Optimizing the expected maximum of two linear functions defined on a multivariate Gaussian distribution

David Bergman
Department of Operations and Information Management, School of Business, University of Connecticut

Carlos Cardonha
Department of Operations and Information Management, School of Business, University of Connecticut

Jason Imbrogno
Department of Finance, Economics, and Data Analytics, University of North Alabama

Leonardo Lozano
Operations, Business Analytics & Information Systems, University of Cincinnati

Abstract

We study stochastic optimization problems with objective function given by the expectation of the maximum of two linear functions defined on the component random variables of a multivariate Gaussian distribution. We consider random variables that are arbitrarily correlated, and we show that the problem is NP-hard even if the space of feasible solutions is unconstrained. We exploit a closed-form expression for the objective function from the literature to construct a cutting-plane algorithm that can be seen as an extension of the integer L-shaped method for a highly nonlinear function, which includes the evaluation of the c.d.f and p.d.f of a standard normal random variable with decision variables as part of the arguments. To exhibit the model’s applicability, we consider two featured applications. The first is daily fantasy sports, where the algorithm identifies entries with positive returns during the 2018-2019 National Football League season. The second is a special case of makespan minimization for two parallel machines and jobs with uncertain processing times; for the special case where the jobs are uncorrelated, we prove the equivalence between its deterministic and stochastic versions and show that our algorithm can deliver a constant-factor approximation guarantee for the problem. The results of our computational evaluation involving synthetic and real-world data suggest that our discretization and upper bounding techniques lead to significant computational improvements and that the proposed algorithm outperforms sub-optimal solutions approaches.

1 Introduction

Consider two zero-mean jointly normal random variables, and the expectation of their maximum. Should they be uncorrelated, the expectation of the maximum is \( \frac{1}{\sqrt{2\pi}} \). With correlation \( \frac{1}{2} \) the expectation is \( \frac{1}{\sqrt{2\pi}} \), and with correlation \( -\frac{1}{2} \) the expectation is \( \frac{\sqrt{3}}{2\sqrt{2\pi}} \). For arbitrary jointly normal random variables \( Z_1, Z_2 \) the maximum is as follows (see Nadarajah and Kotz (2008)):

\[
\mathbb{E} [\max\{Z_1, Z_2\}] = \mathbb{E} [Z_1] \Phi \left( \frac{\mathbb{E}[Z_1] - \mathbb{E}[Z_2]}{\sqrt{\sigma^2(Z_1) + \sigma^2(Z_2) - 2\text{cov}(Z_1, Z_2)}} \right) + \\
\mathbb{E} [Z_2] \Phi \left( \frac{\mathbb{E}[Z_2] - \mathbb{E}[Z_1]}{\sqrt{\sigma^2(Z_2) + \sigma^2(Z_1) - 2\text{cov}(Z_1, Z_2)}} \right) + \\
\sqrt{\sigma^2(Z_1) + \sigma^2(Z_2) - 2\text{cov}(Z_1, Z_2)} \phi \left( \frac{\mathbb{E}[Z_1] - \mathbb{E}[Z_2]}{\sqrt{\sigma^2(Z_1) + \sigma^2(Z_2) - 2\text{cov}(Z_1, Z_2)}} \right),
\]

where \( \Phi \) and \( \phi \) are the standard normal c.d.f and p.d.f, respectively.
where $\Phi$ and $\phi$ are the c.d.f. and p.d.f. of a standard normal random variable, respectively. The evaluation of the expectation requires the calculation of the highly nonlinear c.d.f. and p.d.f. with arguments that are composed of ratios of linear expressions over square roots of linear expressions (and even the product of these expressions). Consider now an optimization problem, where the objective function consists of the expression above, but the selection of $Z_1$ and $Z_2$ are subject to decisions in the problem. This paper studies this decision making setting, describes real-world problems that can be modeled within this context, and explores an efficient algorithm for solving the resulting optimization problems. More formally, we investigate a class of stochastic optimization problems defined over a multivariate Gaussian distribution. In particular, we study the problem of optimizing the expected value of the maximum of two linear functions defined on the component random variables of a multivariate Gaussian distribution, allowing for correlation among the component random variables.

Real-world problems that can be modelled within this context abound, especially if we consider the connection to order statistics, which is used to model types of auctions where the final prices are determined by the first- or second-highest bidder (Brown and Brown 1986); insurance premium determination, where companies use order statistics to determine policies for joint-life insurance (Dimitrova et al. 2018); failure models for wireless communication networks (Yang and Alouini 2011); and risk management (Koutras and Koutras 2018). We study two applications in this paper—(1) expected score maximization in Daily Fantasy Sports (DFS) and (2) makespan minimization for 2 parallel machines with stochastic processing times. These two applications highlight how both objectives (maximizing and minimizing) are important in different contexts.

Exact algorithms for optimizing expressions composed of the maximum of two different stochastic functions is challenging due to the complexity of evaluating or even estimating the expectation of order statistics (David and Nagaraja 2004, Bertsimas et al. 2006, Evans et al. 2006). Specific to the problem class studied in this paper, closed-form expressions for the expected value of the maximum and the minimum are known (see Nadarajah and Kotz (2008)). We show how one can formulate the problem as a binary optimization problem and investigate exact computational approaches for solving the model. Our approach consists of an extension of the integer L-shaped method combined with linearization and discretization techniques tailored for the problem.

Our contributions are the following:

- We show that the underlying optimization problem is NP-hard even in scenarios where the set of solutions is unconstrained;
- We develop an exact optimization algorithm that allows for the incorporation of linear constraints and establish theoretical results on the quality of the upper bounds obtained;
- We study two featured applications, aimed at exhibiting the generality of the class of problems and to showcase the scalability of the proposed algorithms.
- For $P|\rho_i \sim \mathcal{N}(\mu, \Sigma), \rho_{j,a} = 0|\mathbb{E}[C_{\text{max}}]$, a machine scheduling problem investigated in this article, we show that a single iteration of our algorithm can deliver a solution with constant-factor approximation guarantees, and we also prove that this problem is equivalent to its deterministic counterpart (where processing times are replaced by the mean of the respective distributions); and
- We conduct a comprehensive computational study both on synthetic problem instances and real-world data in order to show the computational performance and the quality of the results delivered by our algorithms.

The paper is organized as follows. Section 2 formally defines the class of problems we study. Section 3 presents computational complexity results. In Section 4 we present an exact cutting-plane algorithmic framework. Numerical experiments on synthetic instances are presented and discussed in Section 5. Sections 6 and 7 present our two featured applications. Section 8 concludes the paper and discusses future work.
2 Problem Description and Featured Applications

We consider a collection of random variables $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$ following a multivariate Gaussian distribution $\mathbf{Y} \sim \mathcal{N}(\mathbf{\mu}, \mathbf{\Sigma})$. The random variables $Y_1, \ldots, Y_n$ will be referred to as component random variables which have component-wise expectations $\mu_1, \ldots, \mu_n$. The $n \times n$ covariance matrix $\mathbf{\Sigma}$ has elements $\Sigma_{i,j} = \text{Cov}(Y_i, Y_j)$.

We study the class of problems

$$\max_{x \in \Omega \subseteq \{0,1\}^{2n}} \mathbb{E} \left[ \max \{ Z_1(x), Z_2(x) \} \right], \tag{P}$$

where for $i = 1, 2$, $Z_i(x) = \sum_{j=1}^n Y_j x_{i,j}$ with $x_{i,j} \in \{0,1\}$ for $j = 1, \ldots, n$. The set $\Omega$ defines the feasible region for the decision variables, assumed in this paper to be composed solely of linear constraints. In order to exhibit the expansive generalizability of this class of problems, we present two featured applications.

**Featured Application 1: Daily Fantasy Sports**

In certain DFS competitions, participants select up to 2 fantasy entries, with each entry being composed of a set of players who will participate in upcoming sporting events. The selection of players is subject to roster and budget constraints; more details about these and other constraints are presented in Section 6. Each fantasy entry receives points based on the actual performance of the players selected in the sporting events. Entries are ranked according to the total fantasy points scored, and the payout of each entry depends on its position in this rank. The payout structures are top-heavy, with a small fraction of the entries receiving substantial amounts, including the top-scoring entry receiving approximately 25% of the total entry fees paid.

Suppose there are $n$ players that can be selected for any of a participant’s 2 entries. Each player will score a random number of fantasy points $Y_j$. Letting $x_{i,j}$ indicate the selection of player $j$ for entry $i$, for $i = 1, 2$ and $j = 1, \ldots, n$, one can model the selection problem for a participant as

$$\max \mathbb{E} \left[ \max \{ Z_1(x), Z_2(x) \} \right] \tag{DFS}$$

s.t.

$$Z_i(x) = \sum_{j=1}^n Y_j x_{i,j} \quad i = 1, 2$$

$$x \in \Omega$$

The correlation between the points received for a pair of players can be significant, for example between a quarterback and a wide receiver in football, since most fantasy points that a wide receiver receives will generally be associated with fantasy points for the quarterback on the same team. The constraints in $\Omega$ define conditions which enforce what configurations of players make a legal entry, which will be explained in more detail later and can be modeled through linear constraints.

Algorithmic sports betting recently became a topic of interest in the operations management and operations research literature (Kaplan and Garstka 2001, Clair and Letscher 2007, Haugh and Singal 2021). Some examples involve selecting multiple entries for maximizing the expected score of the maximum scoring entry in both National Football League (NFL) survival pools (Bergman and Imbrogno (2017)) and DFS (Hunter et al. (2016), Haugh and Singal (2021)).

**Featured Application 2: Makespan Minimization with Stochastic Processing Times**

Consider a set of $n$ jobs with processing times $Y_j$ drawn from a multivariate Gaussian distribution to be partitioned for execution on 2 parallel (identical) machines such that the makespan (the time at which the last job completes) is minimized; this problem is represented by $P2|p \sim \mathcal{N}(\mathbf{\mu}, \mathbf{\Sigma})|\mathbb{E}[C_{\text{max}}]$ in the notation of Graham et al. (1979). Letting binary variable $x_{i,j}$ indicate if job $j$ is assigned to machine $i$ for $i = 1, 2$ and $j = 1, \ldots, n$, the problem can be
formulated as
\[
\begin{align*}
\min & \quad \mathbb{E} \left[ \max \{Z_1(x), Z_2(x)\} \right] \\
\text{s.t.} & \quad Z_i(x) = \sum_{j=1}^{n} Y_j x_{i,j} \quad i = 1, 2 \\
& \quad x_{1,j} + x_{2,j} = 1 \quad j = 1, \ldots, n \\
& \quad x_{i,j} \in \{0, 1\} \quad i = 1, 2, \ j = 1, \ldots, n.
\end{align*}
\]

The objective seeks to minimize the maximum expected completion time of all machines. Note that no assumption is presumed on the correlation between the processing time of the jobs. We incorporate $\rho_{j,j'} = 0$ to the notation to represent scenarios where processing times are uncorrelated.

There is a vast literature on stochastic scheduling (see Nino-Mora (2009)). In particular, Coffman Jr et al. (1987) characterize optimal solutions for makespan minimization in scenarios involving 2 or 3 machines and jobs with exponentially distributed processing times. Pinedo (2005) shows the connection between a stochastic flow shop scheduling problem and a deterministic traveling salesman problem; we present a similar result involving $P2|p \sim \mathcal{N}(\mu, \Sigma), \rho_{j,j'} = 0|\mathbb{E}[C_{\text{max}}]$ in this work.

\section{Computational Complexity}

Consider random variables $Y = (Y_1, \ldots, Y_n)^T$ following a multivariate Gaussian distribution $Y \sim \mathcal{N}(\mu, \Sigma)$ and note that all component random variables $Y_j$ are normally distributed and potentially correlated, i.e., $Y_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$, with mean $\mu_j$ and variance $\sigma_j^2 = \Sigma_{j,j}$, for all $j \in \{1, \ldots, n\}$.

\textbf{Theorem 1.} Optimization problem $P$ is NP-hard even if $\Omega = \{0, 1\}^{2^\alpha n}$.

Before proceeding with the proof, we recall some known results that will be relevant in this section and throughout the manuscript. Since $Y$ follows a multivariate Gaussian distribution, $Z_i(x) = \sum_{j=1}^{n} Y_j x_{i,j}$ is normally distributed for all $i = 1, 2$. Moreover, for any $x \in \{0, 1\}^{2^\alpha n}$, the mean and variance of $Z_1(x)$ and $Z_2(x)$, as well as their covariances (and correlations), are given by:

\begin{align*}
\mathbb{E} [Z_i(x)] &= \sum_{j=1}^{n} \mu_j x_{i,j} \quad \forall i \in \{1, 2\} \\
\sigma^2 (Z_i(x)) &= \sum_{j=1}^{n} \sigma_j^2 x_{i,j} + 2 \sum_{1 \leq j < j' \leq n} \text{cov}(Y_j, Y_{j'}) x_{i,j} x_{i,j'} \quad \forall i \in \{1, 2\} \\
\text{cov} (Z_1(x), Z_2(x)) &= \sum_{j=1}^{n} \sum_{j'=1}^{n} \text{cov}(Y_j, Y_{j'}) x_{1,j} x_{2,j'}.
\end{align*}

Let $\phi(\cdot)$ and $\Phi(\cdot)$ be the probability density function (p.d.f.) and the cumulative distribution function (c.d.f.), respectively, of standard normal random variables; i.e., for $w \in (-\infty, \infty)$,

\[
\phi(w) = \frac{1}{\sqrt{2\pi}} e^{-\frac{w^2}{2}}, \quad \Phi(w) = \int_{-\infty}^{w} \phi(u)du.
\]

An exact expression is known for $\mathbb{E} \left[ \max\{Z_1(x), Z_2(x)\} \right]$ (see Nadarajah and Kotz (2008) and Clark (1961)):

\[
\mathbb{E} [Z_1(x)] \Phi \left( \frac{\delta(x)}{\theta(x)} \right) + \mathbb{E} [Z_2(x)] \Phi \left( -\frac{\delta(x)}{\theta(x)} \right) + \theta(x) \phi \left( \frac{\delta(x)}{\theta(x)} \right),
\]

\[
(4)
\]
where
\[ \delta(x) = \mathbb{E}[Z_1(x)] - \mathbb{E}[Z_2(x)]. \]
and
\[ \theta(x) = \sqrt{\sigma^2(Z_1(x)) + \sigma^2(Z_2(x)) - 2\text{cov}(Z_1(x), Z_2(x))}. \] (5)
In order to simplify notation, we assume without loss of generality that \( \mathbb{E}[Z_1(x)] \geq \mathbb{E}[Z_2(x)] \).

