. We next show that (C) is not merely a relaxation of (P), rather it solves (P) exactly.

3.3. Tightness

To show that (C) and (P) are equivalent, we construct a sequence of distributions that satisfies the moment constraints in the limit and achieves the bound. Before proving the result, we review some important properties of the solutions to (C), which have been demonstrated by Burer (2009). For completeness, we outline his relevant proofs in Propositions 3.1 and 3.2. It should be noted that because the feasible region is bounded in our setting, the recession cone only contains the zero vector.

Define

$$F := \{(\mathbf{p}, X) \mid \exists Y \text{ such that } (\mathbf{p}, Y, X) \text{ is feasible to (C)}\}.$$

Let $(\mathbf{p}, X) \in F$, and consider any completely positive decomposition

$$\begin{pmatrix} 1 & \mathbf{p}^T \\ \mathbf{p} & X \end{pmatrix} = \sum_{k \in \mathcal{K}} \begin{pmatrix} \zeta_k \\ \mathbf{z}_k \end{pmatrix} \begin{pmatrix} \zeta_k \\ \mathbf{z}_k \end{pmatrix}^T \quad (2)$$

where $\zeta_k \in \mathbb{R}_+$, $\mathbf{z}_k \in \mathbb{R}_+^n$, $\forall k \in \mathcal{K}$.

PROPOSITION 3.1 (BURER 2009). *For the decomposition (2), define $\mathcal{H}_+ := \{k \in \mathcal{H} \mid \zeta_k > 0\}$, and $\mathcal{H}_0 := \{k \in \mathcal{H} \mid \zeta_k = 0\}$. Then (i) $\mathbf{z}_k/\zeta_k \in \mathcal{L}, \forall k \in \mathcal{H}_+$; (ii) $\mathbf{z}_k = \mathbf{0}, \forall k \in \mathcal{H}_0$.*

PROOF. From the decomposition, we have

$$\mathbf{p} = \sum_{k \in \mathcal{H}} \zeta_k \mathbf{z}_k, \quad \text{and} \quad X = \sum_{k \in \mathcal{H}} \mathbf{z}_k \mathbf{z}_k^T.$$

Then

$$\begin{aligned} \mathbf{a}_i^T \mathbf{p} = b_i &\Rightarrow \sum_{k \in \mathcal{H}} \zeta_k (\mathbf{a}_i^T \mathbf{z}_k) = b_i, \\ \mathbf{a}_i^T X \mathbf{a}_i = b_i^2 &\Rightarrow \sum_{k \in \mathcal{H}} (\mathbf{a}_i^T \mathbf{z}_k)^2 = b_i^2. \end{aligned}$$

From $\sum_{k \in \mathcal{H}} \zeta_k^2 = 1$, we get

$$\left(\sum_{k \in \mathcal{H}} \zeta_k (\mathbf{a}_i^T \mathbf{z}_k) \right)^2 = \left(\sum_{k \in \mathcal{H}} \zeta_k^2 \right) \left(\sum_{k \in \mathcal{H}} (\mathbf{a}_i^T \mathbf{z}_k)^2 \right).$$

By the equality conditions of Cauchy-Schwartz inequality,

$$\exists \delta_i, \text{ such that } \delta_i \zeta_k = \mathbf{a}_i^T \mathbf{z}_k, \quad \forall k \in \mathcal{H}, \forall i = 1, \dots, m.$$

Because $\forall k \in \mathcal{H}_0, \zeta_k = 0$, we have $\mathbf{a}_i^T \mathbf{z}_k = 0$. (A3) implies that $\mathbf{z}_k = \mathbf{0}, \forall k \in \mathcal{H}_0$. Thus, (ii) holds. Furthermore,

$$b_i = \sum_{k \in \mathcal{H}} \zeta_k (\mathbf{a}_i^T \mathbf{z}_k) = \sum_{k \in \mathcal{H}} \zeta_k (\delta_i \zeta_k) = \delta_i \sum_{k \in \mathcal{H}} \zeta_k^2 = \delta_i.$$

Because $\forall k \in \mathcal{H}_+, \zeta_k > 0$, we get $\mathbf{a}_i^T (\mathbf{z}_k/\zeta_k) = \delta_i = b_i$, so by the definition of $\mathcal{L}, \mathbf{z}_k/\zeta_k \in \mathcal{L}, \forall k \in \mathcal{H}_+$. Therefore, (i) holds. \square

Taking $\lambda_k := \zeta_k^2, \mathbf{v}_k := \mathbf{z}_k/\zeta_k, \forall k \in \mathcal{H}_+$, we can rewrite the decomposition (2) as

$$\begin{pmatrix} 1 & \mathbf{p}^T \\ \mathbf{p} & X \end{pmatrix} = \sum_{k \in \mathcal{H}_+} \lambda_k \begin{pmatrix} 1 \\ \mathbf{v}_k \end{pmatrix} \begin{pmatrix} 1 \\ \mathbf{v}_k \end{pmatrix}^T, \quad (3)$$

where $\lambda_k > 0, \forall k \in \mathcal{H}_+, \sum_{k \in \mathcal{H}_+} \lambda_k = 1$, and $\mathbf{v}_k \in \mathcal{L}, \forall k \in \mathcal{H}_+$.

PROPOSITION 3.2 (BURER 2009). *Consider the decomposition (3). Let $\mathbf{v}_k = (v_{k(1)}, \dots, v_{k(n)})^T$, then $v_{k(j)} \in \{0, 1\}, \forall j \in \mathcal{B}, \forall k \in \mathcal{H}_+$.*

PROOF. From the decomposition, we have

$$\mathbf{p} = \sum_{k \in \mathcal{H}_+} \lambda_k \mathbf{v}_k, \quad \text{and} \quad X = \sum_{k \in \mathcal{H}_+} \lambda_k \mathbf{v}_k \mathbf{v}_k^T.$$

Fix any $j \in \mathcal{B}$. By Assumption (A2), we have

$$\mathbf{v}_k \in \mathcal{L} \Rightarrow 0 \leq v_{k(j)} \leq 1, \quad \forall k \in \mathcal{H}_+.$$

Then $v_{k(j)}^2 \leq v_{k(j)}, \forall k \in \mathcal{H}_+$.

$$\begin{aligned} X_{jj} = p_j &\Rightarrow \sum_{k \in \mathcal{H}_+} \lambda_k v_{k(j)}^2 = \sum_{k \in \mathcal{H}_+} \lambda_k v_{k(j)} \\ &\Rightarrow \sum_{k \in \mathcal{H}_+} \lambda_k (v_{k(j)} - v_{k(j)}^2) = 0. \end{aligned}$$

Because $\lambda_k > 0$, and $v_{k(j)} - v_{k(j)}^2 \geq 0, \forall k \in \mathcal{H}_+$, we get $v_{k(j)} - v_{k(j)}^2 = 0, \forall k \in \mathcal{H}_+$. Thus, $v_{k(j)} = 0$ or $1, \forall k \in \mathcal{H}_+$. \square

With the two propositions established, we are ready to prove our main result, which asserts that (C) and (P) are equivalent.

THEOREM 3.3. *Under Assumptions (A1), (A2), and (A3),*

$$Z_P = Z_C.$$

Furthermore, if we let (\mathbf{p}^, Y^*, X^*) be an optimal solution to (C), then there exists a sequence of nonnegative random objective coefficient vectors $\tilde{\mathbf{c}}_\epsilon^*$ and feasible solutions $\mathbf{x}^*(\tilde{\mathbf{c}}_\epsilon^*)$ that converge in moments to this optimal solution, i.e.,*

$$\lim_{\epsilon \downarrow 0} \mathbf{E} \left[\begin{pmatrix} 1 \\ \tilde{\mathbf{c}}_\epsilon^* \\ \mathbf{x}^*(\tilde{\mathbf{c}}_\epsilon^*) \end{pmatrix} \begin{pmatrix} 1 \\ \tilde{\mathbf{c}}_\epsilon^* \\ \mathbf{x}^*(\tilde{\mathbf{c}}_\epsilon^*) \end{pmatrix}^T \right] = \begin{pmatrix} 1 & \boldsymbol{\mu}^T & \mathbf{p}^{*T} \\ \boldsymbol{\mu} & \Sigma & Y^{*T} \\ \mathbf{p}^* & Y^* & X^* \end{pmatrix}.$$

PROOF. *Step 1: Decomposing the matrix.* Consider a completely positive decomposition of the matrix,

$$\begin{pmatrix} 1 & \boldsymbol{\mu}^T & \mathbf{p}^{*T} \\ \boldsymbol{\mu} & \Sigma & Y^{*T} \\ \mathbf{p}^* & Y^* & X^* \end{pmatrix} = \sum_{k \in \mathcal{H}} \begin{pmatrix} \alpha_k \\ \boldsymbol{\beta}_k \\ \boldsymbol{\gamma}_k \end{pmatrix} \begin{pmatrix} \alpha_k \\ \boldsymbol{\beta}_k \\ \boldsymbol{\gamma}_k \end{pmatrix}^T,$$

where $\alpha_k \in \mathbb{R}_+, \boldsymbol{\beta}_k \in \mathbb{R}_+^n, \boldsymbol{\gamma}_k \in \mathbb{R}_+^m, \forall k \in \mathcal{H}$. Define $\mathcal{H}_+ := \{k \in \mathcal{H} \mid \alpha_k > 0\}$, and $\mathcal{H}_0 := \{k \in \mathcal{H} \mid \alpha_k = 0\}$. Then

$$\begin{aligned} \begin{pmatrix} 1 & \boldsymbol{\mu}^T & \mathbf{p}^{*T} \\ \boldsymbol{\mu} & \Sigma & Y^{*T} \\ \mathbf{p}^* & Y^* & X^* \end{pmatrix} &= \sum_{k \in \mathcal{H}_+} \alpha_k^2 \begin{pmatrix} 1 \\ \frac{\boldsymbol{\beta}_k}{\alpha_k} \\ \frac{\boldsymbol{\gamma}_k}{\alpha_k} \end{pmatrix} \begin{pmatrix} 1 \\ \frac{\boldsymbol{\beta}_k}{\alpha_k} \\ \frac{\boldsymbol{\gamma}_k}{\alpha_k} \end{pmatrix}^T \\ &\quad + \sum_{k \in \mathcal{H}_0} \begin{pmatrix} 0 \\ \boldsymbol{\beta}_k \\ \boldsymbol{\gamma}_k \end{pmatrix} \begin{pmatrix} 0 \\ \boldsymbol{\beta}_k \\ \boldsymbol{\gamma}_k \end{pmatrix}^T. \end{aligned}$$

