Noname manuscript No.
(will be inserted by the editor)

Exploiting Partial Correlations in Distributionally Robust Optimization

Divya Padmanabhan · Karthik Natarajan · Karthyek Murthy

Submitted: October 2018, Accepted: November 2019

Abstract In this paper, we identify partial correlation information structures that allow for simpler reformulations in evaluating the maximum expected value of mixed integer linear programs with random objective coefficients. To this end, assuming only the knowledge of the mean and the covariance matrix entries restricted to block-diagonal patterns, we develop a reduced semidefinite programming formulation, the complexity of solving which is related to characterizing a suitable projection of the convex hull of the set $\{(\mathbf{x}, \mathbf{x}\mathbf{x}') : \mathbf{x} \in \mathcal{X}\}$ where $\mathcal{X}$ is the feasible region. In some cases, this lends itself to efficient representations that result in polynomial-time solvable instances, most notably for the distributionally robust appointment scheduling problem with random job durations as well as for computing tight bounds in the newsvendor problem, Project Evaluation and Review Technique (PERT) networks and linear assignment problems. To the best of our knowledge, this is the first example of a distributionally robust optimization formulation for appointment scheduling that permits a tight polynomial-time solvable semidefinite programming reformulation which explicitly captures partially known correlation information between uncertain processing times of the jobs to be scheduled. We also discuss extensions where the random coefficients are assumed to be non-negative and additional overlapping correlation information is available.

Acknowledgements The research of the first and the second author was partially supported by the MOE Academic Research Fund Tier 2 grant T2MOE1706, "On the Interplay of Choice, Robustness and Optimization in Transportation" and the SUTD-MIT International Design Center grant IDG21700101 on "Design of the Last Mile Transportation System: What Does the Customer Really Want?". The authors would like to thank Teo Chung-Piaw (NUS) for providing some useful references on this research. We would also like to thank the editor Alper Atamturk, the two anonymous reviewers and the associate editor for the detailed perusal of the paper and providing useful suggestions towards improving the paper.

Email: divya_padmanabhan@sutd.edu.sg, karthik_natarajan@sutd.edu.sg, karthyek_murthy@sutd.edu.sg. Singapore University of Technology and Design

1 Introduction

We consider decision problems where the objective involves maximizing the expected value of $Z(\tilde{\mathbf{c}})$, where $\tilde{\mathbf{c}} = (\tilde{c}_1, \tilde{c}_2, \ldots, \tilde{c}_n)$ is a n-dimensional real valued random vector and,

$$Z(\tilde{\mathbf{c}}) = \max \left\{ \tilde{\mathbf{c}}'\mathbf{x} : \mathbf{x} \in \mathcal{X} \right\}, \tag{1}$$

and the set $\mathcal{X}$ is the bounded feasible region to a mixed integer linear program (MILP):

$$\mathcal{X} = \left\{ \mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b},\ \mathbf{x} \geq 0,\ x_j \in \mathcal{Z} \text{ for } j \in \mathcal{I} \subseteq [n] \right\}.$$

The set $\mathcal{X}$ includes the feasible region to linear optimization problems as a special case. The distribution θ of $\tilde{\mathbf{c}}$ is not always known explicitly, while many a time, only a set $\mathcal{P}$ of distributions is known such that $\theta \in \mathcal{P}$. In this scenario, we are interested in computing the quantity $\sup\{\mathbb{E}_\theta[Z(\tilde{\mathbf{c}})] : \theta \in \mathcal{P}\}$, referred to as the distributionally robust bound. In this paper, we focus on the case where only the first moment of $\tilde{\mathbf{c}}$ along with some of the second moments are specified. Applications where such bounds have been previously studied include appointment scheduling, portfolio management and the newsvendor problem among others. For more details, the interested reader may refer to [4,6,17,32,43,44,57].

A precise description of the problem is provided next. Suppose that $\mathcal{N}_1, \ldots, \mathcal{N}_R$ form a partition of the set $\mathcal{N} = \{1, \ldots, n\}$, so that $\mathcal{N} = \bigcup_r \mathcal{N}_r$ and $\mathcal{N}_i \cap \mathcal{N}_j = \varnothing$ for $i \neq j$. We use $n_r = |\mathcal{N}_r|$ to denote the size of the subset $\mathcal{N}_r$. For any vector $\mathbf{a} \in \mathbb{R}^n$, let $\mathbf{a}^r \in \mathbb{R}^{n_r}$ be the subvector formed using elements in $\mathcal{N}_r$ as indices. Let $\mathcal{P}(\mathbb{R}^n)$ be the set of probability distributions on $\mathbb{R}^n$. Suppose that the only information we know about the probability distribution of $\tilde{\mathbf{c}}$ is the first moment specified by $\mathbb{E}[\tilde{\mathbf{c}}] = \boldsymbol{\mu}$ and the second moment matrices $\mathbb{E}[\tilde{\mathbf{c}}^r(\tilde{\mathbf{c}}^r)'] = \boldsymbol{\Pi}^r$ for $r \in [R] = \{1, \ldots, R\}$. In this situation, we are interested in:

$$Z^* = \sup \left\{ \mathbb{E}_\theta \left[Z(\tilde{\mathbf{c}}) \right] : \mathbb{E}_\theta[\tilde{\mathbf{c}}] = \boldsymbol{\mu},\ \mathbb{E}_\theta[\tilde{\mathbf{c}}^r(\tilde{\mathbf{c}}^r)'] = \boldsymbol{\Pi}^r \text{ for } r \in [R],\ \theta \in \mathcal{P}(\mathbb{R}^n) \right\}, \tag{2}$$

which quantifies the maximum possible expected value of $Z(\tilde{\mathbf{c}})$ over all probability distributions θ whose first and second moments are consistent with the moment information specified for the random vector $\tilde{\mathbf{c}}$. We assume that all $r \in [R], \boldsymbol{\Pi}^r \succ \boldsymbol{\mu}^r \boldsymbol{\mu}^{r\prime}$, which is sufficient to guarantee that strong duality holds and in the resulting dual formulations, the optimum is attained. Since R denotes the number of non-overlapping subsets, the partition for $R = n$ corresponds to the case where only the mean and diagonal (variance) entries of the covariance matrix are specified. On the other hand, $R = 1$ corresponds to the case where the mean and the entire covariance matrix is specified. Hence, $\boldsymbol{\Pi}^r$'s denote known sub-matrices of $\boldsymbol{\Pi}$ which denotes the matrix of all second moments of $\tilde{\mathbf{c}}$. Thus, we relax the assumption that the complete matrix $\boldsymbol{\Pi}$ is known for a fixed $R > 1$, but only that some entries are known. The model studied in this paper is closest to the model analyzed in [20]. Therein, the authors studied the distributionally robust bound $\sup\{\mathbb{E}_\theta[\max \tilde{\mathbf{c}}'\mathbf{x} : \mathbf{x} \in \mathcal{X} \subset \{0,1\}^n] : \theta \in \mathcal{P}\}$ where multivariate marginal discrete distributions of non-overlapping subsets of random variables are specified. While the bound is NP-hard to compute, [20] identified two instances of the problem for subset selection and Project Evaluation and Review (PERT) networks, where the tight bound is computable in polynomial-time. We build on the model in [20] by allowing for decision variables $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^n$ and considering moment-based ambiguity sets.

Notations. Let $\mathbb{R}^{m\times n}$ be the set of $m \times n$ matrices with real entries, $\mathcal{S}^k$ be the set of $k \times k$ symmetric matrices and $\mathcal{S}^k_+$ be the set of $k \times k$ symmetric positive semidefinite (psd) matrices. We write $\mathbf{A} \succeq 0$ to denote that $\mathbf{A}$ is a psd matrix. For any positive integer k, we use $[k]$ to denote the set $\{1, 2, \ldots, k\}$. For any subset I of $[k]$ and matrix $\mathbf{A} \in \mathbb{R}^{k\times k}$, we use $\mathbf{A}[I]$ to denote the principal submatrix of $\mathbf{A}$ formed by restricting to rows and columns whose indices are elements of the set I. For any two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, we denote by $\mathbf{x} \circ \mathbf{y}$ the vector formed by component-wise multiplication of $\mathbf{x}$ and $\mathbf{y}$. For any set $\mathcal{E}$, we write $conv(\mathcal{E})$ to denote the convex hull of the set $\mathcal{E}$. For a closed convex cone $\mathcal{K}$, the generalized completely positive cone over $\mathcal{K}$ is defined as the set of symmetric matrices that are representable as the sum of rank one matrices of the form:

$$\mathcal{C}(\mathcal{K}) = \{\mathbf{A} \in \mathcal{S}^n : \exists \mathbf{b}_1, \ldots, \mathbf{b}_p \in \mathcal{K} \text{ such that } \mathbf{A} = \sum_{k\in[p]} \mathbf{b}_k \mathbf{b}_k'\}.$$

For $\mathcal{K} = \mathbb{R}^n_+$, $\mathcal{C}(\mathbb{R}^n_+)$ is the cone of completely positive matrices. The dual to this cone is the cone of copositive matrices denoted as $\mathcal{C}^*(\mathbb{R}^n_+)$. More generally for $\mathcal{K} = \mathbb{R}^n \times \mathbb{R}^n_+$, $\mathcal{C}(\mathbb{R}^n \times \mathbb{R}^n_+)$ is given by

$$\mathcal{C}(\mathbb{R}^n \times \mathbb{R}^n_+) = \left\{ \begin{bmatrix} \mathbf{A} & \mathbf{B}' \\ \mathbf{B} & \mathbf{C} \end{bmatrix} \in \mathcal{S}^{n\times n}_+ : \mathbf{C} \in \mathcal{C}(\mathbb{R}^n_+) \right\}. \tag{3}$$

2 Literature review

There is now a fairly significant literature on methods that either compute the tight distributionally robust bound Z^* or weaker upper bounds on Z^* for mixed integer linear optimization problems [6,10,17,28,32,43,44,59]. In general, one of the difficulties that arises in exact formulations to compute Z^* under moment-based ambiguity sets is that it involves optimization over the cone of completely positive matrices, which is typically intractable. This naturally leads to the question of identifying specific instances for which the problem is tractable, which is our focus in this paper. We review some of the key concepts briefly next, before discussing the contributions in this work.

2.1 Exact Reformulations: Completely Positive Matrices and Quadratic Forms

Problem (2) for $R = 1$ corresponds to the case where the mean $\boldsymbol{\mu} \in \mathbb{R}^n$ and the entire second moment matrix $\boldsymbol{\Pi} \in \mathcal{S}^n_+$ is specified. The distributionally robust bound studied in [44] is:

$$Z^*_{\text{full}}(\boldsymbol{\mu}, \boldsymbol{\Pi}) = \sup \left\{ \mathbb{E}_\theta \left[Z(\tilde{\mathbf{c}}) \right] \; : \; \mathbb{E}_\theta[\tilde{\mathbf{c}}] = \boldsymbol{\mu}, \; \mathbb{E}_\theta[\tilde{\mathbf{c}}\tilde{\mathbf{c}}'] = \boldsymbol{\Pi}, \; \theta \in \mathcal{P}(\mathbb{R}^n) \right\}. \tag{4}$$

An exact reformulation of the problem is obtained in [44] by using the expected value of the following random variables as decision variables:

$$\mathbb{E}\left(\begin{bmatrix} 1 \\ \tilde{\mathbf{c}} \\ \mathbf{x}(\tilde{\mathbf{c}}) \end{bmatrix} \begin{bmatrix} 1 \\ \tilde{\mathbf{c}} \\ \mathbf{x}(\tilde{\mathbf{c}}) \end{bmatrix}' \right),$$

where $\mathbf{x}(\tilde{\mathbf{c}})$ is a randomly chosen optimal solution for the objective coefficients $\tilde{\mathbf{c}}$. For the case when the decision variables in the set $\mathcal{I}$ in $\mathcal{X}$ are binary, building on the seminal work in [14], [44] provided an equivalent reformulation of this problem, under mild assumptions on the set $\mathcal{X}$ as a generalized completely positive program of the form:

$$
\begin{aligned}
Z^*_{\text{full}}(\boldsymbol{\mu}, \boldsymbol{\Pi}) = \max_{\mathbf{p},\mathbf{X},\mathbf{Y}} \quad & trace(\mathbf{Y}) \\
\text{s.t} \quad & \begin{bmatrix} 1 & \boldsymbol{\mu}' & \mathbf{p}' \\ \boldsymbol{\mu} & \boldsymbol{\Pi} & \mathbf{Y}' \\ \mathbf{p} & \mathbf{Y} & \mathbf{X} \end{bmatrix} \in \mathcal{C}(\mathbb{R}_+ \times \mathbb{R}^n \times \mathbb{R}^n_+), \\
& \mathbf{a}_k'\mathbf{p} = b_k, \qquad \forall k \in [p], \\
& \mathbf{a}_k'\mathbf{X}\mathbf{a}_k = b_k^2, \qquad \forall k \in [p], \\
& X_{jj} = p_j, \qquad \forall j \in \mathcal{I},
\end{aligned}
\tag{5}
$$

where $\mathbf{a}_k'$ is the kth row of the matrix $\mathbf{A}$. The variables $\mathbf{p}, \mathbf{X}$ and $\mathbf{Y}$ may be interpreted as representing $\mathbb{E}[\mathbf{x}(\tilde{\mathbf{c}})], \mathbb{E}[\mathbf{x}(\tilde{\mathbf{c}})\mathbf{x}(\tilde{\mathbf{c}})']$ and $\mathbb{E}[\tilde{\mathbf{c}}\mathbf{x}(\tilde{\mathbf{c}})']$ respectively. The constraints involving $\mathbf{p}$ correspond to taking expectations on the constraint $\mathbf{a}_k'\mathbf{x}(\tilde{\mathbf{c}}) = b_k$. Taking squares on both sides of $\mathbf{a}_k'\mathbf{x}(\tilde{\mathbf{c}}) = b_k$ followed by expectations gives us the constraint $\mathbf{a}_k'\mathbf{X}\mathbf{a}_k = b_k^2$. Finally all variables $x_j \in \mathcal{I}$ are binary and therefore $X_{jj} = \mathbb{E}[x_j(\tilde{\mathbf{c}})^2] = \mathbb{E}[x_j(\tilde{\mathbf{c}})] = p_j$. Unfortunately, problem (5) is hard to solve due to the difficulty in characterizing the generalized completely positive cone $\mathcal{C}(\mathbb{R}_+ \times \mathbb{R}^n \times \mathbb{R}^n_+)$. For matrices of size $n \geq 5$, testing for membership in the completely positive cone $\mathcal{C}(\mathbb{R}^n_+)$ is known to be NP-hard [19]. However, for $n \leq 4$, the completely positive cone of matrices coincides with the doubly nonnegative cone of matrices $\mathcal{DNN}^n = \mathcal{S}^n_+ \cap \mathcal{N}^n$ where $\mathcal{N}^n$ denote the set of matrices of size n with nonnegative elements. It is straightforward to characterise the doubly nonnegative cone of matrices using psd and nonnegativity conditions and this provides a tractable relaxation to the completely positive cone, since $\mathcal{C}(\mathbb{R}^n_+) \subseteq \mathcal{DNN}^n$ for all n. The doubly nonnegative relaxation thus results in an upper bound on Z^*, which might not be tight. There are several hierarchies of psd and nonnegative cones that have been developed to generate tighter approximations of the completely positive cone and the dual copositive cone including the works of [11], [12], [63], [47]. We note that completely positive and copositive programming representations of distributionally robust optimization problems under alternative ambiguity sets such as Wasserstein-based ambiguity sets have been recently developed in [28] and [59].

A related formulation that builds on characterizing the convex hull of quadratic forms over the feasible region and semidefinite optimization was proposed in [43]. They established an equivalent tight formulation to compute Z^* as follows:

$$
\begin{aligned}
Z^*_{\text{full}}(\boldsymbol{\mu}, \boldsymbol{\Pi}) = \max_{\mathbf{p},\mathbf{X},\mathbf{Y}} \quad & trace(\mathbf{Y}) \\
\text{s.t} \quad & \begin{bmatrix} 1 & \boldsymbol{\mu}' & \mathbf{p}' \\ \boldsymbol{\mu} & \boldsymbol{\Pi} & \mathbf{Y}' \\ \mathbf{p} & \mathbf{Y} & \mathbf{X} \end{bmatrix} \succeq 0, \\
& (\mathbf{p}, \mathbf{X}) \in conv\left\{(\mathbf{x}, \mathbf{x}\mathbf{x}') : \mathbf{x} \in \mathcal{X}\right\}.
\end{aligned}
\tag{6}
$$

This exact formulation requires an explicit characterization of the convex hull of quadratic forms on the feasible region. Characterising this convex hull is known to be NP-hard for sets such as $\mathcal{X} = \{0,1\}^n$ which corresponds to characterizing the Boolean quadric polytope (see [49], [46]). However, the approach allows for the possibility of using valid inequalities that have been developed in deterministic instances for the Boolean quadric polytope, to develop tighter formulations for distributionally robust bounds in applications such as in the newsvendor problem (see [43]). Efficient representations of the convex hull in (6) are known for some special cases of $\mathcal{X}$ in low dimensions as discussed in [2] and in some special cases, in higher dimensions as discussed in [15,60]. Identifying instances where this set is efficiently representable remains an active area of research.

2.2 Contributions

Our contributions in the paper are the following:

1. In Section 3, we study MILPs with random objective coefficients where the first moments are entirely known and only partial information of the second moments is provided, corresponding to non-overlapping subsets of $\mathcal{N}$. We provide a reformulation of the problem in the spirit of formulation (6), building on the results in [43]. However, as we show, this formulation requires psd constraints on smaller matrices and furthermore, it involves characterizing a suitable projection of the convex hull of the set $\{(\mathbf{x}, \mathbf{x}\mathbf{x}') : \mathbf{x} \in \mathcal{X}\}$, rather than the full convex hull. This provides a reduced SDP formulation for the problem under block-diagonal patterns of covariance information.
2. We provide an application of the formulation to appointment scheduling in Section 4.1. In the distributionally robust appointment scheduling problem with moment-based ambiguity sets, tight polynomial-sized formulations exist only for the mean-variance setting which corresponds to $R = n$, to the best of our knowledge. On the other hand, with a full covariance matrix which corresponds to $R = 1$, this problem is known to be hard to solve. By identifying an efficient characterization of projection of the convex hull of the set $\{(\mathbf{x}, \mathbf{x}\mathbf{x}') : \mathbf{x} \in \mathcal{X}\}$ in this example, we identify a new polynomial-time solvable instance of distributionally robust appointment scheduling with partial correlation information when $R = 2$. We also identify polynomial-time solvable instances in the newsvendor problem, longest path problem on PERT networks, and random instances of the assignment problem in Sections 4.2 to 4.4.
3. In Section 5, we perform a detailed computational study of the proposed reformulation in the distributionally robust appointment scheduling application. We compare the results with alternative formulations and help identify specific structures of correlations where the new formulation is most valuable. Finally we study the optimal schedules generated by various formulations including ours.
4. In Section 6 we extend our results to the following settings.
   (a) The incorporation of support information, specifically non-negativity restrictions on the random coefficients is discussed in Section 6.1. Here, we propose inequalities to strengthen the formulation and demonstrate the improvement obtained for appointment scheduling through computations.
   (b) In Section 6.2, we extend our results to incorporate information on additional entries of the second moment matrix. Here we provide an upper bound that is not necessarily tight. Through computations, we also compare the bounds provided by various formulations for appointment scheduling in this setting.

3 Tight bounds in the presence of block-diagonal correlation information

3.1 A reduced semidefinite program

In Theorem 1 below, we identify a reduced semidefinite programming formulation for evaluating Z^* in which the positive semidefinite constraints are imposed only on smaller matrices of dimensions $n_1, \dots, n_r$, instead of a larger matrix of dimension n. Moreover, Theorem 1 asserts that it is sufficient to enforce the $(n^2+3n)/2$ dimensional convex hull constraint (ignoring symmetry) in (6) on a suitable selection involving only $\sum_r (n_r^2 + 3n_r)/2$ variables.

Theorem 1 *Define $\hat{Z}^*$ as the optimal objective value of the following optimization problem:*

$$\begin{aligned}
\hat{Z}^* = \max_{\mathbf{p}, \mathbf{X}^r, \mathbf{Y}^r} & \sum_{r=1}^{R} trace(\mathbf{Y}^r) \\
s.t & \begin{bmatrix} 1 & \boldsymbol{\mu}^{r\prime} & \mathbf{p}^{r\prime} \\ \boldsymbol{\mu}^r & \boldsymbol{\Pi}^r & \mathbf{Y}^{r\prime} \\ \mathbf{p}^r & \mathbf{Y}^r & \mathbf{X}^r \end{bmatrix} \succeq 0, \qquad for\ r \in [R], \\
& \left(\mathbf{p}, \mathbf{X}^1, \dots, \mathbf{X}^R\right) \in conv\left\{ \left(\mathbf{x}, \mathbf{x}^1\mathbf{x}^{1\prime}, \dots, \mathbf{x}^R\mathbf{x}^{R\prime}\right) : \mathbf{x} \in \mathcal{X} \right\}.
\end{aligned} \tag{7}$$

Then, $\hat{Z}^* = Z^*$, where Z^* is defined as in (2).

Before proving this result, which forms the main part of this section, we discuss some implications. In comparison to formulation (6), formulation (7) involves psd constraints on multiple but much smaller matrices when $\max_r n_r$ is smaller than n. Furthermore, the theorem implies that only relevant projections of the convex hull of quadratic forms require to be characterised to compute Z^*, under block-diagonal correlation information. Such sparse characterizations have been previously exploited to identify polynomial-time solvable instances of unconstrained quadratic 0-1 optimization problems using an appropriate projection of the Boolean quadric polytope (see [46]). As we shall see in Section 4, the new formulation allows us to derive compact representations that results in polynomial-time solvable instances for the distributionally robust appointment scheduling problem, as well as for computing worst-case bounds in PERT networks and bounds for the linear assignment problem with random objective.

3.2 On chordal graphs and psd completion

A key element in the proof of Theorem 1 comprises in guaranteeing the existence of a psd matrix whose entries are partially specified. Therefore, as a preparation towards the proof of Theorem 1, we provide a brief review of results on the psd completion problem and a closely related notion of chordal graphs that are relevant for our proofs; see, for example, [26,35] and references therein for a detailed exposition on the psd completion problem.

We call a matrix whose entries are specified only on a subset of its positions as a *partial matrix.* Suppose that $\mathbf{A}$ is a partial matrix. The set of positions corresponding to the specified entries of $\mathbf{A}$ is known as the *pattern* of $\mathbf{A}$. A *completion* of the partial matrix $\mathbf{A}$ is simply a specification of the unspecified entries of $\mathbf{A}$. If $\mathbf{A}$ is a partial symmetric matrix (that is, the entry A_{ji} is specified and is equal to A_{ij} whenever A_{ij} is specified) such that every principal specified submatrix of $\mathbf{A}$ is psd, then $\mathbf{A}$ is said to be *partial psd.* A *psd completion* of the partial psd matrix $\mathbf{A}$ is said to exist if there exists a specification of the unspecified entries of $\mathbf{A}$ such that the fully specified matrix is psd.