As we can see from Expression 4, due to the (possible) dependency among the random variables \( Z_1(x) \) and \( Z_2(x) \), the expectation of their maximum is a highly nonlinear function.

Note that if \( \theta(x) = 0 \), the ratio in the argument of the c.d.f. may be undefined, but the expression holds true if we assume \( \Phi \left( \frac{x}{a} \right) \) to be 0 if \( a < 0 \), \( \frac{1}{2} \) if \( a = 0 \), and 1 if \( a > 0 \).

Equipped with this expression, we proceed with the proof of Theorem 1.

**Proof.** Proof of Theorem 1. The result follows from a reduction from the minimum weighted cut problem, which is known to be NP-hard (McMahan et al. (2003)). Let \( G = (V, E) \) be an undirected graph with integer weights \( w_e \) on each edge \( e \in E \). An \( (S, T) \)-cut of \( G \) is a 2-partition of \( V \), and its weight \( w(S, T) \) is the sum of the weights of the edges “crossing” the cut, i.e., \( w(S, T) = \sum_{e \in E} w_e \).

In the decision version of the problem, we are also given a constant \( K \) and wish to know if there exists a \( (S, T) \)-cut of \( G \) such that \( w(S, T) \leq K \).

We create an instance of \( P \) as follows. Assuming there are \( n \) vertices in \( G \), every vertex \( j \in V \) is associated with a normally distributed random variable \( Y_j \sim \mathcal{N}(0, 1) \), for \( j = 1, \ldots, n \). The covariance between \( Y_j \) and \( Y_{j'} \) is \( \text{cov}(Y_j, Y_{j'}) = \frac{\mathbb{E}[Y_j Y_{j'}]}{\sqrt{\mathbb{E}[Y_j^2] \mathbb{E}[Y_{j'}^2]}} \), where \( M = \sum_{e \in E} |w_e| \). As a result, we have that \( \sum_{j=1}^{2M} \sum_{j' \neq j} \text{cov}(Y_j, Y_{j'}) \leq \frac{2M}{\mathbb{E}[Y_j^2]} < \frac{1}{2} \) and \( \mathbb{E}[Y_j Y_{j'}] \geq \sum_{j' \neq j} \mathbb{E}[Y_j Y_{j'}] \) for all \( j = 1, \ldots, n \). One can then construct a symmetric and diagonally dominant (consequently, positive semi-definite) matrix \( \Sigma \) whose columns and rows are indexed by variables \( Y_j \). It follows that \( \Sigma \) is a valid covariance matrix. Finally, let \( \Omega = \{0, 1\}^{2n \times n} \), i.e., the set of feasible solutions is unconstrained.

By construction, \( \mathbb{E}[Z_1(x)] = \mathbb{E}[Z_2(x)] = 0 \), so the expression for \( \mathbb{E}[Z_2(x)] \) reduces to \( \theta(x) = \frac{1}{\sqrt{2\pi}} \).

As \( n \geq 1 \) and all variances are equal to 1, it follows that at optimality \( \theta(x) \geq 1 \) (this value is achieved if we assign one item to the first set and leave the second set empty), so any \( x \) that maximizes \( \theta(x)^2 \) also maximizes \( \theta(x) \). Therefore, our problem is equivalent to

\[ \max_{x \in \Omega} \theta(x)^2 = \sigma^2(Z_1(x)) + \sigma^2(Z_2(x)) - 2\text{cov}(Z_1(x), Z_2(x)). \]

By expanding the terms of the last expression and replacing all variances for 1, \( \theta(x)^2 \) becomes

\[ \sum_{j=1}^{n} (x_{1,j} + x_{2,j} - 2x_{1,j} x_{2,j}) + 2 \sum_{j=1}^{n} \sum_{j' \neq j} \text{cov}(Y_j, Y_{j'}) (x_{1,j} x_{1,j'} + x_{2,j} x_{2,j'}) - \sum_{j=1}^{n} \sum_{j' \neq j} \text{cov}(Y_j, Y_{j'}) x_{1,j} x_{2,j'} \] (6)

**Claim 1.** In any optimal solution for \( \max_{x \in \Omega} \theta(x)^2 \), \( x_{1,j} + x_{2,j} = 1 \) for \( j = 1, \ldots, n \).

**Proof** Let \( A(x) \) denote the sum within the brackets in 6; \( A(x) \) belongs to the interval defined by \( \pm \sum_{j=1}^{n} \sum_{j' \neq j} \text{cov}(Y_j, Y_{j'}) \), because each covariance term appears at most twice with a positive coefficient and at most twice with a negative coefficient. By construction, \( \sum_{j=1}^{n} \sum_{j' \neq j} \text{cov}(Y_j, Y_{j'}) < \frac{1}{2} \), so we have that \( -1 < 2A(x) < 1 \). Therefore, any optimal solution to \( \max_{x \in \Omega} \theta(x)^2 \) also optimizes

\[ \max_{x \in \Omega} \left\{ \sum_{j=1}^{n} x_{1,j} + \sum_{j=1}^{n} x_{2,j} - 2 \sum_{j=1}^{n} x_{1,j} x_{2,j} \right\}. \]
since the absolute value of each of the coefficients in this expression is greater than or equal to 1. Finally, note that this expression is maximized if and only if \( x_{1,j} + x_{2,j} = 1 \), as desired.

It follows from Claim 1 that

\[
\max_{\mathbf{x} \in \Omega} \sum_{j=1}^{n} \sum_{j'=j+1}^{n-1} \text{cov}(Y_j, Y_{j'}) (x_{1,j} x_{1,j'} + x_{2,j} x_{2,j'}) - \sum_{j=1}^{n} \sum_{j'=j+1}^{n} \text{cov}(Y_j, Y_{j'}) x_{1,j} x_{2,j'}
\]

(7)

s.t. \( x_{1,j} + x_{2,j} = 1 \)

\( j \in [n] \)

(8)

consider the following optimization problem:

\[
\min_{\mathbf{x}} h(x) = \sum_{j=1}^{n} \sum_{j'=j+1}^{n} \text{cov}(Y_j, Y_{j'}) x_{1,j} x_{2,j'}
\]

(10)

s.t. \( x_{1,j} + x_{2,j} = 1 \)

\( j \in [n] \)

\( x_{i,j} \in \{0,1\} \)

\( i \in [2], j \in [n] \).

Claim 2. An optimal solution to optimization problem (10) is also optimal to problem (7).

Proof. Both optimization problems have the same set of feasible solutions \( \Omega' \). Let \( x' \) and \( x'' \) be two feasible solutions with \( h(x') < h(x'') \). Showing that \( \theta(x')^2 > \theta(x'')^2 \) establishes the claim. To show this, we first note that for any feasible solution \( \bar{x} \),

\[
\sum_{j=1}^{n} \sum_{j'=j+1}^{n} \text{cov}(Y_j, Y_{j'}) \bar{x}_{1,j} \bar{x}_{1,j'} + \sum_{j=1}^{n} \sum_{j'=j+1}^{n} \text{cov}(Y_j, Y_{j'}) \bar{x}_{2,j} \bar{x}_{2,j'}
\]

(11)

\[
+ \sum_{j=1}^{n} \sum_{j'=j+1}^{n} \text{cov}(Y_j, Y_{j'}) \bar{x}_{1,j} \bar{x}_{2,j'} = \sum_{j=1}^{n} \sum_{j'=j+1}^{n} \text{cov}(Y_j, Y_{j'}).\]

This follows because for any two indices \( j \neq j' \), the covariance term \( \text{cov}(Y_j, Y_{j'}) \) is counted in exactly one of the three terms in the left-hand size of equation (11):

1. If \( \bar{x}_{1,j} = \bar{x}_{1,j'} = 1 \rightarrow \text{cov}(Y_j, Y_{j'}) \) is counted only in the first term;
2. If \( \bar{x}_{2,j} = \bar{x}_{2,j'} = 1 \rightarrow \text{cov}(Y_j, Y_{j'}) \) is counted only in the second term;
3. If \( \bar{x}_{1,j} = 1, \bar{x}_{2,j'} = 1 \rightarrow \text{cov}(Y_j, Y_{j'}) \) is counted only in the third term; and
4. If \( \bar{x}_{2,j} = 1, \bar{x}_{1,j'} = 1 \rightarrow \text{cov}(Y_j, Y_{j'}) \) is counted only in the third term.

Finally, because \( \bar{x}_{1,j} + \bar{x}_{2,j} = 1 \) and \( \bar{x}_{1,j'} + \bar{x}_{2,j'} = 1 \), it follows that the list above is exhaustive and contains all possible assignments of \( j \) and \( j' \). Therefore,

\[
h(x') < h(x'') \quad \Rightarrow \quad -h(x') > -h(x'') \quad \Rightarrow \quad \sum_{j=1}^{n} \sum_{j'=j+1}^{n} \text{cov}(Y_j, Y_{j'}) - h(x') > \sum_{j=1}^{n} \sum_{j'=j+1}^{n} \text{cov}(Y_j, Y_{j'}) - h(x''),
\]

which implies that \( \theta(x')^2 > \theta(x'')^2 \), as desired.

We conclude that an optimal solution to (10) is also optimal to the original problem. We now show a one-to-one mapping between solutions of (10) and \((S,T)\)-cuts of \( G \). Namely, for a feasible solution \( x' \), we associate it with the \((S(x'),T(x'))\)-cut defined by \( x'_{1,j} = 1 \leftrightarrow j \in S(x') \).
and \( x^2_{i,j} = 1 \iff j \in T(x') \). Additionally, \( w(S(x'),T(x')) \leq K \) if and only if \( h(x') \leq \frac{K}{\ln 2} \). This follows because

\[
w(S(x'),T(x')) = \sum_{j \in S(x')} \sum_{j' \in T(x')} w_{j,j'} = \sum_{j \in S(x')} \sum_{j' \in T(x')} (4M + 1) \text{cov}(Y_j,Y_{j'}) \\
= (4M + 1) \sum_{j=1}^{n} \sum_{j' \neq j} \text{cov}(Y_j,Y_{j'}) = (4M + 1) h(x').
\]

It follows that (10) is equivalent to the minimum weighted cut problem, as desired. ■

4 Exact Solution Algorithm

We propose an exact solution algorithm for problem \( P \) and focus on the maximization of the objective function. We discuss in the Appendix how to adapt the algorithm for the minimization case. Omitted proofs can be found in the Appendix.

4.1 A Cutting-Plane Framework

The objective function presented in (4) is highly nonlinear, thus making the application of direct formulations combined with off-the-shelf solvers unlikely to be successful. We present an exact cutting-plane algorithm based on a mixed-integer linear program (MILP) referred to as the relaxed master problem. This MILP provides upper bounds on the optimal objective value and is iteratively updated through the inclusion of cuts. Lower bounds are obtained through the evaluation of \( \mathbb{E}[\max\{Z_1(x),Z_2(x)\}] \) on the feasible solutions obtained from the optimization of the relaxed master problem. The algorithm stops once it finds a provable optimal solution.

Our approach to solve the problem is presented in Algorithm 1. A key component of our algorithm is the construction of an upper-bounding function for \( \mathbb{E}[\max\{Z_1(x),Z_2(x)\}] \) defined over the set \( \Omega \) of feasible solutions. Namely, we wish to work with a function \( g(x) \) such that \( g(x) \geq \mathbb{E}[\max\{Z_1(x),Z_2(x)\}] \) for each \( x \) in \( \Omega \). Given \( g(x) \), the relaxed master problem can be stated as

\[
\bar{z} = \max_{x \in \Omega} g(x), \quad \text{(RMP)}
\]

and \( \bar{z} \) provides an upper bound on the optimal objective value of problem \( P \).

