

Moment Problems and Semidefinite Optimization

Dimitris Bertsimas ^{*} Ioana Popescu [†] Jay Sethuraman [‡]

April, 2000

Abstract

Problems involving moments of random variables arise naturally in many areas of mathematics, economics, and operations research. How do we obtain optimal bounds on the probability that a random variable belongs in a set, given some of its moments? How do we price financial derivatives without assuming any model for the underlying price dynamics, given only moments of the price of the underlying asset? How do we obtain stronger relaxations for stochastic optimization problems exploiting the knowledge that the decision variables are moments of random variables? Can we generate near optimal solutions for a discrete optimization problem from a semidefinite relaxation by interpreting an optimal solution of the relaxation as a covariance matrix? In this paper, we demonstrate that convex, and in particular semidefinite, optimization methods lead to interesting and often unexpected answers to these questions.

^{*}Dimitris Bertsimas, Sloan School of Management and Operations Research Center, Rm E53-363, MIT, Cambridge, MA 02139, dbertsim@mit.edu. Research supported by NSF grant DMI-9610486 and the Singapore-MIT alliance.

[†]Ioana Popescu, Decision Sciences, INSEAD, Fontainebleau Cedex 77305, FRANCE. ioana.popescu@insead.fr.

[‡]Jay Sethuraman, Columbia University, Department of IEOR, New York, NY 10027, jayc@columbia.edu.

1 Introduction

Problems involving moments of random variables arise naturally in many areas of mathematics, economics, and operations research. Let us give some examples that motivate the present paper.

Moment problems in probability theory

The problem of deriving bounds on the probability that a certain random variable belongs in a set, given information on some of the moments of this random variable, has a rich history, which is very much connected with the development of probability theory in the twentieth century. The inequalities due to Markov, Chebyshev and Chernoff are some of the classical and widely used results of modern probability theory. Natural questions arise, however:

- (a) *Are such bounds “best possible,” i.e., do there exist distributions that match them? A concrete and simple question in the univariate case: Is the Chebyshev inequality “best possible”?*
- (b) *Can such bounds be generalized in multivariate settings?*
- (c) *Can we develop a general theory based on optimization methods to address moment problems in probability theory?*

Moment problems in finance

A central question in financial economics is to find the price of a derivative security given information on the underlying asset. This is exactly the area of the 1997 Nobel prize in economics to Robert Merton and Myron Scholes. Under the assumption that the price of the underlying asset follows a geometric Brownian motion and using the no-arbitrage assumption, the Black-Scholes formula provides an explicit and insightful answer to this question. Natural questions arise, however. Making no assumptions on the underlying price dynamics, but only using the no-arbitrage assumption:

- (a) *What are the best possible bounds for the price of a derivative security based on the first and second moments of the price of the underlying asset?*

- (b) *How can we derive optimal bounds on derivative securities that are based on multiple underlying assets, given the first two moments of the asset prices and their correlations?*
- (c) *Conversely, given observable option prices, what are the best bounds that we can derive on the moments of the underlying asset?*
- (d) *Finally, given observable option prices, what are the best bounds that we can derive on prices of other derivatives on the same asset?*

Moment problems in stochastic optimization

Scheduling a multiclass queueing network, is a central problem in stochastic optimization. Queueing networks represent dynamic and stochastic generalizations of job shops, and have been used in the last thirty years to model communication, computer and manufacturing systems. The central optimization problem is to find a scheduling policy that optimizes a performance cost function $\mathbf{c}'\mathbf{x} + \mathbf{d}'\mathbf{y}$, where $\mathbf{x} = (x_1, \dots, x_N)$, x_j is the mean number of jobs of class j , $\mathbf{y} = (y_1, \dots, y_N)$, and y_j is the second moment of the number of jobs of class j , and \mathbf{c} , \mathbf{d} are N -vectors of nonnegative constants. The design of optimal policies is *EXPTIME*-hard (Papadimitiou and Tsitsiklis [26]), i.e., it provably requires exponential time, as $P \neq EXPTIME$. A natural question that arises:

Can we find strong lower bounds efficiently, exploiting the fact that the performance vectors represent moments of random variables?

Moment problems in discrete optimization

The development of semidefinite relaxations in recent years represents an important advance in discrete optimization. In several problems, semidefinite relaxations are provably closer to the discrete optimization solution value (Goemans and Williamson [13] for the maxcut problem for example) than linear ones. The proof of closeness of the semidefinite relaxation to the discrete optimization solution value involves a randomized argument that exploits the geometry of the semidefinite relaxation. A key question arises:

Is there a general method of generating near optimal integer solutions starting from an optimal solution of the semidefinite relaxation?

We will see that the interpretation of the solution of the semidefinite relaxation as a covariance matrix for a collection of random variables leads to such a method and connects moment problems and discrete optimization.

The central message in this survey paper is to demonstrate that convex and, in particular, semidefinite, optimization methods give interesting and often unexpected answers to moment problems arising in probability, economics, and operations research. We also report new computational results in the area of stochastic optimization that show the effectiveness of semidefinite relaxations.

The key connection

The key connection between moment problems and semidefinite optimization is centered in the notion of a *feasible moment sequence*. Let $\mathbf{k} = (k_1, \dots, k_n)$ be a vector of nonnegative integers.

Definition 1 *A sequence $\bar{\sigma} : (\sigma_{\mathbf{k}})_{k_1 + \dots + k_n \leq k}$ is a feasible (n, k, Ω) -moment vector (or sequence), if there is a multivariate random variable $\mathbf{X} = (X_1, \dots, X_n)$ with domain $\Omega \subseteq R^n$, whose moments are given by $\bar{\sigma}$, that is $\sigma_{\mathbf{k}} = E[X_1^{k_1} \dots X_n^{k_n}]$, $\forall k_1 + \dots + k_n \leq k$. We say that any such multivariate random variable \mathbf{X} has a $\bar{\sigma}$ -feasible distribution and denote this as $\mathbf{X} \sim \bar{\sigma}$.*

We denote by $\mathcal{M} = \mathcal{M}(n, k, \Omega)$ the set of feasible (n, k, Ω) -moment vectors.

The univariate case

For the univariate case ($n = 1$), the problem of deciding if $\bar{\sigma} = (M_1, M_2, \dots, M_k)$ is a feasible $(1, k, \Omega)$ -moment vector is the classical moment problem, which has been completely characterized by necessary and sufficient conditions (see Karlin and Shapley [19], Akhiezer [1], Siu, Sengupta and Lind [27] and Kemperman [20]).

Theorem 1 (a) (Nonnegative random variables) *The vector (M_1, \dots, M_{2n+1}) is a feasible $(1, 2n+1, R^+)$ -moment sequence if and only if the following matrices are semidefi-*

nite:

$$\mathbf{R}_{2n} = \begin{pmatrix} 1 & M_1 & \dots & M_n \\ M_1 & M_2 & \dots & M_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ M_n & M_{n+1} & \dots & M_{2n} \end{pmatrix} \succeq \mathbf{0},$$

$$\mathbf{R}_{2n+1} = \begin{pmatrix} M_1 & M_2 & \dots & M_{n+1} \\ M_2 & M_3 & \dots & M_{n+2} \\ \vdots & \vdots & \ddots & \vdots \\ M_{n+1} & M_{n+2} & \dots & M_{2n+1} \end{pmatrix} \succeq \mathbf{0}.$$

(b) (Arbitrary random variables) *The vector $(M_1, M_2, \dots, M_{2n})$ is a feasible $(1, 2n, R)$ -moment sequence if and only if $\mathbf{R}_{2n} \succeq \mathbf{0}$.*

Proof:

We will only show the necessity of part (b). If $(M_1, M_2, \dots, M_{2n})$ is a feasible $(1, 2n, R)$ -moment sequence, then there exists a probability measure $f(x)$ such that

$$\int_{-\infty}^{\infty} x^k f(x) dx = M_k, \quad k = 0, 1, \dots, 2n,$$

where $M_0 = 1$. Consider the vector $\mathbf{x} = (1, x, x^2, \dots, x^{2n})'$, and the semidefinite matrix $\mathbf{x}\mathbf{x}'$. Since $f(x)$ is nonnegative, the matrix

$$\mathbf{R}_{2n} = \int_{-\infty}^{\infty} \mathbf{x}\mathbf{x}' f(x) dx$$

should be semidefinite. ■

The multivariate case

We consider the question of whether a sequence $\boldsymbol{\sigma} = (\mathbf{M}, \boldsymbol{\Gamma})$ is a feasible $(n, 2, R^n)$ -moment vector, i.e., whether there exists a random vector \mathbf{X} such that $E[\mathbf{X}] = \mathbf{M}$, $E[\mathbf{X}\mathbf{X}'] = \boldsymbol{\Gamma}$.