A few key concepts on graphs will be reviewed next. Consider a graph $G = (V, E)$ where V denote the set of vertices and E denotes the set of edges. Let $|V|$ denote the number of vertices in the graph. Given a cycle in the graph, a *chord* is an edge that is not part of the cycle but connects two vertices of the cycle.

A graph is said to be *chordal* if any cycle of length greater than or equal to four has a chord (see [5]). A set of vertices $S \subseteq V$ is said to form a *clique* if there exists an edge between every pair of vertices in S. A *perfect elimination ordering* is an ordering $\beta_1, \ldots, \beta_{|V|}$ of the vertices such that for every $i \in \{1, \ldots, |V| - 1\}$, the set of vertices $\{\beta_{i+1}, \beta_{i+2}, \ldots, \beta_{|V|}\} \cap \mathcal{N}(\beta_i)$ form a clique, where $\mathcal{N}(v)$ is used to denote the set of vertices adjacent to vertex v.

As we shall note in Lemma 1 below, the existence of a perfect elimination ordering characterizes the chordal property of a graph.

The following well-known results on chordal graphs and psd completion will be useful in proving Theorem 1.

Lemma 1 *([51], Theorem 1) A graph is chordal if and only if it has a perfect elimination ordering.*

Lemma 2 *([26], Proposition 1 and Theorem 7) Every partial positive semidefinite matrix with pattern denoted by a graph G (where the vertices denote the rows (or columns) of the matrix and an edge is present between two vertices if the corresponding entry is specified) has a positive semidefinite completion if and only if G is a chordal graph.*

3.3 Proof of Theorem 1

Step 1: To show $Z^* \leq \hat{Z}^*$. It follows from the definitions of Z^* and $Z^*_{full}(\boldsymbol{\mu}, \boldsymbol{\Pi})$ in (2) and (4) that $Z^* = \max\{Z^*_{\text{full}}(\boldsymbol{\mu}, \boldsymbol{\Delta}) : \boldsymbol{\Delta} \in S_n^+, \boldsymbol{\Delta}[\mathcal{N}_r] = \boldsymbol{\Pi}^r \text{ for } r \in [R]\}$. Therefore we have from [43, Theorem 2] (see formulation (6)) that

$$
\begin{aligned}
Z^* = \max_{\mathbf{p},\mathbf{X},\mathbf{Y},\boldsymbol{\Delta}} \quad & trace(\mathbf{Y}) \\
\text{s.t} \quad & \begin{bmatrix} 1 & \boldsymbol{\mu}' & \mathbf{p}' \\ \boldsymbol{\mu} & \boldsymbol{\Delta} & \mathbf{Y}' \\ \mathbf{p} & \mathbf{Y} & \mathbf{X} \end{bmatrix} \succeq 0, \\
& \boldsymbol{\Delta}[\mathcal{N}_r] = \boldsymbol{\Pi}^r, \quad \text{for } r \in [R], \\
& \left(\mathbf{p}, \mathbf{X}\right) \in conv\left\{\left(\mathbf{x}, \mathbf{x}\mathbf{x}'\right) : \mathbf{x} \in \mathcal{X}\right\}.
\end{aligned} \tag{8}
$$

Consider any $\mathbf{p}, \mathbf{X}, \mathbf{Y}, \boldsymbol{\Delta}$ feasible for (8). Take $\mathbf{X}^r = \mathbf{X}[\mathcal{N}_r]$ and $\mathbf{Y}^r = \mathbf{Y}[\mathcal{N}_r]$. The psd constraint in (8) forces all the principal submatrices to be psd. Given the block-diagonal partition, define $\{\mathcal{V}_r : r \in [R]\}$ to be the following subsets of $\{1, \ldots, 2n+1\}$:

$$
\mathcal{V}_r = \{1\} \cup \{i + 1 : i \in \mathcal{N}_r\} \cup \{n + i + 1 : i \in \mathcal{N}_r\}, \quad \text{for } r \in [R]. \tag{9}
$$

Then, the principal submatrices formed by restricting to entries from the index set $\mathcal{V}_r$, for $r \in [R]$, satisfy,

$$\begin{bmatrix} 1 & \boldsymbol{\mu}' & \mathbf{p}' \\ \boldsymbol{\mu} & \boldsymbol{\Delta} & \mathbf{Y}' \\ \mathbf{p} & \mathbf{Y} & \mathbf{X} \end{bmatrix} [\mathcal{V}_r] = \begin{bmatrix} 1 & \boldsymbol{\mu}^{r\prime} & \mathbf{p}^{r\prime} \\ \boldsymbol{\mu}^r & \boldsymbol{\Delta}[\mathcal{N}_r] & \mathbf{Y}[\mathcal{N}_r]' \\ \mathbf{p}^r & \mathbf{Y}[\mathcal{N}_r] & \mathbf{X}[\mathcal{N}_r] \end{bmatrix} = \begin{bmatrix} 1 & \boldsymbol{\mu}^{r\prime} & \mathbf{p}^{r\prime} \\ \boldsymbol{\mu}^r & \boldsymbol{\Pi}^r & \mathbf{Y}^{r\prime} \\ \mathbf{p}^r & \mathbf{Y}^r & \mathbf{X}^r \end{bmatrix} \succeq 0.$$

In addition, since $(\mathbf{p}, \mathbf{X}) \in conv\{(\mathbf{x}, \mathbf{x}\mathbf{x}') : \mathbf{x} \in \mathcal{X}\}$, it is immediate that the principal submatrices $\mathbf{X}[\mathcal{N}_r] = \mathbf{X}^r$ satisfy the projected convex hull constraint in (7). Furthermore, the objective, $trace(\mathbf{Y}) = \sum_r trace(\mathbf{Y}[\mathcal{N}_r]) = \sum_r trace(\mathbf{Y}^r)$. Thus for every $\mathbf{p}, \mathbf{X}, \mathbf{Y}, \boldsymbol{\Delta}$ feasible for (8), there exist $\{\mathbf{p}, \mathbf{X}^r, \mathbf{Y}^r : r \in [R]\}$ feasible for (7) with the same objective. Therefore $Z^* \le \hat{Z}^*$.

Step 2: To show $Z^* \ge \hat{Z}^*$
Suppose that $\{\mathbf{p}_*, \mathbf{X}_*^r, \mathbf{Y}_*^r : r \in [R]\}$ maximizes (7). We show that $Z^* \ge \hat{Z}^*$ by constructing $\hat{\mathbf{p}}, \hat{\mathbf{X}}, \hat{\mathbf{Y}}, \hat{\boldsymbol{\Delta}}$ feasible to (8) and $trace(\hat{\mathbf{Y}}) = \hat{Z}^* = \sum_r trace(\mathbf{Y}_*^r)$.
Construction of $\hat{\mathbf{p}}$: Simply, take $\hat{\mathbf{p}} = \mathbf{p}_*$.
Construction of $\hat{\mathbf{X}}$: It follows from Carathéodory's theorem and the convex hull constraint in (7) that there exists $\hat{\mathcal{X}}$, a subset of $\mathcal{X}$, containing at most $1 + \sum_r (n_r^2 + 3n_r)/2$ elements such that,

$$\left(\hat{\mathbf{p}}, \mathbf{X}_*^1, \ldots, \mathbf{X}_*^R\right) = \sum_{\mathbf{x} \in \hat{\mathcal{X}}} \alpha_{\mathbf{x}} \left(\mathbf{x},\ \mathbf{x}^1\mathbf{x}^{1\prime}, \ldots, \mathbf{x}^R\mathbf{x}^{R\prime}\right),$$

for some $\{\alpha_{\mathbf{x}} : \mathbf{x} \in \hat{\mathcal{X}}\}$ satisfying $\alpha_{\mathbf{x}} \ge 0$, $\sum_{\mathbf{x} \in \hat{\mathcal{X}}} \alpha_{\mathbf{x}} = 1$. Now take $\hat{\mathbf{X}} = \sum_{\mathbf{x} \in \hat{\mathcal{X}}} \alpha_{\mathbf{x}} \mathbf{x}\mathbf{x}'$. Then,

$$\begin{bmatrix} 1 & \hat{\mathbf{p}} \\ \hat{\mathbf{p}} & \hat{\mathbf{X}} \end{bmatrix} = \sum_{\mathbf{x} \in \hat{\mathcal{X}}} \alpha_{\mathbf{x}} \begin{bmatrix} 1 \\ \mathbf{x} \end{bmatrix} \begin{bmatrix} 1 \\ \mathbf{x} \end{bmatrix}' \quad \text{and} \quad \hat{\mathbf{X}}[\mathcal{N}_r] = \sum_{\mathbf{x} \in \hat{\mathcal{X}}} \alpha_{\mathbf{x}} \mathbf{x}^r \mathbf{x}^{r\prime} = \mathbf{X}_*^r, \text{ for } r \in [R]. \quad (10)$$

Construction of $\hat{\mathbf{Y}}$ *and* $\hat{\boldsymbol{\Delta}}$: Consider $n \times n$ partial matrices $\hat{\mathbf{Y}}_p$ and $\hat{\boldsymbol{\Delta}}_p$ with entries specified only along the following principal submatrices:

$$\hat{\mathbf{Y}}_p[\mathcal{N}_r] = \mathbf{Y}_*^r \quad \text{and} \quad \hat{\boldsymbol{\Delta}}_p[\mathcal{N}_r] = \boldsymbol{\Pi}^r, \quad \text{for } r \in [R]. \quad (11)$$

Next, consider a $(2n+1) \times (2n+1)$ partial symmetric matrix $\mathbf{L}_p$ constructed in terms of the partial matrices $\hat{\mathbf{Y}}_p$, $\hat{\boldsymbol{\Delta}}_p$ and the fully specified matrix $\hat{\mathbf{X}}$ as follows:

$$\mathbf{L}_p = \begin{bmatrix} 1 & \boldsymbol{\mu}' & \hat{\mathbf{p}}' \\ \boldsymbol{\mu} & \hat{\boldsymbol{\Delta}}_p & \hat{\mathbf{Y}}_p \\ \hat{\mathbf{p}} & \hat{\mathbf{Y}}_p' & \hat{\mathbf{X}} \end{bmatrix} = \begin{bmatrix} 1 & \boldsymbol{\mu}^{1\prime} & \ldots & \boldsymbol{\mu}^{R\prime} & \mathbf{p}_*^{1\prime} & \ldots & \mathbf{p}_*^{R\prime} \\ \boldsymbol{\mu}^1 & \boldsymbol{\Pi}^1 & ? & ? & \mathbf{Y}_*^{1\prime} & ? & ? \\ \vdots & ? & \ddots & ? & ? & \ddots & ? \\ \boldsymbol{\mu}^R & ? & ? & \boldsymbol{\Pi}^R & ? & ? & \mathbf{Y}_*^{R\prime} \\ \mathbf{p}_*^1 & \mathbf{Y}_*^1 & ? & ? & & & \\ \vdots & ? & \ddots & ? & & \hat{\mathbf{X}} & \\ \mathbf{p}_*^R & ? & ? & \mathbf{Y}_*^R & & & \end{bmatrix}.$$

The entries marked '?' denote missing entries. By demonstrating that the underlying pattern of $\mathbf{L}_p$ is chordal, Lemma 3 below establishes that there exists a psd completion for the partial matrix $\mathbf{L}_p$.

Lemma 3 *The matrix* $\mathbf{L}_p$ *has a completion* $\mathbf{L}_{comp}$ *such that* $\mathbf{L}_{comp} \succeq 0$.

Proof Consider the following construction of an undirected graph G with vertex set, $V = \{s, c_1, c_2, \ldots, c_n, x_1, \ldots, x_n\}$, comprising $2n+1$ vertices. To define the edge set, identify the vertices $s, c_1, c_2, \ldots, c_n, x_1, \ldots, x_n$, respectively, with the rows (or columns) numbered $1, 2, \ldots, 2n+1$ of the partial matrix $\mathbf{L}_p$. We assign an edge between two vertices of G only if the the respective entry of the partial matrix $\mathbf{L}_p$ is specified. Therefore, graph G represents the pattern of the partial matrix $\mathbf{L}_p$

With the above described construction of graph G, note that the vertices $\{x_1, \ldots, x_n\}$ form a clique in G as the matrix $\hat{\mathbf{X}}$ is specified completely. The edges between the vertices $\{c_1, \ldots, c_n\}$ correspond to the specified entries of the partial matrix $\hat{\boldsymbol{\Delta}}_p$. Likewise, the edges between vertices $\{c_1, \ldots, c_n\}$ and vertices $\{x_1, \ldots, x_n\}$ correspond to the known entries of the partial martial $\hat{\mathbf{Y}}_p$. Thus, for any $r \in [R]$, when restricted to vertices corresponding to $\mathbf{c}^r$ and $\mathbf{x}^r$, we again have a clique (see Figure 1 for an illustration).

Next, consider the ordering of the vertices, $c_1, c_2, \ldots, c_n, s, x_1, x_2, \ldots, x_n$, of G. Since the vertices $\{s, x_1, \ldots, x_n\}$ form a clique, it is immediate that for any x_i, the neighbors of the node that appear after it in the ordering also form a clique. The same reasoning applies for the vertex s. For any $i \in [n]$, let r_i be the unique $r \in [R]$ such that $i \in \mathcal{N}_{r_i}$. Subsequently, the neighbors of c_i that appear after it in the ordering comprises the collection $\{c_j, s, x_k : j, k \in \mathcal{N}_{r_i}, j > i\}$, which again forms a clique. This is because the vertices $\{s, c_j, x_j : j \in \mathcal{N}_r\}$ form a clique, for any $r \in [R]$. Consequently, the ordering $c_1, c_2, \ldots, c_n, s, x_1, x_2, \ldots, x_n$ is a perfect elimination ordering for the graph G. Then due to Lemma 1, G is a chordal graph.

Recalling the definition of $\mathcal{V}_r$ in (9), observe that any fully specified principal submatrix of $\mathbf{L}_p$ is a principal submatrix of

$$\mathbf{L}_p[\mathcal{V}_r] = \begin{bmatrix} 1 & \boldsymbol{\mu}^{r\prime} & (\hat{\mathbf{p}}^r)' \\ \boldsymbol{\mu}^r & \hat{\boldsymbol{\Delta}}_p[\mathcal{N}_r] & \hat{\mathbf{Y}}_p[\mathcal{N}_r]' \\ \hat{\mathbf{p}}^r & \hat{\mathbf{Y}}_p[\mathcal{N}_r] & \hat{\mathbf{X}}[\mathcal{N}_r] \end{bmatrix} = \begin{bmatrix} 1 & \boldsymbol{\mu}^{r\prime} & \mathbf{p}_*^{r\prime} \\ \boldsymbol{\mu}^r & \boldsymbol{\Pi}^r & \mathbf{Y}_*^{r\prime} \\ \mathbf{p}_*^r & \mathbf{Y}_*^r & \mathbf{X}_*^r \end{bmatrix},$$

for some $r \in [R]$. The latter equality follows from (11) and the second observation in (10). Since $\mathbf{p}_*, \mathbf{Y}_*^r, \mathbf{X}_*^r$ are taken to be feasible for (7), we have that $\mathbf{L}_p[\mathcal{V}_r] \succeq 0$ for any $r \in [R]$. With the 'maximal' fully specified principal submatrices $\{\mathbf{L}_p[\mathcal{V}_r] : r \in [R]\}$ being psd, we have that all the fully specified principal submatrices are psd. Therefore $\mathbf{L}_p$ is partial psd.

Finally, with the pattern underlying the partial psd matrix $\mathbf{L}_p$ forming a chordal graph, it follows from Lemma 2 that there exists a psd completion for $\mathbf{L}_p$. This proves Lemma 3.

To complete the proof of Theorem 1, consider the psd completion $\mathbf{L}_{comp}$ of $\mathbf{L}_p$. Take $\hat{\boldsymbol{\Delta}} := \mathbf{L}_{comp}[\{2, \ldots, n+1\}]$ and $\hat{\mathbf{Y}}$ to be the $n \times n$ submatrix of $\mathbf{L}_{comp}$ formed from entries in rows $\{2, \ldots, n+1\}$ and columns $\{n+2, \ldots, 2n+1\}$. Then $\mathbf{L}_{comp} \succeq 0$ allows us to write,

$$\mathbf{L}_{comp} = \begin{bmatrix} 1 & \boldsymbol{\mu}' & \hat{\mathbf{p}}' \\ \boldsymbol{\mu} & \hat{\boldsymbol{\Delta}} & \hat{\mathbf{Y}}' \\ \hat{\mathbf{p}} & \hat{\mathbf{Y}} & \hat{\mathbf{X}} \end{bmatrix} \succeq 0. \tag{12}$$

Since the specified entries of $\mathbf{L}_p$ match with that of $\mathbf{L}_{comp}$, it follows from the construction of $\mathbf{L}_p$ that $\hat{\mathbf{Y}}$ is a completion of $\hat{\mathbf{Y}}_p$ and $\hat{\boldsymbol{\Delta}}$ is a psd completion of $\hat{\boldsymbol{\Delta}}_p$;

the latter completion is psd because the principal submatrices of $\mathbf{L}_{comp}$ are psd. Therefore, we have from (11) that

$$\hat{\boldsymbol{\Delta}}[\mathcal{N}_r] = \boldsymbol{\Pi}^r \quad \text{and} \quad \hat{\mathbf{Y}}[\mathcal{N}_r] = \mathbf{Y}_*^r. \tag{13}$$

Furthermore, as we have taken $\{\mathbf{p}_*, \mathbf{X}_*^r, \mathbf{Y}_*^r : r \in [R]\}$ to maximize (7), we have,

$$\hat{Z}^* = \sum_{r=1}^{R} trace(\mathbf{Y}_*^r) = \sum_{r=1}^{R} trace(\hat{\mathbf{Y}}[\mathcal{N}_r]) = trace(\hat{\mathbf{Y}}). \tag{14}$$

It follows from (12) and the first of the two equations in (10) and (13) that $\hat{\mathbf{p}}, \hat{\boldsymbol{\Delta}}, \hat{\mathbf{X}}, \hat{\mathbf{Y}}$ are feasible for (8). Therefore, the optimal value of (8), denoted by Z^*, satisfies $Z^* \geq trace(\hat{\mathbf{Y}})$. The desired $Z^* \geq \hat{Z}^*$ is now a consequence of (14). This completes Step 2 and the proof of Theorem 1.

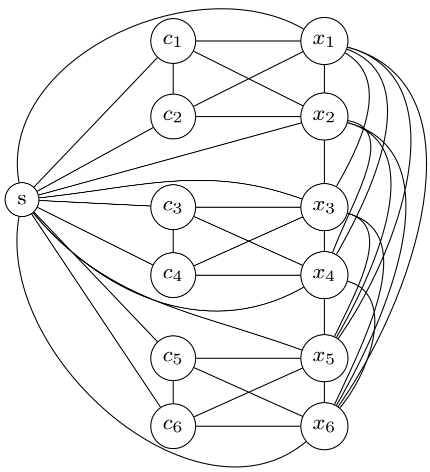


Fig. 1: Illustration for the graph G for the case where $n = 6$ and the partition is given by $\mathcal{N}_1 = \{1, 2\}, \mathcal{N}_2 = \{3, 4\}$ and $\mathcal{N}_3 = \{5, 6\}$. For k odd, c_{k-1} and c_k are not connected and $\{c_k, x_k, c_{k+1}, x_{k+1}, s\}$ form a clique.

The interested reader may refer to Appendix A for details on a worst-case distribution which attains the expected value $\hat{Z}^*$ in (7).

4 Polynomial-time solvable cases

In this section, we identify efficient representations of the convex hull constraint in (7) for three illustrative applications. The common theme in these applications is that the derived efficient characterizations, in turn, result in polynomial-time solvable instances for the partial covariance based distributionally robust formulation in (2). As far as we know, the example we consider in Section 4.1 is the first example of such an approach towards appointment scheduling that results in a polynomial-time solvable tight reformulation in the presence of explicitly known correlation information between uncertain processing times of the jobs to be scheduled.

4.1 Appointment scheduling

4.1.1 Problem description

In the presence of uncertainty in the processing durations of jobs for a sequence of customers, the appointment scheduling problem deals with identifying customer reporting times that minimize the total amount of time spent by customers waiting for service after arrival. As an example, consider n patients who need to meet a doctor. Let $\tilde{u}_i$ be the random service duration of patient $i \in [n]$. Suppose that all patients arrive exactly at the reporting time allotted to them. If we let s_i denote the duration scheduled for patient i, then the reporting time for patient i is $\sum_{j=1}^{i-1} s_j$. We take the waiting time of the first patient to be zero. Then the waiting time for patient i, denoted by w_i, satisfies the well-known Lindley's recursion for the waiting time in single-server queues:

$$w_1 = 0, \quad w_i = \max(w_{i-1} + \tilde{u}_{i-1} - s_{i-1}, 0), \quad i = 2, \ldots, n. \tag{15}$$

The total waiting time for all patients is the sum of waiting times $\sum_{i=1}^{n} w_i$. The overtime of the doctor can be modeled as $w_{n+1} = \max(w_n + \tilde{u}_n - s_n, 0)$. Then the total waiting time of the patients and the overtime of the doctor are cumulatively captured by,

$$f(\tilde{\mathbf{u}}, \mathbf{s}) = \sum_{i=1}^{n} \max(w_i + \tilde{u}_i - s_i, 0). \tag{16}$$

It turns out that $f(\tilde{\mathbf{u}}, \mathbf{s})$ can be computed by solving a network flow problem. More details on this computation can be found in Proposition 1 of [32] and also later in Step 1, proof of Theorem 2. For now, we list the equivalent linear programming formulation as follows:

$$\begin{aligned} f(\tilde{\mathbf{u}}, \mathbf{s}) = \max_{\mathbf{x}} \; & (\tilde{\mathbf{u}} - \mathbf{s})'\mathbf{x} \\ \text{s.t. } & x_i - x_{i-1} \geq -1, \text{ for } i = 2, \ldots, n-1, \\ & x_n \leq 1, \\ & x_i \geq 0, \qquad\qquad \text{for } i = 1, \ldots, n, \end{aligned} \tag{17}$$

Define $S = \{\mathbf{s} \in \mathbb{R}_+^n : s_1 + \ldots + s_n \leq T\}$, where T is a positive upper time limit within which the schedules should be fit. It is then natural to seek a schedule sequence $\mathbf{s} \in S$ that minimizes $\mathbb{E}[f(\tilde{\mathbf{u}}, \mathbf{s})]$.

The described setup is applicable to schedule appointments in various situations where a single server processes the arriving jobs on a first-come-first-serve basis. In settings where the jobs to be processed are dependent and the joint distribution of their processing times $\tilde{\mathbf{u}}$ is difficult to be fully specified, an approach that has gained much attention over the last decade is to seek distributionally robust schedules that minimize the worst case waiting time, $\sup_{\theta \in \mathcal{P}} E_\theta[f(\tilde{\mathbf{u}}, \mathbf{s})]$; here, the set $\mathcal{P}$ is taken to be the family of all probability distributions consistent with the information known about the probability distribution of $\tilde{\mathbf{u}}$. This problem was first studied in [32] where complete information on the first moment $\boldsymbol{\mu}$ and second moment matrix $\boldsymbol{\Pi}$ is assumed to be available on the service time durations. The formulation proposed in [32] builds on the completely positive formulation (5).

