MIXED ZERO-ONE LINEAR PROGRAMS UNDER OBJECTIVE UNCERTAINTY: A COMPLETELY POSITIVE REPRESENTATION

Karthik Natarajan[1] · Chung Piaw Teo[2] · Zhichao Zheng[3]

ABSTRACT. 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 is 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.

KEYWORDS. Mixed 0-1 linear program; Moments; Completely positive program

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,

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

(1.1)

where $\boldsymbol{x} \in \mathbb{R}_+^n$ is the decision vector and $\tilde{\boldsymbol{c}}$ is the random objective coefficient vector. The subset $\mathcal{B} \subseteq \{1, \ldots, n\}$ indexes the 0-1 decision variables and $\{1, \ldots, n\} \backslash \mathcal{B}$ indexes the continuous decision variables. Problem (1.1) includes the class of 0-1 integer programs and the class of linear programs as special cases. Given distributional information on $\tilde{\boldsymbol{c}}$, the object of interest is the expected optimal value $\mathbf{E}[Z(\tilde{\boldsymbol{c}})]$.

[1]Department of Management Sciences, City University of Hong Kong, Hong Kong. Email: knataraj@cityu.edu.hk. Part of this work was supported by the NUS Academic Research Fund R-146-000-112-112.

[2]Department of Decision Sciences, NUS Business School, National University of Singapore, Singapore 117591. Email: bizteocp@nus.edu.sg

[3]Department of Decision Sciences, NUS Business School, National University of Singapore, Singapore 117591. Email: zhichao@nus.edu.sg. Part of this work was done when the author was at the Department of Mathematics, Faculty of Science, National University of Singapore.

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 $\tilde{\boldsymbol{c}}$ 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 $Z(\tilde{\boldsymbol{c}})$. For an excellent review on applications and algorithms for the network reliability analysis problem, the reader is referred to Ball et al. [2] and the references therein. Under the assumption of independence among the random weights, Hagstrom [18] 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 [40] and Provan and Ball [37]. 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 [23] 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. [28] developed a lower bound using a convex quadratic optimization problem,

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

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 [14] proved the following exact result,

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

Similar formulas and asymptotic expressions have been developed for other random optimization problems including the spanning tree [17], assignment [1, 29, 26, 32], traveling salesman [42] and Steiner tree problem [10]. 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 [4, 5, 6, 7, 13, 15, 24, 30, 33, 36, 41]. Links with conic programming has made the approach attractive from a theoretical and computational perspective [7, 34, 38]. In addition, conic programs provide additional insights into the structure of the optimal solution. One such parameter of importance is the *persistency* of a binary variable, which is defined as the probability that the variable takes value 1 in an optimal solution [6]. The notion of persistency generalizes "criticality index" in project networks and "choice probability" in discrete choice models [6, 30, 33].

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

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

where $\tilde{c} \sim (\boldsymbol{\mu}, \Sigma)^+$ denotes the set of feasible multivariate distributions supported on $\mathbb{R}^n_+$ with first moment vector $\boldsymbol{\mu}$ and second moment matrix Σ. In the situation without the support requirement, the corresponding problem with sup replaced by inf in the model reduces to a simple Jensen bound [4]. 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 [28], 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 $tr(A)$, is sum of the diagonal entries of A. The inner product between two matrices of appropriate dimensions A and B is denoted as $A \bullet B = tr(A^T B)$. I_n is used to represent the identity matrix of dimension $n \times n$. For any convex cone $\mathcal{K}$, the dual cone is denoted as $\mathcal{K}^*$ and the closure of the cone is denoted as

$\overline{\mathcal{K}}$. 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,

$$\mathcal{S}_n^+ := \left\{ A \in S_n \,\middle|\, \forall \boldsymbol{v} \in \mathbb{R}^n, \ \boldsymbol{v}^T A \boldsymbol{v} \geq 0 \right\}.$$

$A \succeq 0$ 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

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

or equivalently,

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

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

$$\mathcal{CO}_n := \left\{ A \in \mathcal{S}_n \,\middle|\, \forall \boldsymbol{v} \in \mathbb{R}_+^n, \ \boldsymbol{v}^T A \boldsymbol{v} \geq 0 \right\}.$$

$A \succeq_{cp} (\succeq_{co}) 0$ 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 [38]. 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.

Marginal Moment Model (MMM).

Under the MMM [5, 6], information on $\tilde{\boldsymbol{c}}$ 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. While 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

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

For 0-1 integer programs, Bertsimas, Natarajan and Teo [5, 6] 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 since 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, Song and Teo [33] 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 $\tilde{c}$ 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, Natarajan, Tao and Teo [30] computed the upper bound on the expected maximum of n random variables where the support is in $\mathbb{R}^n$,

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

The SDP formulation developed therein is based on an extreme point enumeration technique. Bertsimas, Vinh, Natarajan and Teo [4] showed that generalizing this model to general linear programs leads to $\mathcal{NP}$-hard problems. In a related vein, Lasserre [25] 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, Boyd and Comanor [41] 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

$$C = \left\{ \boldsymbol{c} \in \mathbb{R}^n \,\middle|\, \boldsymbol{c}^T A_i \boldsymbol{c} + 2\boldsymbol{b}_i^T \boldsymbol{c} + d_i < 0, \, \forall i = 1, \ldots, m \right\},$$

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

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

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 $\tilde{c}$. For other multivariate generalizations of Chebyshev's inequality, the reader is referred to [7, 19, 24, 41, 43].

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)

[11, 12, 22]. 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) [3]. A review on COP and CPP is provided in Appendix II.

Burer [12] recently showed 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,

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

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 [11, 12, 16]. For instance, the problem of testing if a given matrix is copositive is known to be in the co-$\mathcal{NP}$-complete class [31]. 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 [12]. One such result is a well-known hierarchy of linear and semidefinite representable cones that approximate the copositive and completely positive cone [11, 35, 22]. For the numerical experiments, we restrict our attention to the following simple relaxation of CPP,

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

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 (1.1) as

$$
\mathcal{L} := \left\{ \boldsymbol{x} \geq \boldsymbol{0} \,\middle|\, \boldsymbol{a}_i^T \boldsymbol{x} = b_i, \forall i = 1, \dots, m \right\},
$$

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

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

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

Assumptions.

(A1) The set of distributions of $\tilde{\boldsymbol{c}}$ 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) $\boldsymbol{x} \in \mathcal{L} \Rightarrow x_j \leq 1, \forall j \in \mathcal{B}$.

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

The nonnegativity of $\tilde{\boldsymbol{c}}$ 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 [7, 21, 31]. 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

$$\mathcal{M}_2(\Omega) = \left\{ \lambda\left(1, \boldsymbol{\mu}, \Sigma\right) \,\middle|\, \lambda \geq 0, \boldsymbol{\mu} = \mathbf{E}[\tilde{\boldsymbol{c}}], \Sigma = \mathbf{E}[\tilde{\boldsymbol{c}}\tilde{\boldsymbol{c}}^T], \right.$$
$$\left. \text{for some random vector } \tilde{\boldsymbol{c}} \text{ with support } \Omega \right\}.$$

From the theory of moments [20, 21], it is well-known that the dual of this moment cone is given as

$$\mathcal{M}_2(\Omega)^* = \left\{ (w_0, \boldsymbol{w}, \boldsymbol{W}) \,\middle|\, w_0 + \boldsymbol{w}^T\boldsymbol{c} + \boldsymbol{c}^T W \boldsymbol{c} \geq 0 \text{ for all } \boldsymbol{c} \in \Omega \right\}.$$

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 the cone of copositive matrices and the closure of the moment cone is the cone of completely positive matrices. Testing for (A1) is thus a difficult problem since

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

Assumption (A2) is easy to enforce and is based on Burer's paper [12]. 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 $\mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\right]$ is finite and hence the supremum is finite.

3.2. Formulation.

Denote $x_j(\boldsymbol{c})$ to be the value of the variable x_j in an optimal solution to Problem (1.1) obtained under the specific $\boldsymbol{c}$. When $\tilde{\boldsymbol{c}}$ is random, $\boldsymbol{x}(\tilde{\boldsymbol{c}})$ is also random. For continuous distributions, the support of $\tilde{\boldsymbol{c}}$ over which Problem (1.1) has multiple optimal solutions has measure zero. For discrete distributions with possibly multiple optimal solutions in a support of strictly positive measure, we define $\boldsymbol{x}(\boldsymbol{c})$ to an optimal solution randomly selected from the set of optimal solutions at $\boldsymbol{c}$. Next, we define

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

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

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

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

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

Then

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

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

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

$$
\boldsymbol{a}_i^T \boldsymbol{p} = b_i \ \forall i = 1, \ldots, m.
$$

Using a lifting technique, we obtain

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

Taking expectations again,

$$
\boldsymbol{a}_i^T X \boldsymbol{a}_i = b_i^2 \ \forall i = 1, \ldots, m.
$$

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

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

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

(C)
$$Z_C = \begin{array}{ll} \max & I_n \bullet Y \end{array}$$
$$\begin{array}{lll} \text{s.t.} & \boldsymbol{a}_i^T \boldsymbol{p} = b_i & \forall i = 1, \ldots, m \\ & \boldsymbol{a}_i^T X \boldsymbol{a}_i = b_i^2 & \forall i = 1, \ldots, m \\ & X_{jj} = p_j & \forall j \in \mathcal{B} \subseteq \{1, \ldots, n\} \\ & \begin{pmatrix} 1 & \boldsymbol{\mu}^T & \boldsymbol{p}^T \\ \boldsymbol{\mu} & \Sigma & Y^T \\ \boldsymbol{p} & Y & X \end{pmatrix} \succeq_{cp} 0 & \end{array}$$

Note that from Assumption (A3), the variables $\boldsymbol{p}$ 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).