From Propositions 3.1 and 3.2,

$$\frac{\boldsymbol{\gamma}_k}{\alpha_k} \in \mathcal{L}, \quad \forall k \in \mathcal{H}_+ \quad \text{and} \quad \frac{\boldsymbol{\gamma}_{kj}}{\alpha_k} \in \{0, 1\}, \quad \forall j \in \mathcal{B}, \forall k \in \mathcal{H}_+.$$

This implies that $\boldsymbol{\gamma}_k/\alpha_k$ is a feasible solution to the original mixed 0-1 linear program for all $k \in \mathcal{H}_+$. As will be clear in the latter part of the proof, if the random vector $\tilde{\mathbf{c}}$ is realized to be $\boldsymbol{\beta}_k/\alpha_k$, then $\boldsymbol{\gamma}_k/\alpha_k$ is not only feasible, but also optimal, to Problem (1). From Proposition 3.1,

$$\boldsymbol{\gamma}_k = \mathbf{0}, \quad \forall k \in \mathcal{H}_0.$$

Then the decomposition becomes

$$\begin{aligned} \begin{pmatrix} 1 & \boldsymbol{\mu}^T & \mathbf{p}^{*T} \\ \boldsymbol{\mu} & \Sigma & Y^{*T} \\ \mathbf{p}^* & Y^* & X^* \end{pmatrix} &= \sum_{k \in \mathcal{H}_+} \alpha_k^2 \begin{pmatrix} 1 \\ \frac{\boldsymbol{\beta}_k}{\alpha_k} \\ \frac{\boldsymbol{\gamma}_k}{\alpha_k} \end{pmatrix} \begin{pmatrix} 1 \\ \frac{\boldsymbol{\beta}_k}{\alpha_k} \\ \frac{\boldsymbol{\gamma}_k}{\alpha_k} \end{pmatrix}^T \\ &\quad + \sum_{k \in \mathcal{H}_0} \begin{pmatrix} 0 \\ \boldsymbol{\beta}_k \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} 0 \\ \boldsymbol{\beta}_k \\ \mathbf{0} \end{pmatrix}^T. \end{aligned}$$

Step 2: Constructing a sequence of random vectors and feasible solutions. Let $\epsilon \in (0, 1)$. We define a sequence of random vectors $\tilde{\mathbf{c}}_\epsilon^*$ together with their corresponding feasible solutions $\mathbf{x}^*(\tilde{\mathbf{c}}_\epsilon^*)$ as follows,

$$\begin{cases} \mathbf{P}\left(\left(\tilde{\mathbf{c}}_\epsilon^*, \mathbf{x}^*(\tilde{\mathbf{c}}_\epsilon^*)\right) = \left(\frac{\boldsymbol{\beta}_k}{\alpha_k}, \frac{\boldsymbol{\gamma}_k}{\alpha_k}\right)\right) = (1 - \epsilon^2)\alpha_k^2, & \forall k \in \mathcal{H}_+, \\ \mathbf{P}\left(\left(\tilde{\mathbf{c}}_\epsilon^*, \mathbf{x}^*(\tilde{\mathbf{c}}_\epsilon^*)\right) = \left(\frac{\sqrt{|\mathcal{H}_0|}\boldsymbol{\beta}_k}{\epsilon}, \text{any feasible solution } \mathbf{x}\right)\right) \\ = \epsilon^2 \frac{1}{|\mathcal{H}_0|}, & \forall k \in \mathcal{H}_0. \end{cases}$$

This is a valid probability distribution because

$$\begin{aligned} & \sum_{k \in \mathcal{H}_+} (1 - \epsilon^2)\alpha_k^2 + \sum_{k \in \mathcal{H}_0} \epsilon^2 \frac{1}{|\mathcal{H}_0|} \\ &= (1 - \epsilon^2) \sum_{k \in \mathcal{H}_+} \alpha_k^2 + \epsilon^2 \sum_{k \in \mathcal{H}_0} \frac{1}{|\mathcal{H}_0|} \\ &= (1 - \epsilon^2) + \epsilon^2 \\ &= 1. \end{aligned}$$

The mean of the marginal distribution of $\tilde{\mathbf{c}}_\epsilon^*$ satisfies

$$\begin{aligned} \mathbf{E}[\tilde{\mathbf{c}}_\epsilon^*] &= \sum_{k \in \mathcal{H}_+} (1 - \epsilon^2)\alpha_k^2 \frac{\boldsymbol{\beta}_k}{\alpha_k} + \sum_{k \in \mathcal{H}_0} \epsilon^2 \frac{1}{|\mathcal{H}_0|} \frac{\sqrt{|\mathcal{H}_0|}\boldsymbol{\beta}_k}{\epsilon} \\ &= (1 - \epsilon^2) \sum_{k \in \mathcal{H}_+} \alpha_k \boldsymbol{\beta}_k + \epsilon \sum_{k \in \mathcal{H}_0} \frac{\boldsymbol{\beta}_k}{\sqrt{|\mathcal{H}_0|}} \\ &\xrightarrow{\epsilon \downarrow 0} \sum_{k \in \mathcal{H}_+} \alpha_k \boldsymbol{\beta}_k \\ &= \boldsymbol{\mu}. \end{aligned}$$

The second-moment matrix satisfies

$$\begin{aligned} \mathbf{E}[\tilde{\mathbf{c}}_\epsilon^* \tilde{\mathbf{c}}_\epsilon^{*T}] &= \sum_{k \in \mathcal{H}_+} (1 - \epsilon^2)\alpha_k^2 \frac{\boldsymbol{\beta}_k \boldsymbol{\beta}_k^T}{\alpha_k \alpha_k} + \sum_{k \in \mathcal{H}_0} \epsilon^2 \frac{1}{|\mathcal{H}_0|} \\ & \quad \cdot \frac{\sqrt{|\mathcal{H}_0|}\boldsymbol{\beta}_k}{\epsilon} \frac{\sqrt{|\mathcal{H}_0|}\boldsymbol{\beta}_k^T}{\epsilon} \\ &= (1 - \epsilon^2) \sum_{k \in \mathcal{H}_+} \boldsymbol{\beta}_k \boldsymbol{\beta}_k^T + \sum_{k \in \mathcal{H}_0} \boldsymbol{\beta}_k \boldsymbol{\beta}_k^T \\ &\xrightarrow{\epsilon \downarrow 0} \sum_{k \in \mathcal{H}} \boldsymbol{\beta}_k \boldsymbol{\beta}_k^T \\ &= \boldsymbol{\Sigma}. \end{aligned}$$

Similarly, it can be verified that

$$\begin{aligned} \mathbf{E}[\mathbf{x}^*(\tilde{\mathbf{c}}_\epsilon^*)] &\xrightarrow{\epsilon \downarrow 0} \mathbf{p}^*, \\ \mathbf{E}[\mathbf{x}^*(\tilde{\mathbf{c}}_\epsilon^*) \tilde{\mathbf{c}}_\epsilon^{*T}] &\xrightarrow{\epsilon \downarrow 0} \mathbf{Y}^*, \quad \text{and} \\ \mathbf{E}[\mathbf{x}^*(\tilde{\mathbf{c}}_\epsilon^*) \mathbf{x}^{*T}(\tilde{\mathbf{c}}_\epsilon^*)] &\xrightarrow{\epsilon \downarrow 0} \mathbf{X}^*. \end{aligned}$$

Step 3: Evaluating the limit of the sequence of objective values. As $\epsilon \downarrow 0$, the random vectors $(\tilde{\mathbf{c}}_\epsilon^*, \mathbf{x}^*(\tilde{\mathbf{c}}_\epsilon^*))$ converge almost surely (a.s.)¹ to $(\tilde{\mathbf{c}}^*, \mathbf{x}^*(\tilde{\mathbf{c}}^*))$, which is defined as

$$\mathbf{P}\left(\left(\tilde{\mathbf{c}}^*, \mathbf{x}^*(\tilde{\mathbf{c}}^*)\right) = \left(\frac{\boldsymbol{\beta}_k}{\alpha_k}, \frac{\boldsymbol{\gamma}_k}{\alpha_k}\right)\right) = \alpha_k^2, \quad \forall k \in \mathcal{H}_+.$$

From the continuous mapping theorem,

$$\tilde{\mathbf{c}}_\epsilon^* \xrightarrow{\text{a.s.}} \tilde{\mathbf{c}}^* \implies Z(\tilde{\mathbf{c}}_\epsilon^*) \xrightarrow{\text{a.s.}} Z(\tilde{\mathbf{c}}^*).$$

Furthermore, from the boundedness assumption in (A3), every feasible solution $\mathbf{x} \leq u\mathbf{e}$ for some $0 < u < \infty$, where \mathbf{e} is a vector of ones. Hence, the second moment of $Z(\tilde{\mathbf{c}}_\epsilon^*)$ is bounded for all $\epsilon \in (0, 1)$, i.e.,

$$\begin{aligned} \mathbf{E}[Z(\tilde{\mathbf{c}}_\epsilon^*)^2] &\leq \sum_{k \in \mathcal{H}_+} (1 - \epsilon^2)u^2(\boldsymbol{\beta}_k^T \mathbf{e})^2 + \sum_{k \in \mathcal{H}_0} u^2(\boldsymbol{\beta}_k^T \mathbf{e})^2 \\ &\leq \sum_{k \in \mathcal{H}_+} u^2(\boldsymbol{\beta}_k^T \mathbf{e})^2 + \sum_{k \in \mathcal{H}_0} u^2(\boldsymbol{\beta}_k^T \mathbf{e})^2 \\ &< \infty. \end{aligned}$$

The finiteness of the second moment implies that the sequence $Z(\tilde{\mathbf{c}}_\epsilon^*)$ is uniformly integrable. This implies that the sequence of expected optimal objective values converges to the finite value $\mathbf{E}[Z(\tilde{\mathbf{c}}^*)]$ (see Billingsley 1995), i.e.,

$$\lim_{\epsilon \downarrow 0} \mathbf{E}[Z(\tilde{\mathbf{c}}_\epsilon^*)] = \mathbf{E}[Z(\tilde{\mathbf{c}}^*)].$$

Step 4: Testing for tightness.