4.1.2 Polynomial-time solvable instance

To illustrate the applicability of Theorem 1 in this context, suppose that the number of patients, n, is even without loss of generality, and the mean of service times $\tilde{\mathbf{u}}$ is fully specified, and only the entries, $\{\Pi_{ii}, \Pi_{j,j+1}, \Pi_{j+1,j} : i \in [n], j \in \{1, 3, \ldots, n-1\}\}$ of the second moment matrix, $\boldsymbol{\Pi} = [\Pi_{ij}]$, are specified. This corresponds to knowing the correlations among service time durations of adjoining patient pairs. Recalling the definition of the partition $\{\mathcal{N}_r : r \in [R]\}$ of $[n]$, this partial specification of the second moments corresponds to the scenario where,

$$R = n/2, \quad \text{and} \quad \mathcal{N}_r = \{2r-1, 2r\}, \text{ for } r = 1, \ldots, R. \tag{18}$$

For any given schedule $\mathbf{s} \in S$, consider the worst-case expected total waiting time,

$$\begin{aligned} Z^*_{app}(\mathbf{s}) = \sup \big\{ & \mathbb{E}_\theta \left[f(\tilde{\mathbf{u}}, \mathbf{s}) \right] : \mathbb{E}_\theta \left[\tilde{u}_i \right] = \mu_i, \mathbb{E}_\theta[\tilde{u}_i^2] = \Pi_{ii}, \text{ for } i \in [n], \\ & \mathbb{E}_\theta[\tilde{u}_j \tilde{u}_{j+1}] = \Pi_{j,j+1}, \text{ for } j \in \{1, 3, ...n-1\}, \theta \in \mathcal{P}(\mathbb{R}^n) \big\}. \end{aligned} \tag{19}$$

Our key result is that by an appropriate application of Theorem 1, we obtain a polynomial-time solvable formulation for evaluating $Z^*_{app}(\mathbf{s})$ in Theorem 2.

Theorem 2 *Given a schedule* $\mathbf{s} \in S$, *suppose that* $Z^*_{app}(\mathbf{s})$ *is defined as in* (19)*. Then,*

$$\begin{aligned} Z^*_{app}(\mathbf{s}) = \max_{p_i, X_{ij}, Y_{ij}, t_{kj}} \; & \sum_{i=1}^{n} (Y_{ii} - s_i p_i) \\ \text{s.t.} \; & \begin{bmatrix} 1 & \mu_i & \mu_{i+1} & p_i & p_{i+1} \\ \mu_i & \Pi_{ii} & \Pi_{i,i+1} & Y_{ii} & Y_{i,i+1} \\ \mu_{i+1} & \Pi_{i,i+1} & \Pi_{i+1,i+1} & Y_{i+1,i} & Y_{i+1,i+1} \\ p_i & Y_{ii} & Y_{i+1,i} & X_{ii} & X_{i,i+1} \\ p_{i+1} & Y_{i,i+1} & Y_{i+1,i+1} & X_{i,i+1} & X_{i+1,i+1} \end{bmatrix} \succeq 0, \quad \text{for } i \text{ odd, } i \in [n], \\ & p_i = \sum_{k=1}^{i} \sum_{j=i}^{n+1} t_{kj} (j - i), \quad \text{for } i \in [n], \\ & X_{ii} = \sum_{k=1}^{i} \sum_{j=i}^{n+1} t_{kj} (j - i)^2, \quad \text{for } i \in [n], \\ & X_{i,i+1} = X_{i+1,i} = \sum_{k=1}^{i} \sum_{j=i+1}^{n+1} t_{kj} (j - i)(j - (i+1)), \; \text{for } i \text{ odd, } i \in [n], \\ & \sum_{k=1}^{i} \sum_{j=i}^{n+1} t_{kj} = 1, \quad \text{for } i \in [n], \\ & t_{kj} \geq 0, \quad \text{for } 1 \leq k \leq j \leq n+1. \end{aligned}$$

The proof of Theorem 2 is presented in Section 4.1.3. As demonstrated in Corollary 1 below, an optimal schedule that minimizes the worst-case total expected waiting time can be obtained by considering the dual minimization problem of the semidefinite program in Theorem 2.

Corollary 1 *Given $T > 0$, a schedule $s \in S = \{\mathbf{s} \in \mathbb{R}^n_+ : s_1 + \ldots + s_n \leq T\}$ that minimizes $Z^*_{app}(\mathbf{s})$ can be obtained by solving the following semidefinite program:*

$$Z^*_{app} = \min_{\mathbf{s},\boldsymbol{\eta},\boldsymbol{\beta},\boldsymbol{\Gamma},\boldsymbol{\rho},\boldsymbol{\delta},\boldsymbol{\tau},\boldsymbol{\gamma}} \sum_{\substack{i\in[n],\\ i\ odd}} \eta_i + \sum_{i=1}^{n} \beta_i \mu_i + \sum_{\substack{i\in[n],\\ i\ odd}} \sum_{k,l\in\{i,i+1\}} \Gamma_{kl} \Pi_{kl} + \sum_{i=1}^{n+1} \rho_i$$

$$s.t \quad \begin{bmatrix} 2\eta_i & \beta_i & \beta_{i+1} & \delta_i + s_i & \delta_{i+1} + s_{i+1} \\ \beta_i & 2\Gamma_{ii} & \Gamma_{i,i+1} & -1 & 0 \\ \beta_{i+1} & \Gamma_{i,i+1} & 2\Gamma_{i+1,i+1} & 0 & -1 \\ \delta_i + s_i & -1 & 0 & 2\gamma_i & \tau_i \\ \delta_{i+1} + s_{i+1} & 0 & -1 & \tau_i & 2\gamma_{i+1} \end{bmatrix} \succeq 0, \quad for\ i\ odd, i \in [n],$$

$$\sum_{i=k}^{j} \rho_i \geq \sum_{i=k}^{\min\{j,n\}} \delta_i (j-i) + \sum_{i=k}^{\min\{j,n\}} \gamma_i (j-i)^2 + \sum_{\substack{i=k\\ i\ odd}}^{\min\{j,n\}} \tau_i (j-i)(j-(i+1)),$$
$$for\ 1 \leq k \leq j \leq n+1,$$

$$\sum_{i=1}^{n} s_i \leq T, \quad s_i \geq 0,\ for\ i \in [n].$$

Proof The result follows by performing a joint minimization over $\mathbf{s} \in S$ and the objective of the dual of the semidefinite program in Theorem 2. This is because, for any $\mathbf{s} \in S$, the value of the semidefinite program in Theorem 2 is equal to that of its dual minimization problem. Indeed, the existence of an interior feasible point for the dual problem can be exhibited as follows. Given $\mathbf{s} \in S$, set all the variables other than $\eta_i, \gamma_i, \gamma_{i+1}, \Gamma_{ii}, \Gamma_{i+1,i+1}$, for i odd, to zero, and let $\Gamma_{ii}\gamma_i > 1/4$, $\rho_i > \gamma_i(n+1-i)^2$, for every $i \in [n]$; fix η_i, for i odd, to be arbitrarily positive; this assignment results in a dual feasible solution where none of the constraints are active. Moreover, the requirement that $\mathbf{\Pi}^r - \boldsymbol{\mu}^r \boldsymbol{\mu}^{r\prime} \succ 0$, for every $r \in [R]$ is sufficient to guarantee strong duality.

4.1.3 A proof of Theorem 2

Step 1: Recasting the waiting time $f(\tilde{\mathbf{u}}, \mathbf{s})$ in the form of (1). Given a fixed sequence of schedules $\mathbf{s} = (s_1, \ldots, s_n)$, the recursive structure in (15) allows writing the total waiting time, $f(\tilde{\mathbf{u}}, \mathbf{s})$, as the value of the following linear program:

$$\begin{aligned} \min_{\mathbf{w}} \ & \sum_{i=1}^{n+1} w_i \\ \text{s.t} \ & w_i \geq w_{i-1} + \tilde{u}_{i-1} - s_{i-1}, \text{ for } i = 2, \ldots, n+1, \\ & w_i \geq 0, \qquad\qquad\qquad\quad \text{for } i = 1, \ldots, n+1. \end{aligned}$$

Define $\tilde{\mathbf{c}}(\mathbf{s}) := \tilde{\mathbf{u}} - \mathbf{s}$. The dual of this linear program results in,

$$\begin{aligned} f(\tilde{\mathbf{u}}, \mathbf{s}) = \max_{\mathbf{x}} \ & \tilde{\mathbf{c}}(\mathbf{s})'\mathbf{x} \\ \text{s.t. } & x_i - x_{i-1} \geq -1, \text{ for } i = 2, \ldots, n-1, \\ & x_n \leq 1, \\ & x_i \geq 0, \qquad\qquad \text{for } i = 1, \ldots, n, \end{aligned} \tag{20}$$

The constraints in (20) are such that any subset of n active constraints satisfy, for every $i \in [n]$, either $x_i = 0$ or $x_{i-1} = x_i + 1$. It has been shown in [61,62] that any $\mathbf{x} = (x_1, \ldots, x_n)$ with this special structure can be uniquely represented as a partition of intervals of integers in $\{1, \ldots, n+1\}$. This structure was first used in [38] to identify a tractable instance of appointment scheduling with mean-variance information and to the case with no-shows in [30]. Lemma 4 below exploits this representation to characterize the extreme points of the feasible region to (20). Though this representation arises as a consequence of statements in Theorems 1,2,3 in [61], we provide the complete proof here.

Lemma 4 *([61], Theorems 1,2,3) The extreme points of the feasible region in* (20) *is given by,*

$$\mathcal{X}_{app} = \Bigg\{ \mathbf{x} \in \mathbb{R}^n_+ : x_i = \sum_{k=1}^{i} \sum_{j=i}^{n+1} T_{kj}(j-i), \ \textit{for } i \in [n], \sum_{k=1}^{i} \sum_{j=i}^{n+1} T_{kj} = 1, \ \textit{for } i \in [n], \\ T_{kj} \in \{0,1\}, \ \textit{for } 1 \le k \le j \le n+1 \Bigg\}. \quad (21)$$

Proof Recall our observation on the constraints in (20) that any subset of n active constraints must have that, for every $i \in [n]$, either $x_i = 0$ or $x_{i-1} = x_i + 1$. Therefore, any $\mathbf{x}$ in the feasible region to (20) is an extreme point if and only if either $x_i = 0$ or $x_{i-1} = x_i + 1$, for every $i \in [n]$.

Now, for an extreme point $\mathbf{x} = (x_1, \ldots, x_n)$, let $I_{\mathbf{x}}$ be the unique partition of intervals of integers $\{1, 2, \ldots, n, n+1\}$ such that the interval $[k, j] := \{k, k+1, \ldots, j\}$, for $k \le j$, belongs to the partition $I_{\mathbf{x}}$ if and only if $x_j = 0, x_{j-1} = 1, \ldots, x_k = j - k$. Thus there exists a bijection between the extreme points of the feasible region to (20) and the collection of partitions of integer intervals of $\{1, 2, \ldots, n+1\}$. For illustration, if $n = 3$ and $\mathbf{x} = (0, 0, 0)$ then $I_{(0,0,0)} = \{[1], [2], [3], [4]\}$; likewise, the points $(3, 2, 1), (1, 0, 1)$, are identified with their respective partitions given by, $I_{(3,2,1)} = \{[1, 4]\}$ and $I_{(1,0,1)} = \{[1, 2], [3, 4]\}$, and vice versa.

Next, for any extreme point $\mathbf{x}$ (whose unique interval partition representation is $I_{\mathbf{x}}$), consider the following assignment of values to the variables $(T_{kj} : 1 \le k \le j \le n+1)$:

$$T_{kj} = \begin{cases} 1 & \text{if the integer interval } [k, j] \in I_{\mathbf{x}}, \\ 0 & \text{otherwise.} \end{cases} \quad (22)$$

It follows from the very construction of the interval partition $I_{\mathbf{x}}$ that that only one of $\{T_{kj} : k \le i \le j\}$ equals 1, for every $i \in [n]$, and $x_i = \sum_{k=1}^{i} \sum_{j=i}^{n+1} T_{kj}(j-i)$. Therefore any extreme point of the feasible region of (20) lies in $\mathcal{X}_{app}$.

On the other hand, for any $\mathbf{x} = (x_1, \ldots, x_n)$ in $\mathcal{X}_{app}$, we have $x_i = \sum_{k=1}^{i} \sum_{j=i}^{n+1} T_{kj}(j-i)$ satisfying,

$$x_i = \begin{cases} 0 & \text{if } T_{ki} = 1 \text{ for some } k \le i, \\ 1 & \text{if } i = n \text{ and } T_{kn} = 0, \\ x_{i+1} + 1 & \text{otherwise,} \end{cases}$$

for every $i \in [n]$. Here we have again used the observation that for any given assignment of variables $T_{kj} \in \{0,1\}$ satisfying $\sum_{k=1}^{i}\sum_{j=i}^{n+1} T_{kj} = 1$, for every $i \in [n]$, only one of $\{T_{kj} : k \le i \le j\}$ equals one. Since any $\mathbf{x} \in \mathcal{X}_{app}$ satisfies $x_i = 0$ or $x_{i-1} = x_i + 1$ for every $i \in [n]$, we arrive at the conclusion that $\mathcal{X}_{app}$ is indeed the set of extreme points of the feasible region to (20).

□

As the feasible region to the linear program in (20) is bounded, there exists an extreme point at which the maximum is attained. Then as a consequence of Lemma 4, we have that

$$f(\tilde{\mathbf{u}}, \mathbf{s}) = \max\left\{\tilde{\mathbf{c}}(\mathbf{s})'\mathbf{x} : \mathbf{x} \in \mathcal{X}_{app}\right\}. \tag{23}$$

Step 2: Application of Theorem 1. For the partition $\{\mathcal{N}_r : r \in [R]\}$ specified in (18), we use (23) to express $Z^*_{app}(\mathbf{s})$ as,

$$\sup\left\{\mathbb{E}_\theta\left[\max_{\mathbf{x}\in\mathcal{X}_{app}} \tilde{\mathbf{c}}(\mathbf{s})'\mathbf{x}\right] : \mathbb{E}_\theta[\tilde{\mathbf{c}}(\mathbf{s})] = \boldsymbol{\mu} - \mathbf{s}, \mathbb{E}_\theta\left[\tilde{\mathbf{c}}(\mathbf{s})^r\tilde{\mathbf{c}}(\mathbf{s})^{r\prime}\right] = \boldsymbol{\Pi}^r_{\mathbf{s}}\ \forall r \in [R], \theta \in \mathcal{P}(\mathbb{R}^n)\right\},$$

where, for $i = 1, 3, \ldots, n-1$, the second moment of $(\tilde{\mathbf{c}}_i(\mathbf{s}), \tilde{\mathbf{c}}_{i+1}(\mathbf{s}))$ is specified by,

$$\boldsymbol{\Pi}^{\lceil i/2\rceil}_{\mathbf{s}} = \begin{bmatrix} \Pi_{ii} & \Pi_{i,i+1} \\ \Pi_{i,i+1} & \Pi_{i+1,i+1}\end{bmatrix} - \begin{bmatrix} s_i \\ s_{i+1}\end{bmatrix}\begin{bmatrix} \mu_i \\ \mu_{i+1}\end{bmatrix}' - \begin{bmatrix} \mu_i \\ \mu_{i+1}\end{bmatrix}\begin{bmatrix} s_i \\ s_{i+1}\end{bmatrix}' + \begin{bmatrix} s_i \\ s_{i+1}\end{bmatrix}\begin{bmatrix} s_i \\ s_{i+1}\end{bmatrix}'.$$

Then as an application of Theorem 1, we can write $Z^*_{app}(\mathbf{s})$ as the value of the semidefinite program in (7) by replacing parameters $\boldsymbol{\mu}^r, \boldsymbol{\Pi}^r$, respectively, with $\boldsymbol{\mu}^r - \mathbf{s}^r$ and $\boldsymbol{\Pi}^r_{\mathbf{s}}$. Further, changing the variables $\mathbf{Y}^r$ to $\mathbf{Y}^r - \mathbf{p}^r\mathbf{s}^{r\prime}$ for $r \in [R]$, the objective in (7) becomes, $\sum_r trace(\mathbf{Y}^r - \mathbf{p}^r\mathbf{s}^{r\prime})$, and the psd constraints in (7) becomes,

$$\begin{bmatrix} 1 & \boldsymbol{\mu}^{r\prime} - \mathbf{s}^{r\prime} & \mathbf{p}^{r\prime} \\ \boldsymbol{\mu}^r - \mathbf{s}^r & \boldsymbol{\Pi}^r - \mathbf{s}^r\boldsymbol{\mu}^{r\prime} - \boldsymbol{\mu}^r\mathbf{s}^{r\prime} + \mathbf{s}^r\mathbf{s}^{r\prime} & \mathbf{Y}^{r\prime} - \mathbf{s}^r\mathbf{p}^{r\prime} \\ \mathbf{p}^r & \mathbf{Y}^r - \mathbf{p}^r\mathbf{s}^{r\prime} & \mathbf{X}^r \end{bmatrix} \succeq 0.$$

This psd constraint is equivalently written as,

$$\begin{bmatrix} 1 & \boldsymbol{\mu}^{r\prime} & \mathbf{p}^{r\prime} \\ \boldsymbol{\mu}^r & \boldsymbol{\Pi}^r & \mathbf{Y}^{r\prime} \\ \mathbf{p}^r & \mathbf{Y}^r & \mathbf{X}^r \end{bmatrix} \succeq 0,$$

due to the identical constraints that arise as a result of Schur complement conditions (for psd matrices) on both the constraints above. Indeed, block-matrices of the form (52) are psd if and only if both $\mathbf{A}$ and $\mathbf{C} - \mathbf{B}'\mathbf{A}^{-1}\mathbf{B}$ are psd; for the block matrices in the above constraints, take $\mathbf{A} = 1$ to verify the desired equivalence. With these observations, we have

$$\begin{aligned} Z^*_{app}(\mathbf{s}) = \max_{p_i, X_{ij}, Y_{ij}, t_{kj}} & \sum_{i=1}^{n}(Y_{ii} - s_i p_i) \\ \text{s.t.} & \begin{bmatrix} 1 & \mu_i & \mu_{i+1} & p_i & p_{i+1} \\ \mu_i & \Pi_{ii} & \Pi_{i,i+1} & Y_{ii} & Y_{i,i+1} \\ \mu_{i+1} & \Pi_{i,i+1} & \Pi_{i+1,i+1} & Y_{i+1,i} & Y_{i+1,i+1} \\ p_i & Y_{ii} & Y_{i+1,i} & X_{ii} & X_{i,i+1} \\ p_{i+1} & Y_{i,i+1} & Y_{i+1,i+1} & X_{i,i+1} & X_{i+1,i+1} \end{bmatrix} \succeq 0, \quad \text{for } i \text{ odd}, \\ & (p_1, \ldots, p_n, X_{11}, \ldots, X_{nn}, X_{12}, X_{34}, \ldots, X_{n-1,n}) \in C_{app}, \end{aligned} \tag{24}$$

where

$$C_{app} = conv \left\{ \left(x_1, \ldots, x_n, x_1^2, \ldots, x_n^2, x_1x_2, x_3x_4, \ldots, x_{n-1}x_n \right) : \mathbf{x} \in \mathcal{X}_{app} \right\}.$$

Step 3: An efficient representation of the convex hull C_{app}. We now complete the proof of Theorem 2 by identifying a characterization of the convex hull C_{app} that leads to an efficient representation of the last constraint written in (24).

Proposition 1 *The set C_{app} is equivalently written as,*

$$\begin{aligned} C_{app} = \Bigg\{ & (p_1, \ldots, p_n, X_{11}, \ldots, X_{nn}, X_{12}, X_{34}, \ldots, X_{n-1,n}) \in \mathbb{R}^{5n/2} : \\ & p_i = \sum_{k=1}^{i} \sum_{j=i}^{n+1} t_{kj}(j-i), \; X_{ii} = \sum_{k=1}^{i} \sum_{j=i}^{n+1} t_{kj}(j-i)^2, \quad \textit{for } i \in [n], \\ & X_{i,i+1} = \sum_{k=1}^{i} \sum_{j=i+1}^{n+1} t_{kj}(j-i)(j-(i+1)), \quad \textit{for } i \in [n], i \textit{ odd}, \\ & \sum_{k=1}^{i} \sum_{j=i}^{n+1} t_{kj} = 1, \; \textit{for } i \in [n], \quad t_{kj} \geq 0 \textit{ for } 1 \leq k \leq j \leq n+1 \Bigg\}. \end{aligned}$$

Proof Take any $\mathbf{x} \in \mathcal{X}_{app}$. It follows from the characterization in (21) that there exists an assignment for variables $T_{kj} \in \{0,1\}$ such that only one of $\{T_{kj} : k \leq i \leq j\}$ equals one, for every $i \in [n]$, and $x_i = \sum_{k=1}^{i} \sum_{j=i}^{n+1} T_{kj}(j-i)$. Therefore, $x_i = j - i$ and $x_i^2 = (j-i)^2$ for the unique $j \geq i$ such that $T_{kj} = 1$. Equivalently, we have

$$x_i^2 = \sum_{k=1}^{i} \sum_{j=i}^{n+1} T_{kj}(j-i)^2. \tag{25}$$

Again, since only one of $\{T_{kj} : k \leq i \leq j\}$ equals one, for every $i \in [n]$, we have,

$$T_{kj}T_{ab} = 0, \quad \text{when either } k \neq a \text{ or } j \neq b \text{ and } k \leq i \leq j, a \leq i \leq b. \tag{26}$$

Equipped with this observation, consider:

$$\begin{aligned} x_i x_{i+1} &= \left(\sum_{k=1}^{i} \sum_{j=i}^{n+1} T_{kj}(j-i) \right) \left(\sum_{a=1}^{i+1} \sum_{b=i+1}^{n+1} T_{ab}(j-(i+1)) \right) \\ &= \sum_{k=1}^{i} \sum_{j=i}^{n+1} T_{kj}(j-i) \times \sum_{a=1}^{i} \sum_{b=i+1}^{n+1} T_{ab}(b-(i+1)) \\ &\quad + \sum_{k=1}^{i} \sum_{j=i}^{n+1} T_{kj}(j-i) \times \sum_{b=i+1}^{n+1} T_{i+1,b}(b-(i+1)), \end{aligned}$$

where the latter summand is equal to zero because, a) the terms for which $j = i$ are zero due to the appearance of $j - i$, (see Figure 2) and b) the terms for which $j > i$ are zero due to the appearance of $T_{kj}T_{i+1,b}$, which is zero due to (26) (illustrated in Figure 3) . Likewise, in the first summand, the terms for which $k \neq a, j \neq b$ vanish due to (26). As a result,

$$x_i x_{i+1} = \sum_{k=1}^{i} \sum_{j=i+1}^{n+1} T_{kj}(j-i)(j-(i+1)). \tag{27}$$

Fig. 2: Terms involving $T_{ki}T_{i+1,b}$ vanish as $x_i = 0$.