Theorem 2 *A sequence $\boldsymbol{\sigma} = (\mathbf{M}, \boldsymbol{\Gamma})$ is a feasible $(n, 2, R^n)$ -moment vector if and only if the following matrix is semidefinite:*

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & \mathbf{M}' \\ \mathbf{M} & \boldsymbol{\Gamma} \end{bmatrix} \succeq \mathbf{0}.$$

Proof:

Suppose $(\mathbf{M}, \mathbf{\Gamma})$ is a feasible $(n, 2, R^n)$ -moment vector. Then, there exists a random variable \mathbf{X} such that $E[\mathbf{X}] = \mathbf{M}$, $E[\mathbf{X}\mathbf{X}'] = \mathbf{\Gamma}$. The matrix $(\mathbf{X} - \mathbf{M})(\mathbf{X} - \mathbf{M})'$ is semidefinite. Taking expectations, we obtain that

$$E[(\mathbf{X} - \mathbf{M})(\mathbf{X} - \mathbf{M})'] = \mathbf{\Gamma} - \mathbf{M}\mathbf{M}' \succeq \mathbf{0},$$

which expresses the fact that a covariance matrix needs to be semidefinite. It is easy to see that $\mathbf{\Gamma} - \mathbf{M}\mathbf{M}' \succeq \mathbf{0}$ if and only if $\mathbf{\Sigma} \succeq \mathbf{0}$.

Conversely, if $\mathbf{\Sigma} \succeq \mathbf{0}$, then $\mathbf{\Gamma} - \mathbf{M}\mathbf{M}' \succeq \mathbf{0}$. Let \mathbf{X} be a multivariate normal distribution with mean \mathbf{M} and covariance matrix $\mathbf{\Gamma} - \mathbf{M}\mathbf{M}'$. This shows that the vector $\boldsymbol{\sigma} = (\mathbf{M}, \mathbf{\Gamma})$ is a feasible $(n, 2, R^n)$ -moment vector. ■

There are known necessary conditions for a sequence $\boldsymbol{\sigma}$ to be a feasible (n, k, R^n) -moment vector for $k \geq 3$, that also involve the semidefiniteness of a matrix derived from the vector $\boldsymbol{\sigma}$, but these conditions are not known to be sufficient. In general, the complexity of deciding whether a sequence $\boldsymbol{\sigma}$ is a feasible (n, k, R^n) -moment vector has not been resolved.

Structure of the paper

The structure of the paper is as follows. In Section 2, we outline the application of semidefinite programming to stochastic optimization problems. In Section 3, we derive explicit and often surprising optimal bounds in probability theory using convex and semidefinite programming methods. In Section 4, we apply convex and semidefinite programming methods to problems in finance. In Section 5, we illustrate a connection between moment problems and semidefinite relaxations in discrete optimization. Section 6 contains some concluding remarks.

2 Semidefinite Relaxations for Stochastic Optimization Problems

The development of semidefinite relaxations represents an important advance in discrete optimization. In this section, we review a theory for deriving semidefinite relaxations for classical stochastic optimization problems. The idea of deriving semidefinite relaxations

for this class of problems is due to Bertsimas [3]. Our development in this paper follows Bertsimas and Niño-Mora [5], [6]; the interested reader is referred to these papers for further details. We demonstrate the central ideas for the problem of optimizing a multiclass queueing network that represents a stochastic and dynamic generalization of a job shop.

2.1 Model description

We consider a network of queues composed of K single-server stations and populated by N job classes. The set of job classes $\mathcal{N} = \{1, \dots, N\}$ is partitioned into subsets $\mathcal{C}_1, \dots, \mathcal{C}_K$, so that station $m \in \mathcal{K} = \{1, \dots, K\}$ only serves classes in its *constituency* \mathcal{C}_m . We refer to jobs of class i as i -jobs, and we let $s(i)$ be the station that serves i -jobs. The network is *open*, so that jobs arrive from outside, follow a Markovian route through the network (i -jobs wait for service at the i -queue) and eventually exit. External arrivals of i -jobs follow a Poisson process with rate α_i (if class i does not have external arrivals $\alpha_i = 0$). The service times of i -jobs are independent and identically distributed, having an exponential distribution with mean $\beta_i = 1/\mu_i$. Upon completion of service at station $s(i)$, an i -job becomes a j -job (and hence is routed to the j -queue), with probability p_{ij} , or leaves the system, with probability $p_{i0} = 1 - \sum_{j \in \mathcal{N}} p_{ij}$. We assume that the routing matrix $\mathbf{P} = (p_{ij})_{i,j \in \mathcal{N}}$ is such that a single job moving through the network eventually exits, i.e., the matrix $\mathbf{I} - \mathbf{P}$ is invertible. We further assume that all service times and arrival processes are mutually independent.

The network is controlled by a *scheduling policy*, which specifies dynamically how each server is allocated to waiting jobs. Scheduling policies can be either *dynamic* or *static*. In a *dynamic* policy, scheduling decisions may depend on the current or past states of all queues; in a *static* policy, the scheduling decisions of each server are independent of the queue lengths of the job classes. A scheduling policy is *stable* if the queue-length vector process has an equilibrium distribution with finite mean. We allow policies to be *preemptive*, i.e., a job's service may be interrupted and resumed later. Finally, a scheduling policy is *nonidling* if a server cannot idle whenever there is a job waiting for service at that station.

Next, we define other model parameters of interest. The *effective arrival rate* of j -jobs, denoted by λ_j , is the total rate at which both external and internal jobs arrive to the

j -queue. The λ_j 's are computed by solving the system

$$\lambda_j = \alpha_j + \sum_{i \in \mathcal{N}} p_{ij} \lambda_i, \quad \text{for } j \in \mathcal{N}.$$

The *traffic intensity* of j -jobs, denoted by $\rho_j = \lambda_j \beta_j$, is the time-stationary probability that a j -job is in service. The *total traffic intensity* at station m is $\rho(\mathcal{C}_m) = \sum_{j \in \mathcal{C}_m} \rho_j$, and is the time-stationary probability that server m is busy. We note that the condition

$$\rho(\mathcal{C}_m) < 1, \quad \text{for } m \in \mathcal{K}$$

is necessary but not sufficient for guaranteeing the stability of any nonidling policy.

We assume that the system operates in a steady-state regime (under a stable policy), and introduce the following variables:

- $L_i(t)$ = number of i -jobs in system at time t .
- $B_i(t)$ = 1 if an i -job is in service at time t ; 0 otherwise.
- $B^m(t)$ = 1 if server m is busy at time t ; 0 otherwise; notice that $B^m(t) = \sum_{i \in \mathcal{C}_m} B_i(t)$.

In what follows we write, for convenience of notation, $L_i = L_i(0)$, $B_i = B_i(0)$ and $B^m = B^m(0)$.

2.2 The performance optimization problem

The *performance measures* we are interested in are $\mathbf{x} = (x_j)_{j \in \mathcal{N}}$, and $\mathbf{y} = (y_j)_{j \in \mathcal{N}}$, where

$$x_j = E[L_j], \quad y_j = E[L_j^2], \quad \text{for } j \in \mathcal{N},$$

i.e., the vectors whose components are the time-stationary mean and second moment of the number of jobs from each class in the system.

Given a *performance cost function* $\mathbf{c}'\mathbf{x} + \mathbf{d}'\mathbf{y}$, we investigate the following *performance optimization problem*: compute a lower bound $\underline{Z} \leq \mathbf{c}'\mathbf{x} + \mathbf{d}'\mathbf{y}$ that is valid under a given class of admissible policies, and design a policy which nearly minimizes the cost $\mathbf{c}'\mathbf{x} + \mathbf{d}'\mathbf{y}$. For our purposes, any preemptive, nonidling policy is admissible. In this paper, we restrict our attention to the question of computing strong lower bounds. As we mentioned in the Introduction, the design of optimal policies is *EXPTIME*-hard (Papadimitiou and

Tsitsiklis [26]), i.e., it is provably requires exponential time, as $P \neq EXPTIME$. In recent years, progress has been made in designing near-optimal scheduling policies based on the idea of fluid control, in which discrete jobs are replaced by the flow of a fluid; we refer the interested reader to the papers by Avram, Bertsimas and Ricard [2], Weiss [30], Luo and Bertsimas [24], and the references cited therein for details.

We study the problem of computing good lower bounds via the *achievable region* approach. The achievable region (equivalently, performance region) \mathcal{X} is defined as the set of all performance vectors (\mathbf{x}, \mathbf{y}) that can be achieved under admissible policies. Our goal is to derive constraints on the performance vector (\mathbf{x}, \mathbf{y}) that define a relaxation of performance region \mathcal{X} . Since it is not obvious how to derive such constraints directly, we pursue the following plan: (a) identify system *equilibrium relations* and formulate them as constraints involving *auxiliary performance variables*; (b) formulate additional constraints (both *linear* and *positive semidefinite*) on the auxiliary performance variables; (c) formulate constraints that express the original performance vector, (\mathbf{x}, \mathbf{y}) , in terms of the auxiliary variables.

Notice that this approach is fairly standard in the mathematical programming literature and has a clear geometric interpretation: It corresponds to constructing a relaxation of the performance region of the natural variables, (\mathbf{x}, \mathbf{y}) , by (a) *lifting* this region into a higher dimensional space, by means of auxiliary variables, (b) bounding the lifted region through constraints on the auxiliary variables, and (c) *projecting* back into the original space. *Lift and project* techniques have proven powerful tools for constructing tight relaxations for hard discrete optimization problems (see, e.g., Lovász and Schrijver [23]). We have summarized the performance measures considered in this paper (including auxiliary ones) in Table 1.