Since 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 [12]. For completeness, we outline his relevant proofs in Propositions 3.1 and 3.2. It should be noted that since the feasible region is bounded in our setting, the recession cone only contains the zero vector.

Define

$$F := \left\{ (\boldsymbol{p}, X) \,\middle|\, \exists Y, \text{ such that } (\boldsymbol{p}, Y, X) \text{ is feasible to (C)} \right\}.$$

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

(3.1)
$$\begin{pmatrix} 1 & \boldsymbol{p}^T \\ \boldsymbol{p} & X \end{pmatrix} = \sum_{k \in \mathcal{K}} \begin{pmatrix} \zeta_k \\ \boldsymbol{z}_k \end{pmatrix} \begin{pmatrix} \zeta_k \\ \boldsymbol{z}_k \end{pmatrix}^T$$

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

Proposition 3.1. (Burer [12]) *For the decomposition* (3.1)*, define* $\mathcal{K}_+ := \left\{ k \in \mathcal{K} \,\middle|\, \zeta_k > 0 \right\}$, *and* $\mathcal{K}_0 := \left\{ k \in \mathcal{K} \,\middle|\, \zeta_k = 0 \right\}$. *Then (i)* $\boldsymbol{z}_k / \zeta_k \in \mathcal{L}$, $\forall k \in \mathcal{K}_+$; *(ii)* $\boldsymbol{z}_k = \boldsymbol{0}$, $\forall k \in \mathcal{K}_0$.

Proof. From the decomposition, we have

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

Then

$$\boldsymbol{a}_i^T \boldsymbol{p} = b_i \Rightarrow \sum_{k \in \mathcal{K}} \zeta_k (\boldsymbol{a}_i^T \boldsymbol{z}_k) = b_i,$$
$$\boldsymbol{a}_i^T X \boldsymbol{a}_i = b_i^2 \Rightarrow \sum_{k \in \mathcal{K}} (\boldsymbol{a}_i^T \boldsymbol{z}_k)^2 = b_i^2.$$

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

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

By the equality conditions of Cauchy-Schwartz inequality,

$$\exists \delta_i, \text{ such that } \delta_i \zeta_k = \boldsymbol{a}_i^T \boldsymbol{z}_k, \ \forall k \in \mathcal{K}, \ \forall i = 1, \ldots, m.$$

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

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

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

Taking $\lambda_k := \zeta_k^2$, $\boldsymbol{v}_k := \boldsymbol{z}_k / \zeta_k$, $\forall k \in \mathcal{K}_+$, we can rewrite the decomposition (3.1) as:

$$(3.2) \qquad \begin{pmatrix} 1 & \boldsymbol{p}^T \\ \boldsymbol{p} & X \end{pmatrix} = \sum_{k \in \mathcal{K}_+} \lambda_k \begin{pmatrix} 1 \\ \boldsymbol{v}_k \end{pmatrix} \begin{pmatrix} 1 \\ \boldsymbol{v}_k \end{pmatrix}^T,$$

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

Proposition 3.2. (Burer [12]) *Consider the decomposition* (3.2). *Let* $\boldsymbol{v}_k = \left(v_{k(1)}, \ldots, v_{k(n)} \right)^T$, *then* $v_{k(j)} \in \{0, 1\}$, $\forall j \in \mathcal{B}$, $\forall k \in \mathcal{K}_+$.

Proof. From the decomposition, we have

$$\boldsymbol{p} = \sum_{k \in \mathcal{K}^+} \lambda_k \boldsymbol{v}_k, \text{ and } X = \sum_{k \in \mathcal{K}^+} \lambda_k \boldsymbol{v}_k \boldsymbol{v}_k^T.$$

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

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

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

$$X_{jj} = p_j \implies \sum_{k \in \mathcal{K}^+} \lambda_k v_{k(j)}^2 = \sum_{k \in \mathcal{K}^+} \lambda_k v_{k(j)}$$
$$\implies \sum_{k \in \mathcal{K}^+} \lambda_k \left(v_{k(j)} - v_{k(j)}^2 \right) = 0.$$

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

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

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

Proof. **Step 1: Decomposing the matrix.**

Consider a completely positive decomposition of the matrix,

$$\begin{pmatrix} 1 & \mu^T & p^{*T} \\ \mu & \Sigma & Y^{*T} \\ p^* & Y^* & X^* \end{pmatrix} = \sum_{k \in \mathcal{K}} \begin{pmatrix} \alpha_k \\ \beta_k \\ \gamma_k \end{pmatrix} \begin{pmatrix} \alpha_k \\ \beta_k \\ \gamma_k \end{pmatrix}^T,$$

where $\alpha_k \in \mathbb{R}_+$, $\beta_k \in \mathbb{R}_+^n$, $\gamma_k \in \mathbb{R}_+^n$, $\forall k \in \mathcal{K}$. Define $\mathcal{K}_+ := \{k \in \mathcal{K} \,|\, \alpha_k > 0\}$, and $\mathcal{K}_0 := \{k \in \mathcal{K} \,|\, \alpha_k = 0\}$. Then

$$\begin{pmatrix} 1 & \mu^T & p^{*T} \\ \mu & \Sigma & Y^{*T} \\ p^* & Y^* & X^* \end{pmatrix} = \sum_{k \in \mathcal{K}_+} \alpha_k^2 \begin{pmatrix} 1 \\ \frac{\beta_k}{\alpha_k} \\ \frac{\gamma_k}{\alpha_k} \end{pmatrix} \begin{pmatrix} 1 \\ \frac{\beta_k}{\alpha_k} \\ \frac{\gamma_k}{\alpha_k} \end{pmatrix}^T + \sum_{k \in \mathcal{K}_0} \begin{pmatrix} 0 \\ \beta_k \\ \gamma_k \end{pmatrix} \begin{pmatrix} 0 \\ \beta_k \\ \gamma_k \end{pmatrix}^T.$$

From Proposition 3.1 and 3.2,

$$\frac{\gamma_k}{\alpha_k} \in \mathcal{L}, \ \forall k \in \mathcal{K}_+ \text{ and } \frac{\gamma_{kj}}{\alpha_k} \in \{0, 1\}, \ \forall j \in \mathcal{B}, \ \forall k \in \mathcal{K}_+.$$

This implies that γ_k/α_k is a feasible solution to the original mixed 0-1 linear program for all $k \in \mathcal{K}_+$. As will be clear in the latter part of the proof, if the random vector $\tilde{c}$ is realized to be β_k/α_k, then γ_k/α_k is not only feasible but also optimal to Problem (1.1). From Proposition 3.1,

$$\gamma_k = \mathbf{0}, \ \forall k \in \mathcal{K}_0.$$

Then the decomposition becomes

$$\begin{pmatrix} 1 & \mu^T & p^{*T} \\ \mu & \Sigma & Y^{*T} \\ p^* & Y^* & X^* \end{pmatrix} = \sum_{k \in \mathcal{K}_+} \alpha_k^2 \begin{pmatrix} 1 \\ \frac{\beta_k}{\alpha_k} \\ \frac{\gamma_k}{\alpha_k} \end{pmatrix} \begin{pmatrix} 1 \\ \frac{\beta_k}{\alpha_k} \\ \frac{\gamma_k}{\alpha_k} \end{pmatrix}^T + \sum_{k \in \mathcal{K}_0} \begin{pmatrix} 0 \\ \beta_k \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} 0 \\ \beta_k \\ \mathbf{0} \end{pmatrix}^T.$$

Step 2: Constructing a sequence of random vectors and feasible solutions.

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

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

This is a valid probability distribution since

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

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

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

The second moment matrix satisfies

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

Similarly, it can be verified that

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

Step 3: Evaluating the limit of the sequence of objective values.

As $\epsilon \downarrow 0$, the random vectors $(\tilde{\boldsymbol{c}}_\epsilon^*, \boldsymbol{x}^*(\tilde{\boldsymbol{c}}_\epsilon^*))$ converge almost surely (a.s.)[4] to $(\tilde{\boldsymbol{c}}^*, \boldsymbol{x}^*(\tilde{\boldsymbol{c}}^*))$, which is defined as

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

---

[4]Rigorously speaking, the convergenc of $(\tilde{\boldsymbol{c}}_\epsilon^*, \boldsymbol{x}^*(\tilde{\boldsymbol{c}}_\epsilon^*))$ to $(\tilde{\boldsymbol{c}}^*, \boldsymbol{x}^*(\tilde{\boldsymbol{c}}^*))$ is a weak convergence, i.e., convergence in distribution. However, since it is up to our construction on $(\tilde{\boldsymbol{c}}_\epsilon^*, \boldsymbol{x}^*(\tilde{\boldsymbol{c}}_\epsilon^*))$ and $(\tilde{\boldsymbol{c}}^*, \boldsymbol{x}^*(\tilde{\boldsymbol{c}}^*))$, from Skorohod's Theorem, we can construct them in the same probability space with the same probability measure and $(\tilde{\boldsymbol{c}}_\epsilon^*, \boldsymbol{x}^*(\tilde{\boldsymbol{c}}_\epsilon^*))$ converge to $(\tilde{\boldsymbol{c}}^*, \boldsymbol{x}^*(\tilde{\boldsymbol{c}}^*))$ almost surely (see Borkar [9]).