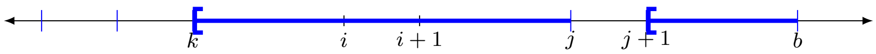


Fig. 3: Terms involving $T_{kj}T_{i+1,b}$ vanish as only one of T_{kj}, $T_{i+1,b}$ can be 1.

Remark 1 The representation in (27) for the cross terms $x_i x_{i+1}$ can be easily understood via the interval partition representation $I_{\mathbf{x}}$ for $\mathbf{x} \in \mathcal{X}_{app}$ described in Lemma 4. For any point $\mathbf{x} \in \mathcal{X}_{app}$, identify the only interval in the partition $I_{\mathbf{x}}$ containing i to be $[k, j]$. Then we have that $x_j = 0, x_{j-1} = 1, \ldots, x_k = j - k$ and $T_{kj} = 1$ (see (22)). If $i + 1 \in [k, j]$, then $x_i = (j - i)$ and $x_{i+1} = j - (i + 1)$ and the product $x_i x_{i+1} = T_{kj}(j - i)(j - (i + 1))$. On the other hand, if $i + 1$ does not belong to the interval $[k, j]$, we have $x_i = 0$; consequently, again $x_i x_{i+1} = T_{kj}(j - i)(j - (i + 1))$. Since only one element, T_{kj}, in the collection $\{T_{ab} : a \leq i \leq b\}$ equals one, the representation in (27) holds. While the representation for the square terms in (25) has been known in the literature (see, for example, [38]), the representation for the specific cross terms in (27) has been explicitly characterized, as far as we know, for the first time in this paper.

Combining the observation in (27) with that in (25) we obtain

$$\begin{aligned}
C_{app} &= conv \left\{ \left(x_1, \ldots, x_n, x_1^2, \ldots, x_n^2, x_1 x_2, x_3 x_4, \ldots, x_{n-1} x_n \right) : \mathbf{x} \in \mathcal{X}_{app} \right\} \\
&= conv \Bigg\{ (p_1, \ldots, p_n, X_{11}, \ldots, X_{nn}, X_{12}, X_{34}, \ldots, X_{n-1,n}) \in \mathbb{R}^{5n/2} : \\
&\qquad p_i = \sum_{k=1}^{i} \sum_{j=i}^{n+1} T_{kj}(j-i), \;\; X_{ii} = \sum_{k=1}^{i} \sum_{j=i}^{n+1} T_{kj}(j-i)^2, \quad \text{for } i \in [n], \\
&\qquad X_{i,i+1} = \sum_{k=1}^{i} \sum_{j=i+1}^{n+1} T_{kj}(j-i)(j-(i+1)), \quad \text{for } i \in [n], i \text{ odd}, \\
&\qquad \sum_{k=1}^{i} \sum_{j=i}^{n+1} T_{kj} = 1, \text{ for } i \in [n], \quad T_{kj} \in \{0, 1\} \text{ for } 1 \leq k \leq j \leq n+1 \Bigg\},
\end{aligned}$$

as a consequence of Lemma 4. Further, total unimodularity of the constraints over T verifies the representation for C_{app} in the statement of Proposition 1. □

With this characterization of the set C_{app} in Proposition 1, observe that the statement of Theorem 2 follows as a consequence of the formulation in (24). This completes the proof of Theorem 2. □

Remark 2 Mean-variance bound: For any given schedule $\mathbf{s} \in S$, the worst-case expected total waiting time,

$$Z^*_{mv}(\mathbf{s}) = \sup \left\{ \mathbb{E}_\theta \left[f(\tilde{\mathbf{u}}, \mathbf{s}) \right] : \ \mathbb{E}_\theta \left[\tilde{u}_i \right] = \mu_i, \mathbb{E}_\theta [\tilde{u}_i^2] = \Pi_{ii}, \ \text{for } i \in [n], \ \theta \in \mathcal{P}(\mathbb{R}^n) \right\}, \tag{28}$$

that is consistent with given mean and variance information of the service times $\{\tilde{u}_i : i \in [n]\}$, can be computed by solving the semidefinite programming formulation below in (29). This formulation results from a similar application of Theorem 1 to the simpler case where $\mathcal{N}_r = \{r\}$, for $r = 1, \ldots, n$.

$$\begin{aligned}
Z^*_{mv}(\mathbf{s}) = \max_{p_i, X_{ii}, Y_{ii}, t_{kj}} \ & \sum_{i=1}^{n} (Y_{ii} - s_i p_i) \\
\text{s.t} \ & \begin{bmatrix} 1 & \mu_i & p_i \\ \mu_i & \Pi_{ii} & Y_{ii} \\ p_i & Y_{ii} & X_{ii} \end{bmatrix} \succeq 0, && \text{for } i \in [n], \\
& p_i = \sum_{k=1}^{i} \sum_{j=i}^{n+1} t_{kj} (j - i), && \text{for } i \in [n], \\
& X_{ii} = \sum_{k=1}^{i} \sum_{j=i}^{n+1} t_{kj} (j - i)^2, && \text{for } i \in [n], \\
& \sum_{k=1}^{i} \sum_{j=i}^{n+1} t_{kj} = 1, && \text{for } i \in [n], \\
& t_{kj} \geq 0, && \text{for } 1 \leq k \leq j \leq n+1.
\end{aligned} \tag{29}$$

Similar to the formulation in Corollary 1, a distributionally robust schedule that minimizes the worst-case total expected waiting time can be found by solving the following semidefinite program:

$$\begin{aligned}
\min_{\mathbf{s}, \boldsymbol{\eta}, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \boldsymbol{\rho}, \boldsymbol{\delta}, \boldsymbol{\gamma}} \ & \sum_{i=1}^{n} \eta_i + \sum_{i=1}^{n} \beta_i \mu_i + \sum_{i=1}^{n} \Gamma_{ii} \Pi_{ii} + \sum_{i=1}^{n+1} \rho_i \\
\text{s.t} \ & \begin{bmatrix} 2\eta_i & \beta_i & \delta_i + s_i \\ \beta_i & 2\Gamma_{ii} & -1 \\ \delta_i + s_i & -1 & 2\gamma_i \end{bmatrix} \succeq 0, && \text{for } i \in [n], \\
& \sum_{i=k}^{j} \rho_i \geq \sum_{i=k}^{\min\{j,n\}} \delta_i (j - i) + \sum_{i=k}^{\min\{j,n\}} \gamma_i (j - i)^2, && \text{for } 1 \leq k \leq j \leq n+1, \\
& \sum_{i=1}^{n} s_i \leq T, \\
& s_i \geq 0, && \text{for } i \in [n].
\end{aligned} \tag{30}$$

The semidefinite program in (30) can be seen as an alternative to the second order conic programming formulation in [38] where the problem of appointment scheduling in the presence of mean and variance information was considered.

Remark 3 A representation for cross terms $x_i x_{i+2}$, similar to that in (27) in terms of variables $T_{kj} \in \{0,1\}$, does not result in linear representation in the variables T_{kj}. To see this, recall the interval partition representation described in Lemma 4. Consider any $\mathbf{x} \in \mathcal{X}_{app}$ such that there exist k, j satisfying $[k, i+1] \in I_{\mathbf{x}}$ and $[i+2, j] \in I_{\mathbf{x}}$. Then $x_i = 1$ and $x_{i+2} = j - (i+2)$, in which case $x_i x_{i+2} = \sum_{k=1}^{i+1} \sum_{j=i+2}^{n+1} T_{k,i+1} T_{i+2,j}(j-(i+2))$, which cannot be reduced in a straightforward manner to a linear representation as in (27).

4.2 The Newsvendor Problem

We next study the version of the newsvendor problem presented in [29,43]. We look at the problem of a newsvendor who sells a product in multiple stores spanning several geographical areas in a city. For a store i, the unit order cost of the product is c_i while the unit selling price is v_i and the quantity to be ordered is q_i while the demand is $\tilde{d}_i$. For each product unsold in store i, a cost of g_i is recovered, while the cost for a unit stock-out is f_i. The decision variable is the optimal order quantity vector $\mathbf{q}$. We consider the budgeted version of the newsvendor problem where the newsvendor has a budget Q for the overall order quantity. The set of all feasible order quantities is therefore, $\mathcal{Q} = \{\mathbf{q} \in \mathbb{R}_+^n : \sum_{i=1}^n q_i \leq Q\}$. For a given order quantity vector and demand vector, the total cost for the newsvendor is,

$$
\begin{aligned}
Z(\mathbf{q}, \tilde{\mathbf{d}}) &= \sum_{i=1}^n c_i q_i - \sum_{i=1}^n v_i \min(q_i, \tilde{d}_i) - \sum_{i=1}^n g_i(q_i - \min(q_i, \tilde{d}_i)) + \sum_{i=1}^n f_i(\tilde{d}_i - \min(q_i, \tilde{d}_i)) \\
&= \mathbf{a}'\mathbf{q} + \mathbf{b}'\tilde{\mathbf{d}} + \sum_{i=1}^n h_i [\tilde{d}_i - q_i]^+ \\
&= \mathbf{a}'\mathbf{q} + \mathbf{b}'\tilde{\mathbf{d}} + \max_{\mathbf{x} \in \{0,1\}^n} \sum_{i=1}^n h_i x_i (d_i - q_i) \qquad (31)
\end{aligned}
$$

where $\mathbf{a} = \mathbf{c} - \mathbf{g} \geq 0, \mathbf{b} = \mathbf{g} - \mathbf{v}, \mathbf{h} = \mathbf{v} - \mathbf{g} + \mathbf{f} \geq 0$ and $[\tilde{d}_i - q_i]^+ = \max(\tilde{d}_i - q_i, 0)$.

The order quantity $\mathbf{q}$ must be decided even before the demand is known and hence the demand vector $\tilde{\mathbf{d}}$ is assumed to be a random variable with a distribution θ. In the distributionally robust version of the newsvendor problem, the distribution θ however is assumed to be unknown and the order quantity is selected in such a way that the worst case expected cost of the newsvendor is minimized. With knowledge of the first moment specified by $\boldsymbol{\mu}$ and the entire second moment matrix $\boldsymbol{\Pi}$, the distributionally robust optimization problem,

$$
\min_{\mathbf{q} \in \mathbb{R}_+^n} \sup \left\{ \mathbb{E}_\theta \left[Z(\mathbf{q}, \tilde{\mathbf{d}}) \right] : \mathbb{E}_\theta[\tilde{\mathbf{d}}] = \boldsymbol{\mu}, \, \mathbb{E}_\theta[\tilde{\mathbf{d}}\tilde{\mathbf{d}}'] = \boldsymbol{\Pi}, \theta \in \mathcal{P}(\mathbb{R}^n) \right\} \qquad (32)
$$

is NP-hard [6,29]. Observe that for a given $\mathbf{q}$ and $\mathbf{d}$, the newsvendor problem in (31) boils down to maximizing a linear function over the hypercube $\{0,1\}^n$. Then, formulation (6) provides an approach to reformulate the second stage problem in (32). However the reformulation thus obtained is exponential sized, as it involves enumerating all the 2^n extreme points of $\{0,1\}^n$, towards capturing the convex hull, $conv\{\mathbf{x}\mathbf{x}', \mathbf{x} \in \{0,1\}^n\}$. This convex hull, commonly referred to as the boolean quadric polytope (BQP), is known to be hard to characterise in general [49]. In order to circumvent the difficulty that arises from the hardness of BQP, [43] propose the following semidefinite programming relaxation for the second stage problem based on approximations for BQP.

$$
\begin{aligned}
\max_{\mathbf{p},\mathbf{X},\mathbf{Y}} \;& trace(diag(\mathbf{h})\mathbf{Y}) - (\mathbf{h}\circ\mathbf{q})'\mathbf{p} + \mathbf{a}'\mathbf{q} + \mathbf{b}'\boldsymbol{\mu} \\
\text{s.t.} \;& \begin{bmatrix} 1 & \boldsymbol{\mu}' & \mathbf{p}' \\ \boldsymbol{\mu} & \boldsymbol{\Pi} & \mathbf{Y}' \\ \mathbf{p} & \mathbf{Y} & \mathbf{X} \end{bmatrix} \succeq 0 \\
& X_{ii} = p_i \text{ for } i \in [n] \\
& X_{ij} \leq \min(p_i, p_j) \text{ for } i,j \in [n] \\
& X_{ij} \geq \max(p_i + p_j - 1, 0) \text{ for } i,j \in [n]
\end{aligned}
\tag{33}
$$

We will now propose an exact reformulation for the distributionally robust newsvendor problem under knowledge of partial correlations. As we will see next, the knowledge of partial correlations occurs naturally in the newsvendor application. Assume that there are a total of R clusters of the n stores. These R clusters can be thought of geographical areas in which the stores are located. Let $\mathcal{N}_1, \ldots, \mathcal{N}_R$ be a partition of the set $\mathcal{N} = \{1, \ldots, n\}$. The set $\mathcal{N}_r$ corresponds to the set of all stores in a geographical area r. The correlation of the demand between the various stores in a particular area is usually available to the newsvendor. Let the first moment be specified by $\mathbb{E}[\tilde{\mathbf{d}}] = \boldsymbol{\mu}$ and the second moment matrices $\mathbb{E}[\tilde{\mathbf{d}}^r(\tilde{\mathbf{d}}^r)'] = \boldsymbol{\Pi}^r$ for $r \in [R] = \{1, \ldots, R\}$. For a fixed order quantity $\mathbf{q}$, denote by $Z^*_{nv}(\mathbf{q})$ the worst-case expected cost to the newsvendor,

$$
Z^*_{nv}(\mathbf{q}) = \sup\left\{\mathbb{E}_\theta\left[Z(\mathbf{q},\tilde{\mathbf{d}})\right] : \mathbb{E}_\theta[\tilde{\mathbf{d}}] = \boldsymbol{\mu},\, \mathbb{E}_\theta[\tilde{\mathbf{d}}^r(\tilde{\mathbf{d}}^r)'] = \boldsymbol{\Pi}^r \text{ for } r \in [R],\, \theta \in \mathcal{P}(\mathbb{R}^n)\right\}
\tag{34}
$$

We are interested in finding the distributionally robust order quantity by solving the following problem:

$$
Z^*_{nv} = \min_{\mathbf{q}\in\mathcal{Q}} Z^*_{nv}(\mathbf{q})
$$

Theorem 3 *For a given order quantity vector $\mathbf{q} \in \mathcal{Q}$, let*

$$
\begin{aligned}
\hat{Z}^*_{nv}(\mathbf{q}) = \max_{\substack{\mathbf{p}^r,\mathbf{X}^r\\ \mathbf{Y}^r,\alpha^r_{\mathbf{x}^r}}} \;& \sum_{r=1}^{R} trace(diag(\mathbf{h}^r)\mathbf{Y}^r) - (\mathbf{h}^r\circ\mathbf{q}^r)'\mathbf{p}^r + \mathbf{a}'\mathbf{q} + \mathbf{b}'\boldsymbol{\mu} \\
s.t. \;& \begin{bmatrix} 1 & \boldsymbol{\mu}^{r\prime} & \mathbf{p}^{r\prime} \\ \boldsymbol{\mu}^r & \boldsymbol{\Pi}^r & \mathbf{Y}^{r\prime} \\ \mathbf{p}^r & \mathbf{Y}^r & \mathbf{X}^r \end{bmatrix} \succeq 0, \qquad \text{for } r \in [R], \\
& \mathbf{p}^r = \textstyle\sum_{\mathbf{x}^r\in\{0,1\}^{n_r}} \alpha^r_{\mathbf{x}^r}\mathbf{x}^r \,\forall r \in [R] \\
& \mathbf{X}^r = \textstyle\sum_{\mathbf{x}^r\in\{0,1\}^{n_r}} \alpha^r_{\mathbf{x}^r}\mathbf{x}^r(\mathbf{x}^r)' \,\forall r \in [R] \\
& \textstyle\sum_{\mathbf{x}^r\in\{0,1\}^{n_r}} \alpha^r_{\mathbf{x}^r} = 1 \,\forall r \in [R] \\
& \alpha^r_{\mathbf{x}^r} \geq 0 \,\forall \mathbf{x}^r \in \{0,1\}^{n_r}, \forall r \in [R]
\end{aligned}
$$

Then $\hat{Z}^*_{nv}(\mathbf{q}) = Z^*_{nv}(\mathbf{q})$.

Proof From (31), $Z(\mathbf{q},\tilde{\mathbf{d}})$ can be re-written as,

$$
\begin{aligned}
Z(\mathbf{q},\tilde{\mathbf{d}}) &= \mathbf{a}'\mathbf{q} + \mathbf{b}'\tilde{\mathbf{d}} + \sum_{r\in[R]} (\mathbf{h}^r)'(\tilde{\mathbf{d}}^r - \mathbf{q}^r)^+ \\
&= \mathbf{a}'\mathbf{q} + \mathbf{b}'\tilde{\mathbf{d}} + \sum_{r\in[R]} \max_{\mathbf{x}^r\in\{0,1\}^{n_r}} (\mathbf{h}^r\circ\tilde{\mathbf{d}}^r)'\mathbf{x}^r - (\mathbf{h}^r\circ\mathbf{q}^r)'\mathbf{x}^r
\end{aligned}
$$

Now, for a fixed order quantity $\mathbf{q}$, an application of Theorem 1 (by setting $\tilde{\mathbf{c}} = \tilde{\mathbf{d}}$), gives us the above formulation for $Z^*_{nv}(\mathbf{q})$. □

We now describe how the optimal first stage decisions can be computed.

Corollary 2 *Given $Q > 0$, an order quantity $\mathbf{q} \in \mathcal{Q} = \{\mathbf{q} \in \mathbb{R}^n_+ : q_1 + \ldots + q_n \leq Q\}$ that minimizes $Z^*_{nv}(\mathbf{q})$ can be obtained by solving the following semidefinite program:*

$$
\begin{aligned}
Z^*_{nv} = & \min_{\mathbf{q},\boldsymbol{\eta},\boldsymbol{\beta},\boldsymbol{\Gamma},\boldsymbol{\delta},\boldsymbol{\tau},\mathbf{W}} \sum_{r\in[R]} (\eta_r + trace(\boldsymbol{\Pi}^r\boldsymbol{\Gamma}^r) + \tau_r) + \boldsymbol{\beta}'\boldsymbol{\mu} + \mathbf{a}'\mathbf{q} + \mathbf{b}'\boldsymbol{\mu} \\
s.t \quad & \begin{bmatrix} 2\eta_r & \boldsymbol{\beta}^{r\prime} & (\boldsymbol{\delta}^r + \mathbf{q}^r \circ \mathbf{h}^r)' \\ \boldsymbol{\beta}^r & 2\boldsymbol{\Gamma}^r & -diag(\mathbf{h}^r) \\ \boldsymbol{\delta}^r + \mathbf{q}^r \circ \mathbf{h}^r & -diag(\mathbf{h}^r) & 2\mathbf{W}^r \end{bmatrix} \succeq 0, \quad for\ r \in [R], \\
& \tau_r \geq \sum_{i\in|\mathcal{N}_r|} (\delta_i + W^r_{ii})\mathbf{x}^r_i + \sum_{i\in|\mathcal{N}_r|} \sum_{\substack{j\in|\mathcal{N}_r| \\ j\neq i}} W^r_{ij}\mathbf{x}^r_i\mathbf{x}^r_j \quad for\ \mathbf{x}^r \in \{0,1\}^{n_r}, r \in [R], \\
& \sum_{i=1}^n q_i \leq Q, \quad q_i \geq 0,\ for\ i \in [n].
\end{aligned}
$$

When $n_r \leq \log n$, for all $r \in [R]$, Z^*_{nv} is polynomial time solvable in n.

Proof Take dual of the semidefinite program in Theorem 3 and perform a joint minimization over the dual variables and $\mathbf{q}$. It can be verified that an interior point exists for the dual as well as primal semidefinite programs if $\boldsymbol{\Pi}^r \succ \boldsymbol{\mu}^r(\boldsymbol{\mu}^r)'$ thereby guaranteeing strong duality. Note that the above dual formulation requires enumeration of all the points in $\{0,1\}^{n_r}$ and therefore involves $\sum_{r\in[R]} 2^{n_r}$ inequality constraints involving variables in R positive semidefinite matrices, each of size $2n_r^2 + 1$. In the special case where $n_r \leq \log n$, the formulation is polynomial sized in n.

4.3 Longest path in directed acyclic graphs

In this section, we examine the problem of computing the expected length of the longest path between a fixed start node and a sink node in a directed acyclic graph whose arc lengths are uncertain. A key application of this longest path problem is to estimate project completion times using Project Evaluation and Review Technique (PERT) networks in project management (see, for example, [55]). A PERT network is a directed acyclic graph representation of a project that consists of several activities with partially specified precedence relationship among the activities. Our objective is to tackle the case where the activity durations (arc lengths) are random, dependent and their joint distribution is not fully known.

Let $V = \{0, \ldots, m-1\}$ denote the set of nodes of a directed acyclic graph G. Suppose that the nodes 0 and $m-1$ represent the start and sink nodes. Let $\mathcal{A}$ denote the set of arcs in G and c_{ij} denote the length of arc (i,j) between nodes i and j. If G is a PERT network, the nodes 0 and $m-1$ represent the start and end of the project; the length of the longest path between nodes 0 and $m-1$ represents the project completion duration. The length of the longest path can be represented as the optimal objective value of the following combinatorial optimization problem:

$$
\begin{aligned}
Z(\mathbf{c}) = \max \sum_{(i,j)\in\mathcal{A}} & c_{ij}x_{ij} \\
\text{s.t} \sum_{j:(i,j)\in\mathcal{A}} x_{ij} - \sum_{j:(j,i)\in\mathcal{A}} x_{ji} &= \begin{cases} 1, & \text{if } i = 0, \\ -1, & \text{if } i = m-1, \\ 0, & \text{otherwise,} \end{cases} \\
x_{ij} \in \{0,1\}, \quad & \text{for } (i,j) \in \mathcal{A}.
\end{aligned} \tag{35}
$$

If the arc lengths $(c_{ij})_{(i,j)\in\mathcal{A}}$ are known, $Z(\mathbf{c})$ can be computed in polynomial-time by solving the linear programming relaxation of the formulation in (35) due to the total unimodularity of the underlying constraint matrix.

On the other hand, if the arc lengths are random, exact computation of the expected length of the longest path is known to be #P-hard even with the assumption of independence among arc lengths (see [27]). For specialized graph structures such as series-parallel graphs, it has been shown in [3,40] that the expected length of the longest path can be computed in time polynomial in the size of the graph and the number of points in the discrete support of the arc lengths.