Algorithm 1 solves problem RMP iteratively, adding no-good constraints (or simply cuts) to a set \( \mathcal{C} \), which are incorporated to RMP in order to prune previously explored solutions (Balas and Jeroslow 1972). RMP(\( \mathcal{C} \)) denotes RMP further constrained by \( \mathcal{C} \), so that a solution \( x \) for RMP(\( \mathcal{C} \)) belongs to \( \Omega \) and satisfies all no-good constraints of \( \mathcal{C} \). In each iteration, after finding an optimal solution \( \hat{x} \) for RMP(\( \mathcal{C} \)), Algorithm 1 adds the following cut \( c(\hat{x}) \) to \( \mathcal{C} \).

\[
\sum_{i \in \{1,2\}} \sum_{j \in \{1,...,n\}} x_{i,j} - \sum_{i \in \{1,2\}} \sum_{j \in \{1,...,n\}} x_{i,j} \leq \sum_{i \in \{1,2\}} \sum_{j \in \{1,...,n\}} \hat{x}_{i,j} - 1. \quad (12)
\]

The only solution in \( \Omega \) that violates \( c(\hat{x}) \) is \( \hat{x} \). Therefore, every solution in \( \Omega \) is explored at most once, and as \( \Omega \) is finite, Algorithm 1 terminates after a finite number of steps.

Our cutting-plane algorithm keeps a lower bound \( LB \) and an upper bound \( UB \) for \( \mathbb{E}[\max\{Z_1(x),Z_2(x)\}] \), which are iteratively updated based on solutions \( \hat{x} \) of RMP(\( \mathcal{C} \)). Upper bounds are given by \( g(\hat{x}) \); these bounds are non-decreasing, as any solution \( x' \) such that \( g(x') > g(\hat{x}) \) must have been explored in a previous iteration of Algorithm 1. Similarly, lower bounds are obtained through the exact evaluation of \( \mathbb{E}[\max\{Z_1(\hat{x}),Z_2(\hat{x})\}] \) (i.e., using Equation (4)). Observe that \( \mathbb{E}[\max\{Z_1(\hat{x}),Z_2(\hat{x})\}] \) may be smaller than \( \mathbb{E}[\max\{Z_1(x'),Z_2(x')\}] \) for some previously explored solution \( x' \), so Algorithm 1 needs to store the largest \( LB \) found in previous iterations. Algorithm 1 terminates when \( LB \) becomes equal to \( UB \) or when RMP(\( \mathcal{C} \)) becomes infeasible.

Algorithm 1 can be seen as an extension of the integer L-shaped method (Laporte and Louveaux 1993), dealing with the added difficulty of a highly nonlinear objective function. Similar cutting-plane algorithms have been extensively used in the context of two-stage stochastic
Algorithm 1 A Cutting-Plane Algorithm

1: Set $LB = -\infty$, $UB = \infty$, $C = \emptyset$, and incumbent solution $\hat{x} = 0$.
2: Optimize RMP($C$) to obtain $\hat{x}$; if the problem is infeasible, go to Step 6.
3: Set $UB = g(\hat{x})$.
4: If $E[\max\{Z_1(\hat{x}), Z_2(\hat{x})\}] > LB$, set $LB = E[\max\{Z_1(\hat{x}), Z_2(\hat{x})\}]$ and update incumbent $\bar{x} = \hat{x}$.
5: If $LB = UB$, go to Step 6. Otherwise, set $C = C \cup \{c(\hat{x})\}$ and return to Step 2.
6: If $LB = -\infty$, original problem is infeasible. Otherwise, terminate with optimal solution $\bar{x}$.

4.2 Baseline Approach for Obtaining Upper Bounds on the Objective Function

The performance of our cutting-plane algorithm is directly tied to the quality of the bounds obtained from $g(x)$ and the difficulty of solving the RMP. A bounding function that delivers accurate overestimations of $E[\max\{Z_1(x), Z_2(x)\}]$ but requires an impractical amount of time to solve the resulting RMP is likely to result in poor performance of the cutting-plane algorithm, as the RMP is solved at every iteration of the algorithm. On the other hand, a bounding function for which the resulting RMP is easily solved but provides poor quality upper bounds is likely to result in a large number of iterations before closing the optimality gap.

We remark that $E[\max\{Z_1(x), Z_2(x)\}] \neq \max\{E[Z_1(x)], E[Z_2(x)]\}$, and in our setting it holds that $E[\max\{Z_1(x), Z_2(x)\}] \geq \max\{E[Z_1(x)], E[Z_2(x)]\}$. As a result, a simple bounding function is given by the following linear expression:

$$\max\{E[Z_1(x)], E[Z_2(x)]\} + M,$$

where $M$ is a sufficiently large constant such that

$$M \geq \theta(x) \phi\left(\frac{\delta(x)}{\theta(x)}\right).$$

We formulate the corresponding RMP as a linear mixed-integer program (MIP) that can be solved with any off-the-shelf optimization software. Unfortunately, such a simple function yields poor quality bounds and virtually requires a complete enumeration of the solution space. In preliminary computational experiments, the cutting-plane algorithm is not able to solve to optimality a single problem instance from our test bed using (13) as the bounding function, resulting in high optimality gaps at the end of the time limit. As a result, we investigate more complex linear bounding functions, aiming to achieve a balance between the difficulty of solving the RMP (which we cast as a linear MIP) and the quality of the upper bounds obtained.

A challenging task involved in obtaining bounds on the optimal objective is the evaluation of $\theta(x)$, a nonlinear expression that appears in all terms of (4), including the denominators of the c.d.f. and the p.d.f. of the standard normal distribution. To avoid the technical issues involved in the evaluation of $\theta(x)$, we propose a baseline approach that evaluates $\theta(x)^2$ exactly and defines an upper bounding function for $E[\max\{Z_1(x), Z_2(x)\}]$, based on a discretization of $\theta(x)^2$, where the value of $\theta(x)^2$ is

$$\theta(x)^2 = \sigma^2(Z_1(x)) + \sigma^2(Z_2(x)) - 2\text{cov}(Z_1(x), Z_2(x)).$$

The exact value of $\theta(x)^2$ is computed via a McCormick linearization technique (McCormick 1976). Let function $u_\theta(x)$ be such that $0 \leq \theta(x) \leq u_\theta(x)$. Our baseline upper-bounding function is given in Proposition 1.
Proposition 1. For every \( x \in \Omega \),
\[
\mathbb{E} \left[ \max\{Z_1(x), Z_2(x)\} \right] = \mathbb{E}[Z_1(x)] \Phi \left( \frac{\delta(x)}{\theta(x)} \right) + \mathbb{E}[Z_2(x)] \Phi \left( -\frac{\delta(x)}{\theta(x)} \right) + \theta(x) \phi \left( \frac{\delta(x)}{\theta(x)} \right)
\leq \mathbb{E} [Z_1(x)] + u_\theta(x) \frac{1}{\sqrt{2\pi}}.
\]

**Proof** From the symmetry of the c.d.f. of the normal distribution follows that
\[
\Phi \left( \frac{\delta(x)}{\theta(x)} \right) + \Phi \left( -\frac{\delta(x)}{\theta(x)} \right) = 1.
\]
Since \( \mathbb{E} [Z_1(x)] \geq \mathbb{E} [Z_2(x)] \) by assumption, we have
\[
\mathbb{E} [Z_1(x)] \geq \Phi \left( \frac{\delta(x)}{\theta(x)} \right) \mathbb{E} [Z_1(x)] + \Phi \left( -\frac{\delta(x)}{\theta(x)} \right) \mathbb{E} [Z_2(x)],
\]
which constitutes an upper bounding expression for the first two terms of \( \mathbb{E} \left[ \max\{Z_1(x), Z_2(x)\} \right] \).
Noting that \( \phi \left( \frac{\delta(x)}{\theta(x)} \right) \leq \frac{1}{\sqrt{2\pi}} \), and \( \theta(x) \leq u_\theta(x) \) by assumption, we have
\[
u_\theta(x) \frac{1}{\sqrt{2\pi}} \geq \theta(x) \phi \left( \frac{\delta(x)}{\theta(x)} \right).
\]
Finally, Inequality (15) follows from the addition of Inequality (16) with Inequality (17). \( \blacksquare \)

We propose a discretization approach to obtain \( u_\theta(x) \) in Section 4.3.1 and present the mathematical model for the baseline RMP in the Appendix. We use this baseline bounding function as a benchmark to measure the improvements in computational performance added by the proposed enhanced bounding techniques in the following sections.

4.3 Enhanced Upper-Bounding Function

Our enhanced approach is based on a joint discretization of \( \theta(x)^2 \) and \( \delta(x) \). The proposed enhanced RMP formulation relies on the following proposition, which is valid for any functions \( u_\theta(x), l_\theta(x), u_\delta(x), \) and \( l_\delta(x) \) such that \( 0 \leq l_\theta(x) \leq \theta(x) \leq u_\theta(x) \) and \( 0 \leq l_\delta(x) \leq \delta(x) \leq u_\delta(x) \):

Proposition 2. For every \( x \in \Omega \),
\[
\mathbb{E} [Z_1(x)] \Phi \left( \frac{u_\delta(x)}{l_\theta(x)} \right) + \mathbb{E} [Z_2(x)] \left( 1 - \Phi \left( \frac{u_\delta(x)}{l_\theta(x)} \right) \right) + u_\theta(x) \phi \left( \frac{l_\delta(x)}{u_\theta(x)} \right) \geq \mathbb{E} \left[ \max\{Z_1(x), Z_2(x)\} \right].
\]

**Proof** This follows because
\[
\mathbb{E} \left[ \max\{Z_1(x), Z_2(x)\} \right] \leq \mathbb{E} [Z_1(x)] \Phi \left( \frac{\delta(x)}{\theta(x)} \right) + \mathbb{E} [Z_2(x)] \left( 1 - \Phi \left( \frac{\delta(x)}{\theta(x)} \right) \right) + u_\theta(x) \phi \left( \frac{\delta(x)}{\theta(x)} \right)
\leq \mathbb{E} [Z_1(x)] \Phi \left( \frac{u_\delta(x)}{l_\theta(x)} \right) + \mathbb{E} [Z_2(x)] \left( 1 - \Phi \left( \frac{u_\delta(x)}{l_\theta(x)} \right) \right) + u_\theta(x) \phi \left( \frac{l_\delta(x)}{u_\theta(x)} \right).
\]
The first inequality follows from the symmetry of the c.d.f. and because \( u_\theta(x) \geq \theta(x) \). For the second inequality, first note that, for any constants \( a, b \) with \( a \geq b, a\lambda_1 + b(1-\lambda_1) \leq a\lambda_2 + b(1-\lambda_2) \), for every \( 0 \leq \lambda_1 \leq \lambda_2 \leq 1 \). As the c.d.f. is non-decreasing on its domain and \( \frac{\delta(x)}{\theta(x)} \leq \frac{u_\delta(x)}{u_\theta(x)} \), and as \( \phi(y) \) is non-increasing for \( y \geq 0 \) and \( \frac{\delta(x)}{\theta(x)} \geq \frac{\delta(x)}{u_\theta(x)} \), the result follows. \( \blacksquare \)

Suppose we are given \( d \) intervals \( \{ [\theta_q^2, \theta_{q+1}^2] \}_{q=1}^d \) and \( l \) intervals \( \{ [\delta_h, \delta_{h+1}] \}_{h=1}^l \), with \( \theta(x)^2 \in [\theta_1^2, \theta_{d+1}^2] \) and \( \delta(x) \in [\delta_1, \delta_{l+1}] \) for every \( x \in \Omega \). Furthermore, let \( \theta_q \) and \( \theta_{q+1} \) denote a lower and
upper bound of $\theta(x)$, respectively, for $\theta(x)^2 \in \left[ \theta_{q}^2, \theta_{q+1}^2 \right]$. Using these intervals we construct the following enhanced RMP formulation:

$$\text{max} \quad U + U'$$

s.t. $u_1 = \sum_{j=1}^{n} \mu_j x_{1,j}; \quad u_2 = \sum_{j=1}^{n} \mu_j x_{2,j}; \quad u_1 \geq u_2$ \hspace{1cm} (19)

$$s = \sum_{i=1}^{3} \left( \sum_{j=1}^{n} \sigma_j^2 x_{i,j} + 2 \sum_{1 \leq j < j' \leq n} \text{cov}(Y_j, Y_{j'}) \nu_{i,j,j'} \right) - 2 \sum_{j=1}^{n} \sum_{j'=1}^{n} \text{cov}(Y_j, Y_{j'}) r_{j,j'}$$ \hspace{1cm} (20)

$$v_{i,j,j'} \leq x_{i,j}; \quad v_{i,j,j'} \leq x_{i,j'} \quad \forall j, j' \in \{1, \ldots, n\}, \quad i \in \{1, 2\}$$ \hspace{1cm} (21)

$$v_{i,j,j'} \geq x_{i,j} + x_{i,j'} - 1 \quad \forall j, j' \in \{1, \ldots, n\}, \quad i \in \{1, 2\}$$ \hspace{1cm} (22)

$$r_{j,j'} \leq x_{1,j}; \quad r_{j,j'} \leq x_{2,j'} \quad \forall j, j' \in \{1, \ldots, n\}$$ \hspace{1cm} (23)

$$r_{j,j'} \geq x_{1,j} + x_{2,j'} - 1 \quad \forall j, j' \in \{1, \ldots, n\}$$ \hspace{1cm} (24)

$$\sum_{q=1}^{d} w_q = 1; \quad \sum_{h=1}^{l} y_h = 1; \quad s' = \sum_{q=1}^{d} \theta_{q+1} w_q$$ \hspace{1cm} (25)

$$\theta_{q}^2 w_q \leq s \leq \theta_{q+1}^2 + \Omega_{q+1} (1 - w_q) \quad q = 1, \ldots, d$$ \hspace{1cm} (26)

$$\delta_h y_h \leq u_1 - u_2 \leq \delta_{h+1} + \Omega_{h+1} (1 - y_h) \quad h = 1, \ldots, l$$ \hspace{1cm} (27)

$$U \leq u_1 \Phi \left( \frac{\delta_{h+1}}{\theta_q} \right) + u_2 \left( 1 - \Phi \left( \frac{\delta_{h+1}}{\theta_q} \right) \right) + M \left( 2 - w_q - y_h \right) \quad q = 1, \ldots, d; \quad h = 1, \ldots, l$$ \hspace{1cm} (28)

$$U' \leq \theta_{q+1} \Phi \left( \frac{\delta_h}{\theta_{q+1}} \right) + M \left( 2 - w_q - y_h \right) \quad q = 1, \ldots, d; \quad h = 1, \ldots, l$$ \hspace{1cm} (29)

$$v \in \{0, 1\}^{n \times n}; \quad r \in \{0, 1\}^{n \times n}; \quad w \in \{0, 1\}^{d}; \quad y \in \{0, 1\}^l; \quad x \in \Omega.$$ \hspace{1cm} (30)

Binary variables $w_q$, $q = 1, \ldots, d$, and $y_h$, $h = 1, \ldots, l$, indicate which interval $\theta(x)^2$ and $\delta(x)$ belong to, respectively. Variable $u_1$ ($u_2$) denotes $\mathbb{E}[Z_1(x)]$ ($\mathbb{E}[Z_2(x)]$) and $s$ ($s'$) represents $\theta(x)^2$ ($\theta_{q+1}^2$). Binary variable $v_{i,j,j'}$ takes a value of 1 iff $x_{i,j} = x_{i,j'} = 1$. Similarly, $r_{j,j'}$ equals 1 iff $x_{1,j} = x_{2,j'} = 1$. Variable $U$ captures the first two terms of expression (18) and $U'$ captures the third term.