From the Continuous Mapping Theorem,

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

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

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

The finiteness of the second moment implies that the sequence $Z(\tilde{\boldsymbol{c}}_\epsilon^*)$ is uniformly integrable. This implies that the sequence of expected optimal objective values converges to the finite value $\mathbf{E}[Z(\tilde{\boldsymbol{c}}^*)]$ (see Billingsley [8]), i.e.,

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

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

$$\mathcal{K}(\mathbb{R}_+^n) = \left\{ \lambda\left(1, \boldsymbol{\mu}', \Sigma', Z'\right) \,\middle|\, \lambda \geq 0, Z' = \mathbf{E}[Z(\tilde{\boldsymbol{c}})] \text{ for some random vector } \boldsymbol{c} \sim (\boldsymbol{\mu}', \Sigma')^+ \right\}.$$

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

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

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

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

or equivalently,

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

The point $(1, \boldsymbol{\mu}, \Sigma, Z_P)$ lies on the boundary of this closed convex cone and hence $Z_P \geq \mathbf{E}[Z(\tilde{\boldsymbol{c}}^*)]$.

Thus,

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

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. $\qquad\square$

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 non-zero for some $k \in \mathcal{K}_0$ and $\boldsymbol{c}^*$ might not be strictly feasible due to the second moment matrix constraint.

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

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

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 $\tilde{\boldsymbol{c}}$ 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 $(\boldsymbol{p}^*, Y^*, X^*)$ be an optimal solution to (C), then there exists a nonnegative random objective coefficient vector $\tilde{\boldsymbol{c}}^*$ and feasible solutions $\boldsymbol{x}^*(\tilde{\boldsymbol{c}}^*)$ that satisfies

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

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, $\forall j \in \mathcal{B}$:

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

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 an limiting distribution $\boldsymbol{c}^*$.

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 $\tilde{\boldsymbol{c}}$ 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.,

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

The problem of interest is

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

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

(CS)
$$\max \quad I_n \bullet Y$$
$$\text{s.t.} \quad \boldsymbol{a}_i^T \boldsymbol{p} = b_i \qquad\qquad \forall i = 1, \ldots, m$$
$$\boldsymbol{a}_i^T X \boldsymbol{a}_i = b_i^2 \qquad\qquad \forall i = 1, \ldots, m$$
$$X_{jj} = p_j \qquad\qquad \forall j \in \mathcal{B} \subseteq \{1, \ldots, n\}$$
$$\begin{pmatrix} 1 & \boldsymbol{\mu}^T & \boldsymbol{p}^T \\ \boldsymbol{\mu} & \Sigma & Y^T \\ \boldsymbol{p} & Y & X \end{pmatrix} \in \overline{\mathcal{M}_2(\mathbb{R}^n \times \mathbb{R}_+^n)}$$

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

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

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),

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

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 $k \in \mathcal{K}_+$. Let the matrix B be defined as

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

Define

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

where $\tilde{\boldsymbol{z}}^* \sim \mathbb{N}(\boldsymbol{0}, B)$ 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 since $\sum_{k \in \mathcal{K}_+} \alpha_k^2 = 1$. Furthermore, the mean of the marginal distribution of $\tilde{\boldsymbol{c}}^*$ satisfies

$$\mathbf{E}[\tilde{\boldsymbol{c}}^*] \;=\; \sum_{k \in \mathcal{K}_+} \alpha_k \boldsymbol{\beta}_k + \mathbf{E}[\tilde{\boldsymbol{z}}^*]$$
$$=\; \boldsymbol{\mu}.$$

Similarly, the second moment matrix satisfies

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

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

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

Since 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

$$
\begin{pmatrix} 1 & \boldsymbol{\mu}^T & \boldsymbol{p}^T \\ \boldsymbol{\mu} & \Sigma & Y^T \\ \boldsymbol{p} & Y & X \end{pmatrix} \in \overline{\mathcal{M}_2(\mathbb{R}^n \times \mathbb{R}_+^n)} \implies \begin{pmatrix} 1 & \boldsymbol{\mu}^T & \boldsymbol{p}^T \\ \boldsymbol{\mu} & \Sigma & Y^T \\ \boldsymbol{p} & Y & X \end{pmatrix} \succeq 0, \; \begin{pmatrix} 1 & \boldsymbol{p}^T \\ \boldsymbol{p} & X \end{pmatrix} \geq 0.
$$

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 [15]. 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 set $\mathbb{U}$. In this case, the problem is to choose the mean and second moment matrix and the corresponding multivariate distribution that provides the tight bound,

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

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

(CU)
$$\max \quad I_n \bullet Y$$
$$\text{s.t.} \quad \boldsymbol{a}_i^T \boldsymbol{p} = b_i \qquad \forall i = 1, \dots, m$$
$$\boldsymbol{a}_i^T X \boldsymbol{a}_i = b_i^2 \qquad \forall i = 1, \dots, m$$
$$X_{jj} = p_j \qquad \forall j \in \mathcal{B} \subseteq \{1, \dots, n\}$$
$$\begin{pmatrix} 1 & \boldsymbol{\mu}^T & \boldsymbol{p}^T \\ \boldsymbol{\mu} & \Sigma & Y^T \\ \boldsymbol{p} & Y & X \end{pmatrix} \succeq_{cp} 0$$
$$(\boldsymbol{\mu}, \Sigma) \in \mathbb{U}$$

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

Two simple uncertainty sets for the moments are indicated next.

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

$$\begin{aligned} \boldsymbol{\mu}_L &\leq \boldsymbol{\mu} \leq \boldsymbol{\mu}_U \\ \boldsymbol{\Sigma}_L &\leq \Sigma \leq \boldsymbol{\Sigma}_U. \end{aligned}$$

(b) Delage and Ye [15] proposed the following uncertainty set in the moments parameterized by $\gamma_1 \geq 0$ and $\gamma_2 \geq 0$,

$$\begin{aligned} (\mathbf{E}[\tilde{\boldsymbol{c}}] - \boldsymbol{\mu}_0)^T Q_0^{-1} (\mathbf{E}[\tilde{\boldsymbol{c}}] - \boldsymbol{\mu}_0) &\leq \gamma_1, \\ \mathbf{E}\left[(\tilde{\boldsymbol{c}} - \boldsymbol{\mu}_0)(\tilde{\boldsymbol{c}} - \boldsymbol{\mu}_0)^T\right] &\preceq \gamma_2 Q_0. \end{aligned}$$

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

$$\begin{aligned} (\boldsymbol{\mu} - \boldsymbol{\mu}_0)^T Q_0^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_0) &\leq \gamma_1, \\ \Sigma - 2\boldsymbol{\mu}\boldsymbol{\mu}_0^T + \boldsymbol{\mu}_0\boldsymbol{\mu}_0^T &\preceq \gamma_2 Q_0. \end{aligned}$$

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

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

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 $\tilde{\boldsymbol{c}}^T = (\tilde{\boldsymbol{c}}_1^T, \bar{\boldsymbol{c}}_2^T)$, where $\tilde{\boldsymbol{c}}_1$ is a random vector of dimension k and $\tilde{\boldsymbol{c}}_1 \sim (\boldsymbol{\mu}_1, \Sigma_1)_+$, while $\bar{\boldsymbol{c}}_2$ 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,

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

If c_j is constant, then $Y_{jl} = \mathbf{E}[c_j x_l(\tilde{\boldsymbol{c}})] = c_j \mathbf{E}[x_l(\tilde{\boldsymbol{c}})] = c_j p_l$, $\forall l = 1, \ldots, n$, 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}, $\forall l = 1, \ldots, n$, $\forall j \geq k$, and adjust the objective function accordingly to obtain a slimmer version of (C),

(CC)
$$\begin{aligned}
\max \quad & \sum_{j=1}^{k} Y_{jj} + \sum_{j=k+1}^{n} \bar{c}_j p_j \\
\text{s.t.} \quad & \boldsymbol{a}_i^T \boldsymbol{p} = b_i && \forall i = 1, \ldots, m \\
& \boldsymbol{a}_i^T X \boldsymbol{a}_i = b_i^2 && \forall i = 1, \ldots, m \\
& X_{jj} = p_j && \forall j \in \mathcal{B} \subseteq \{1, \ldots, n\} \\
& \begin{pmatrix} 1 & \boldsymbol{\mu}_1^T & \boldsymbol{p}^T \\ \boldsymbol{\mu}_1 & \Sigma_1 & Y^T \\ \boldsymbol{p} & Y & X \end{pmatrix} \succeq_{cp} 0
\end{aligned}$$

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 $\tilde{\boldsymbol{c}}$, directly use the values of $\tilde{\boldsymbol{c}}_1$ from the decomposition of the completely matrix and plug in $\bar{\boldsymbol{c}}_2$ to form a complete instance of $\tilde{\boldsymbol{c}}$, i.e.,

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

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 $Z(\tilde{\boldsymbol{c}})$.