In the absence of the knowledge of the entire joint distribution of the arc lengths, the distributionally robust formulations in [7,8] result in polynomial-time solvable bounds for the project duration when the marginal moments of arc lengths are specified. A natural approach to specify correlation information in PERT networks, in order to obtain tighter bounds, is to consider all the activities that enter a node to be related and therefore specify correlation information among all activities that enter a node. Indeed, such a partition formed by sets of incoming arcs into nodes have been considered for specifying marginal distribution information in [20,23,52]. Theorem 4 below identifies a polynomial-time solvable formulation for evaluating the maximum possible (worst-case) expected project duration in the presence of mean and covariance information of activity durations whose arcs enter the same node.

To fix notation, let n be the cardinality of the set $\mathcal{A}$ of arcs and $R = m-1$. For the given directed acyclic graph G, consider the following partition of $\mathcal{A}$,

$$
\mathcal{N}_r = \{i : (i,r) \in \mathcal{A}\}, \qquad \text{for } r = 1, \dots, R,
$$

formed by considering sets of arcs that enter node r, for $r = 1, \dots, m-1$. Let $\tilde{\mathbf{c}} = (\tilde{c}_{ij})_{(i,j)\in\mathcal{A}}$ be the random vector of arc lengths and $\tilde{\mathbf{c}}^r = (\tilde{c}_{ir})_{i:\,(i,r)\in\mathcal{A}}$ be the random subvector of arc lengths of arcs entering node r, for $r = 1, \dots, R$. Given that the expected value of $\tilde{\mathbf{c}}$ is $\boldsymbol{\mu}$ and that of $\tilde{\mathbf{c}}^r(\tilde{\mathbf{c}}^r)'$ is $\mathbf{\Pi}^r$ for every $r \in \{1, \dots, m-1\}$, our objective is to evaluate,

$$
Z^*_{path} = \sup \left\{ \mathbb{E}_\theta \left[Z(\tilde{\mathbf{c}}) \right] : \mathbb{E}_\theta[\tilde{\mathbf{c}}] = \boldsymbol{\mu},\, \mathbb{E}_\theta[\tilde{\mathbf{c}}^r(\tilde{\mathbf{c}}^r)'] = \mathbf{\Pi}^r \text{ for } r \in [R],\, \theta \in \mathcal{P}(\mathbb{R}^n) \right\}, \tag{36}
$$

where $Z(\cdot)$ is specified as in (35).

Theorem 4 Z^*_{path} *can be evaluated as the optimal objective value of the following semidefinite program:*

$$
\begin{aligned}
Z^*_{path} = \max_{\mathbf{p}^r, \mathbf{X}^r, \mathbf{Y}^r} & \sum_{r=1}^{m-1} trace(\mathbf{Y}^r) \\
s.t & \begin{bmatrix} 1 & \boldsymbol{\mu}^{r\prime} & \mathbf{p}^{r\prime} \\ \boldsymbol{\mu}^r & \boldsymbol{\Pi}^r & \mathbf{Y}^{r\prime} \\ \mathbf{p}^r & \mathbf{Y}^r & \mathbf{X}^r \end{bmatrix} \succeq 0, \qquad for\ r \in \{1, \ldots, m-1\}, \\
& \sum_{j:(i,j)\in\mathcal{A}} p_{ij} - \sum_{j:(j,i)\in\mathcal{A}} p_{ji} = \begin{cases} 1, & if\ i = 0, \\ -1, & if\ i = m-1, \\ 0, & otherwise, \end{cases} \\
& X^r_{jk} = \begin{cases} p_{ij}, & if\ j = k, \\ 0, & otherwise, \end{cases} \qquad for\ r = 1, \ldots, m-1, \\
& p_{ij} \geq 0, \quad for\ (i,j) \in \mathcal{A}.
\end{aligned}
\tag{37}
$$

Proof Let us use $\mathcal{X}_{path}$ to denote the feasible region to the formulation (35). Then as an application of Theorem 1, Z^*_{path} can be written as the optimal objective value of the semidefinite program in (7). To efficiently represent the convex hull constraint in (7), observe that for any $\mathbf{x} \in \mathcal{X}_{path}$,

$$
x_{ir}x_{jr} = \begin{cases} x_{ir}, & \text{if } i = j, \\ 0, & \text{if } i \neq j, \end{cases} \tag{38}
$$

for every i, j such that $i, j \in \mathcal{N}_r$. This follows from the observation that any path from 0 to m that passes through r can contain only one of the arcs $\{(k,r) : (k,r) \in \mathcal{A}\}$. To see this explicitly from the constraints in (35), observe that if $\mathbf{x}$ is such that $\sum_{k:(k,r)\in\mathcal{A}} x_{kr} = 1$, then as $x_{kr} \in \{0,1\}$, only one of $\{x_{kr} : (k,r) \in \mathcal{A}\}$ equals 1. Therefore $x_{ir}x_{jr} = 0$ if $i \neq j$. On the other hand, $x^2_{ir} = x_{ir}$ as $x_{ir} \in \{0,1\}$, thus verifying (38). As a result of this and total unimodularity of the constraints in formulation (35),

$$
\begin{aligned}
conv & \left\{ \left(\mathbf{x}, \mathbf{x}^1\mathbf{x}^{1\prime}, \ldots, \mathbf{x}^{m-1}\mathbf{x}^{m-1\prime} \right) : \mathbf{x} \in \mathcal{X}_{path} \right\} \\
& = conv \left\{ \left(\mathbf{x}, \mathrm{Diag}(\mathbf{x}^1), \ldots, \mathrm{Diag}(\mathbf{x}^{m-1}) \right) : \mathbf{x} \in \mathcal{X}_{path} \right\}, \\
& = \left\{ \left(\mathbf{p}, \mathrm{Diag}(\mathbf{p}^1), \ldots, \mathrm{Diag}(\mathbf{p}^{m-1}) \right) : \mathbf{p} \in conv(\mathcal{X}_{path}) \right\},
\end{aligned}
$$

where $\mathrm{Diag}(\mathbf{x}^r)$ denotes the $n_r \times n_r$ diagonal matrix formed with elements from the subvector $\mathbf{x}^r$. Since the convex hull of $\mathcal{X}_{path}$ is simply the collection of points $\mathbf{p} = (p_{ij})_{(i,j)\in\mathcal{A}}$ such that $p_{ij} \geq 0$ and

$$
\sum_{j:(i,j)\in\mathcal{A}} p_{ij} - \sum_{j:(j,i)\in\mathcal{A}} p_{ji} = \begin{cases} 1, & \text{if } i = 0, \\ -1, & \text{if } i = m-1, , \\ 0, & \text{otherwise,} \end{cases}
$$

the constraints in the formulation (37) are equivalent to those in (7). This completes the proof of Theorem 4. □

4.4 Bounds on the expected value of random linear assignment problems

Assignment (or) matching problems constitute a special class of combinatorial optimization problems whose properties of random instances have been studied extensively in the literature. The specific example of linear assignment problem (see [33,41]) corresponds to the setting where m entities belonging to a set V need to be assigned to m entities in a set U such that the total utility is maximized. The entities in the sets U and V can be thought, respectively, as candidates and jobs that need to be performed by the candidates. Each candidate must be assigned to exactly one job and each job must be assigned to exactly one candidate. With c_{ij} representing the utility of assigning the entity i in set U to the entity j in set V, the linear assignment problem which maximizes the total utility is formulated as,

$$\begin{aligned} Z(\mathbf{c}) = \max_{x_{ij}} & \sum_{i \in U} \sum_{j \in V} c_{ij} x_{ij} \\ \text{s.t} & \sum_{j \in V} x_{ij} = 1, \quad \text{for } i \in U, \\ & \sum_{i \in U} x_{ij} = 1, \quad \text{for } j \in V, \\ & x_{ij} \in \{0,1\}, \quad \text{for } i \in U, j \in V. \end{aligned} \tag{39}$$

In graph-theoretic terms, the formulation (39) corresponds to finding a maximum-weight perfect matching in a bipartite graph with edge weights $\mathbf{c} = (c_{ij} : (i,j) \in U \times V)$. Similar to the formulation for computing the length of the longest path in (35), the constraints in the formulation (39) are totally unimodular. Therefore a linear programming relaxation can be used to identify an optimal assignment in polynomial-time.

Studying the distributional properties of random instances of large-scale linear assignment problems has received much attention since the early works of [21,34] in 1960s. Specifically, assuming that the coefficients $\tilde{\mathbf{c}} = (\tilde{c}_{ij} : (i,j) \in U \times V)$ are independent uniform $[0,1]$ random variables, progressively better upper and lower bounds for $E[Z(\tilde{\mathbf{c}})]$ are established in [25,31,36,56] and [45]. In particular, the derivation of the bound in [31] has served as a foundation for analysing a broader class of random linear programs in [22]. One of the most widely known results in this area is the Aldous's proof [1] of the conjecture by Mezard and Parisi [39]; their conjecture is that the limiting expected value of the objective of minimum cost random linear assignment problem equals $\pi^2/6$, when the number of entities $m \to \infty$ and the random coefficients are taken to be independent copies of uniform [0,1] random variable (or) exponential random variable with mean 1. Subsequent non-asymptotic studies that remove the strict assumptions of identical marginal distributions and independence among coefficients can be found, respectively, in [13,58] and [7]. In particular, Bertsimas et. al [7] propose a formulation to compute $\sup \mathbb{E}[Z(\tilde{\mathbf{c}})]$ under knowledge of marginal moments. They show that that the computation can be performed in polynomial time and obtain tight bounds. We will now illustrate how Theorem 1 can serve as a computational tool for arriving at tight bounds for $E[Z(\tilde{\mathbf{c}})]$ when the mean and certain second moments of $\tilde{\mathbf{c}}$ are specified. A precise description of the problem we consider is as follows.

Identifying the entries in U with $\{1,\ldots,m\}$, we take $\mathcal{N}_r = \{(r,j) : j \in V\}$, for $r \in [m]$. This corresponds to the setting where the correlation of utilities between any two jobs for the same candidate is known, but correlation across candidates is not known. Let $\tilde{\mathbf{c}} = (\tilde{c}_{ij})_{i:i\in U, j\in V}$ be the random vector of utilities and $\tilde{\mathbf{c}}^i = (\tilde{c}_{ij})_{j\in V}$ be the random subvector of $\tilde{\mathbf{c}}$ when the indices are restricted to the subset $\mathcal{N}_i$. We aim to evaluate the bound,

$$Z^*_{lap} = \sup\left\{\mathbb{E}_\theta\left[Z(\tilde{\mathbf{c}})\right] : \mathbb{E}_\theta[\tilde{\mathbf{c}}] = \boldsymbol{\mu},\, \mathbb{E}_\theta[\tilde{\mathbf{c}}^r(\tilde{\mathbf{c}}^r)'] = \mathbf{\Pi}^r \text{ for } r \in [R],\, \theta \in \mathcal{P}(\mathbb{R}^n)\right\}, \quad (40)$$

where $Z(\cdot)$ is given by (39). As an alternative to the partition considered, one could consider the partition where we identify the entities in V with $\{1,\dots,m\}$ and take $\mathcal{N}_r = \{(i,r) : i \in U\}$, for $r \in [m]$. In settings where $-c_{ij}$ can be interpreted as the cost for assigning job j to candidate i, this partition corresponds to knowing correlation between costs for the same job when performed by different candidates. The following observation can be replicated for this partition as well.

Theorem 5 *Suppose that Z^*_{lap} is defined as in (40). Then Z^*_{lap} can be evaluated as the optimal objective value of the following semidefinite program:*

$$\begin{aligned}
Z^*_{lap} = \max_{\mathbf{p}^r, \mathbf{X}^r, \mathbf{Y}^r} & \sum_{r=1}^m trace(\mathbf{Y}^r) \\
s.t & \begin{bmatrix} 1 & \boldsymbol{\mu}^{r\prime} & \mathbf{p}^{r\prime} \\ \boldsymbol{\mu}^r & \mathbf{\Pi}^r & \mathbf{Y}^{r\prime} \\ \mathbf{p}^r & \mathbf{Y}^r & \mathbf{X}^r \end{bmatrix} \succeq 0, \ \textit{for } r \in [m], \\
& \sum_{j\in V} p_{ij} = 1, \ \textit{for } i \in U \\
& \sum_{i\in U} p_{ij} = 1, \ \textit{for } j \in V, \\
& X^r_{jk} = \begin{cases} p_{rj}, & \textit{if } j = k, \\ 0, & \textit{otherwise}, \end{cases} \quad \textit{for } r = 1,\dots,m, \\
& p_{ij} \ge 0, \ \textit{for } i \in U, j \in V.
\end{aligned} \quad (41)$$

Proof As in the proof of Theorem 4, let $\mathcal{X}_{lap}$ be the bounded feasible region to the formulation (39). Then as an application of Theorem 1, Z^*_{lap} can be written as the optimal objective value of the semidefinite program in (7). To efficiently represent the convex hull constraint in (7), observe that for any $\mathbf{x} \in \mathcal{X}_{lap}$,

$$x_{ij}x_{ik} = \begin{cases} x_{ij}, & \text{if } j = k, \\ 0, & \text{if } j \neq k, \end{cases}$$

for every $i \in U$, $j,k \in V$. This is because, as in the proof of Theorem 4, only one of $\{x_{ij} : j \in V\}$ equals 1, for every $i \in U$; here recall the constraint, $\sum_{j\in V} x_{ij} = 1$, that dictates that only one entity from V is assigned exactly to every $i \in U$. With $\mathbf{x}^i = (x_{ij})_{j\in V}$, we obtain,

$$\begin{aligned}
& conv\left\{\left(\mathbf{x}, \mathbf{x}^1\mathbf{x}^{1\prime}, \dots, \mathbf{x}^m\mathbf{x}^{m\prime}\right) : \mathbf{x} \in \mathcal{X}_{lap}\right\} \\
&\quad = conv\left\{\left(\mathbf{x}, \mathrm{Diag}(\mathbf{x}^1), \dots, \mathrm{Diag}(\mathbf{x}^m)\right) : \mathbf{x} \in \mathcal{X}_{lap}\right\}, \\
&\quad = \left\{\left(\mathbf{p}, \mathrm{Diag}(\mathbf{p}^1), \dots, \mathrm{Diag}(\mathbf{p}^m)\right) : \mathbf{p} \in conv(\mathcal{X}_{lap})\right\},
\end{aligned}$$

where $\text{Diag}(\mathbf{x}^i)$ denotes the $m \times m$ diagonal matrix formed with elements from the subvector $\mathbf{x}^i$. Since, due to total unimodularity, the convex hull of $\mathcal{X}_{lap}$ is simply the collection of points $\mathbf{p} = (p_{ij})_{i\in U, j\in V}$ such that $p_{ij} \geq 0$, $\sum_{j\in V} p_{ij} = 1$ and $\sum_{i\in U} p_{ij} = 1$, the constraints in the formulation (41) are equivalent to those in (7). This completes the proof of Theorem 5. □

The semidefinite program in (41) can be viewed as a useful computational tool for relaxing the strict independence assumptions which are prevalent in the analysis of random instances of assignment problems.

5 Numerical results

In this section, we report the results of numerical experiments for the appointment scheduling formulation considered in Section 4.1. We compare the performance of the semidefinite programming formulation in Theorem 2 (which we refer to as *Non-overlapping*), with the following three alternatives:

a) The mean-variance formulation is solved using the SOCP reformulation originally proposed in [38]. This approach, addressed "*Mean-Variance*" in the discussions that follow, provides a reformulation for $Z^*_{mv}(\mathbf{s})$ in (28). Since the formulation does not make use of any cross moment information, it provides an upper bound for $Z^*_{app}(\mathbf{s})$ in Theorem 2.
The SOCP formulation in (Page 322, Theorem 1, Formulation (20) in [38]) is provided below for reference, where $\boldsymbol{\sigma^2}$ denotes the vector of variances (diagonal entries of $\mathbf{\Pi} - \boldsymbol{\mu}\boldsymbol{\mu}'$):

$$\begin{aligned} Z_{\text{app}}(\boldsymbol{\mu}, \boldsymbol{\sigma^2}) = \min_{\mathbf{s},\beta>0,\alpha,\lambda} \quad & \sum_{i=1}^{n} \lambda_i + \mu_i\alpha_i + (\mu_i^2 + \sigma_i^2)\beta_i \\ \text{s.t.} \quad & \sum_{i=k}^{\min\{n,j\}} \lambda_i \geq \sum_{i=k}^{\min\{n,j\}} \left(\frac{(\pi_{ij} - \alpha_i)^2}{4\beta_i} - s_i\pi_{ij} \right) \\ & \text{for } 1 \leq k \leq n, k \leq j \leq n+1 \end{aligned} \tag{42}$$

where $\pi_{ij} = j - i, 1 \leq i \leq j \leq n+1$.

b) For the second alternative, we solve for $Z^*_{app}(\mathbf{s})$ using a formulation very similar to the formulation originally proposed in (Page 716, [32], maximization problem in formulation (C)), that assumes the knowledge of the mean and all second moments of $\tilde{\mathbf{u}}$ to compute:

$$Z^*_{cov}(\mathbf{s}) = \sup \left\{ \mathbb{E}_\theta \left[f(\tilde{\mathbf{u}}, \mathbf{s}) \right] : \mathbb{E}_\theta \left[\tilde{\mathbf{u}} \right] = \boldsymbol{\mu}, \mathbb{E}_\theta [\tilde{\mathbf{u}}\tilde{\mathbf{u}}'] = \mathbf{\Pi},\ \theta \in \mathcal{P}(\mathbb{R}^n) \right\},$$

In order to adapt this approach to the non-overlapping moments model, we treat the whole second moment matrix as an additional variable $\mathbf{\Delta}$ and only set the block diagonal elements in $\mathbf{\Delta}$ to be $\mathbf{\Pi}^r$ for all r. This adaptation will now give us an exact formulation for $Z^*_{app}(\mathbf{s})$. However, it still involves a completely positive constraint. We therefore relax the completely positive constraint ($\mathbf{M} \in \mathcal{C}(\mathbb{R}_+^{2n+1})$) with a doubly nonnegative matrix constraint ($\mathbf{M} \succeq 0, \mathbf{M}$ non-negative) for tractability (see [32]). Since the adaptation described above solves a maximization problem, relaxing the feasible region will therefore give us an upper bound for $Z^*_{app}(\mathbf{s})$. In particular, in Section 5.1 where we restrict our attention to the distributionally robust bound, we use formulation (43) where $\mathbf{p} \in \mathbb{R}^{2n}$, $\mathbf{Y} \in \mathbb{R}^{n\times 2n}$ and $\mathbf{X} \in \mathbb{R}^{2n\times 2n}$. The matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is such that $A_{jj} = -1$ for $j \in [n]$, $A_{j+1,j} = 1$ for $j \in [n-1]$, $b_j = -1$ for $j \in [n]$ and $\mathbf{a}'_i$ indicates row i of $\mathbf{A}$ and $\mathbf{e}_i \in \mathbb{R}^n$ denotes a column vector with 1 at position i and zero everywhere else. Note that $Z^*_{app}(\mathbf{s}) \leq Z_{\text{app}}(\boldsymbol{\mu}, \boldsymbol{\Pi}^1, \ldots, \boldsymbol{\Pi}^R, \mathbf{s})$.

$$
\begin{aligned}
Z_{\text{app}}(\boldsymbol{\mu}, \boldsymbol{\Pi}^1, \ldots, \boldsymbol{\Pi}^R, \mathbf{s}) = \max_{\mathbf{p},\mathbf{X},\mathbf{Y},\boldsymbol{\Delta}} \quad & trace(\mathbf{Y}) - \mathbf{s}'\mathbf{p} \\
\text{s.t} \quad & \begin{bmatrix} 1 & \boldsymbol{\mu}' & \mathbf{p}' \\ \boldsymbol{\mu} & \boldsymbol{\Delta} & \mathbf{Y}' \\ \mathbf{p} & \mathbf{Y} & \mathbf{X} \end{bmatrix} \succeq 0, \\
& [\mathbf{A}, -\mathbf{I}_n]\mathbf{p} = \mathbf{b}, \\
& [\mathbf{a}'_i, -\mathbf{e}'_i]\mathbf{X}[\mathbf{a}'_i, -\mathbf{e}'_i]' = b_i^2, \, \forall i \in [n] \\
& \boldsymbol{\Delta}^r = \boldsymbol{\Pi}^r, \, \forall r \in [R] \\
& \mathbf{p}, \mathbf{X} \text{ non-negative.}
\end{aligned} \tag{43}
$$

In Section 5.2, where we analyse the distributionally robust schedules, we adapt the dual formulation (Page 718, [32], $Z_D(\mathbf{s})$ along with constraints in (8)), where instead of the copositivity requirement $\mathbf{M} \in \mathcal{C}^*(\mathbb{R}^{2n+1}_+)$, we use $\mathbf{M} = \mathbf{P} + \mathbf{N}, \mathbf{P} \succeq 0, \mathbf{N}$ non-negative as an approximation. Note that in this dual form, the copositivity constraint arises whereas in the primal form (Page 716, [32], maximization problem in formulation (C)), the completely positivity constraint arises, thereby necessitating different approximations for the two formulations. In particular, the distributionally robust schedules may be obtained by optimizing the dual of formulation (43) over the dual variables as well as $\mathbf{s}$. The relevant formulation is provided in Equation (44) where $\alpha, \gamma \in \mathbb{R}$, $\boldsymbol{\beta}, \mathbf{u}, \mathbf{v}, \boldsymbol{\eta}, \boldsymbol{\chi} \in \mathbb{R}^n$, $\boldsymbol{\Gamma} \in \mathbb{R}^{n\times n}$, $\boldsymbol{\Lambda} \in \mathbb{R}^{2n\times 2n}$. $\mathbf{I}_n$ and $\mathbf{O}_n$ denote the identity matrix and the square matrix of zeros respectively, both of size n. We address the primal (43) and dual formulation (44) as "*DNN-relaxation*" in our discussions.