The rest of this section is organized as follows. In Section 2.3, we include linear constraints that relate the natural performance measures in terms of auxiliary performance variables. Using the fact that our performance measures are expectations of random variables, we describe a set of positive semidefinite constraints in Section 2.4. In Section 2.5, we introduce a linear and a semidefinite relaxation using the constraints of the previous sections. We further present computational results that illustrate that the semidefinite relaxation is substantially stronger than the linear programming relaxation.

Performance variables	Interpretation
$x_j; \mathbf{x} = (x_j)_{j \in \mathcal{N}}$	$E[L_j]$
$x_j^i; \mathbf{X} = (x_j^i)_{i,j \in \mathcal{N}}; \mathbf{x}^i = (x_j^i)_{j \in \mathcal{N}}$	$E[L_j B_i = 1]$
$x_j^{0m}; \mathbf{X}^0 = (x_j^{0m})_{m \in \mathcal{K}, j \in \mathcal{N}}; \mathbf{x}^{0m} = (x_j^{0m})_{j \in \mathcal{N}}$	$E[L_j B^m = 0]$
$r_{ij}; \mathbf{R} = (r_{ij})_{i,j \in \mathcal{N}}$	$E[B_i B_j]$
$r_{ij}^k; \mathbf{R}^k = (r_{ij}^k)_{i,j \in \mathcal{N}}$	$E[B_i B_j B_k = 1]$
$r_{ij}^{0m}; \mathbf{R}^{0m} = (r_{ij}^{0m})_{i,j \in \mathcal{N}}$	$E[B_i B_j B^m = 0]$
$y_{ij}; \mathbf{Y} = (y_{ij})_{i,j \in \mathcal{N}}$	$E[L_i L_j]$
$y_{ij}^k; \mathbf{Y}^k = (y_{ij}^k)_{i,j \in \mathcal{N}}$	$E[L_i L_j B_k = 1]$
$y_{ij}^{0m}; \mathbf{Y}^{0m} = (y_{ij}^{0m})_{i,j \in \mathcal{N}}$	$E[L_i L_j B^m = 0]$

Table 1: Network performance measures.

2.3 Linear constraints

In this section, we present several sets of linear constraints that express natural performance measures in terms of auxiliary ones. The first set of constraints describes constraints that follow from elementary arguments.

Theorem 3 (Elementary constraints) *Under any stable policy, the following equations hold:*

(a) *Projection Constraints:*

$$x_j = \sum_{i \in \mathcal{C}_m} \rho_i x_j^i + (1 - \rho(\mathcal{C}_m)) x_j^{0m}, \quad j \in \mathcal{N}, m \in \mathcal{K}, \quad (1)$$

$$r_{ij} = \sum_{k \in \mathcal{C}_m} \rho_k r_{ij}^k + (1 - \rho(\mathcal{C}_m)) r_{ij}^{0m}, \quad i, j \in \mathcal{N}, m \in \mathcal{K}, \quad (2)$$

$$y_{ij} = \sum_{k \in \mathcal{C}_m} \rho_k y_{ij}^k + (1 - \rho(\mathcal{C}_m)) y_{ij}^{0m}, \quad i, j \in \mathcal{N}, m \in \mathcal{K}. \quad (3)$$

(b) *Definitional Constraints:*

$$r_{ij} = \rho_j r_{ii}^j, \quad i, j \in \mathcal{N}, \quad (4)$$

$$r_{ii} = \rho_i, \quad r_{ii}^i = 1, \quad i \in \mathcal{N}, \quad (5)$$

$$r_{ij} = 0, \quad r_{ij}^k = 0, \quad i, j \in \mathcal{C}_m, \quad (6)$$

$$r_{ij}^k = 0, \quad i, k \in \mathcal{C}_m, \text{ or } j, k \in \mathcal{C}_m, \quad (7)$$

$$r_{ij}^{0m} = 0, \quad i \text{ or } j \in \mathcal{C}_m, \quad (8)$$

$$(9)$$

(c) *Lower bound constraints:*

$$r_{ij} \geq \max(0, \rho_i + \rho_j - 1), \quad i, j \in \mathcal{N}, \quad (10)$$

$$x_j^i \geq \frac{r_{ij}}{\rho_i}, \quad i, j \in \mathcal{N}, \quad (11)$$

$$x_j^{0m} \geq \max\left(0, \frac{\rho_j - \rho(\mathcal{C}_m)}{1 - \rho(\mathcal{C}_m)}\right), \quad m \in \mathcal{K}, j \in \mathcal{N}, \quad (12)$$

$$r_{ij}^k \geq \max\left(0, \frac{r_{ki} + r_{kj}}{\rho_k} - 1\right), \quad i, j, k \in \mathcal{N}, \quad (13)$$

$$r_{ij}^{0m} \geq \max\left(0, \frac{\max(0, \rho_i - \rho(\mathcal{C}_m)) + \max(0, \rho_j - \rho(\mathcal{C}_m))}{1 - \rho(\mathcal{C}_m)} - 1\right), \quad i, j \in \mathcal{N}, m \in \mathcal{K}, \quad (14)$$

$$y_{ij} \geq r_{ij}, \quad i, j \in \mathcal{N}, \quad (15)$$

$$y_{ij}^k \geq r_{ij}^k, \quad i, j, k \in \mathcal{N}, \quad (16)$$

$$y_{ij}^{0m} \geq r_{ij}^{0m}, \quad i, j \in \mathcal{N}, m \in \mathcal{K}. \quad (17)$$

Proof

The constraints in (a) follow by a simple conditioning argument, by noticing that at each time instant, a server is either serving some job class in its constituency or idling. The constraints in (b), (c) follow from elementary arguments. \blacksquare

Flow conservation constraints

We next present a set of linear constraints on performance measures using the classical *flow conservation law* of queueing theory, $L^- = L^+$. We first provide a brief discussion of flow conservation in stochastic systems, and then show how to use these ideas to derive linear relations between time-stationary moments of queue lengths.

The classical flow conservation law of queueing systems states that, under mild restrictions, the stationary state probabilities of the number in system at arrival epochs and that at departure epochs are equal. The key assumption is that jobs arrive to the system and depart from the system one at a time, so that the queue size can change only by unit steps.

Consider a multiclass queueing network operating in a steady state regime, with the number in system process $\{L(t)\}$. We assume that the process $\{L(t)\}$ has right-continuous

sample paths, and we use $L(t^-)$ to denote the left limit of the process at time t . The corresponding right limit $L(t^+) = L(t)$ because of right-continuity of sample paths. Let $A = \{\tau_k^a\}$ and $D = \{\tau_k^d\}$ be the sequences of arrival and departure epochs of jobs respectively. Let $L(\tau_k^{a-})$ be the number of jobs in the system seen by the k^{th} arriving job just before its arrival; similarly, let $L(\tau_k^d)$ be the number of jobs in the system seen by the k^{th} departing job just after its departure. We define

$$L^- = L(\tau_0^{a-}),$$

and

$$L^+ = L(\tau_0^d).$$

Since we assumed the system to be in steady-state, L^- may be interpreted as the number of jobs in the system seen by a typical arrival, while L^+ may be interpreted as the number of jobs in the system seen by a typical departure. By considering any realization, we see that for every upward transition for the number in system from i to $(i + 1)$, there is a corresponding downward transition from $(i + 1)$ to i ; thus every $L(\tau_k^{a-})$ is equal to a distinct $L(\tau_k^d)$ in a sample-path sense. In particular, we have $L^- = L^+$, yielding the following theorem.

Theorem 4 (Flow Conservation Law) *If jobs enter and leave the system one at a time, then*

$$L^- = L^+$$

holds in distribution.

In what follows, we apply the law $L^- = L^+$ to a family of queues obtained by aggregating job classes, as explained next. Let $S \subseteq \mathcal{N}$.

Definition 2 (S -queue) *The S -queue is the queueing system obtained by aggregating job classes in S . The number in system at time t in the S -queue is denoted by $L_S(t) = \sum_{j \in S} L_j(t)$.*

As usual we write $L_S = L_S(0)$, $L_S^- = L_S(0-)$, $L_S^+ = L_S(0+) = L_S(0)$. For convenience of notation we also write

$$p(i, S) = \sum_{j \in S} p_{ij}$$

and

$$\alpha(S) = \sum_{j \in S} \alpha_j.$$

The next theorem formulates the law $L^- = L^+$ as it applies to the S -queue.