Define the space of all feasible first and second moments supported on \mathbb{R}_+^n and the corresponding expected objective value as

$$\mathcal{H}(\mathbb{R}_+^n) = \{\lambda(1, \boldsymbol{\mu}', \boldsymbol{\Sigma}', Z') \mid \lambda \geq 0, Z' = \mathbf{E}[Z(\tilde{\mathbf{c}})]\}$$

for some random vector $\tilde{\mathbf{c}}(\boldsymbol{\mu}', \boldsymbol{\Sigma}')^+$.

$\overline{\mathcal{H}(\mathbb{R}_+^n)}$ is then a closed convex cone. For each $\epsilon \in (0, 1)$, we have

$$(1, \mathbf{E}[\tilde{\mathbf{c}}_\epsilon^*], \mathbf{E}[\tilde{\mathbf{c}}_\epsilon^* \tilde{\mathbf{c}}_\epsilon^{*T}], \mathbf{E}[Z(\tilde{\mathbf{c}}_\epsilon^*)]) \in \overline{\mathcal{H}(\mathbb{R}_+^n)}.$$

Hence, the limit of this sequence of points also lies in the closure, i.e.,

$$\lim_{\epsilon \downarrow 0} (1, \mathbf{E}[\tilde{\mathbf{c}}_\epsilon^*], \mathbf{E}[\tilde{\mathbf{c}}_\epsilon^* \tilde{\mathbf{c}}_\epsilon^{*T}], \mathbf{E}[Z(\tilde{\mathbf{c}}_\epsilon^*)]) \in \overline{\mathcal{H}(\mathbb{R}_+^n)},$$

or equivalently,

$$(1, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{E}[Z(\tilde{\mathbf{c}}^*)]) \in \overline{\mathcal{H}(\mathbb{R}_+^n)}.$$

Consider the line $\{(1, \boldsymbol{\mu}, \boldsymbol{\Sigma}, z)\} \in \overline{\mathcal{H}(\mathbb{R}_+^n)}$. In case z is a constant, the result is obvious.² Otherwise, it is easy to verify that the point $(1, \boldsymbol{\mu}, \boldsymbol{\Sigma}, Z_p)$ is the end point of this line where the line is going to exit the closed convex cone. Hence, we have $Z_p \geq \mathbf{E}[Z(\tilde{\mathbf{c}}^*)]$.

Thus,

$$\begin{aligned} \sup_{\tilde{c} \sim (\mu, \Sigma)^+} \mathbf{E}[Z(\tilde{c})] &= Z_p \\ &\geq \mathbf{E}[Z(\tilde{c}^*)] \\ &\geq \mathbf{E}[\tilde{c}^{*T} \mathbf{x}^*(\tilde{c}^*)] \\ &= \sum_{k \in \mathcal{K}_+} \boldsymbol{\beta}_k^T \boldsymbol{\gamma}_k, \\ &= \sum_{k \in \mathcal{K}_+} \text{tr}(\boldsymbol{\beta}_k \boldsymbol{\gamma}_k^T) \\ &= \text{tr} \left(\sum_{k \in \mathcal{K}_+} \boldsymbol{\beta}_k \boldsymbol{\gamma}_k^T \right) \\ &= \text{tr}(Y^*) \\ &= I_n \cdot Y^*. \end{aligned}$$

The right-hand side is exactly the optimal objective value of (C). Therefore, we have shown that solving (C) provides a lower bound to (P), and hence the two formulations are equivalent. \square

From the construction in Theorem 3.3, it is clear that the moments and the bound are achievable only in a limiting sense. In the completely positive matrix decomposition, $\boldsymbol{\beta}_k$ can be nonzero for some $k \in \mathcal{K}_0$ and \mathbf{c}^* might not be strictly feasible due to the second-moment matrix constraint.

The moments of the limiting random vector \mathbf{c}^* satisfy

$$\begin{aligned} \begin{pmatrix} 1 & \mathbf{E}[\tilde{c}^{*T}] \\ \mathbf{E}[\tilde{c}^*] & \mathbf{E}[\tilde{c}^* \tilde{c}^{*T}] \end{pmatrix} &= \begin{pmatrix} \sum_{k \in \mathcal{K}_+} \alpha_k^2 & \sum_{k \in \mathcal{K}_+} \alpha_k \boldsymbol{\beta}_k^T \\ \sum_{k \in \mathcal{K}_+} \alpha_k \boldsymbol{\beta}_k & \sum_{k \in \mathcal{K}_+} \boldsymbol{\beta}_k \boldsymbol{\beta}_k^T \end{pmatrix} \\ &= \begin{pmatrix} 1 & \boldsymbol{\mu}^T \\ \boldsymbol{\mu} & \Sigma - \sum_{k \in \mathcal{K}_0} \boldsymbol{\beta}_k \boldsymbol{\beta}_k^T \end{pmatrix} \\ &\preceq_{cp} \begin{pmatrix} 1 & \boldsymbol{\mu}^T \\ \boldsymbol{\mu} & \Sigma \end{pmatrix}. \end{aligned}$$

This leads to a corollary to Theorem 3.3 in the case that the second-moment matrix is itself unknown.

Assumption. (A1') The set of distributions of \tilde{c} is defined by the nonnegative support \mathbb{R}_+^n with known finite mean $\boldsymbol{\mu}$. The second-moment matrix Σ' is unknown but satisfies $\Sigma' \preceq_{cp} \Sigma$ where Σ is a known finite second-moment matrix. The set is assumed to be nonempty.

COROLLARY 3.4. *Under Assumptions (A1'), (A2), and (A3),*

$$Z_p = Z_C.$$

Furthermore, if we let (\mathbf{p}^*, Y^*, X^*) be an optimal solution to (C), then there exists a nonnegative random objective

coefficient vector \tilde{c}^* and feasible solutions $\mathbf{x}^*(\tilde{c}^*)$ that satisfy

$$\mathbf{E} \left[\begin{pmatrix} 1 \\ \tilde{c}^* \\ \mathbf{x}^*(\tilde{c}^*) \end{pmatrix} \begin{pmatrix} 1 \\ \tilde{c}^* \\ \mathbf{x}^*(\tilde{c}^*) \end{pmatrix}^T \right] = \begin{pmatrix} 1 & \boldsymbol{\mu}^T & \mathbf{p}^{*T} \\ \boldsymbol{\mu} & \Sigma' & Y^{*T} \\ \mathbf{p}^* & Y^* & X^* \end{pmatrix},$$

where $\Sigma' \preceq_{cp} \Sigma$.

As compared to Theorem 3.3, the bound in Corollary 3.4 is exactly achievable by a feasible distribution.

REMARK. Consider the definition of the variable p_j , $\forall j \in \mathcal{B}$:

$$\begin{aligned} p_j &= \mathbf{E}[x_j(\tilde{c})] \\ &= \mathbf{E}[x_j(\tilde{c}) \mid x_j(\tilde{c}) = 1] \mathbf{P}(x_j(\tilde{c}) = 1) \\ &= \mathbf{P}(x_j(\tilde{c}) = 1). \end{aligned}$$

The optimal solutions p_j^* , $j \in \mathcal{B}$ of (C) give an estimate to the persistency of the variable x_j in the original problem. To be precise, p_j^* is the persistency of x_j under a limiting distribution \mathbf{c}^* .

4. Extensions

4.1. Support in \mathbb{R}^n

As discussed in the previous section, testing for feasibility of distributions with nonnegative support and given mean and second-moment matrix is itself a difficult problem. It is possible to relax this assumption and allow for objective coefficients to possibly take negative values too.

Assumption. (A1'') The set of distributions of \tilde{c} is defined by the support \mathbb{R}^n with known finite mean $\boldsymbol{\mu}$ and known finite second-moment matrix Σ . The set is assumed to be nonempty.

Unlike Assumption (A1), testing for the existence of feasible multivariate distributions in (A1'') is easy. The feasibility condition is equivalent to verifying the positive semidefinite condition, i.e.,

$$(1, \boldsymbol{\mu}, \Sigma) \in \overline{\mathcal{M}_2(\mathbb{R}^n)} \iff \begin{pmatrix} 1 & \boldsymbol{\mu}^T \\ \boldsymbol{\mu} & \Sigma \end{pmatrix} \succeq 0.$$

The problem of interest is

$$(PS) \quad \sup_{\tilde{c} \sim (\mu, \Sigma)} \mathbf{E}[Z(\tilde{c})].$$

Using a constructive approach as in §3, a convex relaxation to (PS) is

$$\begin{aligned} (CS) \quad &\max I_n \cdot Y \\ &\text{s.t. } \mathbf{a}_i^T \mathbf{p} = b_i \quad \forall i = 1, \dots, m \\ &\mathbf{a}_i^T X \mathbf{a}_i = b_i^2 \quad \forall i = 1, \dots, m \\ &X_{jj} = p_j \quad \forall j \in \mathcal{B} \subseteq \{1, \dots, n\} \end{aligned}$$

$$\begin{pmatrix} 1 & \boldsymbol{\mu}^T & \mathbf{p}^T \\ \boldsymbol{\mu} & \Sigma & Y^T \\ \mathbf{p} & Y & X \end{pmatrix} \in \left\{ A \in \mathbb{R}^{(2n+1) \times (2n+1)} \mid \begin{array}{l} V_1 \in \mathbb{R}_+^{1 \times k} \\ V_2 \in \mathbb{R}^{n \times k} \\ V_3 \in \mathbb{R}_+^{n \times k} \end{array} \right. ,$$

$$\text{such that } A = \begin{pmatrix} V_1 \\ V_2 \\ V_3 \end{pmatrix} \begin{pmatrix} V_1 \\ V_2 \\ V_3 \end{pmatrix}^T .$$

The equivalence of the formulations (PS) and (CS) is shown next.