$$
\begin{aligned}
\min_{\substack{\mathbf{s},\alpha,\boldsymbol{\beta},\boldsymbol{\Gamma},\gamma,\\ \mathbf{u},\mathbf{v},\boldsymbol{\Lambda},\boldsymbol{\eta},\boldsymbol{\chi}\\ \mathbf{P},\mathbf{N}}} \quad & \sum_{r=1}^{R} trace((\boldsymbol{\Pi}^r)'\boldsymbol{\Gamma}^r) + \boldsymbol{\mu}'\boldsymbol{\beta} + \alpha + \gamma \\
\text{s.t} \quad & \begin{pmatrix}
\alpha + \sum_{i=1}^{n} v_i - u_i & \frac{\boldsymbol{\beta}'}{2} & \frac{\begin{pmatrix}\mathbf{s}+2\boldsymbol{\eta}\\ 2\boldsymbol{\chi}\end{pmatrix}' - \sum_{i=1}^{n} u_i \begin{pmatrix}\mathbf{a}_i\\ \mathbf{e}_i\end{pmatrix}'}{2} \\
\frac{\boldsymbol{\beta}}{2} & \boldsymbol{\Gamma} & \begin{pmatrix}-0.5\mathbf{I}_n\\ \mathbf{O}_n\end{pmatrix}' \\
\frac{\begin{pmatrix}\mathbf{s}+2\boldsymbol{\eta}\\ 2\boldsymbol{\chi}\end{pmatrix} - \sum_{i=1}^{n} u_i \begin{pmatrix}\mathbf{a}_i\\ \mathbf{e}_i\end{pmatrix}}{2} & \begin{pmatrix}-0.5\mathbf{I}_n\\ \mathbf{O}_n\end{pmatrix} & -\sum_{i=1}^{n} v_i \begin{pmatrix}\mathbf{a}_i\\ -\mathbf{e}_i\end{pmatrix}\begin{pmatrix}\mathbf{a}_i\\ -\mathbf{e}_i\end{pmatrix}' + \boldsymbol{\Lambda}
\end{pmatrix} \succeq 0 \\
& \begin{pmatrix} \gamma & \begin{pmatrix}\boldsymbol{\eta}\\ \boldsymbol{\chi}\end{pmatrix}' \\ \begin{pmatrix}\boldsymbol{\eta}\\ \boldsymbol{\chi}\end{pmatrix} & \boldsymbol{\Lambda} \end{pmatrix} = \mathbf{P} + \mathbf{N} \\
& \mathbf{P} \succeq 0, \mathbf{N} \text{ non-negative}
\end{aligned} \tag{44}
$$

c) The exact value of $Z^*_{app}(\mathbf{s})$ is also computed using an adaptation of the formulation (8) where the extreme points of $\mathcal{X}$ are explicitly enumerated in order to represent the convex hull constraint. The explicit enumeration of the extreme points involves introduction of new scalar variables $\alpha_{\mathbf{x}}$ for each extreme point $\mathbf{x}$ such that $\sum_{\mathbf{x}} \alpha_{\mathbf{x}} = 1$ and $\alpha_{\mathbf{x}} \geq 0$. The convex hull constraint in formulation (8) is captured using the following constraints: $\mathbf{p} = \sum_{\mathbf{x}\in\mathcal{X}_{app}} \alpha_{\mathbf{x}}\mathbf{x}, \mathbf{X} = \sum_{\mathbf{x}\in\mathcal{X}_{app}} \alpha_{\mathbf{x}}\mathbf{x}\mathbf{x}', \sum_{\mathbf{x}\in\mathcal{X}_{app}} \alpha_{\mathbf{x}} = 1, \alpha_{\mathbf{x}} \geq 0 \forall \mathbf{x} \in \mathcal{X}_{app}$. Since the number of extreme points grows exponentially with n, this approach is feasible only for small values of n. This exact approach, labeled as "*Large-SDP*", is feasible in our computational setup only for $n \leq 9$, but nevertheless gives us an exact formulation for $Z^*_{app}(\mathbf{s})$. The formulation is provided below explicitly for a better understanding.

$$\begin{aligned}
\max_{\mathbf{p},\mathbf{X},\mathbf{Y},\boldsymbol{\Delta}} \quad & trace(\mathbf{Y}) - \mathbf{s}'\mathbf{p} \\
\text{s.t} \quad & \begin{bmatrix} 1 & \boldsymbol{\mu}' & \mathbf{p}' \\ \boldsymbol{\mu} & \boldsymbol{\Delta} & \mathbf{Y}' \\ \mathbf{p} & \mathbf{Y} & \mathbf{X} \end{bmatrix} \succeq 0, \\
& \boldsymbol{\Delta}[\mathcal{N}_r] = \boldsymbol{\Pi}^r, \qquad \text{for } r \in [R], \\
& \mathbf{p} = \sum_{\mathbf{x}\in\mathcal{X}_{app}} \alpha_{\mathbf{x}}\mathbf{x} \\
& \mathbf{X} = \sum_{\mathbf{x}\in\mathcal{X}_{app}} \alpha_{\mathbf{x}}\mathbf{x}\mathbf{x}' \\
& \textstyle\sum_{\mathbf{x}\in\mathcal{X}_{app}} \alpha_{\mathbf{x}} = 1, \ \alpha_{\mathbf{x}} \geq 0 \ \forall \mathbf{x} \in \mathcal{X}_{app}
\end{aligned}$$

We also tested a recently proposed alternate approximation scheme proposed in [11] in place of the doubly nonnegative matrix based relaxation for approximating the completely positive constraint in the exact formulation in [32]. The results obtained were identical to the approach labeled above as "*DNN-relaxation*" and hence we only report the results of *DNN-relaxation*. All experiments were run on MATLAB using SDPT3 solver[1] [53,54] and YALMIP interface[2].

5.1 Comparison of worst-case expected total waiting times

Assuming that the correlation coefficient between service times $\tilde{u}_i$ and $\tilde{u}_{i+1}$ equals ρ, for every i in $\{1, 3, \dots, n-1\}$, we compare the objective value of the formulation in Theorem 2 with that of the alternative approaches described above, for various values of ρ in the interval $[-1, 1]$. We report objective values averaged over 50 independent runs, where in each run, the means and variances of $\tilde{c}_i(\mathbf{s}) = \tilde{u}_i - s_i$ are taken to be independent realizations of random variables uniformly distributed in the intervals $[-2, 2]$ and $(0, 5]$ respectively. For all the results in the current subsection, since we are interested in only the bound computation, we set $\mathbf{s} = 0$ in all the formulations.

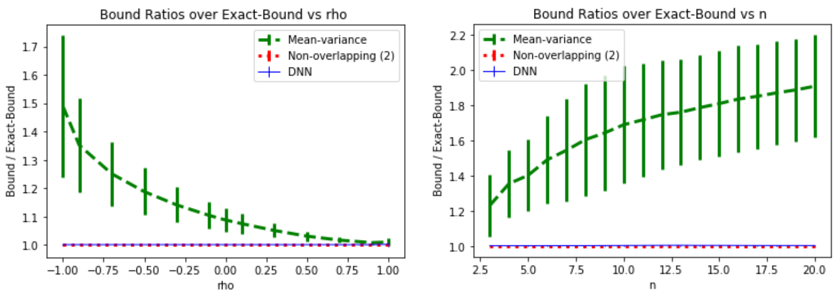
See Figure 4a for a comparison of the ratio of average objective values of our formulation in Theorem 2 and the *Large-SDP* approach for $n = 6$. Table 1 gives the min, max and mean ratios for various formulations. Since our formulation is exact, it is not surprising that the ratio is 1 for all values of ρ. The ratios resulting

[1] http://www.math.nus.edu.sg/ mattohkc/sdpt3.html
[2] https://yalmip.github.io/

by comparing average objective values of *Mean-Variance* and *DNN-relaxation* approaches with the exact *Large-SDP* approach are also reported in Figure 4a. The variability in the ratio for the *Mean-Variance* approach can be inferred from the error bars in Figure 4a. The growth in gap between the objective values of the *Mean-Variance* approach and our partial covariance based approach in (19), as n increases can be inferred from Figure 4b.

Table 1: Bound ratios over Large-SDP bound for various approaches to DR appointment scheduling for various ρ values, n=6. 50 runs were performed with random means in [-2,2] and variances in (0,5].

	Mean-variance			Our Approach			DNN Relaxation		
ρ	mean	min	max	mean	min	max	mean	min	max
-1.0	1.489	1.054	2.028	1	1	1	1.001	1	1.008
-0.7	1.251	1.036	1.492	1	1	1	1.001	1	1.006
-0.3	1.141	1.023	1.285	1	1	1	1.001	1	1.004
0.0	1.088	1.016	1.185	1	1	1	1.001	1.001	1.007
0.3	1.051	1.010	1.111	1	1	1	1.001	1	1.002
0.7	1.017	1.001	1.039	1	1	1	1.001	1	1.001
1.0	1.010	1	1.055	1	1	1	1.002	1	1.056



(a) Ratio of objective value and $Z^*_{app}(\mathbf{s})$ for various values of ρ, $n = 6$

(b) Ratio of objective value and $Z^*_{app}(\mathbf{s})$ for various values of n, $\rho = -1$

Fig. 4: Bound Ratios of various approaches.

It is evident from Figure 4a that the bound resulting from mean-variance formulation in (19) is significantly higher than $Z^*_{app}(\mathbf{s})$ for negative values of ρ. As ρ approaches 1, the bound resulting from *Mean-Variance* approach appears to coincide with $Z^*_{app}(\mathbf{s})$.

While the numerical results appear to suggest that the distributionally robust formulation with partial correlation information offers a behaviour similar to that of the *Mean-Variance* approach as $\rho \to 1$, it is worthwhile to note that the correlation coefficients between $\tilde{c}_i$ and $\tilde{c}_{i+1}$ need not equal 1 for the worst-case distribution that attains the supremum in the *Mean-Variance* formulation (19). Indeed, given marginal distributions, for objective functions that are supermodular in its random variable arguments, it is well known that the comonotone joint distribution maximises the expectation (see, for example, [16,24,37]). However this comonotone joint distribution may very well be such that the correlation coefficients between its components are lesser than 1. This also explains the reason why the mean-variance bound need not exactly match the Large-SDP bound for $\rho = 1$ (see the last row in Table 1).

From Figures 4a and 4b, we also observe that the *DNN-relaxation* approach consistently gives a good approximation ratio (close to 1, see Table 1 for specific values), though it tends to be computationally expensive for large values of n; see Figure 5 for comparison of execution times for the different approaches considered.

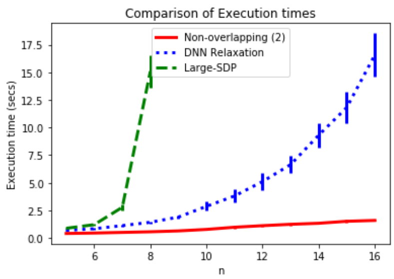


Fig. 5: Execution times in seconds of various approaches with n

n	Mean	Min	Max
30	8.397	8.052	8.835
40	19.565	18.712	21.127
50	41.215	38.515	48.330
60	78.533	75.563	82.552
70	129.533	122.533	142.875
80	227.400	206.607	244.174
90	416.586	343.712	478.861
100	672.803	611.037	716.489

Table 2: Execution times (in sec) for solving the semidefinite program in Theorem 2

It can be inferred from Figure 5 that the *Large-SDP* approach is computationally prohibitive for large values of n. The mean, minimum and maximum of observed execution times of the semidefinite program in Theorem 2 are provided for larger values of n in Table 2. Even for $n = 100$, the average execution time of our approach is only 672 seconds (roughly 11 minutes).

5.2 Comparison of optimal schedules

We next compare the optimal schedule obtained using the semidefinite program in Corollary 1 with those obtained from the *Mean-Variance* and *DNN-relaxation* approaches. For this purpose, we consider $n = 20$ patients, all with mean processing duration 2 and standard deviation 0.5. We take $T = 45$ units to be the time within which the schedules need to be fit. Figure 6a - 6d portray the schedules, respectively, for the cases where the correlation coefficient between $\tilde{u}_i$ and $\tilde{u}_{i+1}$, for $i \in \{1, 3, \ldots, n-1\}$ is given by $\rho = 1, -1, 0$ and -0.5. In order to understand the differences in the optimal schedules when the full covariance matrix is known, we plot the schedules given by the *DNN-relaxation* approach for the specific instance where the covariance entries that are not specified are set to 0. We use the label *DNN-Full covariance* for this scenario and *DNN-Non-overlapping* for the DNN-relaxation with partial moments.

Interestingly, for negative values of ρ, we observe that the inclusion of partial correlation information results in optimal schedules that are considerably different (in the relative durations allotted for earlier and later patients) when compared to those resulting from the *Mean-Variance* approach that assumes only the knowledge of mean and variance (see Figure 6b). For the extreme case where $\rho = -1$, we observe that the worst-case waiting time, Z^*_{app}, in the presence of partial correlation information is 4.116; this quantity is much smaller when compared to the worst-case expected total waiting time of 25.615 for the *Mean-Variance* approach where the partial correlation information is not included in the formulation. Moreover, we observe that employing the optimal schedule resulting from the mean-variance approach increases the worst-case waiting time, $Z^*_{app}(\cdot)$, by nearly 100% over the optimal Z^*_{app}. On the other hand, employing the optimal schedule from our formulation (30) results in about 30% increase in the worst-case waiting time $Z^*_{mv}(\cdot)$. Such stark changes in the structure and objective value for optimal schedules are typically not observed for nonnegative values of ρ (see Table 3).

Table 3: Mean percentage increase in the worst-case waiting time $Z^*_{app}(\cdot)$ when the optimal schedule from *Mean-Variance* approach is used instead of the optimal schedule that minimizes $Z^*_{app}(\mathbf{s})$, and vice versa, for $n = 20$ and cases $\rho = -1, 0$ and 1. The rows indicate schedules and columns indicate the DRO formulation used: M-V for the objective, $Z^*_{mv}(\cdot)$, of the *Mean-Variance* approach and P-C for the objective, $Z^*_{app}(\cdot)$, that also includes the knowledge of partial correlations.

	$\rho = -1$		$\rho = 0$		$\rho = 1$	
Objective / Schedule	M-V	P-C	P-C	P-C	M-V	P-C
M-V optimal	0	98	0	7.9	0	2.8
P-C optimal	34	0	5.2	0	1.9	0

6 Extensions

In this section, we discuss how our approach can be extended to capture additional information such as non-negative support and also additional entries of the second moment matrix. In all these cases, we can get polynomial time computable bounds, but tightness is not necessarily guaranteed.

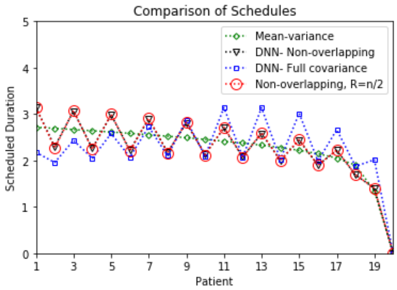
6.1 Non-negativity assumptions on $\tilde{\mathbf{c}}$

In Theorem 1, we assumed that $\tilde{\mathbf{c}} \in \mathbb{R}^n$ and proposed a tight reformulation to compute Z^*. We now discuss how our results can be extended to the case where $\tilde{\mathbf{c}} \in \mathbb{R}^n_+$. In particular, in this subsection we obtain bounds for

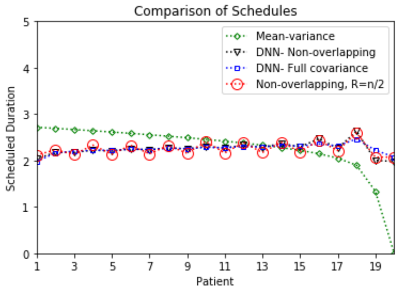
$$Z^*_+ = \sup \left\{ \mathbb{E}_\theta \left[Z(\tilde{\mathbf{c}}) \right] : \mathbb{E}_\theta[\tilde{\mathbf{c}}] = \boldsymbol{\mu},\ \mathbb{E}_\theta[\tilde{\mathbf{c}}^r (\tilde{\mathbf{c}}^r)'] = \mathbf{\Pi}^r \text{ for } r \in [R],\ \theta \in \mathcal{P}(\mathbb{R}^n_+) \right\},$$

In general, it is well known that checking for a feasible distribution with prescribed cross moments and support restricted to $\mathbb{R}^n_+$ is hard [9,42]. Formulations based on copositive optimization (which is again known to be hard [18]) have been provided in [44]. Our formulation in (7) clearly provides an upper bound to Z^*_+. Additionally inequalities may be added to formulation (7) so as to obtain tighter bounds for Z^*_+, as discussed next.

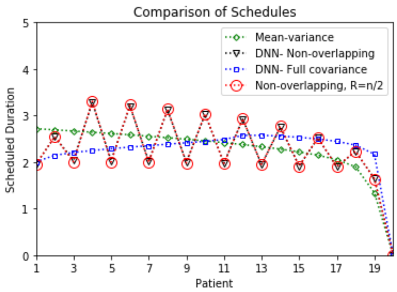
Suppose the sets $\mathcal{N}_1, \ldots, \mathcal{N}_R$ form a partition of $\{1, \ldots, n\}$ as before and without loss of generality assume an order among the elements $\mathcal{N}_r$, for every r.



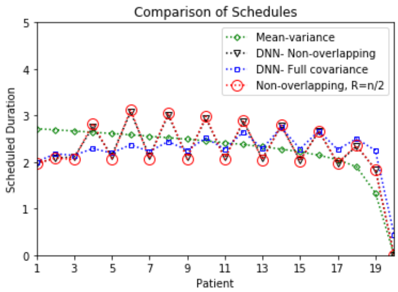
(a) Correlation between patient 1 and 2 = correlation between patients 3 and 4 = ... = $\rho = 1$. For Full-covariance, all other correlations are set to zero. Maximum time available (T) is 45 units. Optimal objective value = 25.0688, *Mean-Variance* bound= 25.6151, DNN Non-overlapping bound= 25.1534, DNN Full covariance bound = 15.9460.



(b) Correlation between patient 1 and 2 = correlation between patients 3 and 4 = ... = $\rho = -1$. For Full-covariance, all other correlations are set to zero. Maximum time available (T) is 45 units. Optimal objective value = 4.1162, *Mean-Variance* bound= 25.6151, DNN Non-overlapping bound= 4.2290, DNN Full covariance bound = 4.2250.



(c) Correlation between patient 1 and 2 = correlation between patients 3 and 4 = ... = $\rho = 0$. For Full-covariance, all other correlations are set to zero. Maximum time available (T) is 45 units. Optimal objective value=19.7474, *Mean-Variance* bound= 25.6151 , DNN Non-overlapping bound= 19.8607, DNN Full covariance bound = 11.4211.



(d) Correlations between patient 1 and 2 = correlations between patients 3 and 4 = ... = $\rho = -0.5$. For Full-covariance, all other correlations are set to zero. Maximum time available (T) is 45 units. Exact bound= 14.6842, *Mean-Variance* bound= 25.6151, DNN Non-overlapping bound= 14.7904, DNN Full covariance bound =9.4101.

Fig. 6: Optimal schedules under knowledge of non-overlapping moments. 20 patients all with mean 2 and standard deviation 0.5.

Lemma 5 *Denote by i_r the i^{th} element in the ordered set $\mathcal{N}_r$. Then,*

$$0 \leq Y_{ij}^r \leq \mu_{i_r} \max_{x \in \mathcal{X}} x_{j_r}, \quad 1 \leq i \leq |\mathcal{N}_r|, 1 \leq j \leq |\mathcal{N}_r|, \forall r \in [R] \tag{45}$$

provide valid inequalities. Inclusion of the above inequalities with formulation (7) *gives an upper bound for Z_+^*, that is at least as tight as $\hat{Z}^*$.*

Proof The variable Y_{ij}^r corresponds to $\mathbb{E}[\tilde{c}_{i_r} x(\tilde{\mathbf{c}})_{j_r}]$. The lemma is a consequence of the fact that $0 \leq \tilde{c}_i x_j(\tilde{\mathbf{c}}) \leq \tilde{c}_i \max_{\mathbf{x} \in \mathcal{X}} x_j$ whenever $\tilde{c}_i \geq 0$ and $\mathcal{X} \subseteq \mathbb{R}_+^n$. Taking expectations gives the inequalities. These inequalities are to be necessarily satisfied and hence the resulting formulation with these inequalities yields an upper bound for Z_+^*.

As an example, let $\mathcal{N}_r = \{3,4,5,6\}$. Note that $\mathbf{Y}^r \in \mathbb{R}^{4\times 4}$. The entry Y_{24}^r corresponds to $\mathbb{E}[\tilde{c}_4 x_6(\tilde{\mathbf{c}})]$. Since $\mathbf{c} \in \mathbb{R}_+^n, \mathbb{E}[\tilde{c}_4 x_6(\tilde{\mathbf{c}})] \leq \mu_4 \max_{\mathbf{x}\in\mathcal{X}} x_6$. We will now demonstrate how these valid inequalities can be used in the appointment scheduling context.

6.1.1 Appointment Scheduling: Including Support Information on $\tilde{\mathbf{u}}$

For a given schedule $\mathbf{s}$ define $Z_{app}^+(\mathbf{s})$ as,

$$Z_{app}^+(\mathbf{s}) = \sup\big\{\mathbb{E}_\theta\left[f(\tilde{\mathbf{u}},\mathbf{s})\right] : \mathbb{E}_\theta\left[\tilde{u}_i\right] = \mu_i, \mathbb{E}_\theta[\tilde{u}_i^2] = \Pi_{ii},\ \text{for } i \in [n], \\ \mathbb{E}_\theta[\tilde{u}_j\tilde{u}_{j+1}] = \Pi_{j,j+1},\ \text{for } j \in \{1,3,...n-1\}, \theta \in \mathcal{P}(\mathbb{R}_+^n)\big\}.$$

where $f(\tilde{\mathbf{u}},\mathbf{s})$ is as defined in Equation (17). Theorem 2 provides an exact formulation to compute $Z_{app}^*(\mathbf{s})$. Note that this formulation already gives an upper bound for $Z_{app}^+(\mathbf{s})$. We now describe how Lemma 5 can be used to obtain valid inequalities. These valid inequalities when used in conjunction with Theorem 2, provide a tighter bound for $Z_{app}^+(\mathbf{s})$.

Lemma 6 *The following are valid inequalities for computing* $Z_{app}^+(\mathbf{s})$*:*

$$Y_{i,i+1} \leq \mu_i(n-i)\ \ \forall i \ \text{odd}, i \in [n]$$
$$Y_{i+1,i} \leq \mu_{i+1}(n+1-i)\ \ \forall i \ \text{odd}, i \in [n]$$

The proof follows by applying Lemma 5 and noting that $x_i \leq n+1-i$ for all $\mathbf{x} \in \mathcal{X}$ (from Lemma 4).

Upon including these valid inequalities we still have a semidefinite program that provides bounds for $Z_{app}^+(\mathbf{s})$. The dual of this semidefinite program can be used to get distributionally robust schedules. However in our numerical results, we are unable to distinguish between the schedules thus obtained and the optimal schedules obtained in Corollary 1 (without enforcing support information).

6.1.2 Numerical Results

We now compare various bounds obtained when the support information $\tilde{\mathbf{u}} \geq 0$ is explicitly taken into account in the context of appointment scheduling. For the experiments in this subsection, for $n = 10$ patients, the mean processing durations μ were generated randomly at uniform in $[5, 10]$. The standard deviation was fixed at 5 for all the patients. The correlation ρ between the patients' processing durations were varied between -1 and 1. All the bound computations were performed for 50 randomly generated instances. We considered the following formulations:

i. Formulation in Theorem 2 (which does not account for support information).
ii. Formulation in Theorem 2 along with valid inequalities from Lemma 6 to take into account the support information.
iii. DNN relaxation approach: Formulation (43) with non-negativity constraint on the whole matrix $\begin{bmatrix} 1 & \boldsymbol{\mu}' & \mathbf{p}' \\ \boldsymbol{\mu} & \boldsymbol{\Delta} & \mathbf{Y}' \\ \mathbf{p} & \mathbf{Y} & \mathbf{X} \end{bmatrix}$.