Theorem 5 (The law $L^- = L^+$ in MQNETs) *Under any dynamic stable policy, and for any subset of job classes $S \subseteq \mathcal{N}$ and nonnegative integer l :*

$$\alpha(S)P(L_S = l) + \sum_{i \in S^c} \lambda_i p(i, S) P(L_S = l \mid B_i = 1) = \sum_{i \in S} \lambda_i (1 - p(i, S)) P(L_S = l + 1 \mid B_i = 1). \quad (18)$$

Proof

By applying Theorem 4 to the S -queue, we have that

$$P(L_S^- = l) = P(L_S^+ = l).$$

An arrival epoch to the S -queue is either an arrival from the outside world (external arrival) that happens with rate $\alpha(S)$, or an internal movement from a class i in S^c to a class in S (internal arrival) that happens only if $B_i = 1$, for $i \in S^c$ with rate

$$\mu_i p(i, S) P(B_i = 1) = \mu_i p(i, S) \rho_i = \lambda_i p(i, S).$$

The total arrival rate to S -queue is

$$\lambda_S = \alpha(S) + \sum_{i \in S^c} \lambda_i p(i, S).$$

Therefore,

$$P(L_S^- = l) = \frac{\alpha(S)}{\lambda_S} P(L_S = l) + \sum_{i \in S^c} \frac{\lambda_i p(i, S)}{\lambda_S} P(L_S = l \mid B_i = 1).$$

A departure epoch from S -queue happens with rate

$$\mu_i (1 - p(i, S)) P(B_i = 1) = \lambda_i (1 - p(i, S)),$$

for all $i \in S$. The total departure rate is:

$$\mu_S = \sum_{i \in S} \lambda_i (1 - p(i, S)).$$

It can be easily checked that the total arrival rate to the S -queue and the total departure rate from the S -queue are equal, i.e., $\lambda_S = \mu_S$. Therefore,

$$P(L_S^+ = l) = \sum_{i \in S} \frac{\lambda_i (1 - p(i, S))}{\mu_S} P(L_S = l + 1 | B_i = 1).$$

By applying $P(L_S^- = l) = P(L_S^+ = l)$, Eq. (18) follows. \blacksquare

Taking expectations in identity (18) we obtain:

Corollary 1 *Under any stable policy, and for any subset of job classes $S \subseteq \mathcal{N}$ and positive integer K for which $E[(L_1 + \dots + L_N)^K] < \infty$,*

$$\alpha(S) E[L_S^K] + \sum_{i \in S^c} \lambda_i p(i, S) E[L_S^K | B_i = 1] = \sum_{i \in S} \lambda_i (1 - p(i, S)) E[(L_S - 1)^K | B_i = 1]. \quad (19)$$

Note that Corollary 1 formulates a linear relation between time-stationary moments of queue lengths. The equilibrium equations in Corollary 1 corresponding to $K = 1, 2$ and $S = \{i\}, \{i, j\}$, for $i, j \in \mathcal{N}$, yield directly the system of linear constraints on performance variables shown next. Let $\mathbf{\Lambda} = \text{Diag}(\boldsymbol{\lambda})$.

Corollary 2 (Flow conservation constraints) *Under any dynamic stable policy, the following linear constraints hold:*

(a)

$$-\boldsymbol{\alpha} \mathbf{x}' - \mathbf{x} \boldsymbol{\alpha}' + (\mathbf{I} - \mathbf{P})' \boldsymbol{\Lambda} \mathbf{X} + \mathbf{X}' \boldsymbol{\Lambda} (\mathbf{I} - \mathbf{P}) = (\mathbf{I} - \mathbf{P})' \boldsymbol{\Lambda} + \boldsymbol{\Lambda} (\mathbf{I} - \mathbf{P}). \quad (20)$$

(b) *If $E[(L_1 + \dots + L_N)^2] < \infty$, then*

$$\begin{aligned} \alpha_j y_{jj} + \sum_{r \in \mathcal{N}} \lambda_r p_{rj} y_{jj}^r - \lambda_j y_{jj}^j + 2\lambda_j (1 - p_{jj}) x_j^j &= \lambda_j (1 - p_{jj}), \quad j \in \mathcal{N} \quad (21) \\ \alpha_i y_{jj} + \alpha_j y_{ii} + 2(\alpha_i + \alpha_j) y_{ij} + \sum_{r \in \mathcal{N}} \lambda_r p_{ri} y_{jj}^r + \sum_{r \in \mathcal{N}} \lambda_r p_{rj} y_{ii}^r \\ + \sum_{r \in \mathcal{N}} 2\lambda_r (p_{ri} + p_{rj}) y_{ij}^r - \lambda_i y_{jj}^i - \lambda_j y_{ii}^j - 2\lambda_i y_{ij}^i - 2\lambda_j y_{ij}^j \\ + 2\lambda_i (1 - p_{ii} - p_{ij}) x_j^i + 2\lambda_j (1 - p_{ji} - p_{jj}) x_i^j &= -\lambda_i p_{ij} - \lambda_j p_{ji}, \quad (22) \\ & i, j \in \mathcal{N}. \end{aligned}$$

The flow conservation constraints were first derived for multi-station MQNETs by Bertsimas, Paschalidis and Tsitsiklis [7], and by Kumar and Kumar [21], using a potential function approach. The derivation we presented is from Bertsimas and Niño-Mora [6].

2.4 Positive semidefinite constraints

We present in this section, a set of *positive semidefinite constraints* that strengthen the formulations obtained through equilibrium relations. Recall that the performance measures \mathbf{x}, \mathbf{y} are moments of random variables. Applying Theorem 2 to the performance variables introduced in Table 1 yields directly the following result.

Theorem 6 *Under any dynamic stable policy, the following semidefinite constraints hold:*

(a) Let $\mathbf{r}^k = (r_{ii}^k)_{i \in \mathcal{N}}$ and $\mathbf{r}^{0m} = (r_{ii}^{0m})_{i \in \mathcal{N}}$.

$$\begin{bmatrix} 1 & \boldsymbol{\rho}' \\ \boldsymbol{\rho} & \mathbf{R} \end{bmatrix} \succeq \mathbf{0}, \quad (23)$$

$$\begin{bmatrix} 1 & \mathbf{r}^{k'} \\ \mathbf{r}^k & \mathbf{R}^k \end{bmatrix} \succeq \mathbf{0}, \quad k \in \mathcal{N}, \quad (24)$$

$$\begin{bmatrix} 1 & \mathbf{r}^{0m'} \\ \mathbf{r}^{0m} & \mathbf{R}^{0m} \end{bmatrix} \succeq \mathbf{0}, \quad m \in \mathcal{K}. \quad (25)$$

(b) If $E \left[\left(\sum_{j \in \mathcal{N}} L_j \right)^2 \right] < \infty$, then

$$\begin{bmatrix} 1 & \mathbf{x}' \\ \mathbf{x} & \mathbf{Y} \end{bmatrix} \succeq \mathbf{0}, \quad (26)$$

$$\begin{bmatrix} 1 & \mathbf{x}^{k'} \\ \mathbf{x}^k & \mathbf{Y}^k \end{bmatrix} \succeq \mathbf{0}, \quad k \in \mathcal{N}, \quad (27)$$

$$\begin{bmatrix} 1 & \mathbf{x}^{0m'} \\ \mathbf{x}^{0m} & \mathbf{Y}^{0m} \end{bmatrix} \succeq \mathbf{0}, \quad m \in \mathcal{K}. \quad (28)$$

2.5 On the power of the semidefinite relaxation

Our objective in this section is to compare computationally the linear and semidefinite relaxations of the multiclass queueing network performance optimization problem. The

linear programming relaxation is defined as follows:

$$\begin{aligned}
Z_{LP} = & \text{minimize } \mathbf{c}'\mathbf{x} + \mathbf{d}'\mathbf{y} \\
& \text{subject to } \text{Projection constraints : (1), (2), (3),} \\
& \text{Definitional constraints : (4), (5), (6), (7), (8),} \\
& \text{Lower bound constraints : (10), (11), (13), (13), (14), (15), (16), (17),} \\
& \text{Flow – conservation constraints : (20), (21),} \\
& \mathbf{x} \geq \mathbf{0}, \mathbf{y} \geq \mathbf{0}.
\end{aligned}$$

The semidefinite relaxation Z_{SD} is obtained by adding the constraints (23), (24), (25) (26), (27), (28).

A multiclass network

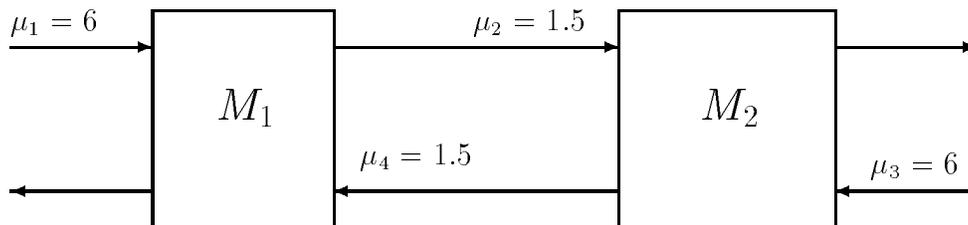


Figure 1: A Multiclass Network.