THEOREM 4.1. *Under Assumptions (A1''), (A2), and (A3), $Z_P = Z_C$.*

Furthermore, if we let (\mathbf{p}^*, Y^*, X^*) be an optimal solution to (CS), then there exists a random objective coefficient vector $\tilde{\mathbf{c}}^*$ supported in \mathbb{R}^n and feasible solutions $\mathbf{x}^*(\tilde{\mathbf{c}}^*)$ that satisfy

$$\mathbf{E} \left[\begin{pmatrix} 1 \\ \tilde{\mathbf{c}}^* \\ \mathbf{x}^*(\tilde{\mathbf{c}}^*) \end{pmatrix} \begin{pmatrix} 1 \\ \tilde{\mathbf{c}}^* \\ \mathbf{x}^*(\tilde{\mathbf{c}}^*) \end{pmatrix}^T \right] = \begin{pmatrix} 1 & \boldsymbol{\mu}^T & \mathbf{p}^{*T} \\ \boldsymbol{\mu} & \Sigma & Y^{*T} \\ \mathbf{p}^* & Y^* & X^* \end{pmatrix} .$$

PROOF. We only outline the key differences from the proof of Theorem 3.3. Consider the decomposition of the matrix in the optimal solution from (CS),

$$\begin{pmatrix} 1 & \boldsymbol{\mu}^T & \mathbf{p}^{*T} \\ \boldsymbol{\mu} & \Sigma & Y^{*T} \\ \mathbf{p}^* & Y^* & X^* \end{pmatrix} = \sum_{k \in \mathcal{H}_+} \alpha_k^2 \begin{pmatrix} 1 \\ \frac{\boldsymbol{\beta}_k}{\alpha_k} \\ \frac{\boldsymbol{\gamma}_k}{\alpha_k} \end{pmatrix} \begin{pmatrix} 1 \\ \frac{\boldsymbol{\beta}_k}{\alpha_k} \\ \frac{\boldsymbol{\gamma}_k}{\alpha_k} \end{pmatrix}^T + \sum_{k \in \mathcal{H}_0} \begin{pmatrix} 0 \\ \boldsymbol{\beta}_k \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} 0 \\ \boldsymbol{\beta}_k \\ \mathbf{0} \end{pmatrix}^T ,$$

where $\alpha_k \in \mathbb{R}_+$, $\boldsymbol{\beta}_k \in \mathbb{R}^n$, $\boldsymbol{\gamma}_k \in \mathbb{R}_+^n$, and $\boldsymbol{\gamma}_k/\alpha_k$ are feasible solutions to the mixed 0-1 linear program for all $k \in \mathcal{H}_+$. Let the matrix B be defined as

$$B = \sum_{k \in \mathcal{H}_0} \boldsymbol{\beta}_k \boldsymbol{\beta}_k^T .$$

Define

$$\mathbf{P} \left((\tilde{\mathbf{c}}^*, \mathbf{x}^*(\tilde{\mathbf{c}}^*)) = \left(\frac{\boldsymbol{\beta}_k}{\alpha_k} + \tilde{\mathbf{z}}^*, \frac{\boldsymbol{\gamma}_k}{\alpha_k} \right) \right) = \alpha_k^2, \quad \forall k \in \mathcal{H}_+,$$

where $\tilde{\mathbf{z}}^* \sim \mathcal{N}(\mathbf{0}, B)$ is a multivariate normal random vector with mean zero and covariance matrix B , generated independent of the scenario k . This is a valid probability distribution because $\sum_{k \in \mathcal{H}_+} \alpha_k^2 = 1$. Furthermore, the mean of the marginal distribution of $\tilde{\mathbf{c}}^*$ satisfies

$$\begin{aligned} \mathbf{E}[\tilde{\mathbf{c}}^*] &= \sum_{k \in \mathcal{H}_+} \alpha_k \boldsymbol{\beta}_k + \mathbf{E}[\tilde{\mathbf{z}}^*] \\ &= \boldsymbol{\mu} . \end{aligned}$$

Similarly, the second-moment matrix satisfies

$$\begin{aligned} \mathbf{E}[\tilde{\mathbf{c}}^* \tilde{\mathbf{c}}^{*T}] &= \sum_{k \in \mathcal{H}_+} \boldsymbol{\beta}_k \boldsymbol{\beta}_k^T + \mathbf{E}[\tilde{\mathbf{z}}^* \tilde{\mathbf{z}}^{*T}] \\ &= \sum_{k \in \mathcal{H}_+} \boldsymbol{\beta}_k \boldsymbol{\beta}_k^T + B \\ &= \Sigma . \end{aligned}$$

Thus, $\tilde{\mathbf{c}}^* \sim (\boldsymbol{\mu}, \Sigma)$. Finally,

$$\begin{aligned} \sup_{\tilde{\mathbf{c}} \sim (\boldsymbol{\mu}, \Sigma)^+} \mathbf{E}[Z(\tilde{\mathbf{c}})] &\geq \mathbf{E}[Z(\tilde{\mathbf{c}}^*)] \\ &\geq \mathbf{E}[\tilde{\mathbf{c}}^{*T} \mathbf{x}^*(\tilde{\mathbf{c}}^*)] \\ &= \sum_{k \in \mathcal{H}_+} \boldsymbol{\beta}_k^T \boldsymbol{\gamma}_k + \sum_{k \in \mathcal{H}_+} \alpha_k \mathbf{E}[\tilde{\mathbf{z}}^{*T}] \boldsymbol{\gamma}_k \\ &= I_n \cdot Y^* . \end{aligned}$$

Because the right-hand side is the optimal objective value of (CS), the two formulations are equivalent. \square

Thus, by relaxing the assumption on the support of the objective coefficients, it is possible to guarantee that the bound is exactly achievable. In computational experiments, a simple relaxation for matrix inequality constraint in (CS) is to use

$$\begin{pmatrix} 1 & \boldsymbol{\mu}^T & \mathbf{p}^T \\ \boldsymbol{\mu} & \Sigma & Y^T \\ \mathbf{p} & Y & X \end{pmatrix} \geq 0 \quad \text{and} \quad \begin{pmatrix} 1 & \mathbf{p}^T \\ \mathbf{p} & X \end{pmatrix} \geq 0 .$$

4.2. Uncertainty in Moment Information

A natural assumption to relax is the exact knowledge of moments and incorporate uncertainty in the moment estimates. This is particularly useful when confidence intervals can be built around the sample moment estimates that are often computed from the empirical distribution (Delage and Ye 2010). In Corollary 3.4, the assumption on the exact knowledge of the second-moment matrix is relaxed. More generally, suppose that the exact values of the mean and the second moments are unknown, i.e., $(\boldsymbol{\mu}, \Sigma)$ lies in a nonempty, closed, and convex set \mathcal{U} . In this case, the problem is to choose the mean and second-moment matrix and the corresponding multivariate distribution that provides the tight bound,

$$\text{(PU)} \quad \sup_{(\boldsymbol{\mu}, \Sigma) \in \mathcal{U}, \tilde{\mathbf{c}} \sim (\boldsymbol{\mu}, \Sigma)^+} \mathbf{E}[Z(\tilde{\mathbf{c}})] .$$

It is easy to modify CPCMM to capture the additional uncertainty,

$$\begin{aligned} \text{(CU)} \quad \max \quad & I_n \cdot Y \\ \text{s.t.} \quad & \mathbf{a}_i^T \mathbf{p} = b_i \quad \forall i = 1, \dots, m \\ & \mathbf{a}_i^T X \mathbf{a}_i = b_i^2 \quad \forall i = 1, \dots, m \end{aligned}$$

$$X_{jj} = p_j \quad \forall j \in \mathcal{B} \subseteq \{1, \dots, n\}$$

$$\begin{pmatrix} 1 & \boldsymbol{\mu}^T & \mathbf{p}^T \\ \boldsymbol{\mu} & \Sigma & Y^T \\ \mathbf{p} & Y & X \end{pmatrix} \succeq_{cp} 0$$

$$(\boldsymbol{\mu}, \Sigma) \in \mathbb{U},$$

where $\boldsymbol{\mu}$, Σ , \mathbf{p} , X , and Y are the decision variables in the formulation.

Two examples of \mathbb{U} are described next.