The objective function (19) maximizes the upper bounding function defined by Proposition 2. Constraints (20) define the $u$-variables according to equation (1) and impose the symmetry breaking condition $u_1 \geq u_2$. Constraint (21) imposes $s = \theta(x)^2$ as described by equation (14), where $\sigma^2(Z_1(x))$, $\sigma^2(Z_2(x))$, and $\text{cov}(Z_1(x), Z_2(x))$ are computed according to equations (2) and (3), respectively. Constraints (22)–(25) are the McCormick linearization constraints. Constraints (26) ensure that exactly one interval is chosen for $\theta(x)^2$ and $\delta(x)$, and set $s'$ equal to the upper bound of $\theta(x)$ for the interval that $\theta(x)^2$ belongs to. Constraints (27)–(28) select the right interval for $\theta(x)^2$ and $\delta(x)$. Constraints (29) are only active for the selected intervals and enforce that $U$ is bounded by a linear combination of $u_1$ and $u_2$, defined by the evaluation of the c.d.f. at appropriately chosen constants associated with the intervals that $\theta(x)^2$ and $\delta(x)$ lie in; where $M$ is a sufficiently large value. Similarly, constraints (30) enforce $U'$ to equal the third term of expression (18) for the corresponding intervals. Constraints (31) define the domains of the variables appropriately.

### 4.3.1 Discretization of $\theta(x)^2$

We obtain an upper bound $\theta_{d+1}^2$ for $\theta(x)^2$ by solving $\max\{s \mid (x, v, r, u_1, u_2, s) \in \Psi\}$, where $\Psi$ is the space defined by constraints (20)–(25) and (31). This problem is computationally challenging
for all $\Phi$ amount of time in order to obtain a relatively refined upper bound.

We define $d$ intervals for $\theta(x)^2$ as follows: $\theta_1^2 = 0$, $\theta_2^2 = 1$, and $\theta_q^2 = \theta_{q-1}^2 + \frac{\theta_{q-1}^2}{d-1}$ for $q = 3, \ldots, d+1$. The first interval is different from the others since $\sqrt{a^2} \geq a^2$ for $0 \leq a \leq 1$. Note that for any $x$ in $\Omega$, $0 \leq \theta(x)^2 \leq \theta_{2q+1}^2$, so $\theta(x)^2$ must belong to $[\theta_q^2, \theta_{q+1}^2]$ for some $q = 1, \ldots, d$. Given these intervals, upper bounds and lower bounds for $\theta(x)$ are given by

$$
\theta_{q+1} = \begin{cases} 
1, & q = 1 \\
\sqrt{\theta_{q-1}^2}, & q = 2, \ldots, d \\
0, & q = 1 \\
\sqrt{\theta_q^2}, & q = 2, \ldots, d.
\end{cases}
$$

### 4.3.2 Discretization of $\delta(x)$

The procedure is analogous to the one described in 4.3.1. Namely, we obtain an upper bound $\delta_{q+1}$ for all $\delta(x)$ by solving problem $\max \{ u_1 - u_2 \mid (x, v, r, u_1, u_2, s) \in \Psi \}$ for a limited amount of time. Given $\delta_{q+1}$, we generate $l$ discretization intervals defined by $\delta_h = \frac{l-1}{l} \delta$ for $h = 1, \ldots, l+1$. By construction, $\delta(x)$ belongs to one of the intervals defined by the values $\delta_h$, as desired.

### 4.4 Tightness of the Upper-Bounding Function

We investigate now $\Delta(x) = g(x) - \mathbb{E} [\max(Z_1(x), Z_2(x))]$, the difference between the upper bound

$$
g(x) = \mathbb{E} [Z_1(x)] \Phi \left( \frac{u_\delta(x)}{l\theta(x)} \right) + \mathbb{E} [Z_2(x)] \left( 1 - \Phi \left( \frac{u_\delta(x)}{l\theta(x)} \right) \right) + u_{\theta}(x) \phi \left( \frac{l_\delta(x)}{u_{\theta}(x)} \right),
$$

and the exact expression for $\mathbb{E} [\max(Z_1(x), Z_2(x))]$, which is given by

$$
\Delta(x) = \mathbb{E} [Z_1(x)] \Phi \left( \frac{u_\delta(x)}{l\theta(x)} \right) - \mathbb{E} [Z_2(x)] \Phi \left( \frac{\delta(x)}{\theta(x)} \right) + u_{\theta}(x) \phi \left( \frac{l_\delta(x)}{u_{\theta}(x)} \right) - \theta(x) \phi \left( \frac{\delta(x)}{\theta(x)} \right).
$$

The inflection points of $\Phi$ and $\phi$ are 0 and $\pm 1$, respectively, so their first derivatives are bounded by $\frac{1}{\sqrt{2\pi}}$ and $\frac{1}{\sqrt{2\pi}}$, respectively. As $\frac{\delta(x)}{\theta(x)} \leq \frac{u_\delta(x)}{l\theta(x)} \leq \frac{\delta(x)}{\theta(x)}$, we have

$$
\frac{\delta(x)}{\theta(x)} \leq \Phi \left( \frac{\delta(x)}{\theta(x)} \right) + \frac{1}{\sqrt{2\pi}} \left( \frac{\delta(x)}{\theta(x)} - \frac{\delta(x)}{\theta(x)} \right).
$$

Similarly,

$$
\phi \left( \frac{l_\delta(x)}{u_{\theta}(x)} \right) \leq \phi \left( \frac{\delta(x)}{\theta(x)} \right) + \frac{1}{\sqrt{2\pi}} \left( \frac{\delta(x)}{\theta(x)} - \frac{l_\delta(x)}{u_{\theta}(x)} \right).
$$

Therefore, we have

$$
\Delta(x) \leq \frac{\delta(x)}{\sqrt{2\pi}} \left( \frac{u_\delta(x)}{l\theta(x)} - \frac{\delta(x)}{\theta(x)} \right) + \frac{\theta(x)}{\sqrt{2e\pi}} \left( \frac{\delta(x)}{\theta(x)} - \frac{l_\delta(x)}{u_{\theta}(x)} \right) + (u_{\theta}(x) - \theta(x)) \phi \left( \frac{\delta(x)}{\theta(x)} \right)
$$

$$
\leq \frac{\delta(x)}{\sqrt{2\pi}} \left( \frac{u_\delta(x)}{l\theta(x)} - \frac{l_\delta(x)}{u_{\theta}(x)} \right) + \frac{\theta(x)}{\sqrt{2e\pi}} \left( \frac{u_\delta(x)}{l\theta(x)} - \frac{l_\delta(x)}{u_{\theta}(x)} \right) + u_{\theta}(x) - \theta(x)
$$

$$
= \frac{\delta(x)}{\sqrt{2\pi}} \left( \frac{u_\delta(x)}{l\theta(x)} - \frac{l_\delta(x)}{u_{\theta}(x)} \right) + \frac{\theta(x)}{\sqrt{2e\pi}} \left( \frac{u_\delta(x)}{l\theta(x)} - \frac{l_\delta(x)}{u_{\theta}(x)} \right) + u_{\theta}(x) - \theta(x) \frac{\sqrt{2\pi}}{\sqrt{2e\pi}}
$$

(33)
First, note that if \( l_\theta(x) = \theta(x) = u_\delta(x) \) and \( l_\delta(x) = \delta(x) = u_\delta(x) \), \( \Delta(x) = 0 \), as expected. For the element of the first expression above that does not depend on the discretization, we have for every \( x \)

\[
\frac{\delta(x)}{\sqrt{2\pi}} + \frac{\theta(x)}{\sqrt{2\pi}} \leq \frac{\mathbb{E} [\max\{Z_1(x), Z_2(x)\}]}{\sqrt{2\pi}} + \frac{\mathbb{E} [\max\{Z_1(x), Z_2(x)\}]}{\sqrt{2\pi}} \leq \frac{\sqrt{2\pi} + \sqrt{2\pi}}{\sqrt{2\pi}} \mathbb{E} [\max\{Z_1(x), Z_2(x)\}] \approx 1.005 \mathbb{E} [\max\{Z_1(x), Z_2(x)\}] \quad (34)
\]

Therefore, the quality of \( g(x) \) as an estimator of \( \mathbb{E} [\max\{Z_1(x), Z_2(x)\}] \) depends on \( \frac{u_\delta(x)}{l_\delta(x)} - \frac{l_\delta(x)}{u_\delta(x)} \) and \( u_\theta(x) - \theta(x) \). We leverage this result in order to obtain an approximation guarantee delivered by the enhanced RMP for a special case of our featured machine scheduling application in Section 7.

### 4.5 Strengthening Inequalities for the RMP

We show next how estimates on \( \delta(x) \) and \( \theta(x) \) can be coupled via supervalid inequalities (SVIs), which potentially eliminate integer solutions without removing all the optimal ones (Israeli and Wood 2002). We propose SVIs of the form

\[
\delta(x) \in [\delta_h, \delta_{h+1}] \implies \theta(x) \geq \theta^h \quad \forall h = 1, \ldots, l, \quad (35)
\]

where \( \theta^h \) establishes a lower bound on \( \theta(x) \) for every \( x \) such that \( \delta(x) \in [\delta_h, \delta_{h+1}] \). First, we show in Proposition 3 how to obtain a lower bound \( \theta(\delta, z^{LB}) \) for \( \theta(x) \) for all solutions \( x \) such that \( \delta(x) = \delta \) when \( z^{LB} \) is a known lower bound for the exact problem, which can be obtained using any primal heuristic. Proposition 4 extends this result to the case in which \( \delta(x) \) belongs to a given interval. Both propositions use value \( \mu = \max_{x \in \Omega \delta} \mathbb{E}[Z_1(x)] \), which typically can be quickly computed.

**Proposition 3.** Given \( \delta \) and \( z^{LB} \), let \( \Omega(\delta, z^{LB}) \) be the set of solutions \( x \) such that \( \delta(x) = \delta \) and \( \mathbb{E}[\max\{Z_1(x), Z_2(x)\}] \geq z^{LB} \). A lower bound \( \theta(\delta, z^{LB}) \) of \( \theta(x) \) for all \( x \) in \( \Omega(\delta, z^{LB}) \) is given by

\[
\theta(\delta, z^{LB}) = \min_{\theta \geq 0} \left\{ \theta \mid \bar{u} + \theta \left( \frac{\delta}{\theta} \right) \geq z^{LB} \right\}.
\]

**Proposition 4.** Given \( \delta_1, \delta_2 \), and \( z^{LB} \) such that \( \delta_1 \leq \delta_2 \), we have \( \theta(\delta_1, z^{LB}) \leq \theta(\delta_2, z^{LB}) \).

**Corollary 1.** If \( \delta(x) \in [\delta_h, \delta_{h+1}] \) and \( \mathbb{E}[\max\{Z_1(x), Z_2(x)\}] \geq z^{LB} \), we have \( \theta(\delta_h, z^{LB}) \leq \theta(x) \).

By setting \( \theta^h = \theta(\delta_h, z^{LB}) \), we obtain the following valid inequality to RMP:

\[
x \geq \left( \theta^h \right)^2 y_h, \quad \forall h = 1, \ldots, l.
\]

These inequalities tighten the bounds by relating the choices of \( h \) for \( y_h = 1 \) to the best-known solution. Similarly to \( z^{LB} \), each \( \theta^h \) is computed at the initialization of Algorithm 1.