We consider the network of Figure 1. In this network external arrivals come into either class 1 or class 3, and so $\alpha_2 = \alpha_4 = 0$. In our computations we fix the service times as shown in the figure, and vary only the arrival rates. We maintain the symmetry between classes, and so we set $\alpha_1 = \alpha_3 = \alpha$, where α varies from 0.1 to 1.18. We select $c_i = 1$ and $d_i = 0$, i.e., we are interested in minimizing the expected number of jobs in the system in steady-state. We present below the optimal values Z_{LP} and Z_{SD} . The SDP relaxation has 283 variables (including slack variables) and 259 constraints. We solve the semidefinite relaxation using the package SDPA developed by Fujisawa, Kojima and Nakata [12]. In certain cases SDPA was unable to solve the relaxation to the desired accuracy, but returned primal and dual feasible solutions; in such cases we report the cost of the best primal and dual feasible solutions obtained by SDPA. This has nothing to do with the size of the SDP relaxation, but perhaps something to do with the particular values of the constraint matrices. All of

ρ	Z_{LP}	Z_{SD} (primal)	Z_{SD} (dual)	$E[Z_{LBFS-B}]$	Best B
0.083	0.170	0.174	0.163	0.180	0
0.167	0.347	0.351	0.336	0.391	0
0.250	0.532	0.538	0.538	0.645	0
0.333	0.724	0.796	0.786	0.955	1
0.417	0.926	1.138	1.028	1.342	1
0.500	1.136	1.540	1.495	1.844	1
0.583	1.393	2.117	2.061	2.527	1
0.667	1.928	2.959	2.923	3.516	1
0.750	2.799	4.406	4.325	5.120	1
0.833	4.566	7.206	7.096	8.220	2
0.875	6.325	9.968	9.894	11.242	2
0.917	9.835	15.461	15.341	17.087	2
0.958	20.346	31.818	31.765	34.421	2
0.983	51.854	80.854	80.570	85.643	3

Table 2: Comparison of LP and SDP relaxations for the network of Figure 1.

these instances were solved in less than one minute by SDPA on a Pentium II workstation as the SDP relaxation has a simple block structure.

For comparison purposes, we also report simulation results for a particular policy that was derived from fluid optimal control (see Avram et. al. [2]): When both $L_4(t), L_2(t) > B$, the first station gives preemptive priority to class 4 and the second station gives preemptive priority to class 2. When $L_4(t) \leq B$, class 3 has preemptive priority over class 2. Similarly, when $L_2(t) \leq B$, class 1 has preemptive priority over class 4. We call this policy last-buffer-first-served with a threshold B , denoted by $LBFS - B$. We let $E[Z_{LBFS-B}]$ denote the expected number of jobs under this policy. We select the value of B optimally using simulation.

In Table 2, we report the values Z_{LP} , a primal feasible value for Z_{SD} , a dual feasible value for Z_{SD} , the simulation value $E[Z_{LBFS-B}]$, and the value of the threshold B that

gives the optimal performance. The computational results suggest the following:

- (a) The semidefinite relaxation substantially improves the linear programming relaxation. In our experiments the improvement is in the range of 55%–60%. The improvement is more substantial as the traffic intensity ρ increases.
- (b) The value of the semidefinite relaxation is close to the expected value of the policy *LBFS-B*. This shows that not only the semidefinite relaxation produces near optimal bounds, but the particular policy we constructed is a near optimal.

A multiclass single queue

We consider a single station network with four classes. Our objective here is to minimize $\sum_{i=1}^4 x_i + y_{ii}$. For the case that we do not include terms involving y_{ii} in the objective function, the LP relaxation is exact (see Bertsimas and Niño-Mora [4]).

We assume that the arrival rate for each class is the same, and that the mean service times for the job classes are 0.05, 0.1, 0.2, and 0.4 respectively. The results of the LP and SDP relaxations are tabulated in Table 3. In this experiment the SDP relaxation has 234 variables and 220 constraints. All of these instances were solved in less than two minutes by SDPA on a Pentium II workstation.

For comparison purposes we have simulated the following dynamic priority policy P : At every service completion time t , we give priority to the class that has the highest index $\mu_i L_i(t)$. The policy was derived from fluid optimal control (see Avram et. al. [2]).

We observe again that the semidefinite relaxation provides a substantial improvement over the value of the LP relaxation often by an order of magnitude.

Both computational experiments demonstrate that, unlike the LP relaxation, the semidefinite relaxation provides practically useful suboptimality guarantees that can be used to assess the closeness to optimality of heuristic policies. We believe that the combination of fluid optimal control methods to generate near optimal policies for large scale problems (see Luo and Bertsimas [24]), and semidefinite relaxation to provide near optimal bounds is perhaps the most promising methodology to address the multiclass queueing network optimization problem.

ρ	Z_{LP}	Z_{SD} (primal)	Z_{SD} (dual)	$E[Z_P]$
0.075	0.162	0.163	0.163	0.165
0.150	0.352	0.358	0.357	0.365
0.225	0.578	0.597	0.596	0.616
0.300	0.854	0.900	0.900	0.940
0.375	1.198	1.300	1.300	1.374
0.450	1.639	1.856	1.856	1.978
0.525	2.227	2.675	2.673	2.872
0.600	3.047	3.954	3.954	4.294
0.675	4.270	6.131	6.131	6.740
0.750	6.269	10.330	10.328	11.655
0.825	10.072	20.200	20.200	24.227
0.900	19.814	55.603	55.602	74.020
0.975	89.332	696.646	696.644	1166.362

Table 3: Comparison of LP and SDP relaxations for a multiclass queue.

3 Optimal Bounds in Probability

In this section, we review the work of Bertsimas and Popescu [9]. Suppose that $\bar{\sigma}$ is a feasible moment sequence and \mathbf{X} has a $\bar{\sigma}$ -feasible distribution. We now define the central problem we address in this section:

The (n, k, Ω) -bound problem

Given a sequence $\bar{\sigma}$ of up to k th order moments

$$\sigma_{\mathbf{k}} = E[X_1^{k_1} X_2^{k_2} \cdots X_n^{k_n}], \quad k_1 + k_2 + \cdots + k_n \leq k,$$

of a multivariate random variable $\mathbf{X} = (X_1, X_2, \dots, X_n)$ on $\Omega \subseteq R^n$, find the “best possible” or “tight” upper and lower bounds on $P(\mathbf{X} \in S)$, for arbitrary events $S \subseteq \Omega$.

The term “best possible” or “tight” upper (and by analogy lower) bound above is defined as follows.

Definition 3 We say that α is a tight upper bound on $P(\mathbf{X} \in S)$, and we will denote it by

$\sup_{\mathbf{X} \sim \bar{\sigma}} P(\mathbf{X} \in S)$ if :

- (a) it is an upper bound, i.e., $P(\mathbf{X} \in S) \leq \alpha$ for all random variables $\mathbf{X} \sim \bar{\sigma}$;
- (b) it cannot be improved, i.e., for any $\epsilon > 0$ there is a random variable $\mathbf{X}_\epsilon \sim \bar{\sigma}$ for which $P(\mathbf{X}_\epsilon \in S) > \alpha - \epsilon$.

The well known inequalities due to Markov, Chebyshev and Chernoff, which are widely used if we know the first moment, the first two moments, and all moments (i.e., the generating function) of a random variable, respectively, are feasible but not necessarily optimal solutions to the $(1, k, \Omega)$ -bound problem, i.e., they are not necessarily tight bounds.

In the univariate case, the idea that optimization methods and duality theory can be used to address these type of questions is due to Isii [17]. Thirty years later, Smith [29] has generalized this work to the multivariate case, and proposed interesting applications in decision analysis, dynamic programming, statistics and finance. He also introduces a computational procedure for the (n, k, R^n) -bound problem, although he does not refer to it in this way. Unfortunately, the procedure is far from an actual algorithm, as there is no proof of convergence, and no investigation (theoretical or experimental) of its efficiency.

Bertsimas and Popescu [9], whose work we survey in this paper, have proposed convex optimization algorithms to address the (n, k, Ω) -bound problem.

We examine the existence of an algorithm that on input $\langle n, k, \bar{\sigma}, \epsilon \rangle$ computes a value $\alpha \in [\alpha^* - \epsilon, \alpha^* + \epsilon]$, where $\alpha^* = \sup_{\mathbf{X} \sim \bar{\sigma}} P(\mathbf{X} \in S)$, and runs in time polynomial in $n, k, \log \sigma_{max}$ and $\log \frac{1}{\epsilon}$, where $\sigma_{max} = \max(\sigma_{\mathbf{k}})$. We assume the availability of an oracle to test membership in $S \subset \Omega$, and we allow our algorithm to make oracle queries of the type “is \mathbf{x} in S ?”.

Primal and dual formulations

The (n, k, Ω) -upper bound problem can be formulated as the following optimization problem (P) :

$$\begin{aligned} (P) \quad Z_P = & \text{maximize} \quad \int_S f(\mathbf{z}) d\mathbf{z} \\ & \text{subject to} \quad \int_{\Omega} z_1^{k_1} \cdots z_n^{k_n} f(\mathbf{z}) d\mathbf{z} = \sigma_{\mathbf{k}}, \quad \forall k_1 + \cdots + k_n \leq k, \\ & \quad \quad \quad f(\mathbf{z}) = f(z_1, \dots, z_n) \geq 0, \quad \forall \mathbf{z} = (z_1, \dots, z_n) \in \Omega. \end{aligned}$$

Notice that if Problem (P) is feasible, then $\bar{\sigma}$ is a feasible moment sequence, and any feasible distribution $f(\mathbf{z})$ is a $\bar{\sigma}$ -feasible distribution.