(a) Lower and upper bounds on the mean and second-moment matrix can be incorporated using simple linear inequalities, i.e.,

$$\mathbb{U} = \{(\boldsymbol{\mu}, \Sigma) : \boldsymbol{\mu}_L \leq \boldsymbol{\mu} \leq \boldsymbol{\mu}_U, \Sigma_L \leq \Sigma \leq \Sigma_U\}.$$

(b) Delage and Ye (2010) proposed the following uncertainty set in the moments parameterized by $\gamma_1 \geq 0$ and $\gamma_2 \geq 0$,

$$(\mathbf{E}[\tilde{\mathbf{c}}] - \boldsymbol{\mu}_0)^T Q_0^{-1} (\mathbf{E}[\tilde{\mathbf{c}}] - \boldsymbol{\mu}_0) \leq \gamma_1,$$

$$\mathbf{E}[(\tilde{\mathbf{c}} - \boldsymbol{\mu}_0)(\tilde{\mathbf{c}} - \boldsymbol{\mu}_0)^T] \preceq \gamma_2 Q_0.$$

The first constraint models the mean of $\tilde{\mathbf{c}}$ lying in an ellipsoid of size γ_1 centered at the mean estimate $\boldsymbol{\mu}_0$, and the second constraint models the matrix $\mathbf{E}[(\tilde{\mathbf{c}} - \boldsymbol{\mu}_0)(\tilde{\mathbf{c}} - \boldsymbol{\mu}_0)^T]$ lying in a positive semidefinite cone bounded by a matrix inequality. This uncertainty set is characterized using the variables $\boldsymbol{\mu}$ and Σ in (CU) as follows,

$$(\boldsymbol{\mu} - \boldsymbol{\mu}_0)^T Q_0^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_0) \leq \gamma_1,$$

$$\Sigma - 2\boldsymbol{\mu}\boldsymbol{\mu}_0^T + \boldsymbol{\mu}_0\boldsymbol{\mu}_0^T \preceq \gamma_2 Q_0.$$

These two constraints are semidefinite constraints, where the first one can be rewritten as

$$\begin{pmatrix} \gamma_1 & \boldsymbol{\mu}^T - \boldsymbol{\mu}_0^T \\ \boldsymbol{\mu} - \boldsymbol{\mu}_0 & Q_0 \end{pmatrix} \succeq 0.$$

Then the corresponding \mathbb{U} is defined as

$$\mathbb{U} = \left\{ (\boldsymbol{\mu}, \Sigma) : \begin{pmatrix} \gamma_1 & \boldsymbol{\mu}^T - \boldsymbol{\mu}_0^T \\ \boldsymbol{\mu} - \boldsymbol{\mu}_0 & Q_0 \end{pmatrix} \succeq 0, \right. \\ \left. \Sigma - 2\boldsymbol{\mu}\boldsymbol{\mu}_0^T + \boldsymbol{\mu}_0\boldsymbol{\mu}_0^T \preceq \gamma_2 Q_0 \right\}.$$

4.3. Dimension Reduction for Constant Coefficients

In CPCMM, the size of the completely positive matrix is an obvious bottleneck in computation. One would want to reduce the dimension of the matrix as much as possible. The size of the completely positive matrix in (C) is $(2n + 1) \times (2n + 1)$. However, we do not need such a large

matrix if there are some variables in (P) having constant objective coefficients. Without loss of generality, assume $\tilde{\mathbf{c}}^T = (\tilde{\mathbf{c}}_1^T, \tilde{\mathbf{c}}_2^T)$, where $\tilde{\mathbf{c}}_1$ is a random vector of dimension k and $\tilde{\mathbf{c}}_1 \sim (\boldsymbol{\mu}_1, \Sigma_1)_+$, whereas $\tilde{\mathbf{c}}_2$ is a constant vector of dimension $n - k$. In this case, it is possible to reduce the dimension of the completely positive matrix to $(k + n + 1) \times (k + n + 1)$.

Recall the definition of the variables,

$$\mathbf{p} := \mathbf{E}[\mathbf{x}(\tilde{\mathbf{c}})],$$

$$Y := \mathbf{E}[\mathbf{x}(\tilde{\mathbf{c}})\tilde{\mathbf{c}}^T].$$

If c_j is constant, then $Y_{jl} = \mathbf{E}[c_j x_l(\tilde{\mathbf{c}})] = c_j \mathbf{E}[x_l(\tilde{\mathbf{c}})] = c_j p_l, \forall l = 1, \dots, n$, which indicates Y_{jl} and p_l are linearly dependent, and consequently one of them is redundant in the formulation. Thus, we can safely drop these redundant variables, $Y_{jl}, \forall l = 1, \dots, n, \forall j \geq k$, and adjust the objective function accordingly to obtain a slimmer version of (C),

$$(CC) \quad \max \sum_{j=1}^k Y_{jj} + \sum_{j=k+1}^n \bar{c}_j p_j$$

s.t. $\mathbf{a}_i^T \mathbf{p} = b_i \quad \forall i = 1, \dots, m$

$$\mathbf{a}_i^T X \mathbf{a}_i = b_i^2 \quad \forall i = 1, \dots, m$$

$$X_{jj} = p_j \quad \forall j \in \mathcal{B} \subseteq \{1, \dots, n\}$$

$$\begin{pmatrix} 1 & \boldsymbol{\mu}_1^T & \mathbf{p}^T \\ \boldsymbol{\mu}_1 & \Sigma_1 & Y^T \\ \mathbf{p} & Y & X \end{pmatrix} \succeq_{cp} 0,$$

where $Y \in \mathbb{R}^{n \times k}$ is the matrix after removing the redundant parts. To show (CC) solves (P) with some constant coefficients, we only need to modify one step in the proof of Theorem 3.3: when constructing the limiting distribution for $\tilde{\mathbf{c}}$, use the values of $\tilde{\mathbf{c}}_1$ directly from the decomposition of the completely matrix and plug in $\tilde{\mathbf{c}}_2$ to form a complete instance of $\tilde{\mathbf{c}}$, i.e.,

$$\tilde{\mathbf{c}}^* = \begin{pmatrix} \tilde{\mathbf{c}}_1^* \\ \tilde{\mathbf{c}}_2 \end{pmatrix}.$$

The rest of the proof follows easily.

The advantage of this reduction is significant when one has to add in many slack variables to ensure Assumption (A2) of CPCMM. This might be necessary to ensure that the linear equality constraints in the problem bound the binary variables in $[0, 1]$. If this requirement is not met for some binary variable x_j , one needs to add in a constraint $x_j + s_j = 1$, where $s_j \geq 0$ is a slack variable. Then the objective coefficient for s_j would be 0. The cost of adding slack variables is the increased size of the completely positive matrix, which can be reduced with (CC).

4.4. Increasing Convex Piecewise-Linear Function of $Z(\tilde{\mathbf{c}})$

It is possible to extend the results to bound $\mathbf{E}[f(Z(\tilde{\mathbf{c}}))]$, where $f(x) = \max_{k \in \mathcal{K}} \{\alpha_k x + \beta_k\}$. The function $f(\cdot)$ is a nondecreasing convex piecewise-linear function with $\alpha_k \geq 0, \forall k \in \mathcal{K}$. Then Problem (P) becomes

$$(PF) \quad \sup_{\tilde{\mathbf{c}} \sim (\boldsymbol{\mu}, \boldsymbol{\Sigma})^+} \mathbf{E} \left[\max_{k \in \mathcal{K}} \{\alpha_k Z(\tilde{\mathbf{c}}) + \beta_k\} \right].$$

To obtain the corresponding CPCMM for (PF), we first partition the set of $\mathbf{c} \in \mathbb{R}_+^n$ into K sets with

$$\mathcal{S}_k := \left\{ \mathbf{c} \mid \mathbf{c} \geq \mathbf{0}, \text{ and } \alpha_k Z(\mathbf{c}) + \beta_k \geq \max_{k' \in \mathcal{K}} \{\alpha_{k'} Z(\mathbf{c}) + \beta_{k'}\} \right\}.$$

Define $|\mathcal{H}|$ sets of variables as follows,

$$q^{(k)} := \mathbf{P}(\tilde{\mathbf{c}} \in \mathcal{S}_k),$$

$$\boldsymbol{\mu}^{(k)} := \mathbf{E}[\tilde{\mathbf{c}} \mid \tilde{\mathbf{c}} \in \mathcal{S}_k] \mathbf{P}(\tilde{\mathbf{c}} \in \mathcal{S}_k),$$

$$\boldsymbol{\Sigma}^{(k)} := \mathbf{E}[\tilde{\mathbf{c}} \tilde{\mathbf{c}}^T \mid \tilde{\mathbf{c}} \in \mathcal{S}_k] \mathbf{P}(\tilde{\mathbf{c}} \in \mathcal{S}_k),$$

$$\mathbf{p}^{(k)} := \mathbf{E}[\mathbf{x}(\tilde{\mathbf{c}}) \mid \tilde{\mathbf{c}} \in \mathcal{S}_k] \mathbf{P}(\tilde{\mathbf{c}} \in \mathcal{S}_k),$$

$$Y^{(k)} := \mathbf{E}[\mathbf{x}(\tilde{\mathbf{c}}) \tilde{\mathbf{c}}^T \mid \tilde{\mathbf{c}} \in \mathcal{S}_k] \mathbf{P}(\tilde{\mathbf{c}} \in \mathcal{S}_k),$$

$$X^{(k)} := \mathbf{E}[\mathbf{x}(\tilde{\mathbf{c}}) \mathbf{x}(\tilde{\mathbf{c}}) \mid \tilde{\mathbf{c}} \in \mathcal{S}_k] \mathbf{P}(\tilde{\mathbf{c}} \in \mathcal{S}_k).$$

Using a similar argument as in constructing (C), we formulate the completely positive program,

$$(CF) \quad \max \sum_{k \in \mathcal{K}} (\alpha_k I_n \cdot Y^{(k)} + \beta_k q^{(k)})$$

s.t. $\mathbf{a}_i^T \mathbf{p}^{(k)} = b_i q^{(k)} \quad \forall i = 1, \dots, m, \forall k \in \mathcal{K}$

$$\mathbf{a}_i^T X^{(k)} \mathbf{a}_i = b_i^2 q^{(k)} \quad \forall i = 1, \dots, m, \forall k \in \mathcal{K}$$

$$X_{jj}^{(k)} = p_j^{(k)} \quad \forall j \in \mathcal{B}, \forall k \in \mathcal{K}$$

$$\begin{pmatrix} q^{(k)} & \boldsymbol{\mu}^{(k)T} & \mathbf{p}^{(k)T} \\ \boldsymbol{\mu}^{(k)} & \boldsymbol{\Sigma}^{(k)} & Y^{(k)T} \\ \mathbf{p}^{(k)} & Y^{(k)} & X^{(k)} \end{pmatrix} \succeq_{cp} \mathbf{0} \quad \forall k \in \mathcal{K}$$

$$\sum_{k \in \mathcal{K}} \begin{pmatrix} q^{(k)} & \boldsymbol{\mu}^{(k)T} \\ \boldsymbol{\mu}^{(k)} & \boldsymbol{\Sigma}^{(k)} \end{pmatrix} = \begin{pmatrix} 1 & \boldsymbol{\mu}^T \\ \boldsymbol{\mu} & \boldsymbol{\Sigma} \end{pmatrix}$$

Proving (PF) is solvable as (CF) is very similar to what we have done for Theorem 3.3, and requires only minor modifications. The key steps of the proof can be summarized as follows.