It is possible to extend the results to bound $\mathbf{E}\left[f\left(Z(\tilde{\boldsymbol{c}})\right)\right]$, 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)
$$\sup_{\tilde{\boldsymbol{c}} \sim (\boldsymbol{\mu}, \Sigma)^+} \mathbf{E}\left[\max_{k \in \mathcal{K}} \{\alpha_k Z(\tilde{\boldsymbol{c}}) + \beta_k\}\right].$$

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

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

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

$$
\begin{aligned}
q^{(k)} &:= \boldsymbol{P}\left(\tilde{\boldsymbol{c}} \in \mathcal{S}_k\right), \\
\boldsymbol{\mu}^{(k)} &:= \mathbf{E}\left[\tilde{\boldsymbol{c}} | \tilde{\boldsymbol{c}} \in \mathcal{S}_k\right] \boldsymbol{P}\left(\tilde{\boldsymbol{c}} \in \mathcal{S}_k\right), \\
\Sigma^{(k)} &:= \mathbf{E}\left[\tilde{\boldsymbol{c}} \tilde{\boldsymbol{c}}^T | \tilde{\boldsymbol{c}} \in \mathcal{S}_k\right] \boldsymbol{P}\left(\tilde{\boldsymbol{c}} \in \mathcal{S}_k\right), \\
\boldsymbol{p}^{(k)} &:= \mathbf{E}[\boldsymbol{x}(\tilde{\boldsymbol{c}}) | \tilde{\boldsymbol{c}} \in \mathcal{S}_k] \boldsymbol{P}(\tilde{\boldsymbol{c}} \in \mathcal{S}_k), \\
Y^{(k)} &:= \mathbf{E}[\boldsymbol{x}(\tilde{\boldsymbol{c}}) \tilde{\boldsymbol{c}}^T | \tilde{\boldsymbol{c}} \in \mathcal{S}_k] \boldsymbol{P}(\tilde{\boldsymbol{c}} \in \mathcal{S}_k), \\
X^{(k)} &:= \mathbf{E}[\boldsymbol{x}(\tilde{\boldsymbol{c}}) \boldsymbol{x}(\tilde{\boldsymbol{c}}) | \tilde{\boldsymbol{c}} \in \mathcal{S}_k] \boldsymbol{P}(\tilde{\boldsymbol{c}} \in \mathcal{S}_k).
\end{aligned}
$$

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

$$
\begin{aligned}
(\text{CF}) \quad \max \quad & \sum_{k \in \mathcal{K}} \left(\alpha_k I_n \bullet Y^{(k)} + \beta_k q^{(k)} \right) \\
\text{s.t.} \quad & \boldsymbol{a}_i^T \boldsymbol{p}^{(k)} = b_i q^{(k)} && \forall i = 1, \ldots, m, \forall k \in \mathcal{K} \\
& \boldsymbol{a}_i^T X^{(k)} \boldsymbol{a}_i = b_i^2 q^{(k)} && \forall i = 1, \ldots, m, \forall k \in \mathcal{K} \\
& X_{jj}^{(k)} = p_j^{(k)} && \forall j \in \mathcal{B}, \forall k \in \mathcal{K} \\
& \begin{pmatrix} q^{(k)} & \boldsymbol{\mu}^{(k)T} & \boldsymbol{p}^{(k)T} \\ \boldsymbol{\mu}^{(k)} & \Sigma^{(k)} & Y^{(k)T} \\ \boldsymbol{p}^{(k)} & Y^{(k)} & X^{(k)} \end{pmatrix} \succeq_{cp} 0 && \forall k \in \mathcal{K} \\
& \sum_{k \in \mathcal{K}} \begin{pmatrix} q^{(k)} & \boldsymbol{\mu}^{(k)T} \\ \boldsymbol{\mu}^{(k)} & \Sigma^{(k)} \end{pmatrix} = \begin{pmatrix} 1 & \boldsymbol{\mu}^T \\ \boldsymbol{\mu} & \Sigma \end{pmatrix}
\end{aligned}
$$

Proving (PF) is solvable as (CF) is very similar to what we have done for Theorem 3.3, and only requires 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 $\tilde{\boldsymbol{c}}$. With probability $q^{(k)*}$ (the value of $q^{(k)}$ in the optimal solution to (CF)), construct $\tilde{\boldsymbol{c}}$ using the completely positive decomposition of the kth matrix as in the proof of Theorem 3.3. The final limiting distribution for $\tilde{\boldsymbol{c}}$ would be a mixture distribution of $|\mathcal{K}|$ types and satisfy the moment conditions. The decomposition also provides the $|\mathcal{K}|$ 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 $\boldsymbol{c} = (c_1, c_2, \ldots, c_n)$ of n numbers can be formulated as an optimization problem as follows,

$$
\text{(OS)} \qquad\qquad
\begin{aligned}
\max \quad & \boldsymbol{c}^T \boldsymbol{x} \\
\text{s.t.} \quad & \sum_{j=1}^{n} x_j = 1 \\
& \boldsymbol{x} \geq \boldsymbol{0}
\end{aligned}
$$

Suppose $c_1 > c_2 > \cdots > c_n$. Then the optimal solution to (OS) is $x_1^* = 1$, $x_j^* = 0$, $\forall j = 2, \ldots, n$. For the sensitivity analysis problem, consider a perturbation in the objective coefficients. Let $\tilde{c}_j = c_j + \epsilon \tilde{\delta}_j$, $\forall j = 1, \ldots, n$, 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 jth 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 $\mathbf{E}[\tilde{\boldsymbol{\delta}} \tilde{\boldsymbol{\delta}}^T] = \Sigma_{\boldsymbol{\delta}}$. The random vector $\tilde{\boldsymbol{c}} \sim (\boldsymbol{c}, \Sigma)_+$, and $\Sigma = \boldsymbol{c}\boldsymbol{c}^T + \epsilon^2 I_n$ for the independent case, while $\Sigma = \boldsymbol{c}\boldsymbol{c}^T + \epsilon^2 \Sigma_{\boldsymbol{\delta}}$ for the correlated case. The problem is to identify the probability that the original optimal solution still remains optimal i.e., $\boldsymbol{P}(\tilde{c}_1 \geq \tilde{c}_j, \forall j = 1, \ldots, n)$ as the value of ϵ increases. Moreover, $\boldsymbol{P}(\tilde{c}_1 \geq \tilde{c}_j, \forall j = 1, \ldots, n)$ 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 $\tilde{\boldsymbol{c}}$ used in both cases, is randomly generated as

$$\boldsymbol{c} = (19.7196, 19.0026, 17.8260, 16.4281, 15.2419, 12.1369, 9.1293, 8.8941, 4.6228, 0.3701)^T,$$

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

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

While we have carried out the tests on many different values of $\boldsymbol{c}$ and $\Sigma_{\boldsymbol{\delta}}$, 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 1000-sized sample for $\tilde{\boldsymbol{\delta}}$ satisfying the moment conditions.[5] Then we solve these samples (i.e., 1000 deterministic problems) to estimate $\boldsymbol{P}(\tilde{c}_1 \geq \tilde{c}_j, \forall j = 1, \ldots, n)$.

Figure 5.1 and 5.2 show the results for case (1) and (2) respectively.

From Figure 5.1, 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 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 $\boldsymbol{P}(\tilde{c}_1 > \tilde{c}_j, \forall j = 2, \ldots, n)$ which can be supported by an extreme distribution [41]. Secondly, it bounds $\boldsymbol{P}(\tilde{c}_1 > \tilde{c}_j, \forall j = 2, \ldots, n)$ rather than $\boldsymbol{P}(\tilde{c}_1 \geq \tilde{c}_j, \forall j = 2, \ldots, n)$, 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.