In the spirit of linear programming duality theory, we associate a dual variable $u_{\mathbf{k}}$ with each equality constraint of the primal. We can identify the vector of dual variables with a k -degree, n -variate dual polynomial:

$$g(x_1, \dots, x_n) = \sum_{k_1 + \cdots + k_n \leq k} u_{\mathbf{k}} x_1^{k_1} \cdots x_n^{k_n}.$$

We refer to such a polynomial as a k -degree, n -variate polynomial. The dual objective translates to finding the smallest value of:

$$\sum_{\mathbf{k}} u_{\mathbf{k}} \sigma_{\mathbf{k}} = \sum_{\mathbf{k}} u_{\mathbf{k}} E[X_1^{k_1} \cdots X_n^{k_n}] = E[g(\mathbf{X})],$$

where the expected value is taken over any $\bar{\sigma}$ -feasible distribution. In this framework, the Dual Problem (D) corresponding to Problem (P) can be written as:

$$\begin{aligned} (D) \quad Z_D = & \text{minimize} \quad E[g(\mathbf{X})] \\ & \text{subject to} \quad g(\mathbf{x}) \text{ } k\text{-degree, } n\text{-variate polynomial,} \\ & \quad \quad \quad g(\mathbf{x}) \geq \chi_S(\mathbf{x}), \quad \forall \mathbf{x} \in \Omega, \end{aligned}$$

where $\chi_S(\mathbf{x})$ is the indicator function of the set S , defined by:

$$\chi_S(\mathbf{x}) = \begin{cases} 1, & \text{if } \mathbf{x} \in S, \\ 0, & \text{otherwise.} \end{cases}$$

Notice that in general the optimum may not be achievable. Whenever the primal optimum is achieved, we call the corresponding distribution **an extremal distribution**. We next establish weak duality.

Theorem 7 (Weak duality) $Z_P \leq Z_D$.

Proof

Let $f(\mathbf{z})$ be a primal optimal solution and let $g(\mathbf{z})$ be any dual feasible solution. Then:

$$Z_P = \int_S f(\mathbf{z}) d\mathbf{z} = \int_{\Omega} \chi_S(\mathbf{z}) f(\mathbf{z}) d\mathbf{z} \leq \int_{\Omega} g(\mathbf{z}) f(\mathbf{z}) d\mathbf{z} = E[g(\mathbf{X})],$$

and hence $Z_P \leq \inf_{g(\cdot) \leq \chi_S(\cdot)} E[g(\mathbf{X})] = Z_D$. ■

Theorem 7 indicates that by solving the Dual Problem (D) we obtain an upper bound on the primal objective and hence on the probability we are trying to bound. Under some mild restrictions on the moment vector $\bar{\sigma}$, the dual bound turns out to be tight. This strong duality result follows from a more general theorem first proved in one dimension by Isii [17], and in arbitrary dimensions by Smith [29]. The following theorem is a consequence of their work:

Theorem 8 (Strong Duality and Complementary Slackness) *If the moment vector $\bar{\sigma}$ is an interior point of the set \mathcal{M} of feasible moment vectors, then the following results hold:*

- (a) *Strong Duality:* $Z_P = Z_D$.
- (b) *Complementary Slackness:* *If the dual is bounded, there exists a dual optimal solution $g_{opt}(\cdot)$ and a discrete extremal distribution concentrated on points \mathbf{x} , where $g_{opt}(\mathbf{x}) = \chi_S(\mathbf{x})$, that achieves the bound.*

It can also be shown that if the dual is unbounded, then the primal is infeasible, i.e., the multidimensional moment problem is infeasible. Moreover, if $\bar{\sigma}$ is a boundary point of \mathcal{M} , then it can be shown that the $\bar{\sigma}$ -feasible distributions are concentrated on a subset Ω_0

of Ω , and strong duality holds provided we relax the dual to Ω_0 (see Smith [29], p. 824). In the univariate case, Isii [17] proves that if $\bar{\sigma}$ is a boundary point of \mathcal{M} , then exactly one $\bar{\sigma}$ -feasible distribution exists.

If strong duality holds, then by optimizing over Problem (D) we obtain a **tight** bound on $P(\mathbf{X} \in S)$. On the other hand, solving Problem (D) is equivalent to solving the corresponding separation problem, under certain technical conditions (see Grötschel, Lovász and Schrijver [15]).

3.1 Explicit bounds for the $(n, 1, \Omega)$, $(n, 2, \mathbf{R}^n)$ -bound problems

In this section, we present tight bounds as solutions to n convex optimization problems for the $(n, 1, R_+^n)$ -bound problems, and as a solution to a single convex optimization problem for the $(n, 2, R^n)$ -bound problem for the case when the event S is a convex set. The proof of the theorem uses duality for convex optimization problems.

Theorem 9 (a) *The tight $(n, 1, R_+^n)$ -upper bound for an arbitrary convex event S is given by:*

$$\sup_{\mathbf{X} \sim \mathbf{M}} P(\mathbf{X} \in S) = \min \left(1, \max_{i=1, \dots, n} \frac{M_i}{\inf_{x \in S_i} x_i} \right), \quad (29)$$

where $S_i = S \cap (\cap_{j \neq i} \{\mathbf{x} \mid M_i x_j - M_j x_i \leq 0\})$.

(b) *If the Bound (29) is achievable, then there is an extremal distribution that exactly achieves it; otherwise, there is a sequence of distributions with mean \mathbf{M} , that asymptotically achieve it.*

Theorem 9 constitutes a multivariate generalization of Markov's inequality. We denote by

$$P(\mathbf{X} > \mathbf{M}_{\mathbf{e}+\boldsymbol{\delta}}) = P(X_i > (1 + \delta_i)M_i, \forall i = 1, \dots, n),$$

where $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)'$, $\mathbf{e} = (1, 1, \dots, 1)'$ and $\mathbf{M}_{\boldsymbol{\delta}} = (\delta_1 M_1, \dots, \delta_n M_n)'$. Then, applying Theorem 9 leads to:

$$\sup_{\mathbf{X} \sim \mathbf{M}} P(\mathbf{X} > \mathbf{M}_{\mathbf{e}+\boldsymbol{\delta}}) = \min_{i=1, \dots, n} \frac{1}{1 + \delta_i}.$$

The $(n, 2, \mathbf{R}^n)$ -bound problem for convex sets

We first rewrite the $(n, 2, \mathbf{R}^n)$ -bound problem in a more convenient form. Rather than assuming that $E[\mathbf{X}]$ and $E[\mathbf{X}\mathbf{X}']$ are known, we assume equivalently that the vector $\mathbf{M} = E[\mathbf{X}]$ and the covariance matrix $\mathbf{\Gamma} = E[(\mathbf{X} - \mathbf{M})(\mathbf{X} - \mathbf{M})']$ are known. Given a set $S \subset \mathbf{R}^n$, we find tight upper bounds, denoted by $\sup_{\mathbf{X} \sim (\mathbf{M}, \mathbf{\Gamma})} P(\mathbf{X} \in S)$, on the probability $P(\mathbf{X} \in S)$ for all multivariate random variables \mathbf{X} defined on \mathbf{R}^n with mean $\mathbf{M} = E[\mathbf{X}]$ and covariance matrix $\mathbf{\Gamma} = E[(\mathbf{X} - \mathbf{M})(\mathbf{X} - \mathbf{M})']$.

First, notice that a necessary and sufficient condition for the existence of such a random variable \mathbf{X} , is that the covariance matrix $\mathbf{\Gamma}$ is symmetric and positive semidefinite. Indeed, given \mathbf{X} , for an arbitrary vector \mathbf{a} we have:

$$0 \leq E[(\mathbf{a}'(\mathbf{X} - \mathbf{M}))^2] = \mathbf{a}'E[(\mathbf{X} - \mathbf{M})(\mathbf{X} - \mathbf{M})']\mathbf{a} = \mathbf{a}'\mathbf{\Gamma}\mathbf{a},$$

so $\mathbf{\Gamma}$ must be positive semidefinite. Conversely, given a symmetric semidefinite matrix $\mathbf{\Gamma}$ and a mean vector \mathbf{M} , we can define a multivariate normal distribution with mean \mathbf{M} and covariance $\mathbf{\Gamma}$. Moreover, notice that $\mathbf{\Gamma}$ is positive definite if and only if the components of $\mathbf{X} - \mathbf{M}$ are linearly independent. Indeed, the only way that $0 = \mathbf{a}'\mathbf{\Gamma}\mathbf{a} = E[(\mathbf{a}'(\mathbf{X} - \mathbf{M}))^2]$ for a nonzero vector \mathbf{a} is that $\mathbf{a}'(\mathbf{X} - \mathbf{M}) = 0$.

We assume that $\mathbf{\Gamma}$ has full rank and is positive definite. This does not reduce the generality of the problem, it just eliminates redundant constraints, and thereby insures that Theorem 8 holds. Indeed, the tightness of the bound is guaranteed by Theorem 8 whenever the moment vector is interior to \mathcal{M} . If the moment vector is on the boundary, it means that the covariance matrix of X is not of full rank, implying that the components of X are linearly dependent. By eliminating the dependent components, we reduce without loss of generality the problem to one of smaller dimension for which strong duality holds. Hence, the primal and the dual problems (P) and (D) satisfy $Z_P = Z_D$.