(1) (CF) gives an upper bound to (PF).

(2) Construct the extremal distribution from the optimal solution to (CF) based on the partitions of $\tilde{\mathbf{c}}$. With probability $q^{(k)*}$ (the value of $q^{(k)}$ in the optimal solution to (CF)), construct $\tilde{\mathbf{c}}$ using the completely positive decomposition of the k th matrix as in the proof of Theorem 3.3. The final limiting distribution for $\tilde{\mathbf{c}}$ would be a mixture distribution of $|\mathcal{H}|$ types and satisfy the moment conditions. The decomposition also provides the $|\mathcal{H}|$ sets of feasible solutions.

(3) Under the limiting distribution constructed in Step 2, the feasible solutions identified achieve the upper bound in the limiting case. The nondecreasing condition for function $f(\cdot)$ is required in this step.

5. Applications

In this section, we present two applications of our model and discuss some implementation issues. These applications demonstrate the usefulness and flexibility of CPCMM in dealing with random optimization problems. The first example deals with stochastic sensitivity analysis for the highest-order statistic problem. The second example is a project management problem where the CPCMM results are compared to MMM in particular.

We also compare our results with a Monte Carlo simulation-based approach. In the applications we consider, the deterministic problems are linear programs with the simulation approach needing solutions to multiple linear programs. When the deterministic problem is \mathcal{NP} -hard, implementing the simulation method would require the solution to a number of \mathcal{NP} -hard problems. On the other hand, the CPP model requires the solution to one \mathcal{NP} -hard problem.

5.1. Stochastic Sensitivity Analysis of Highest-Order Statistic

The problem of finding the maximum value from a set $\mathbf{c} = (c_1, c_2, \dots, c_n)$ of n numbers can be formulated as an optimization problem as follows:

$$(OS) \quad \max \mathbf{c}^T \mathbf{x}$$

s.t. $\sum_{j=1}^n x_j = 1$

$$\mathbf{x} \geq \mathbf{0}.$$

Suppose $c_1 > c_2 > \dots > c_n$. Then the optimal solution to (OS) is $x_1^* = 1, x_j^* = 0, \forall j = 2, \dots, n$. For the sensitivity analysis problem, consider a perturbation in the objective coefficients. Let $\tilde{c}_j = c_j + \epsilon \tilde{\delta}_j, \forall j = 1, \dots, n$, where each $\tilde{\delta}_j$ is a random variable with mean 0 and standard deviation 1, and $\epsilon \in \mathbb{R}_+$ is a factor that adjusts the degree of the perturbation. Then the resulting j th objective coefficient \tilde{c}_j has mean c_j and standard deviation ϵ . We vary ϵ to see how the optimal solution changes with different degrees of variation in the objective coefficients.

We consider two cases: (1) Independent $\tilde{\delta}_j$; (2) Correlated $\tilde{\delta}_j$ with $\mathbf{E}[\tilde{\boldsymbol{\delta}} \tilde{\boldsymbol{\delta}}^T] = \boldsymbol{\Sigma}_\delta$. The random vector $\tilde{\mathbf{c}} \sim (\mathbf{c}, \boldsymbol{\Sigma})_+$, and $\boldsymbol{\Sigma} = \mathbf{c} \mathbf{c}^T + \epsilon^2 I_n$ for the independent case, whereas $\boldsymbol{\Sigma} = \mathbf{c} \mathbf{c}^T + \epsilon^2 \boldsymbol{\Sigma}_\delta$ for the correlated case. The problem is to identify the probability that the original optimal solution still remains optimal, i.e., $\mathbf{P}(\tilde{c}_1 \geq \tilde{c}_j, \forall j = 1, \dots, n)$ as the value of ϵ increases. Moreover, $\mathbf{P}(\tilde{c}_1 \geq \tilde{c}_j, \forall j = 1, \dots, n)$ is just the persistency of x_1 . We use MMM, CMM, VBC, and CPCMM to estimate the probability, and then compare their estimates against the simulation results.

Computational Results. The mean for \tilde{c} used in both cases is randomly generated as

$$c = (19.7196, 19.0026, 17.8260, 16.4281, 15.2419, 12.1369, 9.1293, 8.8941, 4.6228, 0.3701)^T;$$

and for case (2), the correlation matrix of $\tilde{\delta}$, which is equal to Σ_{δ} , is

$$\Sigma_{\delta} = \begin{pmatrix} 1 & -0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.5 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0.5 & 0.4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 1 & 0.8 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.4 & 0.8 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Although we have carried out the tests on many different values of c and Σ_{δ} , the results are similar to what is shown here with this example.

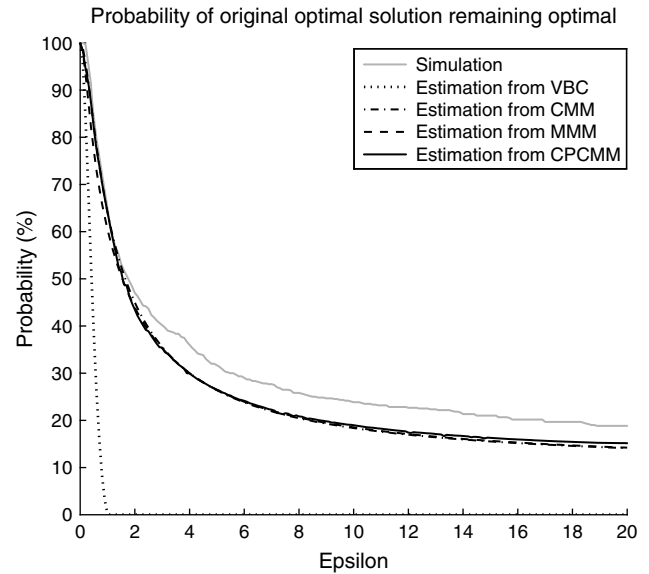
We let ϵ increase from 0 to 20 at an increment of 0.1, and solve all the moment models for each ϵ to obtain the persistency. In simulation, for each ϵ , we generate a 1,000-sized sample for $\tilde{\delta}$ satisfying the moment conditions.³ Then we solve these samples (i.e., 1,000 deterministic problems) to estimate $P(\tilde{c}_1 \geq \tilde{c}_j, \forall j = 1, \dots, n)$.

Figures 1 and 2 show the results for cases (1) and (2), respectively.

From Figure 1, we observe that the probability estimates from CPCMM and the SDP models, except VBC, are almost the same. When ϵ is small, like $\epsilon < 1$, the estimated values are almost the same as the true probabilities obtained from the simulation, and even when ϵ is large, the difference is not significant and the estimated curves look to have the same trend as the simulation curve, i.e., the rate of decrease in probability is well captured by the estimates.

Another observation is that the VBC approach gives estimates that are far away from the true values. The probability curve given by the VBC model drops much faster and gets close to zero quickly. There are two possible reasons. Firstly, the VBC model finds a lower bound on the probability $P(\tilde{c}_1 > \tilde{c}_j, \forall j = 2, \dots, n)$, which can be supported by an extreme distribution (Vandenberghe et al. 2007). Secondly, it bounds $P(\tilde{c}_1 > \tilde{c}_j, \forall j = 2, \dots, n)$ rather than $P(\tilde{c}_1 \geq \tilde{c}_j, \forall j = 2, \dots, n)$, which would be a larger number. The results suggest that the Chebyshev-type bounds computed using the VBC approach might be too conservative in practice.

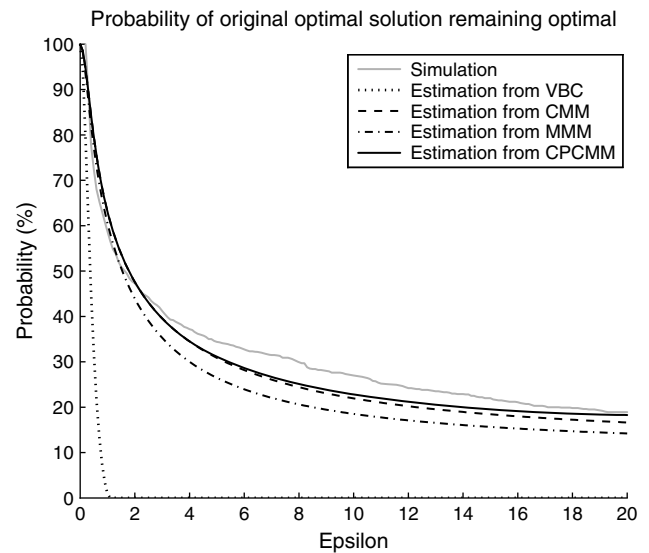
Figure 1. Results for stochastic sensitivity analysis with independent perturbation.



When correlation is added to the analysis, those models that can capture the cross-moment information gain some advantages in terms of improved precision. Figure 2 shows that CMM and CPCMM produce better estimates than MMM that ignores the correlation. However, when ϵ is small—for example, $\epsilon < 1$ —the three models give almost the same estimates that are also close to the true value. Again the VBC approach provides a very conservative estimate on the probability.

From this application, we show that for simple problems with the basic approximation for the completely positivity

Figure 2. Results for stochastic sensitivity analysis with correlated perturbation.



condition, CPCMM can perform at least as well as the other existing SDP models.

5.2. Project Network Problem

In this section, we apply our model on a project management problem to estimate the expected completion time of the project and the persistency for each activity. Then we compare the results with MMM that ignores the cross moment information. The exponential-sized formulation of CMM is based on the number of extreme points, and thus becomes impractical for medium to large projects.