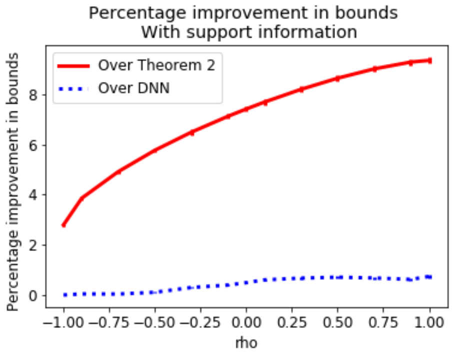


Fig. 7: Improvement in bounds by including support information

In all the examples we focus only on the second stage bound.

In Figure 7, we plot the error bars of percentage improvement in the bounds obtained by (ii.) over (i.) (the red curve), and also the improvement in the bounds obtained by (ii.) over (iii.) (the blue curve), where the percentage improvement in bound X over bound Y is $(X - Y) \times 100/Y$. The standard deviation of the percentage improvement in bounds is quite low for all cases and hence the error bars are not clearly visible. The plot suggests that the valid inequalities, when included with the formulation in Theorem 2, improve the bounds in Theorem 2 significantly. For larger values of correlations, the improvement is more pronounced. An improvement of about 1% is observed over DNN, again for larger values of correlations.

We also computed the schedules by minimizing the dual formulation for (ii.) over the schedules as well as the dual variables. However the schedules obtained were almost identical to the schedules obtained without including the support information via the formulation in Corollary 1.

6.2 The Case of Overlapping Moments

In all the sections above, we described an approach to compute $\sup_{\theta} \mathbb{E}_\theta[Z(\tilde{\mathbf{c}})]$ over all distributions consistent with known first moments and second moments corresponding to a partition of $\{1, \ldots, n\}$. We next show how the approach can be used to obtain bounds when additional entries of the second moment matrix are known. We illustrate this using an example.

Suppose that the only information we know about the probability distribution of $\tilde{\mathbf{c}}$ is the first moment specified by $\mathbb{E}[\tilde{\mathbf{c}}] = \boldsymbol{\mu}$ and the partial second moments $\mathbb{E}[\tilde{c}_i^2] = \Pi_{ii}$, $\mathbb{E}[\tilde{c}_i\tilde{c}_{i+1}] = \Pi_{i,i+1}$ for $1 \leq i \leq n-1$. In this section, we are interested to compute:

$$Z^*_{\text{series}} = \sup \left\{ \mathbb{E}_\theta \left[Z(\tilde{\mathbf{c}})\right] : \begin{array}{l} \mathbb{E}_\theta[\tilde{\mathbf{c}}] = \boldsymbol{\mu},\ \mathbb{E}_\theta[\tilde{c}_i^2] = \Pi_{ii} \text{ for } i \in [n], \\ \mathbb{E}_\theta[\tilde{c}_i\tilde{c}_{i+1}] = \Pi_{i,i+1} \text{ for } i \in [n-1],\ \theta \in \mathcal{P}(\mathbb{R}^n) \end{array} \right\}, \quad (46)$$

Note that this situation is equivalent to knowing the means and second moments corresponding to the sets $\mathcal{S} = \{\{1,2\},\{2,3\},\{3,4\},\ldots,\{n-1,n\}\}$. $\mathcal{S}$ can be interpreted as the set of edges in a series graph on n nodes and hence we refer to the bound in (46) using the subscript 'series'. Our exposition so far had been focussed on incorporating entries on second moment matrix corresponding to non-overlapping blocks (e.g. $\mathcal{N}_1 = \{1,2\}, \mathcal{N}_2 = \{3,4\}, \ldots \mathcal{N}_{n/2} = \{n-1, n\}$). In contrast, we now possess information of $n/2-1$ more entries of the second moment matrix.

We have the following tight bound from formulation (6),

$$\begin{aligned} Z^*_{\text{series}} = \max_{\mathbf{p},\mathbf{X},\mathbf{Y},\boldsymbol{\Delta}} \quad & trace(\mathbf{Y}) \\ \text{s.t} \quad & \begin{bmatrix} 1 & \boldsymbol{\mu}' & \mathbf{p}' \\ \boldsymbol{\mu} & \boldsymbol{\Delta} & \mathbf{Y}' \\ \mathbf{p} & \mathbf{Y} & \mathbf{X} \end{bmatrix} \succeq 0, \\ & \Delta_{ii} = \Pi_{ii}, \quad \text{for } i \in [n], \\ & \Delta_{i,i+1} = \Pi_{i,i+1}, \quad \text{for } i \in [n-1], \\ & (\mathbf{p}, \mathbf{X}) \in conv\left\{ (\mathbf{x}, \mathbf{x}\mathbf{x}') : \mathbf{x} \in \mathcal{X} \right\}. \end{aligned} \tag{47}$$

We next show that a similar approach to our formulation in Theorem 1 (capturing only a subset of the variables $\mathbf{p}, \mathbf{X}$ and $\mathbf{Y}$) to this particular scenario of overlapping moments results in an upper bound for Z^*_{series}.

Lemma 7 *Define $\hat{Z}^*_{series}$ as the optimal value of the optimization problem:*

$$\begin{aligned} \hat{Z}^*_{series} = \max_{\substack{\mathbf{p}, X_{ii}, X_{i,i+1} \\ Y_{ii}, Y_{i,i+1}}} \quad & \sum_{i=1}^{n} Y_{ii} \\ s.t \quad & \begin{bmatrix} 1 & \mu_i & \mu_{i+1} & p_i & p_{i+1} \\ \mu_i & \Pi_{ii} & \Pi_{i,i+1} & Y_{ii} & Y_{i,i+1} \\ \mu_{i+1} & \Pi_{i,i+1} & \Pi_{i+1,i+1} & Y_{i+1,i} & Y_{i+1,i+1} \\ p_i & Y_{ii} & Y_{i+1,i} & X_{ii} & X_{i,i+1} \\ p_{i+1} & Y_{i,i+1} & Y_{i+1,i+1} & X_{i,i+1} & X_{i+1,i+1} \end{bmatrix} \succeq 0, \quad for\ i \in [n-1], \\ & \big(\mathbf{p}, (X_{ii} : i \in [n]), (X_{i,i+1} : i \in [n-1])\big) \in \\ & \quad conv\left\{ \big(\mathbf{x}, (x_i^2 : i \in [n]), (x_i x_{i+1} : i \in [n-1])\big) : \mathbf{x} \in \mathcal{X} \right\}. \end{aligned} \tag{48}$$

Then $Z^*_{series} \leq \hat{Z}^*_{series}$.

The proof follows by using a similar reasoning in proof of Theorem 1, Step 1. However the bound is not guaranteed to be tight in general. In Appendix B, we investigate why the bound provided above is not necessarily tight always.

6.2.1 Numerical Results

Taking the number of patients $n = 6$ for the appointment scheduling application and setting the correlation coefficient between service times $\tilde{u}_i$ and $\tilde{u}_{i+1}$ to ρ, for every i in $\{1, \ldots, n-1\}$, we vary ρ in the interval $[-1, 1]$. We compute the expected total waiting times averaged over 50 independent runs, where in each run, the means and variances of $\tilde{c}_i(\mathbf{s}) = \tilde{u}_i - s_i$ are taken to be independent realizations of random variables uniformly distributed in the intervals $[-2, 2]$ and $(0, 5]$ respectively. We are interested in only the bound computation and we set $\mathbf{s} = 0$ in all the formulations. We compute the following bounds.

(a) Formulation (47) applied to appointment scheduling by explicitly enumerating the extreme points of $\mathcal{X}_{app}$. This gives the exact value of Z^*_{series}. The formulation is provided next for reference.

$$\begin{aligned}
\max_{\mathbf{p},\mathbf{X},\mathbf{Y},\boldsymbol{\Delta}} \quad & trace(\mathbf{Y}) - \mathbf{s}'\mathbf{p} \\
\text{s.t} \quad & \begin{bmatrix} 1 & \boldsymbol{\mu}' & \mathbf{p}' \\ \boldsymbol{\mu} & \boldsymbol{\Delta} & \mathbf{Y}' \\ \mathbf{p} & \mathbf{Y} & \mathbf{X} \end{bmatrix} \succeq 0, \\
& \Delta_{ii} = \Pi_{ii}, \quad \text{for } i \in [n], \\
& \Delta_{i,i+1} = \Pi_{i,i+1}, \quad \text{for } i \in [n-1], \\
& \mathbf{p} = \sum_{\mathbf{x}\in\mathcal{X}_{app}} \alpha_{\mathbf{x}} \mathbf{x} \\
& \mathbf{X} = \sum_{\mathbf{x}\in\mathcal{X}_{app}} \alpha_{\mathbf{x}} \mathbf{x}\mathbf{x}' \\
& \textstyle\sum_{\mathbf{x}\in\mathcal{X}_{app}} \alpha_{\mathbf{x}} = 1, \; \alpha_{\mathbf{x}} \geq 0 \, \forall \mathbf{x} \in \mathcal{X}_{app}
\end{aligned} \tag{49}$$

(b) The upper bound $\hat{Z}^*_{series}$ from (48), which when applied to appointment scheduling (by using Lemma 4, (25),(27) in formulation (48)), boils down to,

$$\begin{aligned}
\max_{p_i, X_{ij}, Y_{ij}, t_{kj}} \; & \sum_{i=1}^{n} (Y_{ii} - s_i p_i) \\
\text{s.t.} \; & \begin{bmatrix} 1 & \mu_i & \mu_{i+1} & p_i & p_{i+1} \\ \mu_i & \Pi_{ii} & \Pi_{i,i+1} & Y_{ii} & Y_{i,i+1} \\ \mu_{i+1} & \Pi_{i,i+1} & \Pi_{i+1,i+1} & Y_{i+1,i} & Y_{i+1,i+1} \\ p_i & Y_{ii} & Y_{i+1,i} & X_{ii} & X_{i,i+1} \\ p_{i+1} & Y_{i,i+1} & Y_{i+1,i+1} & X_{i,i+1} & X_{i+1,i+1} \end{bmatrix} \succeq 0, \quad \text{for } i \in [n-1], \\
& p_i = \sum_{k=1}^{i} \sum_{j=i}^{n+1} t_{kj}(j-i), \quad \text{for } i \in [n], \\
& X_{ii} = \sum_{k=1}^{i} \sum_{j=i}^{n+1} t_{kj}(j-i)^2, \quad \text{for } i \in [n], \\
& X_{i,i+1} = X_{i+1,i} = \sum_{k=1}^{i} \sum_{j=i+1}^{n+1} t_{kj}(j-i)(j-(i+1)), \; \text{for } i \in [n-1], \\
& \sum_{k=1}^{i} \sum_{j=i}^{n+1} t_{kj} = 1, \quad \text{for } i \in [n], \\
& t_{kj} \geq 0, \quad \text{for } 1 \leq k \leq j \leq n+1.
\end{aligned} \tag{50}$$

The difference from the formulation in Theorem 2 is that there is a positive semidefinite constraint on $n-1$ matrices, each of size 5×5 here. In Theorem 2, the positive semidefinite constraint appears for only $n/2$ matrices. Similarly here the constraints on $X_{i,i+1}$ appear for all values of $i \leq n-1$ whereas in the formulation in Theorem 2, the constraint appears for odd values of i alone. This formulation is referred to as '*Overlapping*' in the plots.

(c) The following upper bound from a DNN relaxation to the completely positive reformulation in [32] for the scenario of overlapping moments,

$$\begin{aligned}
\max_{\mathbf{p},\mathbf{X},\mathbf{Y},\boldsymbol{\Delta}} \quad & trace(\mathbf{Y}) - \mathbf{s}'\mathbf{p} \\
\text{s.t} \quad & \begin{bmatrix} 1 & \boldsymbol{\mu}' & \mathbf{p}' \\ \boldsymbol{\mu} & \boldsymbol{\Delta} & \mathbf{Y}' \\ \mathbf{p} & \mathbf{Y} & \mathbf{X} \end{bmatrix} \succeq 0, \\
& [\mathbf{A}, -\mathbf{I}_n]\mathbf{p} = \mathbf{b}, \\
& [\mathbf{a}_i', -\mathbf{e}_i']\mathbf{X}[\mathbf{a}_i', -\mathbf{e}_i']' = b_i^2, \qquad \forall i \in [n] \\
& \boldsymbol{\Delta}_{i,i+1} = \boldsymbol{\Pi}_{i,i+1}, \ \forall i \in [n-1] \\
& \mathbf{p}, \mathbf{X} \text{ non-negative.}
\end{aligned} \tag{51}$$

In the above formulation, $\mathbf{p} \in \mathbb{R}^{2n}$, $\mathbf{Y} \in \mathbb{R}^{n \times 2n}$ and $\mathbf{X} \in \mathbb{R}^{2n \times 2n}$. The matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is such that $A_{jj} = -1$ for $j \in [n]$, $A_{j+1,j} = 1$ for $j \in [n-1]$, $b_j = -1$ for $j \in [n]$ and $\mathbf{a}_i'$ indicates row i of $\mathbf{A}$ and $\mathbf{e}_i \in \mathbb{R}^n$ denotes a column vector with 1 at position i and zero everywhere else. This formulation is referred to as '*DNN-Overlapping*' in the plots.

(d) We compute $\hat{Z}^*_{app}(\mathbf{s})$ from Theorem 2 which makes use of information only on non-overlapping blocks of the second moment matrix. Only the cross moments $\Pi_{12}, \Pi_{34}, \Pi_{56}$ are used in this formulation while Π_{23} and Π_{45} are unused. Therefore note that this formulation is an upper bound for Z^*_{series}. This formulation is referred to as '*Non-overlapping (2)*' in the plots.

(e) Finally we compute the mean-variance bound from formulation (42). Note that this bound does not make use of any off-diagonal entries in $\boldsymbol{\Pi}$.

In Figure 8, the percentage gap in the bounds over (a) is plotted in brown for (b), blue for (c), red for (d) and pink for (e). The bound from formulation (50), labelled *Overlapping* is not tight always. The gap can be seen especially for negative correlations. For larger correlations however, the gap is close to zero. In spite of the gap in bounds, this formulation offers an advantage in terms of speed (see Figure 9). *Non-Overlapping (2)* shows a larger gap, as expected, as the entries Π_{23}, Π_{45} are not used here. *Mean-variance* demonstrates the largest gap as it does not make use of any of the off-diagonal entries in $\boldsymbol{\Pi}$. *DNN-Overlapping* proves to be the best out of the four. However from the execution time analysis in Figure 9, *Overlapping* offers a significant advantage compared to *DNN-Overlapping*. *Non-Overlapping (2)* is faster than *Overlapping* due to fewer number of variables and positive semidefinite constraints in the formulation inTheorem 2 over (50). The performance of the mean-variance formulation is very close to *Non-Overlapping (2)* in terms of execution time. The exact formulation (shown in green) is labelled as '*Large SDP(overlapping)*'. Due to explicit enumeration of the extreme points, this formulation is feasible in our computational setup only for $n \leq 9$. Note that the execution time for this formulation increases rapidly.

References

1. David J. Aldous. The $\zeta(2)$ limit in the random assignment problem. *Random Structures & Algorithms*, 18(4):381–418, 2001.
2. Kurt M. Anstreicher and Samuel Burer. Computable representations for convex hulls of low-dimensional quadratic forms. *Mathematical Programming*, 124:33–43, 2010.
3. Michael O Ball, Charles J Colbourn, and J Scott Provan. Network reliability. *Handbooks in operations research and management science*, 7:673–762, 1995.

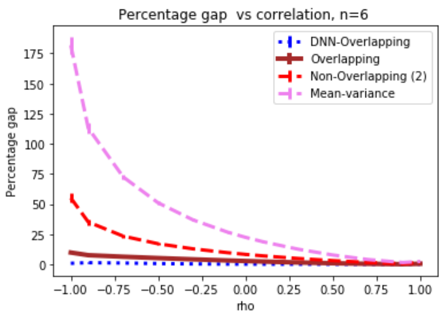


Fig. 8: Percentage gap in bounds for overlapping moment information

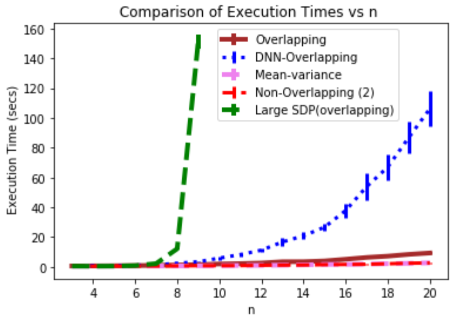


Fig. 9: Execution Time Analysis for the case of overlapping moments

4. Aharon Ben-Tal, Laurent El Ghaoui, and Arkadi Nemirovski. *Robust Optimization*. Princeton University Press, 2009.
5. Claude Berge. Some classes of perfect graphs. *Graph Theory and Theoretical Physics*, F. Harary Ed:155–166, 1967.
6. Dimitris Bertsimas, Xuan Vinh Doan, Karthik Natarajan, and Chung-Piaw Teo. Models for minimax stochastic linear optimization problems with risk aversion. *Mathematics of Operations Research*, 35(3):580–602, 2010.
7. Dimitris Bertsimas, Karthik Natarajan, and Chung-Piaw Teo. Probabilistic combinatorial optimization: Moments, semidefinite programming, and asymptotic bounds. *SIAM Journal on Optimization*, 15(1):185–209, 2004.
8. Dimitris Bertsimas, Karthik Natarajan, and Chung-Piaw Teo. Persistence in discrete optimization under data uncertainty. *Mathematical Programming*, 108(2):251–274, Sep 2006.
9. Dimitris Bertsimas and Ioana Popescu. Optimal inequalities in probability theory: A convex optimization approach. *SIAM J. on Optimization*, 15(3):780–804, March 2005.
10. Dimitris Bertsimas, Melvyn Sim, and Meilin Zhang. Adaptive distributionally robust optimization. *o appear in Management Science*, 2018.
11. Immanuel M. Bomze, Jianqiang Cheng, Peter J. C. Dickinson, and Abdel Lisser. A fresh CP look at mixed-binary QPs: new formulations and relaxations. *Math. Program.*, 166(1-2):159–184, 2017.
12. Immanuel M. Bomze and Etienne De Klerk. Solving standard quadratic optimization problems via linear, semidefinite and copositive programming. *J. of Global Optimization*, 24(2):163–185, October 2002.
13. Marshall W. Buck, Clara S. Chan, and David P. Robbins. On the expected value of the minimum assignment. *Random Structures & Algorithms*, 21(1):33–58, 2002.
14. Samuel Burer. On the copositive representation of binary and continuous nonconvex quadratic programs. *Mathematical Programming*, 120(2):479–495, 2010.
15. Samuel Burer. A gentle, geometric introduction to copositive optimization. *Mathematical Programming*, 151(1):89–116, Jun 2015.
16. Stamatis Cambanis, Gordon Simons, and William Stout. Inequalities for E k (x, y) when the marginals are fixed. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 36(4):285–294, 1976.
17. Erick Delage and Yinyu Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612, 2010.
18. Peter J. C. Dickinson. Geometry of the copositive and completely positive cones. *Journal of Mathematical Analysis and Applications*, 380(1):377 – 395, 2011.
19. Peter J. C. Dickinson and Luuk Gijben. On the computational complexity of membership problems for the completely positive cone and its dual. *Computational Optimization and Applications*, 57(2):403–415, Mar 2014.
20. Xuan Vinh Doan and Karthik Natarajan. On the complexity of nonoverlapping multivariate marginal bounds for probabilistic combinatorial optimization problems. *Operations Research*, 60(1):138–149, 2012.

21. W. E. Donath. Algorithm and average-value bounds for assignment problems. *IBM Journal of Research and Development*, 13(4):380–386, July 1969.
22. M. E. Dyer, A. M. Frieze, and C. J. H. Mcdiarmid. On linear programs with random costs. *Mathematical Programming*, 35(1):3–16, May 1986.
23. D. R. Fulkerson. Expected critical path lengths in pert networks. *Operations Research*, 10(6):808–817, 1962.
24. A. Galichon. *Optimal Transport Methods in Economics*. Princeton University Press, 2016.
25. Michel X. Goemans and Muralidharan S. Kodialam. A lower bound on the expected cost of an optimal assignment. *Mathematics of Operations Research*, 18(2):267–274, 1993.
26. Robert Grone, Charles R. Johnson, Eduardo M. S, and Henry Wolkowicz. Positive definite completions of partial hermitian matrices. *Linear Algebra and its Applications*, 58:109 – 124, 1984.
27. Jane N. Hagstrom. Computational complexity of pert problems. *Networks*, 18(2):139–147, 1988.
28. Grani A. Hanasusanto and Daniel Kuhn. Conic programming reformulations of two-stage distributionally robust linear programs over wasserstein balls. *Operations Research*, 66(3):849–869, 2018.
29. Grani A. Hanasusanto, Daniel Kuhn, Stein W. Wallace, and Steve Zymler. Distributionally robust multi-item newsvendor problems with multimodal demand distributions. *Mathematical Programming*, 152(1):1–32, Aug 2015.
30. Ruiwei Jiang, Siqian Shen, and Yiling Zhang. Integer programming approaches for appointment scheduling with random no-shows and service durations. *Operations Research*, 65(6):1638–1656, 2017.
31. Richard M. Karp. An upper bound on the expected cost of an optimal assignment. In David S. Johnson, Takao Nishizeki, Akihiro Nozaki, and Herbert S. Wilf, editors, *Discrete Algorithms and Complexity*, pages 1 – 4. Academic Press, 1987.
32. Qingxia Kong, Chung-Yee Lee, Chung-Piaw Teo, and Zhichao Zheng. Scheduling arrivals to a stochastic service delivery system using copositive cones. *Operations Research*, 61(3):711–726, 2013.
33. H. W. Kuhn. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(12):83–97, 1955.
34. Jerome M. Kurtzberg. On approximation methods for the assignment problem. *J. ACM*, 9(4):419–439, October 1962.
35. Monique Laurent. *Matrix completion problems Matrix Completion Problems*, pages 1967–1975. Springer US, Boston, MA, 2009.
36. Andrew J. Lazarus. Certain expected values in the random assignment problem. *Operations Research Letters*, 14(4):207 – 214, 1993.
37. G. G. Lorentz. An inequality for rearrangements. *The American Mathematical Monthly*, 60(3):176–179, 1953.
38. Ho-Yin Mak, Ying Rong, and Jiawei Zhang. Appointment scheduling with limited distributional information. *Management Science*, 61(2):316–334, 2015.
39. Marc Mézard and Giorgio Parisi. On the solution of the random link matching problems. *Journal de Physique*, 48(9):1451–1459, 1987.
40. Rolf H Möhring. Scheduling under uncertainty: Bounding the makespan distribution. In *Computational Discrete Mathematics*, pages 79–97. Springer, 2001.
41. James Munkres. Algorithms for the assignment and transportation problems. *Journal of the Society for Industrial and Applied Mathematics*, 5(1):32–38, 1957.
42. Katta G. Murty and Santosh N. Kabadi. Some np-complete problems in quadratic and nonlinear programming. *Math. Program.*, 39(2):117–129, November 1987.
43. Karthik Natarajan and Chung-Piaw Teo. On reduced semidefinite programs for second order moment bounds with applications. *Mathematical Programming*, 161(1-2):487–518, 2017.
44. Karthik Natarajan, Chung-Piaw Teo, and Zhichao Zheng. Mixed 0-1 linear programs under objective uncertainty: A completely positive representation. *Operations Research*, 59(3):713–728, 2011.
45. B. Olin. *Asymptotic Properties of Random Assignment Problems*. PhD thesis, Division ofOptimization and Systems Theory, Department of Mathematics, Royal Institute of Technology, Stockholm, Sweden, 1992.
46. M. Padberg. The boolean quadric polytope: Some characteristics, facets and relatives. *Math. Program.*, 45(1):139–172, August 1989.