---

[5]To be precise, we first generate a sample for each entry of $\boldsymbol{\delta}$ independently with univariate uniform distribution of zero-mean and unit standard deviation. Then we apply the variance and correlation requirements to the sample using Cholesky Decomposition method, i.e. multiplying the sample with the lower triangular matrix obtained from the Cholesky Decomposition of the second moment matrix, $\Sigma_{\boldsymbol{\delta}}$. Hence, the resulting distribution for $\boldsymbol{\delta}$ will be 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 $\tilde{\boldsymbol{c}}$ is nonpositive, since 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.
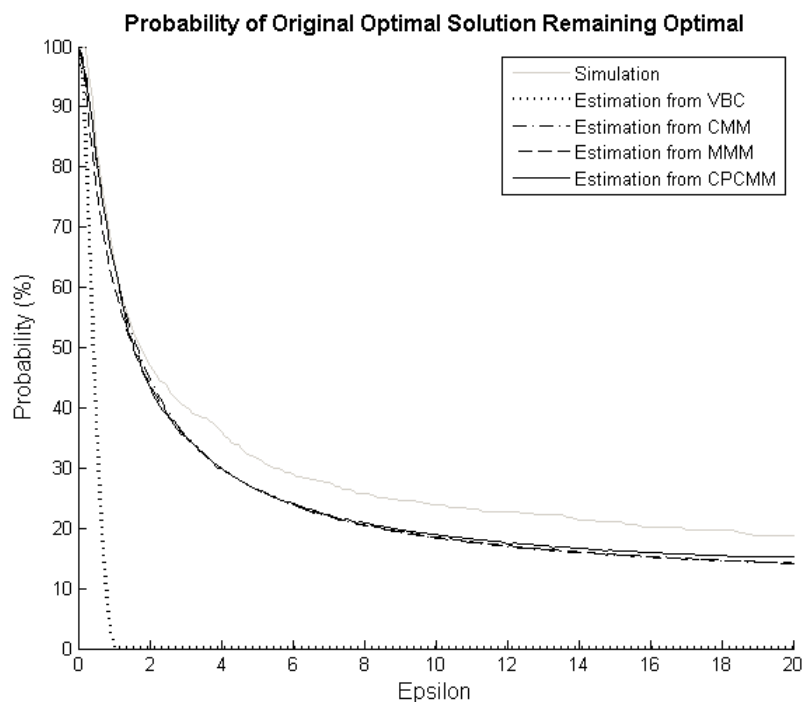
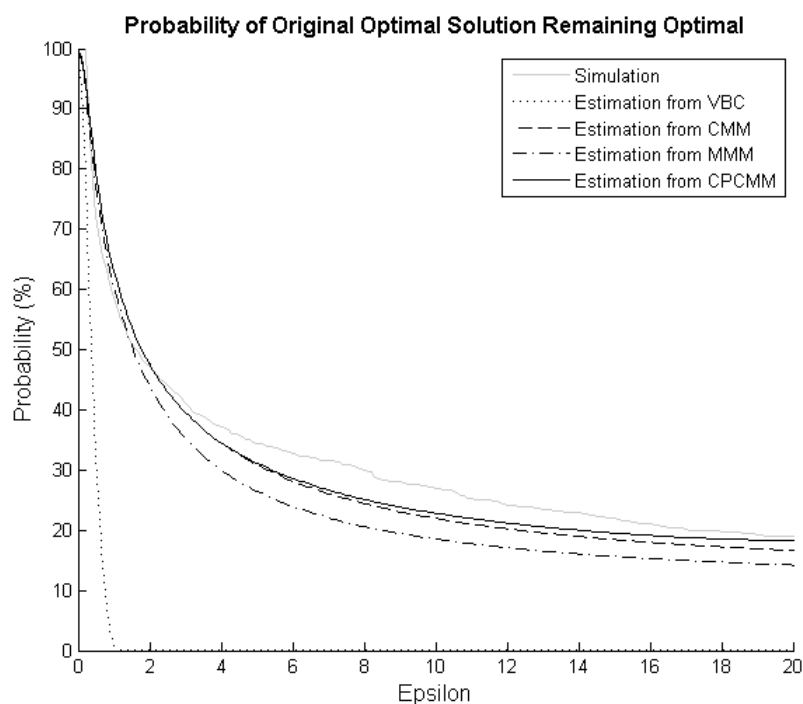FIGURE 5.1. Results for stochastic sensitivity analysis with independent perturbation

FIGURE 5.2. Results for stochastic sensitivity analysis with correlated 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 5.2 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 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,

$$\max \sum_{(i,j)\in \mathcal{ARCS}} c_{ij}x_{ij}$$

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

$$x_{ij} \geq 0, \, \forall (i,j) \in \mathcal{ARCS}$$

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 5.3. 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 ($\sim$Uniform(0,5)), standard deviations ($\sim$Uniform(0,2)) and correlations for the arc lengths.[6] 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 1000 instances of arc lengths were generated for each

---

[6]We used MATLAB function "gallery('randcor',n)" to generate random correlation matrices.
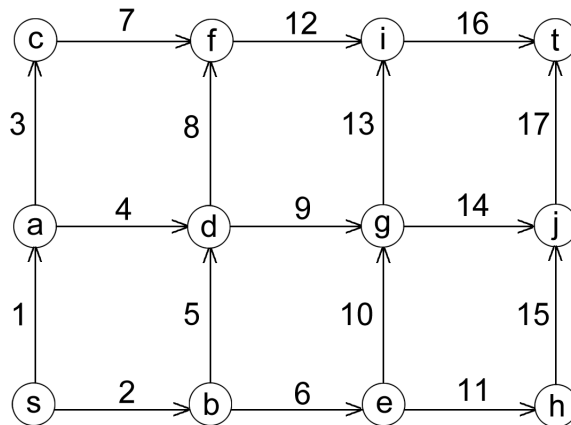
FIGURE 5.3. Project network

data set [7] satisfying the moment conditions. In simulation, the linear program was solved for each sample to assess the persistency values.

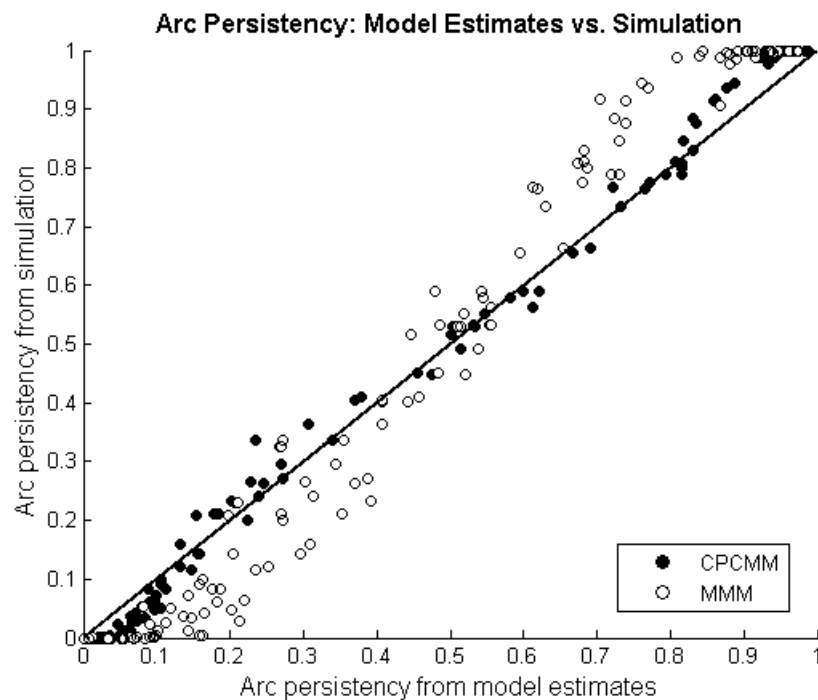Figure 5.4 and 5.5 show the results of arc persistency estimates for the 100 instances.

FIGURE 5.4. Arc persistency values for 100 instances: CPCMM vs. MMM

To interpret Figure 5.4: 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)*. Since there are 17 arcs for one instance in our problem, there are 3400 ($17 \times 2 \times 100$) points in Figure 5.4.

---

[7]The method for generating the sample is the same as the one used for previous application shown in §5.1.

Over 94% of the persistency values are less than 0.1 which means that around 3200 points are clustered in the bottom left corner of Figure 5.4. 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, while the error in the estimates is presented as a deviation from the line.

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.
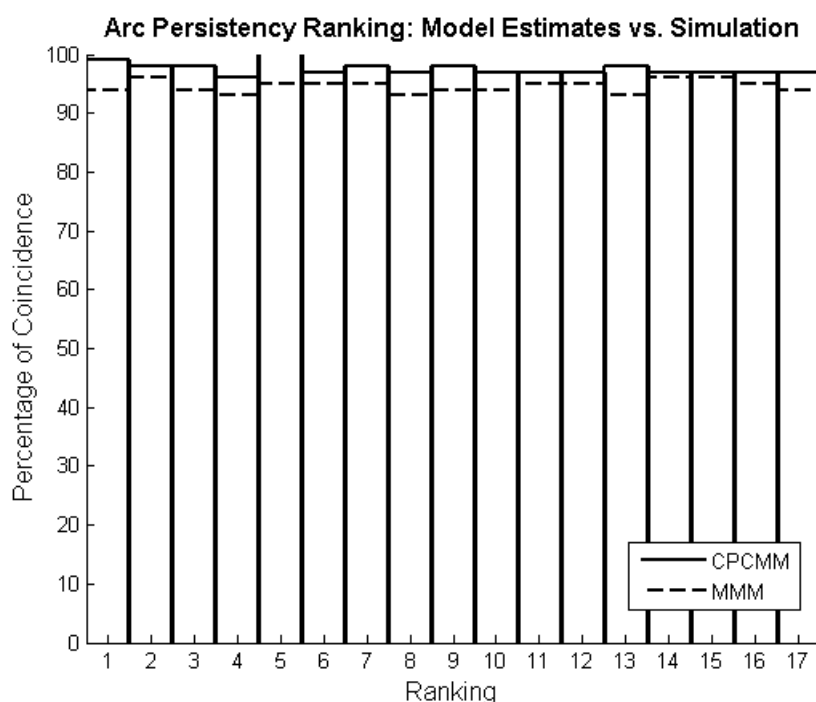