The project management problem can be formulated as a longest-path problem on a directed acyclic graph. The arcs denote activities and nodes denote completion of a set of activities. Arc lengths denote the time to complete the activities. Thus, the longest path from the starting node s to ending node t gives the time needed to complete the project. Let c_{ij} be the length (time) of arc (activity) (i, j) . The problem can be solved as a linear program due to the network flow structure,

$$\begin{aligned} \max \quad & \sum_{(i,j) \in \mathcal{ARCS}} c_{ij} x_{ij} \\ \text{s.t.} \quad & \sum_{i:(i,j) \in \mathcal{ARCS}} x_{ij} - \sum_{j:(i,j) \in \mathcal{ARCS}} x_{ji} \\ & = \begin{cases} 1, & \text{if } i = s \\ 0, & \text{if } i \in \mathcal{NODS}, \text{ and } i \neq s, t \\ -1, & \text{if } i = t \end{cases} \\ & x_{ij} \geq 0, \quad \forall (i, j) \in \mathcal{ARCS}. \end{aligned}$$

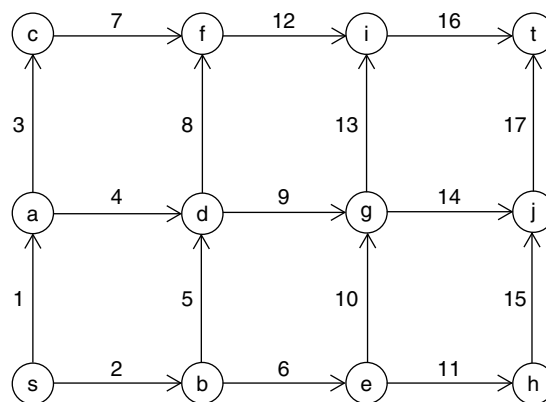
For the stochastic project management problem, the activity times are random. In such cases, due to the resource allocation and management issues, the project manager would like to focus on the activities (arcs) that are critical (lie on the longest path) with high probabilities. Next, we demonstrate how to use CPCMM to help managers identify these activities.

Computational Results. Consider the project network with the graphical representation shown in Figure 3. The directed arcs are either upwards or to the right. This network consists of 17 arcs and 10 paths.

We generate 100 data sets with each data set having a set of randomly generated means ($\sim \text{Uniform}(0, 5)$), standard deviations ($\sim \text{Uniform}(0, 2)$), and correlations for the arc lengths.⁴ CPCMM and MMM are used to estimate the persistency of each activity and expected completion time. We resort to extensive simulations to assess the accuracy of the results numerically. A sample of 1,000 instances of arc lengths were generated for each data set⁵ satisfying the moment conditions. In simulation, the linear program was solved for each sample to assess the persistency values.

Figures 4 and 5 show the results of arc persistency estimates for the 100 instances.

Figure 3. Project network.



To interpret Figure 4: for each arc, we obtain three persistency values (i.e., from CPCMM, MMM, and simulation, respectively), and we plot two points for every arc: one with coordinates (*persistency from CPCMM, persistency from simulation*), and the other with coordinates (*persistency from MMM, persistency from simulation*). Because there are 17 arcs for one instance in our problem, there are 3,400 ($17 \times 2 \times 100$) points in Figure 4. Over 94% of the persistency values are less than 0.1, which means that around 3,200 points are clustered in the bottom-left corner of Figure 4. The line represents the function “ $y = x$.” Roughly, it can be interpreted as the closeness of the results from the SDP models to simulation. If the models are perfect and the simulation is carried out with the particular extremal distribution, all the points should roughly lie on the line, whereas the error in the estimates is presented as a deviation from the line.

Figure 4. Arc persistency values for 100 instances: CPCMM vs. MMM

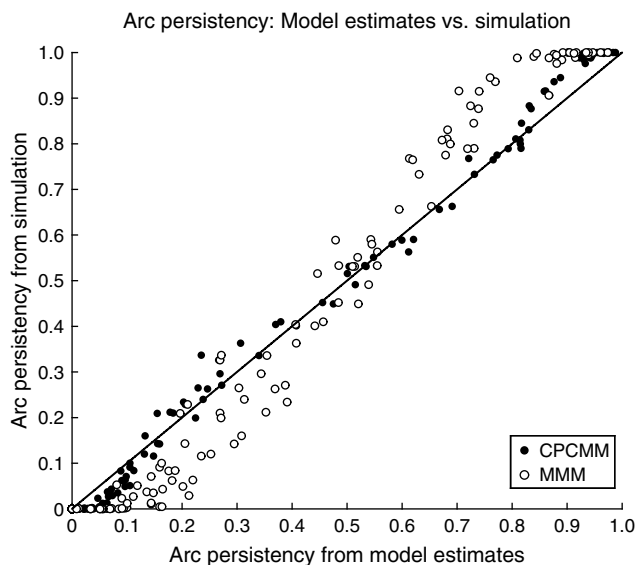
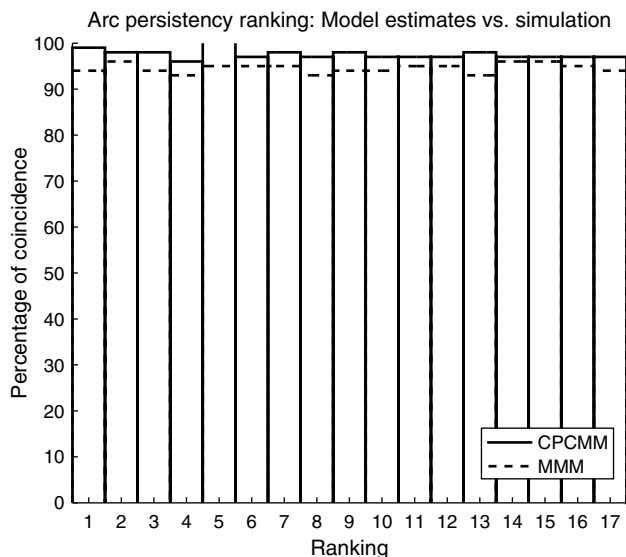


Figure 5. Arc persistency ranking for 100 instances: CPCMM vs. MMM.

From the figure, we observe that CPCMM outperforms MMM in estimating the persistency values. Furthermore, we observe an “S”-shaped pattern for the points plotted for both models, which means the models tend to overestimate the small persistency values and underestimate the large persistency values.

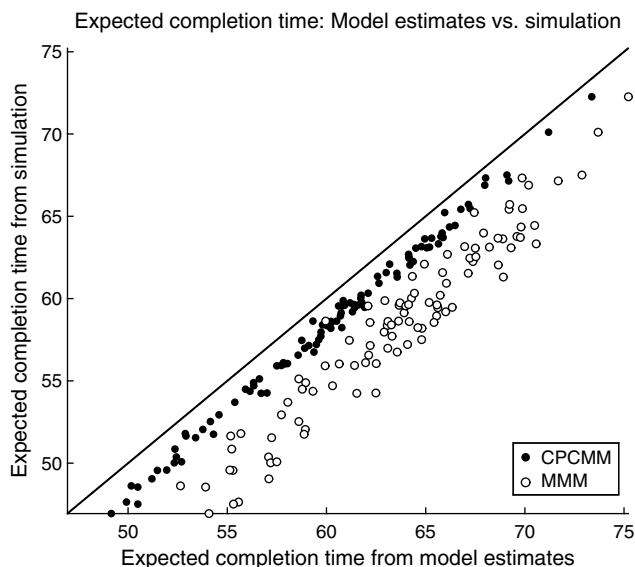
For Figure 5, we first obtain the rankings for the persistency values from the two models and simulation, and then record the number of times that the rank from the models coincide with the rank from simulation. For example, Figure 5 tells us that out of 100 instances, there are 99 instances for which CPCMM identifies the arc with largest persistency (rank 1), whereas MMM identifies that in 94 instances.

Although CPCMM performs better than MMM in giving the persistency ranking, the difference is surprisingly small, as seen in the figure. This seems to suggest that MMM is good in determining the ranking of the persistency for the activities, and probably capturing the cross-moment information has very little additional value on estimating the relative criticality of the activities.

Figure 6 shows the results of expected completion time for these 100 instances. The way to construct the figure is similar to the one for Figure 4.

The results from CPCMM are clearly better than MMM, whose results could be off for certain instances. The maximum difference between the simulation results and CPCMM results is 2.9738, and the average difference is only 1.7597. Considering that the expected completion time from simulation ranges from 46.9413 to 72.2714, the average percentage difference is less than 3%. Hence, the optimal objective values obtained from CPCMM seem to be a good estimate to the project’s expected completion time.

To conclude, CPCMM outperforms MMM in estimating both the arc persistency and the expected completion time,

Figure 6. Expected completion time for 100 instances: CPCMM vs. MMM.**Table 1.** Computation time for CPCMM and MMM.

| Model | Average CPU time (secs) |
|-------|-------------------------|
| CPCMM | 7.89 |
| MMM | 1.31 |

but MMM is good enough if only the arc persistency ranking is concerned. However, the improved accuracy comes with a cost, one of which is the increased computation time. The average computation times of solving one instance for CPCMM and MMM are summarized in Table 1. The machine used to perform all the computation is Acer TravelMate 2,350 with Intel® Celeron® M 1.40 GHz, RAM 768 MB, Microsoft Windows XP Professional SP3. SDPT3 is used to solve the conic programs in MATLAB environment with YALMIP as the user interface (Löfberg 2004, Toh et al. 1999).

6. Conclusion

Although CPCMM is functionally powerful, it is a challenging problem to solve computationally. Compared to MMM, one major drawback of CPCMM is the size of the completely positive matrix, which makes it more difficult to solve computationally. When there are n variables in the original problem, i.e., $Z(\tilde{c})$, the completely positive matrix in (C) is of the dimension $(2n + 1) \times (2n + 1)$. MMM, on the other hand, has $2n$ matrices of dimension 2×2 . Large-scale solvers might be an approach for solving the semidefinite relaxations in these cases.