### 5 Performance Evaluation

We present results of an extensive evaluation on synthetic instances to assess the performance of the enhanced algorithm compared to the baseline approach. The models and algorithms are implemented in CPLEX version 12.7.1 (ILOG 2018) through the Java API. We utilize Python 3.6 with Scikit-learn functions (Pedregosa et al. 2011) to generate the random instances. All experiments are conducted on an Intel(R) Xeon(R) CPU E5-1650 v4 at 3.60GHz with 32GB of memory, and we impose a maximum time limit of 10 minutes. Source code and synthetic instances are available upon request.
The problem used for evaluation is the following:

$$\min \mathbb{E} [\max \{Z_1(x), Z_2(x)\}] \quad \text{(KP)}$$

$$\text{s.t.} \quad Z_i(x) = \sum_{j=1}^{n} Y_j x_{i,j} \quad i = 1, 2$$

$$\sum_{j=1}^{n} a_j x_{i,j} \leq b_i \quad i = 1, 2$$

$$x_{1,j} + x_{2,j} \leq 1 \quad j = 1, \ldots, n$$

$$x_{i,j} \in \{0, 1\} \quad i = 1, 2, \quad j = 1, \ldots, n.$$ 

We generate random problem instances by considering two knapsack constraints defined over a set of $n$ items. Each item has an integer weight $a_j$, which is drawn independently from $\mathcal{U}(1,19)$, and the profit associated to each item, $Y_j$, follows a normal distribution, with mean sampled also from $\mathcal{U}(15,25)$. The right-hand side of the constraints is fixed to $b_1 = b_2 = 40$. The objective function is to maximize the expected value of the knapsack with maximum profit, where the component random variables $Y_j$ correspond to the uncertain profit of each item.

To generate variances and correlations for the profits, we use Scikit-learn functions to generate a random, positive semidefinite matrix (PSD), which is then multiplied by a constant factor $\alpha$. We generate 5 instances for each configuration of $(n, \alpha)$, where $n \in \{15, 20, 25, 30\}$ and $\alpha$ in $\{50, 100, 150, 200, 250\}$, for a total of 100 instances. The $\alpha$ multiplier allows us to evaluate how sensitive the algorithms are to increasing variance and correlation.

Figure 1 presents a performance plot comparing solution times and optimality gaps for the baseline and enhanced algorithm. On the left half of the graph, the line plots report the number of instances solved by a given time limit, up to 600 seconds. On the right half of the graph, the line plots report the number of instances solved to within an optimality gap threshold, up to 18.3%.

Figure 1: Cumulative distribution plot of performance comparing the baseline and the enhanced optimization algorithms.

Figure 1 provides strong evidence for (1) the efficacy of the proposed enhancements and (2) the complexity of solving these problems to optimality. The baseline model solves only 3 instances within 10 minutes, and results in significant optimality gaps at time limit for the remaining 96 instances, at 9.3% on average and up to 18.3%. With our proposed enhancements, 55 instances are solved within 10 minutes, with 32 solved within 100 seconds, and much smaller optimality gaps for those instances that are not solved to optimality, at 0.9% on average and
3% at most. Regarding the number of cuts required to prove optimality, the baseline algorithm generates on average 270.5 cuts, while the enhanced algorithm generates 18.3 cuts, thus demonstrating the superior quality of the upper bounds obtained by the enhanced bounding function.

Table 1: Tabular comparison of solution times and gaps for baseline and enhanced algorithm.

| n  | α  | Baseline Time (s) Gap (%) | Enhanced Time (s) Gap (%) |
|----|----|---------------------------|---------------------------|
| 15 | 50 | 254(2) 1.7 | 28(5) 0.0 |
| 15 | 100 | 81(1) 2.6 | 44(5) 0.0 |
| 15 | 150 | - (0) 0.0 | 102(5) 0.0 |
| 15 | 200 | - (0) 0.0 | 49(5) 0.0 |
| 15 | 250 | - (0) 0.0 | 28(5) 0.0 |
| 20 | 50 | - (0) 0.0 | 132(5) 0.0 |
| 20 | 100 | - (0) 0.0 | 87(2) 0.5 |
| 20 | 150 | - (0) 0.0 | 109(4) 0.2 |
| 20 | 200 | - (0) 0.0 | 201(4) 0.1 |
| 20 | 250 | - (0) 0.0 | 111(5) 0.0 |
| 25 | 50 | - (0) 0.0 | - (0) 0.0 |
| 25 | 100 | - (0) 0.0 | - (0) 0.0 |
| 25 | 150 | - (0) 0.0 | - (0) 0.0 |
| 25 | 200 | - (0) 0.0 | - (0) 0.0 |
| 25 | 250 | - (0) 0.0 | - (0) 0.0 |
| 30 | 50 | - (0) 0.0 | - (0) 0.0 |
| 30 | 100 | - (0) 0.0 | - (0) 0.0 |
| 30 | 150 | - (0) 0.0 | - (0) 0.0 |
| 30 | 200 | - (0) 0.0 | - (0) 0.0 |
| 30 | 250 | - (0) 0.0 | - (0) 0.0 |

Table 1 reports more detailed information concerning the comparison of the baseline and enhanced algorithms. For each configuration of \((n, α)\) and for both the baseline and enhanced algorithms, we report the average solution time for those instances solved to optimality within 10 minutes, with the number of instances solved to optimality in superscript, and the average percent gap over all instances. This table makes it even more clear that the enhanced algorithm provides significant improvement and suggests that instances become significantly harder to solve for larger values of the variance and correlation.

6 Daily Fantasy Sports

In daily fantasy football, contests are arranged based on the starting times of each week’s slate of NFL games, and only players in those games are eligible for inclusion on a fantasy roster. Further, not all players are eligible for rosters because the fantasy scoring system only rewards points for specific tasks. Namely, only the quarterbacks (QB), kickers (K), and offensive “skill position” players—wide receivers (WR), running backs (RB), and tight ends (TE)—are eligible as single players; other players can be selected collectively as “team” defenses (DEF).

Generally the individual players receive points for gaining yardage and scoring points in the actual game (via touchdowns, extra points, two-point conversions, and field goals) and the team defense earns points by preventing the opposing team from scoring or by scoring game points itself.

DraftKings is one of the two major DFS providers. Different types of contests are offered on the betting platform, including showdows, classics, tiers, and others. We focus exclusively on showdown contests in this application. Showdown contests only include the players in a
single NFL game, and entries consist of six players, regardless of position, with one designated as the captain. The captain costs 1.5 times the normal salary and earns 1.5 times the normal points. Each player may appear no more than once on a given entry, and different contests allow a different number of entries per participant (some as high as 150). Additionally, each entry requires at least one player from each team in the contest. We focus here on smaller, high-entry-fee contests that limit each participant to two or three entries (we always enter just two).

DFS contests align directly with the focus of this paper. First, when assembling a collection of entries, a natural goal for a participant would be to have one of his entries score very high. This is because of the payout structure—most of the payout goes to the highest few entries in the competitions, as we will discuss below. This therefore can be modeled as a problem of maximizing the expected value of the higher of the two entries.

6.1 Problem definition

We model the two-entry selection problem in a showdown contest as a special case of Problem P. Let $n'$ be the number of players, and $n = 2n'$. The first $n'$ players represent standard versions of the players (“flex” in the lingo used by DraftKings), whereas the next $n'$ represent their “captain” versions, in a common order. As shown in formulation (DFS), the uncertain player scores correspond to the random variables in P. We define the feasible space $\Omega$ via the following constraints. Let $I_1$ and $I_2$ be the sets containing the players from the first and second team, respectively. For $i \in \{1, 2\}$ and $j \in [n]$, binary variable $x_{i,j}$ indicates the selection of player $j$ for roster $i$. The following constraints apply to each roster:

- Exactly 5 flex players and 1 captain must be selected:

$$\sum_{j=1}^{n'} x_{i,j} = 5, \sum_{j=n'+1}^{n} x_{i,j} = 1, \quad i = 1, 2.$$ 

- The same player cannot be selected both for a flex and the captain positions:

$$x_{i,j} + x_{i,j+n'} \leq 1, \quad i = 1, 2, \quad j = 1, \ldots, n'.$$

- At least one player from each team appears on each roster:

$$\sum_{j \in I_i} x_{i,j} \geq 1, \quad i = 1, 2.$$

We also limit the set of players under consideration to include only those with an expected score of at least 5 fantasy points. Players with expectation below this threshold are never selected by our algorithm anyway. However, as previously mentioned, the assumption of normality in player scores becomes less likely to be rejected as projected scores increase.

6.2 Data Sources and Estimation

There are several parameters that need to be estimated—in particular, $\forall j \in [n']$, parameters $\mu_j$ and $\sigma_j$, defining the normal distribution for the points scored by player $j$; and, $\forall j, j' \in [n']$, the covariance $\rho_{j,j'}$ of the performance of players $j$ and $j'$. We used a training set consisting of historical data from four NFL seasons (2014–2017) to estimate those parameters. We briefly discuss the process for each in turn below. We then compare our algorithm’s performance against a heuristic over 16 competitions in the 2018 season. More details on the evaluation can be found in the Appendix.
6.2.1 Expected Value Estimation

Due to the growth of the fantasy sports industry, estimated DFS points for players is the topic of many non-academic articles and websites (e.g., https://rotogrinders.com). For the purposes of this paper, we do not generate our own player points projections, but rather use the data from https://fantasydata.com. Their projections are consistent and reliable across the years analyzed.

6.2.2 Variance, Correlation, and Covariance Estimation

We also used the data from https://fantasydata.com in order to learn variances, correlations, and covariances for player scores using a nearest-neighbor-like algorithm. For a player $j$, his variance is estimated as the variance of the actual scores of the 50 players that share a common position with player $j$ in the training set data and whose expected value is as close as possible (in terms of squared difference) to player $j$. For example, if a RB is projected to score 20.5 fantasy points, we select the 50 RBs in the training set with projected fantasy score as close as possible to 20.5 measured by squared difference, and use the actual fantasy scores of those 50 players to calculate the variance of that player’s fantasy score.

We estimate the correlation of players $j$ and $j'$ in a similar way to our single-player variance estimates. Since we are only reporting results for showdown contests, all available player pairs are either on the same or opposing teams. For teammates, we find the 50 pairs of players and games in the training set for which the players play on the same team, played the same positions as $j$ and $j'$, and have expected values as close as possible to that of $j$ and $j'$, and we use the sample correlation of their actual game scores to estimate their correlation. For example, if a QB and WR pair are on the same team and are expected to score 30 and 15 points, respectively, we find the 50 instances in the training set of QB and WR teammates with the sum of squared differences from 30 and 15 in expected values as low as possible. We follow the same process for players on opposing teams. Finally, a small correction is made to ensure that the covariance matrix is PSD, if needed.

6.2.3 Normality Assumption

The assumption for normality of fantasy points production for players is well-grounded. Using the Shapiro-Wilk test (Shapiro and Wilk 1965), the null hypothesis of a normal distribution of player scores cannot be rejected for 76% of the QBs (by far the most valuable position) from the 2016-2018 seasons. In addition, the null hypothesis of a normal distribution cannot be rejected nearly half the time when considering players from any position over that same time frame with an expected score of at least 10 points. As the expected score increases, the likelihood of rejecting the assumption of normally distributed actual scores falls across all players. So while some of the less important positions, which score fewer points, may not follow a normal distribution, we cannot reject the normal distribution for the vast majority of the QBs and in general for the players projected to score a reasonably high number of points. These and other results regarding the normal distribution of player scores are available from the authors, by request.

6.3 Benchmark Heuristic

We used entries generated by a simple heuristic that would select a first entry with maximum expected value and a second entry that differs from the first by at least one player and otherwise again maximizes the expected value. The set of players available was the same for both our algorithm and the heuristic. Note that this heuristic is similar to the online tools many fantasy participants pay to use, such as https://www.fantasycruncher.com/.
6.4 Results

Over a collection of 16 contests in the 2018 NFL season, the net realized profit of employing our exact two-entry model using the inputs described above would have been over $5,000. On the same collection of contests, employing the heuristic and the same inputs would have resulted in a net loss of over $4,000. This application therefore highlights how important the joint decision-making over the entries is in order to get a high scoring entry.

We provide detailed results in the Appendix. One important takeaway is that the two entries selected often score on opposite sides of their expectation, showing how correlation is exploited to elevate the expected score of the maximum. As an example, in the game Redskins vs. Saints played on 10-8 the two entries selected had expectation of 104.74 and 102.74, respectively. Their actual scores were 65.55 and 140.80, where the second entry would have won the competition and resulted in substantial payout. When we use the heuristic, the two entries selected have expectation 109.51 and 108.88, and scored 78.95 and 114.20. The score of 114.20 would have resulted in a positive payout, but a very marginal one. This example shows how important optimizing exactly can be in practice.

7 Makespan Minimization with Stochastic Processing Times

We applied our theoretical and algorithmic results to \( P2|\mu \sim \mathcal{N}(\mu, \Sigma), \mathbb{E}[C_{\text{max}}] \), a machine scheduling problem involving the minimization of makespan on two parallel machines for jobs with processing times drawn from a multivariate gaussian distribution. We present two theoretical results for the case where the processing times of the jobs are uncorrelated. First, we show that \( P2|\mu \sim \mathcal{N}(\mu, \Sigma), \rho_{j,j'} = 0 | \mathbb{E}[C_{\text{max}}] \) is equivalent to \( P2|\mu = \mu_j | C_{\text{max}} \), the deterministic version of the problem where the means are used as processing times.