Theorem 10 (a) *The tight $(n, 2, \mathbf{R}^n)$ -upper bound for an arbitrary convex event S is given by:*

$$\sup_{\mathbf{X} \sim (\mathbf{M}, \mathbf{\Gamma})} P(\mathbf{X} \in S) = \frac{1}{1 + d^2}, \quad (30)$$

where $d^2 = \inf_{\mathbf{x} \in S} (\mathbf{x} - \mathbf{M})'\mathbf{\Gamma}^{-1}(\mathbf{x} - \mathbf{M})$, is the squared distance from \mathbf{M} to the set S , under the norm induced by the matrix $\mathbf{\Gamma}^{-1}$.

(b) If $\mathbf{M} \notin S$ and if $d^2 = \inf_{\mathbf{x} \in S} (\mathbf{x} - \mathbf{M})' \mathbf{\Gamma}^{-1} (\mathbf{x} - \mathbf{M})$ is achievable, then there is an extremal distribution that exactly achieves the Bound (30); otherwise, if $\mathbf{M} \in S$ or if d^2 is not achievable, then there is a sequence of $(\mathbf{M}, \mathbf{\Gamma})$ -feasible distributions that asymptotically approach the Bound (30).

Theorem 10 constitutes a multivariate generalization of Chebyshev's inequality. The tight multivariate one-sided Chebyshev bound is

$$\sup_{\mathbf{X} \sim (\mathbf{M}, \mathbf{\Gamma})} P(\mathbf{X} > \mathbf{M}_{e+\delta}) = \frac{1}{1 + d^2}, \quad (31)$$

where d^2 is given by:

$$\begin{aligned} d^2 = \text{minimize} \quad & \mathbf{x}' \mathbf{\Gamma}^{-1} \mathbf{x} \\ \text{subject to} \quad & \mathbf{x} \geq \mathbf{M}_{\delta}, \end{aligned} \quad (32)$$

or alternatively d^2 is given by the Gauge dual problem of (32):

$$\begin{aligned} \frac{1}{d^2} = \text{minimize} \quad & \mathbf{x}' \mathbf{\Gamma} \mathbf{x} \\ \text{subject to} \quad & \mathbf{x}' \mathbf{M}_{\delta} = 1 \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned} \quad (33)$$

If $\mathbf{\Gamma}^{-1} \mathbf{M}_{\delta} \geq \mathbf{0}$, then the tight bound is expressible in closed form:

$$\sup_{\mathbf{X} \sim (\mathbf{M}, \mathbf{\Gamma})} P(\mathbf{X} > \mathbf{M}_{e+\delta}) = \frac{1}{1 + \mathbf{M}'_{\delta} \mathbf{\Gamma}^{-1} \mathbf{M}_{\delta}}. \quad (34)$$

Surprisingly, the bound (34) improves upon the Chebyshev's inequality for scalar random variables. In order to express Chebyshev's inequality we define the squared coefficient of variation: $C_M^2 = \frac{M_2 - M_1^2}{M_1^2}$. Chebyshev's inequality is given by:

$$P(X > (1 + \delta)M_1) \leq \frac{C_M^2}{\delta^2},$$

where as bound (34) is stronger:

$$P(X > (1 + \delta)M_1) \leq \frac{C_M^2}{C_M^2 + \delta^2}.$$

In Section 4.1, we review a polynomial time algorithm for the $(n, 2, R^n)$ -upper bound problem, when the set S is a disjoint union of convex sets.

3.2 The complexity of the $(n, 2, \mathbf{R}_+^n)$, (n, k, \mathbf{R}^n) -bound problems

In this section, we show that the separation problem associated with Problem (D) for the cases $(n, 2, \mathbf{R}_+^n)$, (n, k, \mathbf{R}^n) -bound problems are NP-hard for $k \geq 3$. By the equivalence of optimization and separation (see Grötschel, Lovász and Schrijver [15]), solving Problem (D) is NP-hard as well. Finally, because of Theorem 2, solving the $(n, 2, \mathbf{R}_+^n)$, (n, k, \mathbf{R}^n) -bound problems with $k \geq 3$ is NP-hard.

The complexity of the $(n, 2, \mathbf{R}_+^n)$ -bound problem

The separation problem can be formulated as follows in this case:

Problem 2SEP: Given a multivariate polynomial $g(\mathbf{x}) = \mathbf{x}'\mathbf{H}\mathbf{x} + \mathbf{c}'\mathbf{x} + d$, and a set $S \subseteq \mathbf{R}_+^n$, does there exist $\mathbf{x} \in S$ such that $g(\mathbf{x}) < 0$?

If we consider the special case $\mathbf{c} = \mathbf{0}$, $d = 0$, and $S = \mathbf{R}_+^n$, Problem 2SEP reduces to the question whether a given matrix H is co-positive, which is NP-hard (see Murty and Kabadi [25]).

The complexity of the (n, k, \mathbf{R}^n) -bound problem for $k \geq 3$

For $k \geq 3$, the separation problem can be formulated as follows:

Problem 3SEP: Given a multivariate polynomial $g(\mathbf{x})$ of degree $k \geq 3$, and a set $S \subseteq \mathbf{R}^n$, does there exist $\mathbf{x} \in S$ such that $g(\mathbf{x}) < 0$?

Bertsimas and Popescu [9] show that problem 3SEP is NP-hard by performing a reduction from 3SAT.

4 Moment Problems in Finance

The idea of investigating the relation of option and stock prices just based on the no-arbitrage assumption, but without assuming any model for the underlying price dynamics has a long history in the financial economics literature. Cox and Ross [11] and Harrison and Kreps [16] show that the no-arbitrage assumption is equivalent with the existence of a probability distribution π (the so-called martingale measure) such that that option prices become martingales under π . In this section, we survey some recent work of Bertsimas and Popescu [8] that sheds new light to the relation of option and stock prices, and shows that

the natural way to address this relation, without making distributional assumptions for the underlying price dynamics, but only using the no-arbitrage assumption, is the use of convex optimization methods.

In order to motivate the overall approach we formulate the problem of deriving optimal bounds on the price of a European call option given the mean and variance of the underlying stock price solved by Lo [22]. A call option on a certain stock with maturity T and strike k gives the owner of the option the right to buy the underlying stock at time T at price k . If X is the price of the stock at time T , then the payoff of such an option is zero if $X < k$ (the owner will not exercise the option), and $X - k$ if $X \geq k$, i.e., it is $\max(0, X - k)$. Following Cox and Ross [11] and Harrison and Kreps [16], the no-arbitrage assumption is equivalent with the existence of a probability distribution π of the stock price X , such that the price of any European call option with strike price k is given by

$$q(k) = E[\max(0, X - k)],$$

where the expectation is taken over the unknown distribution π . Note that we have assumed, without loss of generality, that the risk free interest rate is zero. Moreover, given that the mean and variance of the underlying asset are observable:

$$E[X] = \mu, \quad \text{and} \quad \text{Var}[X] = \sigma^2,$$

the problem of finding the best possible upper bound on the call price, written as

$$Z^* = \sup_{X \sim (\mu, \sigma^2)^+} E[\max(0, X - k)],$$

can be formulated as follows:

$$Z^* = \sup E[\max(0, X - k)]$$

$$\text{subject to } E[X] = \mu$$

$$\text{Var}[X] = \sigma^2.$$

Conversely, the problem of finding sharp upper and lower bounds on the moments of the stock price using known option prices, can be formulated as follows:

$$\text{sup / inf } E[X], \text{ or } E[X^2], \text{ or } E[\max(0, X - k)]$$

$$\text{subject to } E[\max(0, X - k_i)] = q_i, \quad i = 1, \dots, n.$$

These formulations naturally lead us to the following general optimization problem:

$$\begin{aligned} & \sup / \inf \quad E[\phi(\mathbf{X})] \\ & \text{subject to} \quad E[f_i(\mathbf{X})] = q_i, \quad i = 0, 1, \dots, n, \end{aligned} \tag{35}$$

where $\mathbf{X} = (X_1, \dots, X_m)$ is a multivariate random variable, and $\phi : R^m \rightarrow R$ is a real-valued objective function, $f_i : R^m \rightarrow R$, $i = 1, \dots, n$ are also real-valued, so-called *moment functions* whose expectations $q_i \in R$, referred to as *moments*, are known and finite. We assume that $f_0(\mathbf{x}) = 1$ and $q_0 = E[f_0(\mathbf{X})] = 1$, corresponding to the implied probability-mass constraint.

Note that the problem of finding optimal bounds on $P(\mathbf{X} \in S)$ of the previous section given moments of \mathbf{X} can be formulated as a special case of Problem (35) with $\phi(\mathbf{x}) = \chi_S(\mathbf{x})$.