Besides the size of the matrix, capturing the complete positivity condition is the key difficulty. Currently we implement CPCMM using the basic relaxation. Although

it works quite well in simulations, it is conceivable that the gap could be large in specially constructed scenarios. Higher-order relaxations are computationally much more tedious and almost impossible to use for medium to large problems under normal computational power. This basically tells us that CPCMM does not resolve the difficulty of original problems; it shifts the difficulty into the completely positive cones with the hope of better understanding on these cones in the future.

Despite the computational difficulties, CPCMM holds a very strong theoretical foundation, and its flexibility in handling various situations has been demonstrated through the extensions and applications.

7. Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at <http://or.journal.informs.org/>.

Endnotes

1. Rigorously speaking, the convergence of $(\tilde{\mathbf{c}}_\epsilon^*, \mathbf{x}^*(\tilde{\mathbf{c}}_\epsilon^*))$ to $(\tilde{\mathbf{c}}^*, \mathbf{x}^*(\tilde{\mathbf{c}}^*))$ is a weak convergence, i.e., convergence in distribution. However, because it is up to our construction on $(\tilde{\mathbf{c}}_\epsilon^*, \mathbf{x}^*(\tilde{\mathbf{c}}_\epsilon^*))$ and $(\tilde{\mathbf{c}}^*, \mathbf{x}^*(\tilde{\mathbf{c}}^*))$, from Skorohod's Theorem, we can construct them in the same probability space with the same probability measure and $(\tilde{\mathbf{c}}_\epsilon^*, \mathbf{x}^*(\tilde{\mathbf{c}}_\epsilon^*))$ converge to $(\tilde{\mathbf{c}}^*, \mathbf{x}^*(\tilde{\mathbf{c}}^*))$ almost surely (see Borkar 1995).
2. We would like to thank Professor Immanuel Bomze for pointing out this special case and making the proof complete.
3. To be precise, we first generate a sample for each entry of δ independently with univariate uniform distribution of zero-mean and unit standard deviation. Then we apply the variance and correlation requirements to the sample using the Cholesky Decomposition method, i.e., multiplying the sample with the lower triangular matrix obtained from the Cholesky Decomposition of the second-moment matrix, Σ_δ . Hence, the resulting distribution for δ will be a linear combination of uniform distributions and is close to a normal distribution. The reason for using the uniform distribution is to rule out the instances where $\tilde{\mathbf{c}}$ is nonpositive, because Assumption (A1) for the CPCMM requires the cost coefficients to be nonnegative. Moreover, we also test the problem with a truncated multivariate normal distribution, which gives a curve very close to the one given by the uniform distribution. Thus, we omit it from this paper.
4. We used MATLAB function "gallery('randcor', n)" to generate random correlation matrices.
5. The method for generating the sample is the same as the one used for the previous application shown in §5.1.

Acknowledgments

The authors thank Alexander Shapiro and two anonymous reviewers for their valuable comments and suggestions on improving this manuscript.

References

- Aldous, D. 2001. The $\zeta(2)$ limit in the random assignment problem. *Random Structures Algorithms* **18**(4) 381–418.
- Ball, M. O., C. J. Colbourn, J. S. Provan. 1995. Network reliability. M. O. Ball, T. L. Magnanti, C. L. Monma, G. L. Nemhauser. *Network Models*. Handbooks in Operations Research and Management Science, Vol. 7. Elsevier, Amsterdam, 673–762.
- Berman, A., N. Shaked-Monderer. 2003. *Completely Positive Matrices*. World Scientific, Singapore.
- Bertsimas, D., I. Popescu. 2005. Optimal inequalities in probability theory: A convex optimization approach. *SIAM J. Optim.* **15**(3) 780–804.
- Bertsimas, D., K. Natarajan, C. P. Teo. 2004. Probabilistic combinatorial optimization: Moments, semidefinite programming and asymptotic bounds. *SIAM J. Optim.* **15**(1) 185–209.
- Bertsimas, D., K. Natarajan, C. P. Teo. 2006. Persistence in discrete optimization under data uncertainty. *Math. Programming, Ser. B, Issue Optim. Under Uncertainty* **108**(2–3) 251–274.
- Bertsimas, D., X. V. Doan, K. Natarajan, C. P. Teo. 2010. Models for minimax stochastic linear optimization problems with risk aversion. *Math. Oper. Res.* **35**(3) 580–602.
- Billingsley, P. 1995. *Probability and Measure*, 3rd ed. John Wiley & Sons, New York.
- Bollobás, B., D. Gamarnik, O. Riordan, B. Sudakov. 2004. On the value of a random minimum length Steiner tree. *Combinatorica* **24**(2) 187–207.
- Bomze, I. M., M. Dür, E. D. Klerk, C. Roos, A. J. Quist, T. Terlaky. 2000. On copositive programming and standard quadratic optimization problems. *J. Global Optim.* **18**(4) 301–320.
- Borkar, V. 1995. Spaces of probability measure. *Probability Theory: An Advanced Course*. S. Axler, F. W. Gehring, P. R. Halmos, eds. Springer, New York.
- Burer, S. 2009. On the copositive representation of binary and continuous nonconvex quadratic programs. *Math. Programming* **120**(2) 479–495.
- Calafiore, G., L. E. Ghaoui. 2006. On distributionally robust chance-constrained linear programs. *Optim. Theory Appl.* **130**(1) 1–22.
- Davis, R., A. Prieditis. 1993. The expected length of a shortest path. *Inform. Processing Lett.* **46**(3) 135–141.
- de Klerk, E., D. V. Pasechnik. 2002. Approximation of the stability number of a graph via copositive programming. *SIAM J. Optim.* **12**(4) 875–892.
- Delage, E., Y. Ye. 2010. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Oper. Res.* **55**(3) 596–612.
- Dür, M., G. Still. 2007. Interior points of the completely positive cone. *Electronic J. Linear Algebra* **17** 48–53.
- Frieze, A. M. 1985. On the value of a random minimum spanning tree problem. *Discrete Appl. Math.* **10**(1) 47–56.
- Hagstrom, J. N. 1988. Computational complexity of PERT problems. *Networks* **18**(2) 139–147.
- Ishii, K. 1959. On a method for generalizations of Tchebycheff's inequality. *Ann. Instit. Statist. Math.* **10**(2) 65–88.
- Karlin, S., S. Studden. 1966. *Tchebycheff Systems: With Applications in Analysis and Statistics*. Interscience Publishers, New York.
- Kemperman, J. H. B., M. Skibinsky. 1993. Covariance spaces for measures on polyhedral sets. *Stochastic Inequalities*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, Vol. 22, 182–195.
- Kulkarni, V. G. 1986. Shortest paths in networks with exponentially distributed arc lengths. *Networks* **16**(3) 255–274.
- Lasserre, J. 2002. Bounds on measures satisfying moment conditions. *Ann. Appl. Probab.* **12**(3) 1114–1137.
- Lasserre, J. 2010. A "Joint+Marginal" approach to parametric polynomial optimization. *SIAM J. Optim.* **20**(4) 1995–2022.
- Linusson, S., J. Wästlund. 2004. A proof of Parisi's conjecture on the random assignment problem. *Probab. Theory Related Fields* **128**(3) 419–440.
- Löfberg, J. 2004. YALMIP: A toolbox for modeling and optimization in MATLAB. *Proc. CACSD Conf., Taipei, Taiwan*. <http://users.isy.liu.se/johan/yalmip>.

- Lyons, R., R. Pemantle, Y. Peres. 1999. Resistance bounds for first-passage percolation and maximum flow. *J. Combin. Theory Ser. A* **86**(1) 158–168.
- Mézard, M., G. Parisi. 1987. On the solution of the random link matching problems. *J. de Physique Lettres* **48** 1451–1459.
- Mishra, V. K., K. Natarajan, H. Tao, C.-P. Teo. 2008. Choice modeling with semidefinite optimization when utilities are correlated. Submitted.
- Murty, K. G., S. N. Kabadi. 1987. Some \mathcal{NP} -complete problems in quadratic and nonlinear programming. *Math. Programming* **39**(2) 117–129.
- Nair, C., B. Prabhakar, M. Sharma. 2006. Proofs of the Parisi and Coppersmith-Sorkin random assignment conjectures. *Random Structures Algorithms* **27**(4) 413–444.
- Natarajan, K., M. Song, C. P. Teo. 2009. Persistency model and its applications in choice modeling. *Management Sci.* **55**(3) 453–469.
- Nesterov, Y. 2000. Squared functional systems and optimization problems. H. Frenk, K. Roos, T. Terlaky, S. Zhang, eds. *High Performance Optimization*. Kluwer Academic Press, Dordrecht, The Netherlands, 405–440.
- Parrilo, P. A. 2000. Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization. Ph.D. dissertation, California Institute of Technology, Pasadena.
- Popescu, I. 2007. Robust mean-covariance solutions for stochastic optimization. *Oper. Res.* **55**(1) 98–112.
- Provan, J. S., M. O. Ball. 1983. The complexity of counting cuts and of computing the probability that a graph is connected. *SIAM J. Comput.* **12**(4) 777–788.
- Todd, M. J. 2001. Semidefinite optimization. *Acta Numerica* **10** 515–560.
- Toh, K. C., M. J. Todd, R. H. Tutuncu. 1999. SDPT3—A Matlab software package for semidefinite programming. *Optim. Methods Software* **11**(1–4) 545–581.
- Valiant, L. G. 1979. The complexity of enumeration and reliability problems. *SIAM J. Comput.* **8**(3) 410–442.
- Vandenberghe, L., S. Boyd, K. Comanor. 2007. Generalized Chebyshev bounds via semidefinite programming. *SIAM Rev.* **49**(1) 52–64.
- Wästlund, J. 2010. The mean field traveling salesman and related problems. *Acta Math.* **204**(1) 91–150.
- Zuluaga, L., J. F. Pena. 2005. A conic programming approach to generalized Tchebycheff inequalities. *Math. Oper. Res.* **30**(2) 369–388.