**Theorem 2.** \( P2|\mu \sim \mathcal{N}(\mu, \Sigma), \rho_{j,j'} = 0 | \mathbb{E}[C_{\text{max}}] \) is equivalent to \( P2|\mu = \mu_j | C_{\text{max}} \).

**Proof** Because there is no correlation between the random variables, we have \( \theta(x) = \sqrt{\sigma^2(Z_1(x)) + \sigma^2(Z_2(x))} \).

Similarly, we have \( \sigma^2(Z_i) = \sum_{j=1}^{n} \sigma_{ij}^2, i = 1, 2 \), and as \( x_{1,j} + x_{2,j} = 1 \) for each \( j \in [n] \), we have \( \theta(x) = \sqrt{\sum_{j=1}^{n} \sigma_{ij}^2} \), i.e., \( \theta(x) \) is actually constant and can simply be written as \( \theta \). If we set \( \delta_\theta = \delta_\theta(x) = \frac{\delta(x)}{\theta} \), we can rewrite \( \mathbb{E}[\max\{Z_1(x), Z_2(x)\}] \) as follows:

\[
\mathbb{E}[\max\{Z_1(x), Z_2(x)\}] = \mathbb{E}[Z_1(x)] \Phi(\delta_\theta) + \mathbb{E}[Z_2(x)] \Phi(\delta_\theta) + \theta \phi(\delta_\theta).
\]

The c.d.f. of the standard normal distribution can be written as

\[
\Phi(x) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \left[ x + \frac{x^3}{3} + \frac{x^5}{5 \cdot 3} + \ldots + \frac{x^{2n+1}}{(2n+1)!!} + \ldots \right],
\]

where \( n!! = n(n-2)(n-4)\ldots((n-1) \mod 2) + 1 \) is the double factorial of \( n \). Therefore,

\[
\mathbb{E}[\max\{Z_1(x), Z_2(x)\}] = \mathbb{E}[Z_1(x)] \left( \frac{1}{2} + \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \left[ \delta_\theta + \frac{\delta_\theta^3}{3!!} + \frac{\delta_\theta^5}{5!!} + \ldots \right] \right) + \\
\mathbb{E}[Z_2(x)] \left( \frac{1}{2} - \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \left[ \delta_\theta + \frac{\delta_\theta^3}{3!!} + \frac{\delta_\theta^5}{5!!} + \ldots \right] \right) + \frac{\theta}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}
= \frac{\mathbb{E}[Z_1(x)] + \mathbb{E}[Z_2(x)]}{2} + \frac{\theta d_\theta \frac{\delta_\theta}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \left[ \delta_\theta + \frac{\delta_\theta^3}{3!!} + \frac{\delta_\theta^5}{5!!} + \ldots \right] + \frac{\theta}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}}{2}
= \frac{\mathbb{E}[Z_1(x)] + \mathbb{E}[Z_2(x)]}{2} + \left( \frac{\theta \sqrt{2\pi}}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \left[ 1 + \frac{\delta_\theta^2}{2!!} + \frac{\delta_\theta^4}{4!!} + \frac{\delta_\theta^6}{6!!} + \ldots \right] \right). 
\]
By taking the first derivative in $\delta_\theta$, we obtain

$$\frac{d\mathbb{E}[\max\{Z_1(x), Z_2(x)\}]}{d\delta_\theta} = \left( \frac{\theta}{\sqrt{2\pi}} e^{-\frac{\delta_\theta^2}{2}} \left[ 2\delta_\theta + \frac{4\delta_\theta^3}{3!} + \frac{6\delta_\theta^5}{5!} + \ldots \right] \right) - \left( \frac{\theta}{\sqrt{2\pi}} e^{-\frac{\delta_\theta^2}{2}} \left[ \delta_\theta + \frac{\delta_\theta^3}{3!} + \frac{\delta_\theta^5}{5!} + \ldots \right] \right)$$

which is strictly positive for $\delta_\theta \geq 0$. Therefore, we conclude that $\mathbb{E}[\max\{Z_1(x), Z_2(x)\}]$ attains its minimum (maximum) for every $x$ such that $\delta_\theta(x)$ is minimum (maximum). ■

This framing of $P2|p \sim N(\mu, \Sigma), \rho_{i,j} = 0|\mathbb{E}[C_{max}]$ allows one to see that the minimization of makespan is equivalent to maximization of the idlest machine’s load. Similar arguments hold if one wishes to optimize $\mathbb{E}[\min\{Z_1(x), Z_2(x)\}]$; in the machine scheduling setting, this problem consists of minimizing the load of the idlest machine (or maximizing the load of the busiest machine), which admits a trivial optimal solution, where all jobs are assigned to one machine.

The next result leverages our results in Section 4.4 involving the tightness of the upper-bound function to show that RMP delivers solutions with constant-factor approximation guarantees for $P2|p \sim N(\mu, \Sigma), \rho_{i,j} = 0|\mathbb{E}[C_{max}]$.

**Theorem 3.** RMP delivers a 2.005-approximation for $P2|p \sim N(\mu, \Sigma), \rho_{i,j} = 0|\mathbb{E}[C_{max}]$.

**Proof** As $\theta(x) = \theta$ if variables are uncorrelated, the second term in (33) vanishes and we have

$$\Delta(x) = \frac{\delta(x)}{\sqrt{2\pi}} + \frac{\theta}{\sqrt{2\pi e^\pi}} \left( \frac{u_\delta(x) - l_\delta(x)}{\theta} \right).$$

Also, because variances play no role in this setting (from Theorem 2), one may set all discretization intervals of $\delta$ to have the same length and scale the variances such that $\theta = u_\delta(x) - l_\delta(x)$ for every $x$. Therefore, $g(x) \leq \mathbb{E}[\max\{Z_1(x), Z_2(x)\}] \left( 1 + \frac{\sqrt{2\pi e^\pi}}{\sqrt{2\pi e}} \right) \approx 2.005\mathbb{E}[\max\{Z_1(x), Z_2(x)\}]$. ■

We also conducted a computational evaluation of our algorithm using synthetic instances of $P2|p \sim N(\mu, \Sigma)|\mathbb{E}[C_{max}]$ that were generated using the same procedures as Ranjbar et al. (2012) and Stec et al. (2019); details about the generation of these instances are presented in the Appendix (see Section B). Overall, our algorithm had a solid performance, achieving an average optimality gap of approximately 0.12% within the time limit of 10 minutes; in particular, 98 out of 180 instances were solved to optimality. In a comparison with a deterministic heuristic that ignores correlations and uses the averages as the execution times of the jobs, the average improvements are roughly 1.7%; slightly larger differences were observed when the processing times had higher variances.

8 Conclusion

We investigate a class of challenging discrete stochastic optimization problems for which the objective function is given as the expected value of the maximum of two functions of the component random variables of a multivariate Gaussian distribution. We show that our problem is NP-hard and provide two real-world applications that can be modeled within our settings.

From a computational perspective, the main difficulty for solving these problems comes from the highly nonlinear expression describing the objective function, which contains the evaluation of both the c.d.f and p.d.f of a standard normal distribution with arguments given as functions of the decision variables. We propose an exact cutting-plane algorithm based on a linear function that provides upper bounds on the nonlinear objective. We investigate strengthening techniques for the bounding function and our computational results show a considerable improvement in performance as a result of the proposed techniques.
For the featured applications, computational results show that our algorithm provides a clear advantage over deterministic heuristics that do not consider correlations and covariances. We also present results of theoretical relevance for an stochastic makespan minimization problem with two machines. We prove the equivalence of $2|\mu_{i,j} \sim N(\mu_{i,j}, \sigma^2_{i,j}), \rho_{j,j} = 0|\mathbb{E}[C_{\text{max}}]$ and its deterministic counterpart and show that optimizing over our bounding function gives a 2.005-approximation for the stochastic version. Real-world settings of the problem, such as DFS, allow for scenarios where three or more functions (entries, in this case) can be selected. The resulting problems are challenging from both a mathematical and computational perspective, and investigating them is an exciting possibility for future work.

References

Gustavo Angulo, Shabbir Ahmed, and Santanu S Dey. Improving the integer l-shaped method. *INFORMS Journal on Computing*, 28(3):483–499, 2016.

Egon Balas and Robert Jeroslow. Canonical cuts on the unit hypercube. *SIAM Journal on Applied Mathematics*, 23(1):61–69, 1972.

Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen. Pearson correlation coefficient. In *Noise reduction in speech processing*, pages 37–40. Springer, 2009.

David Bergman and Jason Imbrogno. Surviving a national football league survivor pool. *Operations Research*, 65(5):1343–1354, 2017.

Dimitris Bertsimas, Karthik Natarajan, and Chung-Piaw Teo. Tight bounds on expected order statistics. *Probab. Eng. Inf. Sci.*, 20(4):667–686, October 2006. ISSN 0269-9648. doi: 10.1017/S0269964806060414. URL http://dx.doi.org/10.1017/S0269964806060414.

Keith C Brown and Deborah J Brown. Using order statistics to estimate real estate bid distributions. *Management science*, 32(3):289–297, 1986.

Bryan Clair and David Letscher. Optimal Strategies for Sports Betting Pools. *Operations Research*, 55(6):1163–1177, 2007. doi: 10.1287/opre.1070.0448. URL http://dx.doi.org/10.1287/opre.1070.0448.

Charles E Clark. The greatest of a finite set of random variables. *Operations Research*, 9(2): 145–162, 1961.

EG Coffman Jr, L Flatto, MR Garey, and RR Weber. Minimizing expected makespans on uniform processor systems. *Advances in Applied Probability*, pages 177–201, 1987.

H.A. David and H.N. Nagaraja. *Order Statistics*. Wiley Series in Probability and Statistics. Wiley, 2004.

D. S. Dimitrova, V. K. Kaishev, and Z. G. Ignatov. Ruin and deficit under claim arrivals with the order statistics property. *Methodology and Computing in Applied Probability*, September 2018.

Diane L Evans, Lawrence M Leemis, and John H Drew. The distribution of order statistics for discrete random variables with applications to bootstrapping. *INFORMS Journal on Computing*, 18(1):19–30, 2006.

Ronald L Graham, Eugene L Lawler, Jan Karel Lenstra, and AHG Rinnooy Kan. Optimization and approximation in deterministic sequencing and scheduling: a survey. In *Annals of discrete mathematics*, volume 5, pages 287–326. Elsevier, 1979.

Martin B Haugh and Raghav Singal. How to play fantasy sports strategically (and win). *Management Science*, 67(1):72–92, 2021.

John N Hooker and Greger Ottosson. Logic-based benders decomposition. *Mathematical Programming*, 96(1):33–60, 2003.

David Scott Hunter, Juan Pablo Vielma, and Tauhid Zaman. Picking winners in daily fantasy sports using integer programming. *arXiv preprint arXiv:1604.01455*, 2016.

IBM ILOG. Cplex optimization studio, 2018. URL http://www.cplex.com.
E. Israeli and R. K. Wood. Shortest-path network interdiction. *Networks*, 40(2):97–111, 2002.

Edward H. Kaplan and Stanley J. Garstka. March Madness and the Office Pool. *Management Science*, 47(3):pp. 369–382, 2001. ISSN 00251909. URL http://www.jstor.org/stable/2661505.

Vasileios M. Koutras and Markos V. Koutras. Exact distribution of random order statistics and applications in risk management. *Methodology and Computing in Applied Probability*, Aug 2018. ISSN 1573-7713.

Gilbert Laporte and François V Louveaux. The integer l-shaped method for stochastic integer programs with complete recourse. *Operations research letters*, 13(3):133–142, 1993.

Garth P. McCormick. Computability of global solutions to factorable nonconvex programs: Part i — convex underestimating problems. *Mathematical Programming*, 10(1):147–175, Dec 1976. ISSN 1436-4646.

S Thomas McCormick, M Rammohan Rao, and Giovanni Rinaldi. Easy and difficult objective functions for max cut. *Mathematical programming*, 94(2-3):459–466, 2003.

Saralees Nadarajah and Samuel Kotz. Exact distribution of the max/min of two gaussian random variables. *IEEE Transactions on very large scale integration (VLSI) systems*, 16(2):210–212, 2008.

José Nino-Mora. Stochastic scheduling. *Encyclopedia of optimization*, 5:367–372, 2009.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Michael L Pinedo. *Planning and scheduling in manufacturing and services*. Springer, 2005.

Mohammad Ranjbar, Morteza Davari, and Roel Leus. Two branch-and-bound algorithms for the robust parallel machine scheduling problem. *Computers & Operations Research*, 39(7):1652–1660, 2012.

Skipper Seabold and Josef Perktold. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*, 2010.

Suvrajeet Sen and Hanif D Sherali. Decomposition with branch-and-cut approaches for two-stage stochastic mixed-integer programming. *Mathematical Programming*, 106(2):203–223, 2006.

S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611, Dec. 1965.

Richard Stec, Antonin Novak, Premysl Sucha, and Zdenek Hanzalek. Scheduling jobs with stochastic processing time on parallel identical machines. In *International Joint Conference on Artificial Intelligence*, pages 5628–5634. AAAI Press, 2019.

Hong-Chuan Yang and Mohamed-Slim Alouini. *Distributions of order statistics*, page 40–71. Cambridge University Press, 2011.
A Daily Fantasy Sports Estimation Details and Details of Results

Further details of Section 6 are shown below.