FIGURE 5.5. Arc persistency ranking for 100 instances: CPCMM vs. MMM

For Figure 5.5, 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 5.5 tells us that out of 100 instances, there are 99 instances for which CPCMM identifies the arc with largest persistency (rank 1), while 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 5.6 on the facing page shows the results of expected completion time for these 100 instances. The way to construct the figure is similar to the one for Figure 5.4.
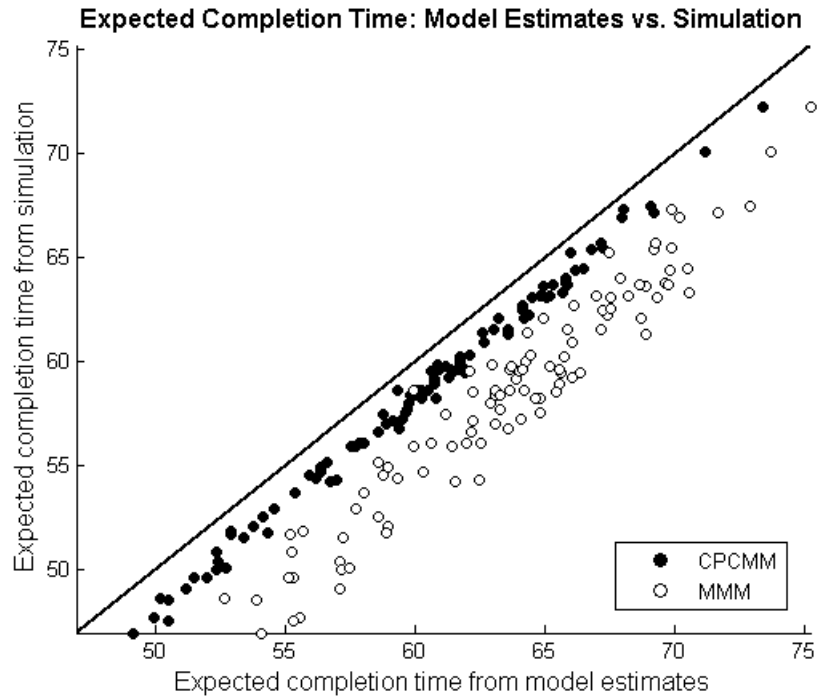
FIGURE 5.6. Expected completion time for 100 instances: CPCMM vs. MMM

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, 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 time of solving one instance for CPCMM and MMM are summarized in Table 1. The machine used to perform all the computation is Acer TravelMate 2350 with Intel® Celeron® M 1.40GHz, RAM 768MB, Microsoft Windows XP Professional SP3. SDPT3 is used to solve the conic programs in MATLAB environment with YALMIP as the user interface [27, 39].

Model	Average CPU Time (secs)
CPCMM	7.89
MMM	1.31

TABLE 1. Computation time for CPCMM and MMM

6. **Conclusion**

Though 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 completely 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 function 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.

References

[1] D. Aldous (2001) *The $\zeta(2)$ limit in the random assignment problem*, Random Structures and Algorithms. **18**, No. 4, pp. 381-418.

[2] M. O. Ball, C. J. Colbourn and J. S. Provan (1995) *Network reliability*, Handbook in Operations Research and Management Science: Network Models, **7**, pp. 673-762.

[3] A. Berman, N. Shaked-Monderer (2003) *Completely Positive Matrices*, World Scientific.

[4] D. Bertsimas, X. V. Doan, K. Natarajan, C. P. Teo (2008) *Models for minimax stochastic linear optimization problems with risk aversion*, To appear in Mathematics of Operations Research.

[5] D. Bertsimas, K. Natarajan, C. P. Teo (2004) *Probabilistic combinatorial optimization: moments, semidefinite programming and asymptotic bounds*, SIAM Journal of Optimization, **15**, No. 1, pp. 185-209.

[6] D. Bertsimas, K. Natarajan, C. P. Teo (2006) *Persistence in discrete optimization under data uncertainty*, Mathematical Programming, Series B, Issue on Optimization under Uncertainty, **108**, No. 2-3, pp. 251-274.

[7] D. Bertsimas, I. Popescu (2005) *Optimal inequalities in probability theory: A convex optimization approach*, SIAM Journal of Optimization, **15**, No. 3, pp. 780-804.

[8] P. Billingsley (1995) *Probability and Measure*, 3rd edition, John Wiley & Sons, New York.

[9] V. Borkar (1995) *Probability Theory: An Advanced Course*, Editorial Board: S. Axler, F. W. Gehring, P. R. Halmos, Springer, New York.

[10] B. Bollobás, D. Gamarnik, O. Riordan and B. Sudakov (2004) *On the value of a random minimum length Steiner Tree*, Combinatorica, **24**, No. 2, pp. 187-207.

[11] I. M. Bomze, M. Dür, E. D. Klerk, C. Roos, A. J. Quist, T. Terlaky (2000) *On copositive programming and standard quadratic optimization problems*, Journal of Global Optimization, **18**, No. 4, pp. 301-320.

[12] S. Burer (2009) *On the copositive representation of binary and continuous nonconvex quadratic programs,* Mathematical Programming, **120**, No. 2, pp. 479-495.

[13] G. Calafiore, L. E. Ghaoui (2006) *On distributionally robust chance-constrained linear programs*, Optimization Theory and Applications, **130**, No. 1, pp. 1-22.

[14] R. Davis, A. Prieditis (1993) *The expected length of a shortest path*, Information Processing Letters, **46**, No. 3, pp. 135-141.

[15] E. Delage, Y. Ye (2008) *Distributionally robust optimization under moment uncertainty with application to data-driven problems*, To appear in Operations Research.

[16] M. Dür, G. Still (2007) *Interior points of the completely positive cone*, Electronic Journal of Linear Algebra, **17**, pp. 48-53.

[17] A. M. Frieze (1985) *On the value of a random minimum spanning tree problem*, Discrete Applied Mathematics, **10**, No. 1, pp. 47-56.

[18] J. N. Hagstrom (1988) *Computational complexity of PERT problems*, Networks, **18**, No. 2, pp. 139-147.

[19] K. Isii (1959) *On a method for generalizations of Tchebycheff's inequality*, Annals of the Institute of Statistical Mathematics, **10**, pp. 65-88.

[20] S. Karlin, S. Studden (1966) *Tchebycheff Systems: with Applications in Analysis and Statistics*, John Wiley & Sons.

[21] J. H. B. Kemperman, M. Skibinsky (1993) *Covariance spaces for measures on polyhedral sets*, Stochastic Inequalities, Institute of Mathematical Statistics Lecture Notes - Monograph Series, **22**, pp. 182-195.

[22] E. de Klerk, D. V. Pasechnik (2002) *Approximation of the stability number of a graph via copositive programming*, SIAM Journal on Optimization, **12**, No. 4, pp. 875-892.

[23] V. G. Kulkarni (1986) *Shortest paths in networks with exponentially distributed arc capacities*, Networks, **18**, pp. 111-124.

[24] J. Lasserre (2002) *Bounds on measures satisfying moment conditions*, The Annals of Applied Probability, **12**, No. 3, pp. 1114-1137.

[25] J. Lasserre (2010) *A "Joint+Marginal" approach to parametric polynomial optimization*, To appear in SIAM Journal of Optimization.

[26] S. Linusson, J. Wästlund (2004) *A proof of Parisi's conjecture on the random assignment problem*, Probability Theory and Related Fields, **128**, 419-440.

[27] J. Löfberg (2004) *YALMIP: A Toolbox for Modeling and Optimization in MATLAB*, In Proceedings of the CACSD Conference, Taipei, Taiwan, available at http://control.ee.ethz.ch/ joloef/yalmip.php.

[28] R. Lyons, R. Pemantle, Y. Peres (1999) *Resistance bounds for first-passage percolation and maximum flow*, Journal of Combinatorial Theory Series A, **86**, No. 1, pp. 158-168.

[29] M. Mézard, G. Parisi (1987) *On the solution of the random link matching problems*, Journal de Physique Lettres, **48**, pp. 1451-1459.

[30] V. K. Mishra, K. Natarajan, H. Tao, C-P. Teo (2008) *Choice modeling with semidefinite optimization when utilities are correlated*, Submitted.

[31] K. G. Murty, S. N. Kabadi (1987) *Some $\mathcal{NP}$-complete problems in quadratic and nonlinear programming*, Mathematical Programming, **39**, No. 2, pp. 117-129.

[32] C. Nair, B. Prabhakar, M. Sharma (2006) *Proofs of the Parisi and Coppersmith-Sorkin random assignment conjectures*, Random Structures and Algorithms, **27**, No. 4, pp. 413-444.

[33] K. Natarajan, M. Song, C. P. Teo (2009) *Persistency model and its applications in choice modeling*, Management Science, **55**, No. 3, pp. 453-469.

[34] Y. Nesterov (2000) *Squared functional systems and optimization problems*, In: High Performance Optimization, Kluwer Academic Press, pp. 405-440.

[35] P. A. Parrilo (2000) *Structured Semidefinite Programs and Semi-algebraic Geometry Methods in Robustness and Optimization*, Ph.D. thesis, California Institute of Technology, Pasadena, CA, available online at: http://www.cds.caltech.edu/~pablo/.

[36] I. Popescu (2007) *Robust mean-covariance solutions for stochastic optimization*, Operations Research, **55**, No. 1, pp. 98-112.

[37] J. S. Provan, M. O. Ball (1983) *The complexity of counting cuts and of computing the probability that a graph is connected*, SIAM Journal on Computing, **12**, No. 4, pp. 777-788.

[38] M. J. Todd (2001) *Semidefinite optimization*, Acta Numerica, **10**, pp. 515-560.

[39] K. C. Toh, M. J. Todd, R. H. Tutuncu (1999) *SDPT3 — a Matlab software package for semidefinite programming*, Optimization Methods and Software, **11**, pp. 545-581.

[40] L. G. Valiant (1979) *The complexity of enumeration and reliability problems*, SIAM Journal on Computing, **8**, pp. 410-442.

[41] L. Vandenberghe, S. Boyd, K. Comanor (2007) *Generalized Chebyshev bounds via semidefinite programming*, SIAM Review, **49**, No. 1, pp. 52-64.

[42] J. Wästlund (2009) *The mean field traveling salesman and related problems*

[43] L. Zuluaga, J. F. Pena (2005) *A conic programming approach to generalized Tchebycheff inequalities*, Mathematics of Operations Research, **30**, No. 2, pp. 369-388.