47. Pablo A. Parrillo. *Structured Semidefinite Programs and Semi-algebraic Geometry Methods in Robustness*. PhD thesis, California Institute of Technology, 2000.
48. R. Penrose. A generalized inverse for matrices. *Mathematical Proceedings of the Cambridge Philosophical Society*, 51(3):406413, 1955.
49. Itamar Pitowsky. Correlation polytopes: Their geometry and complexity. *Mathematical Programming*, 50(1):395–414, 1991.
50. C. Radhakrishna Rao and Sujit Kumar Mitra. Generalized inverse of a matrix and its applications. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Theory of Statistics*, pages 601–620, Berkeley, Calif., 1972. University of California Press.
51. Donald J. Rose. Triangulated graphs and the elimination process. *Journal of Mathematical Analysis and Applications*, 32(3):597 – 609, 1970.
52. A. W. Shogan. Bounding distributions for a stochastic pert network. *Networks*, 7(4):359–381, 1977.
53. K.C. Toh, M.J. Todd, and R.H. Tutuncu. SDPT3 — a matlab software package for semidefinite programming. *Optimization Methods and Software*, 11:545–581, 1999.
54. R.H Tutuncu, K.C. Toh, and M.J. Todd. Solving semidefinite-quadratic-linear programs using sdpt3. *Mathematical Programming Ser.*, B(95):189–217, 2003.
55. Richard M. Van Slyke. Monte carlo methods and the pert problem. *Operations Research*, 11(5):839–860, 1963.
56. David Walkup. On the expected value of a random assignment problem. *SIAM Journal on Computing*, 8(3):440–442, 1979.
57. Wolfram Wiesemann, Daniel Kuhn, and Melvyn Sim. Distributionally robust convex optimization. *Operations Research*, 62(6):1358–1376, 2014.
58. Johan Wstlund. A proof of a conjecture of buck, chan, and robbins on the expected value of the minimum assignment. *Random Structures & Algorithms*, 26(12):237–251, 2005.
59. Guanglin Xu and Samuel Burer. A data-driven distributionally robust bound on the expected optimal value of uncertain mixed 0-1 linear programming. *Comput. Manag. Science*, 15(1):111–134, 2018.
60. Anstreicher Kurt Yang Boshi and Samuel Burer. Quadratic programs with hollows. *Mathematical Programming*, 170(2):541–552, 2018.
61. Willard I. Zangwill. A deterministic multi-period production scheduling model with backlogging. *Management Science*, 13(1):105–119, 1966.
62. Willard I. Zangwill. A backlogging model and a multi-echelon model of a dynamic economic lot size production systema network approach. *Management Science*, 15(9):506–527, 1969.
63. Luis F. Zuluaga and Javier F. Pea. A conic programming approach to generalized tchebycheff inequalities. *Mathematics of Operations Research*, 30(2):369–388, 2005.

Appendix A On the structure of a worst-case distribution

In this section, we exhibit a probability distribution for $\tilde{\mathbf{c}}$ that attains the optimal value Z^* of (7). The construction is along the lines of the worst case distribution proposed in proof of Theorem 1 - step 2 of [43]. The worst-case distribution we identify in particular is a mixture of normal distributions. Each of these normal distributions is in turn constructed by first constructing suitable marginal distributions and then applying conditional independence.

We begin with a result on psd matrix factorization in [43]. The following definition of Moore-Penrose pseudoinverse (see [48,50]) is useful in stating the psd matrix factorization in Lemma 6. Let $\mathbf{X}$ be a matrix of dimension $k_1 \times k_2$. Then the Moore-Penrose pseudoinverse of $\mathbf{X}$ is a matrix $\mathbf{X}^\dagger$ of dimension $k_2 \times k_1$ and is defined as a unique solution to the set of four equations:

$$\mathbf{X}\mathbf{X}^\dagger\mathbf{X} = \mathbf{X}, \quad \mathbf{X}^\dagger\mathbf{X}\mathbf{X}^\dagger = \mathbf{X}^\dagger, \quad \mathbf{X}\mathbf{X}^\dagger = (\mathbf{X}\mathbf{X}^\dagger)', \quad \text{and} \quad \mathbf{X}^\dagger\mathbf{X} = (\mathbf{X}^\dagger\mathbf{X})'.$$

Theorem 6 *[43, Theorem 1] Suppose that* $\mathbf{L}$ *is a* $(k_1+k_2)\times(k_1+k_2)$ *positive semidefinite block matrix of the form,*

$$\mathbf{L} = \begin{bmatrix} \mathbf{A} & \mathbf{B}' \\ \mathbf{B} & \mathbf{C} \end{bmatrix} \succeq 0, \tag{52}$$

where the matrices $\mathbf{A} \in \mathbb{R}^{k_1 \times k_1}, \mathbf{C} \in \mathbb{R}^{k_2 \times k_2}$ *are symmetric and the matrix* $\mathbf{C}$ *admits an explicit factorization given by* $\mathbf{C} = \mathbf{V}\mathbf{V}'$*. Then* $\mathbf{L}$ *admits the following factorization:*

$$\mathbf{L} = \begin{bmatrix} \mathbf{B}'(\mathbf{V}^\dagger)' \\ \mathbf{V} \end{bmatrix} \begin{bmatrix} \mathbf{B}'(\mathbf{V}^\dagger)' \\ \mathbf{V} \end{bmatrix}' + \begin{bmatrix} \mathbf{U} \\ 0 \end{bmatrix} \begin{bmatrix} \mathbf{U} \\ 0 \end{bmatrix}', \tag{53}$$

where the matrix $\mathbf{U}$ *is defined such that* $\mathbf{A} - \mathbf{B}'\mathbf{C}^\dagger\mathbf{B} = \mathbf{U}\mathbf{U}' \succeq 0$.

For a given partition $\{\mathcal{N}_r : r \in [R]\}$ and projected covariance matrices $\{\mathbf{\Pi}^r : r \in [R]\}$, suppose that $\{\mathbf{p}_*, \mathbf{X}_*^r, \mathbf{Y}_*^r : r \in [R]\}$ maximizes (7). As in the proof of Theorem 1, it follows from Carathéodory's theorem and the convex hull constraint in (7) that there exists $\hat{\mathcal{X}}$, a subset of $\mathcal{X}$, containing at most $1 + \sum_r (n_r^2 + 3n_r)/2$ elements such that,

$$\left(\mathbf{p}_*, \mathbf{X}_*^1, \ldots, \mathbf{X}_*^R\right) = \sum_{\mathbf{x} \in \hat{\mathcal{X}}} \alpha_{\mathbf{x}} \left(\mathbf{x},\ \mathbf{x}^1\mathbf{x}^{1'}, \ldots, \mathbf{x}^R\mathbf{x}^{R'}\right),$$

for some $\{\alpha_{\mathbf{x}} : \mathbf{x} \in \hat{\mathcal{X}}\}$ satisfying $\alpha_{\mathbf{x}} \geq 0$, $\sum_{\mathbf{x} \in \hat{\mathcal{X}}} \alpha_{\mathbf{x}} = 1$. Consequently, for any $r \in [R]$, we have from Lemma 6 that,

$$\begin{aligned} \begin{bmatrix} \mathbf{\Pi}^r & \boldsymbol{\mu}^r & \mathbf{Y}_*^{r'} \\ \boldsymbol{\mu}^{r'} & 1 & \mathbf{p}_*^{r'} \\ \mathbf{Y}_*^r & \mathbf{p}_*^r & \mathbf{X}_*^r \end{bmatrix} &= \sum_{\mathbf{x} \in \hat{\mathcal{X}}} \alpha_{\mathbf{x}} \begin{bmatrix} \mathbf{\Pi}^r & \boldsymbol{\mu}^r & \mathbf{Y}_*^{r'} \\ \boldsymbol{\mu}^{r'} & 1 & \mathbf{x}^{r'} \\ \mathbf{Y}_*^r & \mathbf{x}^r & \mathbf{x}^r\mathbf{x}^{r'} \end{bmatrix} \\ &= \sum_{\mathbf{x} \in \hat{\mathcal{X}}} \alpha_{\mathbf{x}} \begin{bmatrix} \mathbf{d}_r(\mathbf{x}^r) \\ 1 \\ \mathbf{x}^r \end{bmatrix} \begin{bmatrix} \mathbf{d}_r(\mathbf{x}^r) \\ 1 \\ \mathbf{x}^r \end{bmatrix}' + \begin{bmatrix} \mathbf{\Phi}_r & \mathbf{0}_{n_r,1} & \mathbf{0}_{n_r,n_r} \\ \mathbf{0}_{1,n_r} & 0 & \mathbf{0}_{1,n_r} \\ \mathbf{0}_{n_r,n_r} & 0 & \mathbf{0}_{n_r,n_r} \end{bmatrix}, \end{aligned} \tag{54}$$

where $\mathbf{d}_r(\mathbf{x}^r) \in \mathbb{R}^{n_r}$ and $\mathbf{\Phi}_r \in \mathcal{S}_{n_r}^+$ for every $r \in [R]$. From the above factorization, observe that,

$$trace(\mathbf{Y}_*^r) = trace\left(\sum_{x \in \hat{\mathcal{X}}} \alpha_{\mathbf{x}} \mathbf{x}^r \mathbf{d}_r(\mathbf{x}^r)'\right) = \sum_{x \in \hat{\mathcal{X}}} \alpha_{\mathbf{x}} \sum_{r=1}^R \mathbf{d}_r(\mathbf{x}^r)'\mathbf{x}^r. \tag{55}$$

For completeness, we will now list explicitly the expressions for the means $\mathbf{d}_r(\mathbf{x}^r)$ and $\mathbf{\Phi_r}$. These expressions are obtained by making appropriate substitutions as per Theorem 6. For every $r \in [R]$, define the matrix $\mathbf{V}_r$ of size $(n_r+1) \times m_r$ where m_r is the number of points in the projected space $\mathcal{X}^r$ as follows.

$$\mathbf{V}_r = \begin{bmatrix} \cdots & \sqrt{\alpha_r(\mathbf{x}^r)} & \cdots \\ \vdots & \sqrt{\alpha_r(\mathbf{x}^r)}\mathbf{x}^r & \vdots \end{bmatrix}$$

Each column of $\mathbf{V}_r$ corresponds to an element $\mathbf{x}^r$ of $\mathcal{X}^r$ and is of the form $\begin{bmatrix} \sqrt{\alpha_r(\mathbf{x}^r)} \\ \sqrt{\alpha_r(\mathbf{x}^r)}\mathbf{x}^r \end{bmatrix}$. Define $\mathbf{\Phi}_r$ of size $n_r \times n_r$ as:

$$\mathbf{\Phi}_r = \mathbf{\Pi}^r - [\boldsymbol{\mu}^r \quad \hat{\mathbf{Y}}_*^r] \begin{bmatrix} 1 & \sum_{\mathbf{x}^r} \alpha_r(\mathbf{x}^r)(\mathbf{x}^r)' \\ \sum_{\mathbf{x}^r} \alpha_r(\mathbf{x}^r)\mathbf{x}^r & \sum_{\mathbf{x}^r} \alpha_r(\mathbf{x}^r)\mathbf{x}^r(\mathbf{x}^r)' \end{bmatrix}^\dagger \begin{bmatrix} \boldsymbol{\mu}^r \\ \hat{\mathbf{Y}}_*^r \end{bmatrix}$$

The mean vector $\mathbf{d}_r(\mathbf{x}_r)$ is set to be the column vector of the matrix $\left[\boldsymbol{\mu}^r \ \hat{\mathbf{Y}}_*^r\right](\mathbf{V}_r^\dagger)' \times 1/\sqrt{\alpha_r(\mathbf{x}^r)}$ corresponding to where $\mathbf{x}_r$ occurs in $\mathbf{V}_r$.

Proposition 2 *Suppose that* $\{\mathbf{p}_*, \mathbf{X}_*^r, \mathbf{Y}_*^r : r \in [R]\}$ *maximizes* (7). *Let* $\hat{\mathcal{X}} \subseteq \mathcal{X}$ *be a finite subset and* $\{\alpha_{\mathbf{x}} : \mathbf{x} \in \hat{\mathcal{X}}\}$ *satisfy* (54). *Let* θ^* *be the distribution of* $\tilde{\mathbf{c}}$ *generated as follows:*

Step 1: Generate a random vector $\tilde{\mathbf{x}} \in \hat{\mathcal{X}} \subseteq \mathcal{X}$ *such that* $P(\tilde{\mathbf{x}} = \mathbf{x}) = \alpha_{\mathbf{x}}$.
Step 2: For every $r \in [R]$, *independently generate a normally distributed random vector* $\tilde{\mathbf{z}}_r \in \mathbb{R}^{n_r}$, *conditionally on* $\mathbf{x}$, *with mean* $\mathbf{d}_r(\mathbf{x}^r)$ *and covariance* $\mathbf{\Phi_r}$. *Set* $\tilde{\mathbf{c}}^r = \tilde{\mathbf{z}}_r$.

Then θ^* *attains the maximum in* (2).

Proof Consider $(\tilde{\mathbf{x}}, \tilde{\mathbf{c}})$ generated jointly according to the described steps. Then it follows from the law of iterated expectations that, $\mathbb{E}[f(\tilde{\mathbf{c}})] = \mathbb{E}[\mathbb{E}[f(\tilde{\mathbf{c}})|\tilde{\mathbf{x}}] = \sum_{\mathbf{x}\in\hat{\mathcal{X}}} \alpha_{\mathbf{x}}\mathbb{E}[f(\tilde{\mathbf{c}})|\tilde{\mathbf{x}} = \mathbf{x}]$, for any function f. As a result, we have from (54) that for any $r \in [R]$,

$$\mathbb{E}\begin{bmatrix} \tilde{\mathbf{c}}^r \\ 1 \end{bmatrix}\begin{bmatrix} \tilde{\mathbf{c}}^r \\ 1 \end{bmatrix}' = \sum_{\mathbf{x}\in\hat{\mathcal{X}}} \alpha_{\mathbf{x}} \begin{bmatrix} \mathbf{d}_r(\mathbf{x}^r)\mathbf{d}_r(\mathbf{x}^r)' + \mathbf{\Phi}_r & \mathbf{d}_r(\mathbf{x}^r) \\ \mathbf{d}_r(\mathbf{x}^r)' & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{\Pi}^r & \boldsymbol{\mu}^r \\ \boldsymbol{\mu}^{r\prime} & 1 \end{bmatrix}. \tag{56}$$

Moreover, as $\tilde{\mathbf{x}} \in \mathcal{X}$, the objective $\mathbb{E}[\max_{\mathbf{x}\in\mathcal{X}} \tilde{\mathbf{c}}'\mathbf{x}]$ satisfies,

$$\begin{aligned} Z^* &\geq \mathbb{E}\left[\max_{\mathbf{x}\in\mathcal{X}} \tilde{\mathbf{c}}'\mathbf{x}\right] \geq \mathbb{E}[\tilde{\mathbf{c}}'\tilde{\mathbf{x}}] = \mathbb{E}\left[\mathbb{E}\left[\tilde{\mathbf{c}} \,|\, \tilde{\mathbf{x}}\right]' \tilde{\mathbf{x}}\right] = \mathbb{E}\left[\sum_{r=1}^{R} \mathbb{E}\left[\tilde{\mathbf{c}}^r \,|\, \tilde{\mathbf{x}}\right]' \tilde{\mathbf{x}}^r\right] \\ &= \mathbb{E}\left[\sum_{r=1}^{R} \mathbf{d}_r(\tilde{\mathbf{x}}^r)'\tilde{\mathbf{x}}^r\right] = \sum_{\mathbf{x}\in\mathcal{X}} \alpha_{\mathbf{x}} \sum_{r=1}^{R} \mathbf{d}_r(\mathbf{x}^r)'\mathbf{x}^r = \sum_{r=1}^{R} trace(\hat{\mathbf{Y}}_*^r) = \hat{Z}^* = Z^*, \end{aligned}$$

where the last three equalities follow, respectively, from (55), the optimality of $\{\mathbf{p}_*, \mathbf{X}_*^r, \mathbf{Y}_*^r : r \in [R]\}$ for (7), and Theorem 1. Combining this observation with (56), we have that the distribution of $\tilde{\mathbf{c}}$, denoted by θ^*, is feasible and it attains the maximum in (2). □

The generation of the normal distributions for each of $\tilde{\mathbf{c}}^r$and the mixture proportions $\alpha_{\mathbf{x}}$ are both identical to [43]. The difference is that in step 2 above, the joint distributions over the whole vector $\tilde{\mathbf{c}}$ is the independent distribution on $\tilde{\mathbf{c}}^r$, $r \in [R]$ conditional on $\mathbf{x}$. Note that in [43], this additional step of constructing a joint distribution was not required as the whole vector $\tilde{\mathbf{c}}$ was entirely generated at once.

Appendix B Reasoning for Gap in Bounds Produced by Formulation (48)

In this section we investigate why the formulation (48) does not necessarily provide tight bounds for Z^*_{series}.

Using a similar reasoning in proof of Theorem 1, Step 1, we can show that $Z^*_{\text{series}} \leq \hat{Z}^*_{\text{series}}$. However a similar adoption of Step 2, proof of Theorem 1 to check $Z^*_{\text{series}} \geq \hat{Z}_{\text{series}}$ does not go through, unfortunately. To see this, let $\mathbf{p}^*, \mathbf{X}^*, \mathbf{Y}^*$ be an optimal solution to formulation (48) and let us attempt to construct a solution $\bar{\mathbf{p}}, \bar{\mathbf{X}}, \bar{\mathbf{Y}}, \bar{\mathbf{\Delta}}$ feasible to (47) such that $trace(\bar{\mathbf{Y}}) = \hat{Z}^*_{\text{series}} = \sum_{i=1}^{n} trace(\mathbf{Y}^*_{ii})$.
Construction of $\bar{\mathbf{p}}$ *and* $\bar{\mathbf{X}}$: Analogous to proof of Theorem 1, step 2.
Construction of $\bar{\mathbf{Y}}$ *and* $\bar{\mathbf{\Delta}}$:
Set $\bar{Y}_{ii} = Y_{ii}, \bar{Y}_{i,i+1} = Y_{i,i+1}, \bar{Y}_{i+1,i} = Y_{i+1,i}$ and $\bar{\Delta}_{ii} = \Pi_{ii},\ \bar{\Delta}_{i,i+1} = \Pi_{i,i+1}$.

As before, consider a $(2n+1)\times(2n+1)$ partial symmetric matrix $\mathbf{L}_p$ constructed using the analogous partial matrices $\bar{\mathbf{Y}}_p$ (with entries $\bar{Y}_{ii}$ and $\bar{Y}_{i,i+1}$ described above), $\bar{\mathbf{\Delta}}_p$(with entries $\bar{\Delta}_{ii}$ and $\bar{\Delta}_{i,i+1}$ described above) and the fully specified matrix $\bar{\mathbf{X}}$ as follows:

$$\bar{\mathbf{L}}_p = \begin{bmatrix} 1 & \boldsymbol{\mu}' & \bar{\mathbf{p}}' \\ \boldsymbol{\mu} & \bar{\mathbf{\Delta}}_p & \bar{\mathbf{Y}}_p \\ \bar{\mathbf{p}} & \bar{\mathbf{Y}}_p' & \bar{\mathbf{X}} \end{bmatrix} = \begin{bmatrix} 1 & \mu_1 & \mu_2 & \mu_3 & \mu_4 & \mu_5 & \dots & p_1^* & p_2^* & p_3^* & p_4^* & \dots \\ \mu_1 & \Pi_{11} & \Pi_{12} & ? & ? & ? & \dots & Y_{11}^* & Y_{12}^* & ? & ? & \dots \\ \mu_2 & \Pi_{21} & \Pi_{22} & \Pi_{23} & ? & ? & \dots & Y_{21}^* & Y_{12}^* & Y_{23}^* & ? & \\ \mu_3 & ? & \Pi_{32} & \Pi_{33} & \Pi_{34} & ? & \dots & ? & Y_{32}^* & Y_{33}^* & Y_{34}^* & ?\dots \\ \mu_4 & ? & ? & \Pi_{43} & \Pi_{44} & \Pi_{45} & ?\dots & ? & ? & Y_{43}^* & Y_{44}^* & \dots \\ & \vdots & & \vdots & & & \vdots & & & \vdots & & \\ p_1^* & Y_{11}^* & Y_{21}^* & ? & ? & ?\dots & & & & & & \\ p_2^* & Y_{12}^* & Y_{22}^* & Y_{32}^* & ? & ?\dots & & & & & & \\ p_3^* & ? & Y_{23}^* & Y_{33}^* & Y_{43}^* & ?\dots & & & \bar{\mathbf{X}} & & & \\ p_4^* & ? & ? & Y_{34}^* & Y_{44}^* & Y_{54}^*\dots & & & & & & \\ & & & & & & & & & & & \\ & \vdots & & \vdots & & & \vdots & & & & & \end{bmatrix}$$

We note that the fully specified principal submatrices of $\bar{\mathbf{L}}_p$ are exactly the matrices that appear in the positive semidefinite constraints in formulation (48) and are therefore guaranteed to be positive-semidefinite. Similar to proof of Theorem 1, if the partial matrix $\bar{\mathbf{L}}_p$ can be shown to admit a completion $\bar{\mathbf{L}}_{\text{comp}}$ such that $\bar{\mathbf{L}}_{\text{comp}} \succeq 0$ then the rest of the entries in $\bar{\mathbf{Y}}$ and $\bar{\mathbf{\Delta}}$ can be computed. However in this stage, the analogous graph constructed as in Lemma 3 is not chordal (refer to Figure 10). Since the graph constructed is not chordal, the construction of a positive semidefinite completion cannot be guaranteed. Therefore the bound $\hat{Z}^*_{\text{series}}$ is not necessarily tight. However it may be used as a polynomial time computable upper bound for Z^*_{series}. Whether Z^*_{series} can be solved in polynomial time or not, is an interesting open question.

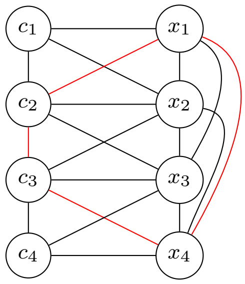


Fig. 10: Illustration of a graph G for the case where $n = 4$ and the moments corresponding to $\{\{1, 2\}, \{2, 3\}, \{3, 4\}\}$ are known. The whole graph additionally includes a vertex s connected to all the nodes in the graph and is omitted here for clarity. Consider a subgraph formed by the vertices $\{c_2, x_1, x_4, c_3\}$. The edges induced by this subset of vertices are shown in red. These edges form a chordless cycle and therefore the graph is not chordal.