A.1 Expected Value Estimation

The community of DFS participants puts tremendous resources into calculating reliable and accurate estimates for how a player will perform. Although there is potentially room to improve upon published estimates, for this paper we use estimated player data from https://fantasydata.com, which is updated frequently for upcoming games and contains historical NFL game data since 2014, including projected fantasy points for players, actual fantasy points for players, and player salaries on the DraftKings platform, all for a monthly fee.

The projected points estimates used for this paper are the final https://fantasydata.com estimates set just before game time (changes leading up to game time may occur due to weather, injury updates, or a myriad of other reasons). This data therefore provides $\mu_j$, for $j = 1, \ldots, p_0$, and, because $E[1.5X] = 1.5E[X]$, we have $\mu_j = 1.5\mu_{j-p'}$, for $j \geq p' + 1$. These https://fantasydata.com estimates are consistent, and therefore reliable for future use, from the 2014-2017 seasons. We fail to reject the null hypothesis that the difference between the https://fantasydata.com estimates and the actual fantasy scores across all players (with projected points above 5) and all games from 2014-2017 is equal to 0 (the 95% confidence interval on those 11,785 observations is [-0.211, 0.071]).

A.2 Variance, Correlation, and Covariance Estimation

Due to the nature of the sport itself and the way fantasy scoring works, players on the same (or opposing) teams often have correlated scoring. For example, if a QB throws a touchdown to a WR, both players receive a number of fantasy points (and the opposing DEF would possibly lose some). But since that QB had to have completed the pass to someone, fantasy points for individual players rarely exist in a vacuum. Rather, we would expect that players at certain positions would have significantly correlated scores with teammates, and even opponents, at other positions. Therefore, given the heavily skewed payoff structure of most DraftKings showdown contests, when trying to maximize the expected value of the maximum score of the entries, participants must take into account these correlations. The DFS betting community is well aware of this strategy of choosing teammates (or opponents) whose scores should correlate by design; the strategy is termed “stacking.” It should also be noted that opponents may see their actual scores correlate even if the players are not on the field at the same time due to the nature of the game of football. For example, the actual fantasy scores of the QBs on opposing teams (who are never on the field at the same time) are positively correlated because higher-scoring NFL games tend to generate more fantasy points for the QBs. As one QB scores fantasy points (as his on-field team scores lots of points in the game itself), the other team tends to throw more to try and catch up, and this tends to raise the opponent QB’s fantasy scores as a result.

Once the correlations for players on the same and opposing teams have been estimated as previously described, one can then set the correlation to 0 if the significance of the Pearson’s correlation test (Benesty et al. (2009)) is above some threshold $\eta$. Further, there are instances in which the correlation identifies values that are incompatible because of the way the estimation is done and the actual scores achieved by certain player pairs. Formally, the estimated covariance matrix $\Sigma$ for a multivariate normal distribution must be PSD. Correlation estimates, multiplied by the corresponding standard deviations, are found only to provide an estimate for $\Sigma$, and the estimation procedure previously described often yields correlations for which $\Sigma$ is not PSD. There are several ways to correct for this. One is to use packages available in common statistical software to adjust the covariances, like the function $\text{cov-nearest}$ in Python’s statsmodels module (Seabold and Perktold (2010)). Another is to dampen the correlations by a fixed constant (i.e., set $\rho_{j,j'} := \frac{\rho_{j,j'}}{\chi}$, with $\chi > 0$, for all $j,j'$). Yet another way is to apply
the PSD condition and, whenever a solution is found for which $\theta < 0$, redefine $\theta := 0$. We tried several procedures, though our reported results use the \texttt{cov nearest} function, using $\eta = 0.25$ and \texttt{cov nearest} for correcting $\Sigma$; i.e., setting to 0 any correlation whose $p$-value is above 0.25.

### A.3 Results

In Table 2, we report results from actual competitions on the DraftKings platform that would have been obtained on 16 showdown contests from the 2018 season by our exact algorithm with the parameters estimated as described in the Appendix. In Table 3, we report the same information for the heuristic. These 16 specific contests were selected because we have actual output from real contests for these competitions. They include 2-entry and 3-entry competitions.

The content of the first seven columns of Tables 2 and 3 is as follows. The first column indicates the game and the date when it took place. The second indicates the total number of entries in the contest. The third indicates the price paid by participants per entry. The fourth and fifth indicate the results obtained by showing the winnings and the profit that the entries identified by the algorithm would have yielded, respectively. Columns six and seven indicate the best score and the minimum score for entries in the money (i.e., those obtaining any payout from the competition). Columns eight and nine contain the expected values of the first and the second entry obtained, respectively. Finally, columns ten and eleven represent the actual scores of the selected entries. The values in columns ten and eleven are \textbf{bolded} if the entry would have finished in a position receiving payout in that particular contest. The total entry fees, winnings, and profit for each method are computed in the last row of each table. Note, however, that the money won shown in Tables 2 and 3 is an approximation in the sense that if our two entries from the algorithms in this paper were \textit{actually} included in the contests, two other entries that were in the contest necessarily would have been excluded, since these contests filled up to capacity. If, for example, the two “eliminated” entries were top-scoring ones, the payoffs for the remaining entries, and ours, would necessarily be the same or larger. There is no way to account for this game of elimination in a reasonable way, so we report results as if our entries were simply added to the contest without changing any other details about other entries and the payout structure.

There are two initial takeaways from Tables 2 and 3. First, the contests differ in cost and number of entries considerably, though the most common version of this contest features 100 total entries at a cost of $444 per entry. Second, the scores needed to win, or even place in the money, differ tremendously by game. Namely, higher-scoring NFL games result in more fantasy points for the involved players; for instance, in the game Vikings vs. Rams played on 9-27, which finished with a score of 38-31, the \textit{minimum payout score} would have been the \textit{winner score} in every single other contest shown. Because of the vast discrepancy in contest results, it is difficult to say \textit{a priori} what score is going to be necessary to win or place in a given contest. The comparison of overall results across the methods reveals the power of our algorithm, so the more important takeaway is the following: our exact algorithm would result in a positive return of over 50% and the heuristic in a loss of nearly 50%.

In nearly all contests, across both specifications, the winner score is significantly higher than the expected value scores of our entries (the only major exception being the contest featuring the game Falcons vs. Saints played on 11-22). The players on the winning entry, as a whole, significantly outperform their projections in nearly every contest. Something akin to the heuristic method is what we expect many participants in these contests are using as their DFS strategy. Although it unsurprisingly outperforms our entries in \textit{expected value}, its actual results are inferior. As we have observed in our experiments with the other featured applications, in general, the maximum expected value entry does not necessarily belong to an optimal configuration of entries, and it also will not necessarily result in a payout, so bettors need to exploit joint decision-making across multiple entries in order to elevate their actual scores above the threshold needed to win money.

Table 4 further illustrates the power of our algorithm for $P$. We again compare to the heuristic by evaluating Expression (4) using the solution that the heuristic obtains with the
same covariance matrix we use. In this case, “EV” represents the single highest expected value entry, “AV” represents the single highest actual score of the entries, and “OF”, the evaluation of the objective function at the solution obtained. Notice that while EV for the heuristic is in most cases higher than EV for the exact approach, the evaluation of the objective function is considerably higher for the exact method in all the instances. The last row of the table considers the average values above it. Over these contests, the average EV is 1.19 points higher for the heuristic but the average objective function is 6.68 points higher for the exact method, once again showing that our approach can exploit correlation in a way that a common heuristic cannot.
Table 2: Results from real-world betting scenarios using \( P \) and corrected covariance

| Game                                      | Entrants | Entry Fee | Winnings | Profit | Winner Score | Min. Pay Score | EV E1 | EV E2 | AV E1 | AV E2 |
|-------------------------------------------|----------|-----------|----------|--------|--------------|----------------|-------|-------|-------|-------|
| Jets vs. Browns on 9-20                   | 100      | $600      | $0       | -666   | 96.51        | 67.96          | 77.97 | 76.55 | 64.07 | 42.02 |
| Seahawks vs. Cowboys on 9-23              | 1,189    | $8        | $0       | -8     | 109.78       | 76.78          | 79.63 | 77.97 | 75.22 | 61.57 |
| Vikings vs. Rams on 9-27                  | 83       | $800      | $0       | -800   | 205.28       | 175.13         | 98.50 | 95.90 | 159.28 | 150.13 |
| Ravens vs. Steelers on 9-30               | 294      | $40       | $0       | -40    | 111.01       | 96.73          | 93.40 | 89.16 | 84.66 | 86.17 |
| Broncos vs. Chiefs on 10-1                | 69       | $800      | $0       | -800   | 110.36       | 90.69          | 106.88 | 106.46 | 79.69 | 71.50 |
| Colts vs. Patriots on 10-4                | 100      | $888      | $700     | -188   | 148.89       | 131.85         | 93.13 | 92.75 | 114.19 | 132.34 |
| Redskins vs. Saints on 10-8               | 100      | $888      | $10,000  | 9,112  | 134.97       | 99.47          | 104.74 | 102.74 | 65.55 | 140.80 |
| Giants vs. Falcons on 10-22               | 100      | $888      | $2,700   | 1,812  | 143.56       | 120.70         | 103.03 | 101.08 | 133.76 | 123.26 |
| Raiders vs. 49ers on 11-1                 | 100      | $888      | $1,200   | 312    | 95.81        | 78.61          | 77.54 | 77.35 | 86.72 | 43.61 |
| Panthers vs. Steelers on 11-8             | 100      | $888      | $0       | -888   | 149.93       | 118.70         | 103.03 | 99.76 | 107.00 | 115.49 |
| Giants vs. 49ers on 11-12                 | 100      | $888      | $0       | -888   | 120.60       | 95.18          | 91.00 | 90.61 | 81.52 | 81.37 |
| Packers vs. Seahawks on 11-15             | 100      | $888      | $0       | -888   | 126.53       | 114.43         | 91.12 | 88.11 | 107.43 | 96.68 |
| Steelers vs. Jaguars on 11-18             | 70       | $66       | $0       | -66    | 115.86       | 96.49          | 92.96 | 88.87 | 94.45 | 81.31 |
| Falcons vs. Saints on 11-22               | 151      | $150      | $450     | 300    | 122.22       | 106.52         | 130.90 | 129.62 | 110.94 | 114.12 |
| Redskins vs. Eagles on 12-3                | 100      | $888      | $0       | -888   | 111.36       | 84.29          | 88.60 | 83.49 | 82.74 | 64.14 |
| Jaguars vs. Titans on 12-6                 | 882      | $40       | $0       | -40    | 134.60       | 74.83          | 78.30 | 77.25 | 54.93 | 47.83 |

TOTALS                                     | $9,674   | $15,050   | $5,376   |
| Game                                | Entrants | Entry Fee | Winnings | Profit | Winner Score | Min. Pay Score | EV E1   | EV E2   | AV E1   | AV E2   |
|-------------------------------------|----------|-----------|----------|--------|--------------|----------------|---------|---------|---------|---------|
| Jets vs. Browns on 9-20             | 100      | $666      | $0       | -$666  | 96.51        | 67.96          | 78.79   | 78.50   | 63.82   | 60.81   |
| Seahawks vs. Cowboys on 9-23       | 1,189    | $8        | $0       | -$8    | 109.78       | 76.78          | 80.93   | 80.91   | 64.20   | 66.70   |
| Vikings vs. Rams on 9-27           | 83       | $800      | $0       | -$800  | 205.28       | 175.13         | 98.50   | 97.91   | 159.28  | 147.68  |
| Ravens vs. Steelers on 9-30        | 294      | $40       | $0       | -$40   | 111.01       | 96.73          | 94.07   | 93.40   | 73.80   | 84.66   |
| Broncos vs. Chiefs on 10-1          | 69       | $800      | $1,200   | $400   | 110.36       | 90.69          | 109.91  | 109.85  | 64.20   | 66.70   |
| Colts vs. Patriots on 10-4         | 100      | $888      | $0       | -$888  | 148.09       | 131.85         | 94.25   | 93.59   | 119.44  | 112.11  |
| Redskins vs. Saints on 10-8        | 100      | $888      | $1,000   | $112   | 134.97       | 99.67          | 109.51  | 108.88  | 78.95   | 114.20  |
| Giants vs. Falcons on 10-22        | 100      | $888      | $2,700   | $1,812 | 143.56       | 120.70         | 103.03  | 103.03  | 133.76  | 123.96  |
| Raiders vs.49ers on 11-1           | 100      | $888      | $0       | -$888  | 95.81        | 78.61          | 80.21   | 80.21   | 71.49   | 76.79   |
| Panthers vs. Steelers on 11-8      | 100      | $888      | $0       | -$888  | 149.93       | 118.70         | 103.93  | 103.59  | 118.50  | 99.00   |
| Giants vs. 49ers on 11-12          | 100      | $888      | $0       | -$888  | 120.60       | 95.18          | 93.45   | 93.30   | 80.72   | 88.72   |
| Packers vs. Seahawks on 11-15      | 100      | $888      | $0       | -$888  | 126.53       | 114.43         | 92.54   | 92.26   | 105.43  | 105.08  |
| Steelers vs. Jaguars on 11-18      | 70       | $66       | $0       | -$66   | 115.86       | 96.49          | 95.26   | 94.92   | 80.25   | 82.67   |
| Falcons vs. Saints on 11-22        | 151      | $150      | $400     | $250   | 122.22       | 106.52         | 130.90  | 130.02  | 110.94  | 113.06  |
| Redskins vs. Eagles on 12-3         | 100      | $888      | $0       | -$888  | 111.36       | 84.29          | 88.70   | 88.60   | 78.94   | 82.74   |
| Jaguars vs. Titans on 12-6         | 882      | $40       | $0       | -$40   | 134.60       | 74.83          | 79.53   | 79.38   | 49.17   | 36.73   |