APPENDIX I. FORMULATIONS OF RELATED MOMENT MODELS

Marginal Moment Model (MMM).

Consider the special case of $Z(\tilde{c})$ with $\mathcal{B} = \{1, \ldots, n\}$, and denote it as $Z_{01}(\tilde{c})$:

$$\begin{aligned} (6.1) \qquad Z_{01}(\tilde{c}) = \quad \max \quad & \tilde{c}^T x \\ \text{s.t.} \quad & a_i^T x = b_i \quad \forall i = 1, \ldots, m \\ & x_j \in \{0, 1\} \quad \forall j = 1, \ldots, n \end{aligned}$$

Bertsimas, Natarajan and Teo [6] solve the following problem:

$$\sup_{\tilde{c}_j \sim (\mu_j, \Sigma_{jj})^+, \, \forall j=1,\ldots,n} \mathbf{E}\left[Z_{01}(\tilde{c})\right]$$

under the assumption that the convex hull of the 0-1 problem is given by the linear constraints

$$\left\{ x \mid a_i^T x = b_i, \forall i = 1, \ldots, m, \, 0 \le x_j \le 1, \forall j = 1, \ldots, n \right\}.$$

The SDP formulation they developed for this problem is:

$$\begin{aligned} \text{(MMM)} \qquad \sup \quad & \sum_{j=1}^{n} y_j \\ \text{s.t.} \quad & a_i^T p = b_i & \forall i = 1, \ldots, m \\ & \mu_j \ge y_j \ge 0 & \forall j = 1, \ldots, n \\ & \begin{pmatrix} 1 & \mu_j \\ \mu_j & \Sigma_{jj} \end{pmatrix} \succeq \begin{pmatrix} p_j & y_j \\ y_j & z_j \end{pmatrix} \succeq 0 & \forall j = 1, \ldots, n \end{aligned}$$

The variables in this formulation can be interpreted as

$$\begin{aligned} p_j &= \mathbf{P}(x_j(\tilde{c}) = 1), \\ y_j &= \mathbf{E}[\tilde{c}_j | x_j(\tilde{c}) = 1] \mathbf{P}(x_j(\tilde{c}) = 1), \\ z_j &= \mathbf{E}[\tilde{c}_j^2 | x_j(\tilde{c}) = 1] \mathbf{P}(x_j(\tilde{c}) = 1). \end{aligned}$$

where $x_j(c)$ is the optimal value of x_j under objective c. The value p_j in the optimal solution is the persistency of corresponding variable under the extremal distribution.

Cross Moment Model (CMM).

Consider the special case of $Z(\tilde{c})$ with $\mathcal{B} = \emptyset$ and denote it as $Z_{LP}(\tilde{c})$:

$$\begin{aligned} (6.2) \qquad Z_{LP}(\tilde{c}) = \quad \max \quad & \tilde{c}^T x \\ \text{s.t.} \quad & A x = b \\ & x \ge 0 \end{aligned}$$

Let $\mathcal{BASIS}$ index the set of all basic feasible solutions to this linear program and $x^{(j)}$ be the jth basic feasible solution. Bertsimas, Vinh, Natarajan and Teo [4] solve the following problem:

$$\sup_{\tilde{c} \sim (\mu, \Sigma)} \mathbf{E}\left[Z_{LP}(\tilde{c})\right]$$

The SDP formulation they developed for this problem is:

(CMM) $\quad\quad \sup \quad \sum\limits_{j\in\mathcal{BASIS}} \boldsymbol{x}^{(j)T}\boldsymbol{y}_j$

$\quad\quad\quad$ s.t. $\quad \sum\limits_{j\in\mathcal{BASIS}} \begin{pmatrix} p_j & \boldsymbol{y}_j^T \\ \boldsymbol{y}_j & Z_j \end{pmatrix} = \begin{pmatrix} 1 & \boldsymbol{\mu}^T \\ \boldsymbol{\mu} & \Sigma \end{pmatrix}$

$\quad\quad\quad\quad\quad\quad \begin{pmatrix} p_j & \boldsymbol{y}_j^T \\ \boldsymbol{y}_j & Z_j \end{pmatrix} \succeq 0 \quad\quad\quad\quad\quad\quad \forall j \in \mathcal{BASIS}$

The variables $p_j \in \mathbb{R}$, $\boldsymbol{y}_j \in \mathbb{R}^n$ and $Z_j \in \mathcal{S}_n$ in this formulation can be interpreted as

$$\begin{aligned} p_j &= \boldsymbol{P}(\text{Basis } j \text{ is optimal}), \\ \boldsymbol{y}_j &= \mathbf{E}[\tilde{\boldsymbol{c}}|\text{Basis } j \text{ is optimal}], \\ Z_j &= \mathbf{E}[\tilde{\boldsymbol{c}}\tilde{\boldsymbol{c}}^T|\text{Basis } j \text{ is optimal}]. \end{aligned}$$

The exponential number of basic feasible solutions for linear programs makes this formulation very large and difficult to use for general linear programs.

Generalized Chebyshev Bounds.

Vandenberghe, Boyd and Comanor [41] consider a generalization of Chebyshev's inequality:

$$\inf_{\tilde{\boldsymbol{c}}\sim(\boldsymbol{\mu},\Sigma)} \boldsymbol{P}\left(\boldsymbol{c}^T A_i \boldsymbol{c} + 2\boldsymbol{b}_i^T \boldsymbol{c} + d_i < 0, \forall i = 1, \dots, m\right)$$

The SDP formulation they proposed for this problem is:

$$\min \quad 1 - \sum_{i=1}^m \lambda_i$$

$\quad\quad$ s.t. $\quad tr(A_i Z_i) + 2\boldsymbol{b}_i^T \boldsymbol{z}_i + d_i \lambda_i \geq 0 \quad\quad \forall i = 1, \dots, m$

$\quad\quad\quad \sum\limits_{i=1}^m \begin{pmatrix} \lambda_i & \boldsymbol{z}_i^T \\ \boldsymbol{z}_i & Z_i \end{pmatrix} \preceq \begin{pmatrix} 1 & \boldsymbol{\mu}^T \\ \boldsymbol{\mu} & \Sigma \end{pmatrix}$

$\quad\quad\quad \begin{pmatrix} \lambda_i & \boldsymbol{z}_i^T \\ \boldsymbol{z}_i & Z_i \end{pmatrix} \succeq 0 \quad\quad\quad\quad\quad\quad \forall i = 1, \dots, m$

where the variables are $Z_i \in \mathcal{S}_n$, $\boldsymbol{z}_i \in \mathbb{R}^n$, and $\lambda_i \in \mathbb{R}$, $\forall i = 1, \dots, m$. The VBC approach can be used in stochastic sensitivity analysis for linear programming problems. Let $\mathcal{B}$ be the index set of the basic variables in a basic feasible solution. The reduced cost $\bar{c}_j$ of the variable x_j is defined as

$$\bar{c}_j := c_j - \boldsymbol{c}_{\mathcal{B}}^T A_{\mathcal{B}}^{-1} A_j,$$

where $A_{\mathcal{B}}$ is the columns of A indexed by $\mathcal{B}$, and A_j is the jth column of A. Let $\mathcal{N} = \{1, \dots, n\} \setminus \mathcal{B}$ be the index set of the nonbasic variables. In order for a given feasible solution of the linear program to be optimal, the reduced cost of all the nonbasic variables must be nonpositive (for maximization problems). This defines a set of linear inequalities on $\tilde{\boldsymbol{c}}$ that we can exploit in the VBC approach, i.e. $\forall j \in \mathcal{N}$,

$$\begin{aligned} \bar{\tilde{c}}_j \leq 0 &\iff \tilde{c}_j - \tilde{\boldsymbol{c}}_{\mathcal{B}}^T A_{\mathcal{B}}^{-1} A_j \leq 0 \\ &\iff \mathbf{E}_j^T \tilde{\boldsymbol{c}}_{\mathcal{N}} - \left(A_{\mathcal{B}}^{-1} A_j\right)^T \tilde{\boldsymbol{c}}_{\mathcal{B}} \leq 0 \\ &\iff \begin{pmatrix} -A_{\mathcal{B}}^{-1} A_j \\ \mathbf{E}_j \end{pmatrix}^T \begin{pmatrix} \tilde{\boldsymbol{c}}_{\mathcal{B}} \\ \tilde{\boldsymbol{c}}_{\mathcal{N}} \end{pmatrix} \leq 0 \end{aligned}$$

Hence, these $|\mathcal{N}|$ inequalities can be viewed as a set constraining $\tilde{\boldsymbol{c}}$,

$$C_{LP} = \left\{ \boldsymbol{c} \in \mathbb{R}^n \left| \left(\begin{array}{c} -A_{\mathcal{B}}^{-1} A_j \\ \mathbf{E}_j \end{array} \right)^T \boldsymbol{c} \leq 0, \ \forall j \in \mathcal{N} \right. \right\}.$$

Then any realization of $\tilde{\boldsymbol{c}}$ falling in this set (i.e. $\tilde{\boldsymbol{c}} \in C_{LP}$) will make the pre-given feasible solution optimal. Therefore, the probability that $\tilde{\boldsymbol{c}}$ lies in the set is just the probability that the given feasible solution is optimal. Furthermore, when A is of rank one, i.e. there is only one basic variable in any feasible solution, that probability is just the persistency of that particular basic variable. One difference between C_{LP} and C in the VBC approach is that all the inequalities in C are strict, so when we apply the VBC model to this problem, the interpretation of the resulting probability has to be changed to the probability that the given feasible solution is the unique optimum. Using the VBC approach on the LP problem, we obtain:

(VBC)
$$\min \quad 1 - \sum_{j \in \mathcal{N}} \lambda_j$$

$$\text{s.t.} \quad \left(\begin{array}{c} -A_{\mathcal{B}}^{-1} A_j \\ \mathbf{E}_j \end{array} \right)^T \boldsymbol{z}_j \geq 0 \qquad \forall j \in \mathcal{N}$$

$$\sum_{j \in \mathcal{N}} \left(\begin{array}{cc} Z_j & \boldsymbol{z}_j \\ \boldsymbol{z}_j^T & \lambda_j \end{array} \right) \preceq \left(\begin{array}{cc} \Sigma & \boldsymbol{\mu} \\ \boldsymbol{\mu}^T & 1 \end{array} \right)$$

$$\left(\begin{array}{cc} Z_j & \boldsymbol{z}_j \\ \boldsymbol{z}_j^T & \lambda_j \end{array} \right) \succeq 0 \qquad \forall j \in \mathcal{N}$$

Thus, for any given feasible solution, we can solve the corresponding (VBC) and obtain the optimal objective value, which is the tightest lower bound on the probability that the given feasible solution is the unique optimal solution to Problem (6.2).

APPENDIX II. COPOSITIVE AND COMPLETELY POSITIVE PROGRAMS

The materials in this section are based on [3, 11, 16].

Properties.

The cone of completely positive matrices, positive semidefinite matrices, copositive matrices and symmetric matrices satisfy

$$\mathcal{CP}_n \subsetneq \mathcal{S}_n^+ \subsetneq \mathcal{CO}_n \subsetneq \mathcal{S}_n,$$

and all these cones are pointed and closed convex.

Proposition. *A is completely positive if and only if there exist vectors $\boldsymbol{v}_1, \boldsymbol{v}_2 \ldots, \boldsymbol{v}_k \in \mathbb{R}_+^n$ such that*

(6.3)
$$A = \sum_{i=1}^k \boldsymbol{v}_i \boldsymbol{v}_i^T.$$

Clearly, the factorization of a completely positive matrix is not unique. The representation (6.3) is called a *rank 1 representation* of A. The decomposition of A into the sum of rank 1 matrices is referred to as a *completely positive decomposition*. The minimal k for which there exists a rank 1 representation is called the *cp-rank* of A. From the definition, it is clear that $cp - rankA \geq rankA$ for every CP matrix A. Equality holds when $n \leq 3$, or when $rankA \leq 2$. An upper bound on cp-rank of A in terms of rank A (when $rankA \geq 2$) is

$$cp - rankA \leq \frac{rankA(rankA + 1)}{2} - 1.$$

Duality.

For a pointed closed convex cone $\mathcal{K}$, its *dual cone* $\mathcal{K}^*$ is defined as:

$$\mathcal{K}^* := \left\{ A \in \mathcal{S}_n \,\middle|\, \forall B \in \mathcal{K}, \, tr(AB) \geq 0 \right\}.$$

$\mathcal{K}^*$ is also a pointed closed convex cone. The following shows a pair of primal-dual conic optimization problems:

(6.4)
$$
\begin{aligned}
\text{(Primal)} \quad &\min \quad C \bullet X \\
&\text{s.t.} \quad A_i \bullet X = b_i \quad i = 1, \ldots, m \\
&\qquad\;\; X \in \mathcal{K}
\end{aligned}
$$

(6.5)
$$
\begin{aligned}
\text{(Dual)} \quad &\max \quad b^T y \\
&\text{s.t.} \quad S = C - \sum_{i=1}^m y_i A_i \\
&\qquad\;\; S \in \mathcal{K}^*
\end{aligned}
$$

If both problems have a strictly feasible point, i.e. $\exists X \in int(\mathcal{K})$, such that

$$A_i \bullet X = b_i, \, \forall i = 1, \ldots, m,$$

and $\exists S \in int(\mathcal{K}^*)$, such that

$$S = C - \sum_{i=1}^{m} y_i A_i,$$

then Problem (6.4) and (6.5) are equivalent, i.e. the duality gap $C \bullet X^* - b^T y^*$ is 0 at optimality, where X^* and y^* are the optimal solutions to Problem (6.4) and (6.5) respectively. We verify next that the cones $\mathcal{CO}_n$ and $\mathcal{CP}_n$ are dual cones in $\mathcal{S}_n$.

Theorem. $\mathcal{CP}_n^* = \mathcal{CO}_n$ and $\mathcal{CO}_n^* = \mathcal{CP}_n$.

Proof. We first prove $\mathcal{CP}_n^* = \mathcal{CO}_n$, and then $\mathcal{CO}_n^* = \mathcal{CP}_n$ will follow since both cones are closed. Let $A \in \mathcal{S}_n$. Then

$$
\begin{aligned}
A \in \mathcal{CP}_n^* &\iff \forall B \in \mathcal{CP}_n,\ tr(AB) \geq 0 \\
&\iff \forall V \in \mathbb{R}_+^{n \times k},\ tr(AVV^T) \geq 0 \\
&\iff \forall V \in \mathbb{R}_+^{n \times k},\ tr(V^T AV) \geq 0 \\
&\iff \forall \boldsymbol{v} \in \mathbb{R}_+^n,\ \boldsymbol{v}^T A \boldsymbol{v} \geq 0 \\
&\iff A \in \mathcal{CO}_n.
\end{aligned}
$$

Thus, $\mathcal{CP}_n^* = \mathcal{CO}_n$. $\qquad\square$

Similarly, it can be shown that $\mathcal{S}_n^{+*} = \mathcal{S}_n^+$, i.e. $\mathcal{S}_n^+$ is self-dual. The interior of the completely positive and copositive cone is characterized as:

$$int(\mathcal{CP}_n) = \left\{ A \in \mathcal{S}_n \,\middle|\, \exists V_1 > 0 \text{ nonsingular }, V_2 \geq 0, \text{ such that } A = [V_1|V_2][V_1|V_2]^T \right\}.$$

$$int(\mathcal{CO}_n) = \left\{ A \in \mathcal{S}_n \,\middle|\, \forall \boldsymbol{v} \in \mathbb{R}_+^n, \boldsymbol{v} \neq \boldsymbol{0},\ \boldsymbol{v}^T A \boldsymbol{v} > 0 \right\}.$$

The notation $[V_1|V_2]$ describes the matrix whose columns are the columns of V_1 augmented with the columns of V_2.

Approximating the Copositive Cone and the Completely Positive Cone.

Klerk and Pasechnik [22] show that there exists a series of linear and semidefinite representable cones approximating the copositive cone $\mathcal{CO}_n$ from the inside, i.e.

$$\exists \text{ closed convex cones } \{\mathcal{K}_n^r : r = 0, 1, 2, \dots\}$$
$$\text{such that } \mathcal{K}_n^r \subseteq \mathcal{K}_n^{r+1},\ \forall r \geq 0 \text{ and } \overline{\bigcup_{r \geq 0} \mathcal{K}_n^r} = \mathcal{CO}_n.$$

The dual cones $\{(\mathcal{K}_n^r)^* : r = 0, 1, 2, \dots\}$ approximate the completely positive cones $\mathcal{CP}_n$ from outside, i.e.

$$(\mathcal{K}_n^r)^* \supseteq \left(\mathcal{K}_n^{r+1}\right)^*,\ \forall r \geq 0 \text{ and } \bigcap_{r \geq 0} (\mathcal{K}_n^r)^* = \mathcal{CO}_n^* = \mathcal{CP}_n.$$

For example, when $r = 0$, we have

$$\mathcal{K}_n^0 = \left\{ A \in \mathcal{S}_n \,\middle|\, \exists X \in \mathcal{S}_n^+, \exists Y \in \mathbb{R}_+^{n \times n}, A = X + Y \right\},$$
$$(\mathcal{K}_n^0)^* = \left\{ A \in \mathcal{S}_n^+ \,\middle|\, A \in \mathbb{R}_+^{n \times n} \right\}.$$

It can be shown that when $n \leq 4$, the above two approximations are exact, i.e. $\mathcal{K}_n^0 = \mathcal{CO}_n$ and $(\mathcal{K}_n^0)^* = \mathcal{CP}_n$.

The higher order approximation ($r \geq 1$) becomes much more complicated. For instance, when $r = 1$, Parrilo (2000, [35]) showed that

$$\mathcal{K}_n^1 = \left\{ A \in \mathcal{S}_n \;\middle|\; \begin{array}{l} \exists M^{(i)} \in \mathcal{S}_n, \; i = 1, 2 \ldots, n \\ \qquad\qquad \text{such that} \end{array} \; \left\{ \begin{array}{l} A - M^{(i)} \succeq 0, \; i = 1, 2 \ldots, n \\ M_{ii}^{(i)} = 0, \; i = 1, 2 \ldots, n \\ M_{jj}^{(i)} + M_{ij}^{(j)} + M_{ji}^{(j)} = 0, \; i \neq j \\ M_{jk}^{(i)} + M_{ik}^{(j)} + M_{ij}^{(k)} \geq 0, \; i \neq j \neq k \end{array} \right. \right\}.$$