**Table 3: Results from real-world betting scenarios using heuristic**

| Game                                | Entrants | Entry Fee | Winnings | Profit | Winner Score | Min. Pay Score | EV E1   | EV E2   | AV E1   | AV E2   |
|-------------------------------------|----------|-----------|----------|--------|--------------|----------------|---------|---------|---------|---------|
| **TOTALS**                          |          |           | $9,674   | $5,300 | -$4,374      |                |         |         |         |         |
Table 4: Comparing the exact approach with the benchmark heuristic on 16 DFS competitions.

| Game                                      | Exact | Heuristic |
|-------------------------------------------|-------|-----------|
| Jets vs. Browns on 9-20                   | 77.97 | 93.30     |
| Seahawks vs. Cowboys on 9-23              | 79.63 | 96.81     |
| Vikings vs. Rams on 9-27                  | 98.50 | 110.91    |
| Ravens vs. Steelers on 9-30               | 93.40 | 104.07    |
| Broncos vs. Chiefs on 10-1                 | 106.88| 109.91    |
| Colts vs. Patriots on 10-4                 | 93.13 | 94.25     |
| Redskins vs. Saints on 10-8                | 104.74| 109.51    |
| Giants vs. Falcons on 10-22               | 103.03| 103.03    |
| Raiders vs. 49ers on 11-1                 | 77.54 | 80.21     |
| Panthers vs. Steelers on 11-8             | 103.03| 103.93    |
| Giants vs. 49ers on 11-12                 | 91.00 | 93.45     |
| Packers vs. Seahawks on 11-15             | 91.12 | 92.54     |
| Steelers vs. Jaguars on 11-18             | 92.96 | 95.26     |
| Falcons vs. Saints on 11-22               | 130.90| 130.90    |
| Redskins vs. Eagles on 12-3                | 88.60 | 88.70     |
| Jaguars vs. Titans on 12-6                | 78.30 | 79.53     |
| AVERAGES                                  | 94.65 | 95.84     |

B Instances used in the Experiments of the Makespan Minimization Application

We evaluate our algorithms for $P2|p \sim N(\mu, \Sigma)|\mathbb{E}[C_{\text{max}}]$ using a set of synthetic instances that follow the same generation procedure proposed by Ranjbar et al. (2012) and Stec et al. (2019). These papers consider a related objective of maximizing the number of jobs completed within a pre-specified time limit and assume pairwise independence between the processing time of the jobs.

We generated instances with $n = 15, 20, 25,$ and $30$, where $n$ denotes the number of jobs. The processing time of each job $j$ follows a normal distribution $\mathcal{N}(\mu_j, \sigma_j^2)$, where $\mu_j$ is a positive value drawn from $\mathcal{N}(20, 9)$ and $\sigma_j^2$ is drawn from $\mathcal{U}(0, 0.1 \cdot \mu_j^2 \cdot \eta)$ for $\eta \in \{0.25, 0.5, 0.75\}$. To generate covariance matrices, for each job, we assign uniformly at random a number from $\{1, 2, 3\}$, representing one of three clusters, and we assume that the execution time of jobs assigned to the same cluster have a correlation of one; otherwise, jobs in different clusters have correlation of zero. These correlation matrices are suitable for scenarios where processing times are strongly affected by features shared by several jobs; for example, processing times of computational experiments originating from a same batch are expected to be highly correlated. We generated 5 instances for each combination of values of $n$ and $\eta$, so our data set for this application consists of 180 instances.

C Adapting the Optimization Algorithm for Minimization

In the minimization case, the objective function is still given by expression (4); however, in contrast to the maximization case, our RMP should yield lower bounds on the objective. Proposition 5 presents our proposed enhanced lower-bounding function $g$.

**Proposition 5.** For every $x \in \Omega$,

$$\mathbb{E}[Z_1(x)] \Phi \left( \frac{l_{\phi}(x)}{u_{\phi}(x)} \right) + \mathbb{E}[Z_2(x)] \left( 1 - \Phi \left( \frac{l_{\phi}(x)}{u_{\phi}(x)} \right) \right) + l_{\phi}(x) \phi \left( \frac{u_{\phi}(x)}{l_{\phi}(x)} \right) \leq \mathbb{E} \left[ \max\{Z_1(x), Z_2(x)\} \right]. \quad (37)$$

**Proof** The proof follows the same arguments as the proof of Proposition 2.

Using this lower bounding function it is straightforward to modify the RMP to yield lower bounds. The modified cutting algorithm for the minimization case is as follows:
Suppose we are given $d$ intervals $\{[\theta_q^2, \theta_{q+1}^2]\}_{q=1}^d$ with $\theta(x)^2 \in [\theta_1^2, \theta_{d+1}^2]$ for every $x \in \Omega$. Furthermore, let $\theta_q$ and $\theta_{q+1}$ denote a lower and upper bound of $\theta(x)$, respectively, for $\theta(x)^2 \in [\theta_q^2, \theta_{q+1}^2]$. Using these intervals we construct the following RMP formulation, where binary variables $w_q$, $q = 1, \ldots, d$ indicate which interval $\theta(x)^2$. Variable $u_1$ ($u_2$) denotes $\mathbb{E}[Z_1(x)]$ ($\mathbb{E}[Z_2(x)]$) and $s$ ($s'$) represents $\theta(x)^2$ ($u_q(x)$). Binary variable $v_{i,j,j'}$ takes a value of 1 iff $x_{i,j} = x_{i,j'} = 1$. Similarly, $r_{j,j'}$ equals 1 iff $x_{i,j} = x_{i,j'} = 1$. We formulate the baseline RMP as:

$$\begin{align*}
\text{max} & \quad u_1 + s' \frac{1}{\sqrt{2\pi}} \\
\text{s.t.} & \quad u_1 = \sum_{j=1}^n \mu_j x_{1,j}; \quad u_2 = \sum_{j=1}^n \mu_j x_{2,j}; \quad u_1 \geq u_2 \\
& \quad s = 2 \sum_{i=1}^n \left( \sum_{j=1}^n \sigma_i^2 x_{i,j} + 2 \sum_{1 \leq j < j' \leq n} \text{cov}(Y_j, Y_{j'}) v_{i,j,j'} \right) - 2 \sum_{j=1}^n \sum_{j'=1}^n \text{cov}(Y_j, Y_{j'}) r_{j,j'} \\
& \quad v_{i,j,j'} \leq x_{i,j'}; \quad v_{i,j,j'} \leq x_{i,j'} \quad \forall j, j' \in \{1, \ldots, n\}, i \in \{1, 2\} \\
& \quad v_{i,j,j'} \geq x_{i,j} + x_{i,j'} - 1 \quad \forall j, j' \in \{1, \ldots, n\}, i \in \{1, 2\} \\
& \quad r_{j,j'} \leq x_{1,j'}; \quad r_{j,j'} \leq x_{2,j'} \quad \forall j, j' \in \{1, \ldots, n\} \\
& \quad r_{j,j'} \geq x_{1,j} + x_{2,j'} - 1 \quad \forall j, j' \in \{1, \ldots, n\} \\
& \quad \sum_{q=1}^d w_q = 1; \quad s' = \sum_{q=1}^d \theta_{q+1} w_q \\
& \quad \theta_q^2 w_q \leq s \leq \theta_{q+1}^2 + \theta_{d+1}^2 (1 - w_q) \quad q = 1, \ldots, d \\
& \quad v \in \{0, 1\}^{n \times n}; \quad r \in \{0, 1\}^{n \times n}; \quad w \in \{0, 1\}^d; \quad x \in \Omega.
\end{align*}$$

The objective function (38) maximizes the upper bounding function defined by Proposition 1. Constraints (39) define the $u$-variables according to equation (1) and impose the symmetry breaking condition $u_1 \geq u_2$. Constraint (40) imposes $s = \theta(x)^2$ as described by equation (14). Constraints (41)–(44) are the McCormick linearization constraints. Constraints (45) ensure that exactly one interval is chosen for $\theta(x)^2$ and set $s'$ equal to the upper bound of $\theta(x)$ for the interval that $\theta(x)^2$ belongs to. Constraints (46) select the right interval for $\theta(x)^2$. Constraints (47) define the domains of the variables appropriately.

## E Proofs of Propositions for the Strong Valid Inequalities

The statements of the propositions are reproduced here for ease of reference.
Proposition 2. Given $\delta$ and $z^{LB}$, let $\Omega(\delta, z^{LB})$ be the set of solutions $x$ such that $\delta(x) = \delta$ and $E[\max\{Z_1(x), Z_2(x)\}] \geq z^{LB}$. A lower bound $\theta(\delta, z^{LB})$ of $\theta(x)$ for all $x \in \Omega(\delta, z^{LB})$ is given by

$$
\theta(\delta, z^{LB}) = \min_{\theta \geq 0} \left\{ \theta \mid \bar{u} + \theta \phi \left( \frac{\delta}{\theta(x)} \right) \geq z^{LB} \right\}.
$$

Proof Assume by contradiction that there exists an $x \in \Omega(\delta, z^{LB})$ such that $\theta(x) < \theta(\delta, z^{LB})$. Since $E[\max\{Z_1(x), Z_2(x)\}] \geq z^{LB}$ we have that

$$
E[Z_1(x)] \Phi \left( \frac{\delta}{\theta(x)} \right) + E[Z_2(x)] \Phi \left( \frac{-\delta}{\theta(x)} \right) + \theta(x) \phi \left( \frac{\delta}{\theta(x)} \right) \geq z^{LB}.
$$

Because $\bar{u} = \max_{x \in \Omega} E[Z_1(x)]$, we have

$$
E[Z_1(x)] \Phi \left( \frac{\delta}{\theta(x)} \right) + E[Z_2(x)] \Phi \left( \frac{-\delta}{\theta(x)} \right) \leq \bar{u}.
$$

As a result,

$$
\bar{u} + \theta(x) \phi \left( \frac{\delta}{\theta(x)} \right) \geq z^{LB},
$$

thus contradicting the optimality of $\theta(\delta, z^{LB})$. ■

Proposition 3. Given $\delta_1$, $\delta_2$, and $z^{LB}$ such that $\delta_1 \leq \delta_2$, we have $\theta(\delta_1, z^{LB}) \leq \theta(\delta_2, z^{LB})$.

Proof For $\theta \geq 0$ and fixed $\delta$, both $\theta$ and $\phi \left( \frac{\delta}{\theta} \right)$ are continuous and non-decreasing functions of $\theta$, so we have that $\theta \phi \left( \frac{\delta}{\theta} \right)$ is also a continuous and non-decreasing function of $\theta$. Therefore, at optimality we have $\theta(\delta_1, z^{LB}) \phi \left( \frac{\delta_1}{\theta(\delta_1, z^{LB})} \right) = \theta(\delta_2, z^{LB}) \phi \left( \frac{\delta_2}{\theta(\delta_2, z^{LB})} \right) = z^{LB} - \bar{u}$. Assume by contradiction that $\theta(\delta_2, z^{LB}) < \theta(\delta_1, z^{LB})$; if this holds, we must also have $\phi \left( \frac{\delta_1}{\theta(\delta_1, z^{LB})} \right) < \phi \left( \frac{\delta_2}{\theta(\delta_2, z^{LB})} \right)$, and, consequently, that $\theta(\delta_2, z^{LB}) \phi \left( \frac{\delta_2}{\theta(\delta_2, z^{LB})} \right) < \theta(\delta_1, z^{LB}) \phi \left( \frac{\delta_1}{\theta(\delta_1, z^{LB})} \right)$, thus contradicting the optimality of $\theta(\delta_1, z^{LB})$. ■

F Parameters Used in the Computational Implementation

The following are the parameters used in our implementation for the exact algorithm, after fine-tuning with a small subset of instances:

- The number of intervals for the discretizations of $\theta(x)$ and $\delta(x)$ are $d = 25$ and $l = 15$ for the synthetic instances, $d = l = 50$ for the machine scheduling application, and $d = 50$ and $l = 10$ for the DFS application.
- To obtain upper bounds on $\theta(x)$ and $\delta(x)$ we use the MILP described in Section 4 with a time limit of 120 seconds.
- The large $M$ constant used for constraints (29) is the expected makespan of assigning all the jobs to one machine and leaving the other one idle for the scheduling problem. For the other applications we use the maximum profit (score) that can be obtained by a single knapsack (team entry), which we exactly compute via an MILP that solves in less than a second over our testbed.
- The large $M$ constant used for constraints (30) is $\theta^2_{d+1} \left( \frac{1}{\sqrt{2\pi}} \right)$. Recall that $\theta^2_{d+1}$ is an upper bound on $\theta(x)$.
- The primal heuristic used to obtain a starting feasible solution and lower bound on the objective is as follows. We use the RMP formulation with an additional constraint that requires the model to only select items within the top 90% expected profit or score. We run this MILP for 60 seconds and recover the best feasible solution found within the time limit.
• We include the proposed SVIs for solving the synthetic instances and the fantasy football problem. We do not include them in the scheduling application.

• We stop the algorithm with a tolerance of 0.001, i.e., once \( \frac{UB-LB}{UB} < 0.001 \) for maximization or \( \frac{UB-LB}{LB} < 0.001 \) for minimization.