4.1 Efficient algorithms

In this section, we propose a polynomial time algorithm for Problem (35). We analyze the upper bound problem, since we can solve the lower bound problem by changing the sign of the objective function ϕ :

$$\begin{aligned} (P) \quad & Z_P = \sup \quad E[\phi(X)] \\ & \text{subject to} \quad E[f_i(X)] = q_i, \quad i = 0, 1, \dots, n. \end{aligned} \tag{36}$$

We define the vector of moment functions $\mathbf{f} = (f_0, f_1, \dots, f_n)$ and the corresponding vector of moments $\mathbf{q} = (q_0, q_1, \dots, q_n)$. Without loss of generality, we assume that the moment functions are linearly independent on R^m , meaning that there is no nonzero vector \mathbf{y} so that $\mathbf{y}'f(\mathbf{x}) = 0$ for all $\mathbf{x} \in R^m$.

The dual problem can be written as

$$\begin{aligned} (D) \quad & Z_D = \inf \quad E[\mathbf{y}'f(\mathbf{X})] = \inf \quad \mathbf{y}'\mathbf{q} \\ & \text{subject to} \quad \mathbf{y}'f(\mathbf{x}) \geq \phi(\mathbf{x}), \quad \forall \mathbf{x} \in R^m. \end{aligned}$$

Smith [29] shows that if the vector of moments \mathbf{q} is interior to the feasible moment set $\mathcal{M} = \{E[f(\mathbf{X})] \mid \mathbf{X} \text{ arbitrary multivariate distribution}\}$, then strong duality holds: $Z_P = Z_D$.

Thus by solving Problem (D) we obtain the desired sharp bounds. On the other hand, under certain technical conditions (see Grötschel, Lovász and Schrijver [14]), solving Problem (D) is equivalent to solving the corresponding separation problem (S):

The Separation Problem (S):

Given an arbitrary $\mathbf{y} = (y_0, y_1, \dots, y_n)$, check whether $g(\mathbf{x}) = \mathbf{y}'(\mathbf{x}) - \phi(\mathbf{x}) \geq \mathbf{0}$, for all $\mathbf{x} \in R^m$, and if not find a violated inequality.

The following theorem shows that for important special cases Problem (36) is solvable in polynomial time.

Theorem 11 *If ϕ and f_i , $i = 1, \dots, n$ are quadratic or piecewise linear functions (or piecewise constant) over d convex sets, and d is a polynomial in n, m , then Problem (36) can be solved in polynomial time.*

Proof:

We consider first the case when all functions are quadratic or linear: $\phi(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x} + \mathbf{b}'\mathbf{x} + c$ and $f_i(\mathbf{x}) = \mathbf{x}'\mathbf{A}_i\mathbf{x} + \mathbf{b}'_i\mathbf{x} + c_i$, $i = 1, \dots, n$. Then,

$$g(\mathbf{x}) = \mathbf{y}'f(\mathbf{x}) - \phi(\mathbf{x}) = y_0 + \sum_{i=1}^n y_i f_i(\mathbf{x}) - \phi(\mathbf{x}) = \mathbf{x}'\hat{\mathbf{A}}\mathbf{x} + \hat{\mathbf{b}}'\mathbf{x} + \hat{c},$$

where $\hat{\mathbf{A}} = \sum_{i=1}^n y_i \mathbf{A}_i - \mathbf{A}$, $\hat{\mathbf{b}} = \sum_{i=1}^n y_i \mathbf{b}_i - \mathbf{b}$, $\hat{c} = \sum_{i=1}^n y_i c_i + y_0 - c$. We show that solving the separation problem “ $\inf_{\mathbf{x} \in R^m} g(\mathbf{x}) \geq 0$?” reduces to checking whether the matrix $\hat{\mathbf{A}} = \sum_{i=1}^n y_i \mathbf{A}_i - \mathbf{A}$ is positive semidefinite, and in that case, solving the corresponding convex quadratic optimization problem.

The following algorithm solves the separation problem in polynomial time:

Algorithm A:

- (a) If $\hat{\mathbf{A}}$ is not positive semidefinite, then we find a vector \mathbf{x}_0 so that $g(\mathbf{x}_0) < 0$. We decompose $\hat{\mathbf{A}} = \mathbf{Q}'\mathbf{\Lambda}\mathbf{Q}$, where $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_n)$ is the diagonal matrix of eigenvalues of $\hat{\mathbf{A}}$. Let $\lambda_i < 0$ be a negative eigenvalue of $\hat{\mathbf{A}}$. Let \mathbf{u} be a vector with $u_j = 0$,

for all $j \neq i$, and u_i large enough so that $\lambda_i u_i^2 + (\mathbf{Q}\hat{\mathbf{b}})_i u_i + \hat{c} < 0$. Let $\mathbf{x}_0 = \mathbf{Q}'\mathbf{u}$. Then,

$$\begin{aligned}
g(\mathbf{x}_0) &= \mathbf{x}_0' \hat{\mathbf{A}} \mathbf{x}_0 + \hat{\mathbf{b}}' \mathbf{x}_0 + \hat{c} \\
&= \mathbf{u}' \mathbf{Q} \mathbf{Q}' \hat{\mathbf{A}} \mathbf{Q} \mathbf{Q}' \mathbf{u} + \hat{\mathbf{b}}' \mathbf{Q}' \mathbf{u} + \hat{c} \\
&= \mathbf{u}' \hat{\mathbf{A}} \mathbf{u} + \hat{\mathbf{b}}' \mathbf{Q}' \mathbf{u} + \hat{c} \\
&= \sum_{j=1}^n \lambda_j u_j^2 + \sum_{j=1}^n (\mathbf{Q}\hat{\mathbf{b}})_j u_j + \hat{c} \\
&= \lambda_i u_i^2 + (\mathbf{Q}\hat{\mathbf{b}})_i u_i + \hat{c} < 0.
\end{aligned}$$

This produces a violated inequality.

- (b) Otherwise, if $\hat{\mathbf{A}}$ is positive semidefinite, then we test if $g(\mathbf{x}) \geq 0, \forall \mathbf{x} \in R^m$ by solving the convex quadratic optimization problem:

$$\inf_{\mathbf{x} \in R^m} g(\mathbf{x}).$$

If the optimal value is $z_0 < 0$, we find \mathbf{x}_0 such that $g(\mathbf{x}_0) < 0$, which represents a violated inequality.

We next examine the case that some of the functions ϕ or $f_i, i = 1, \dots, n$ are piecewise linear over d convex sets, with d being a polynomial in n, m . In this case, the function $g(\mathbf{x}) = \mathbf{y}'(\mathbf{x}) - \phi(\mathbf{x})$ can be written as $g(\mathbf{x}) = \mathbf{x}' \hat{\mathbf{A}} \mathbf{x} + \hat{\mathbf{b}}'_k \mathbf{x} + \hat{c}_k$, for all $\mathbf{x} \in D_k, k = 1, \dots, d$, where the sets D_k form a convex partition of R^m and $d = \text{poly}(n, m)$. We show that in this case, again, solving the separation problem reduces to checking whether the matrix $\hat{\mathbf{A}}$ is positive semidefinite, and in that case solving the convex quadratic problems $\inf_{\mathbf{x} \in D_k} g(\mathbf{x}), k = 1, \dots, d$. This can be done in polynomial time using ellipsoid algorithm (see [14]). The following algorithm solves the separation problem in polynomial time:

Algorithm B:

- (a) If $\hat{\mathbf{A}}$ is not positive semidefinite, we find a violated inequality in exactly the same way as part (a) of Algorithm A.
- (b) Otherwise, if $\hat{\mathbf{A}}$ is positive semidefinite, then we test if $g(\mathbf{x}) \geq 0, \forall \mathbf{x} \in R^m$ by solving a polynomial number d of convex quadratic optimization problems:

$$\inf_{\mathbf{x} \in D_k} \mathbf{x}' \hat{\mathbf{A}} \mathbf{x} + \hat{\mathbf{b}}'_k \mathbf{x} + \hat{c}_k, \text{ for } k = 1, \dots, d.$$

If the optimal value in any of these problems is $z_0^k < 0$, we find \mathbf{x}_0^k such that $g(\mathbf{x}_0^k) < 0$, which represents a violated inequality. ■

Remarks:

- (a) Theorem 11 covers the case, in which we would like to find optimal bounds for an option, given prices for other options and the first two moments of the underlying stock price. In these cases the functions $\mathbf{f}(\mathbf{x})$ and $\phi(\mathbf{x})$ are quadratic or piecewise linear. We will see in the next section that we can derive an explicit answer in the univariate case.
- (b) If $\phi(\mathbf{x}) = \chi_S(\mathbf{x})$ where S is a disjoint union of d convex sets, (and thus $\phi(\mathbf{x})$ is piecewise constant), $\mathbf{f}(\mathbf{x}) = (\mathbf{x}, (\mathbf{x} - \mathbf{M})(\mathbf{x} - \mathbf{M})')$, $\mathbf{q} = (\mathbf{M}, \mathbf{\Gamma})$ and $n = m$ then Problem (36) is the $(n, 2, R^n)$ -upper bound problem $\sup_{\mathbf{x} \sim \mathbf{M}} P(\mathbf{x} \in S)$, where S is a disjoint union of d convex sets. We have seen in the previous section that when S is a convex set, then we can find an explicit answer for the problem (Theorem 10). Theorem 11 provides an algorithmic solution if the set S is a disjoint union of a polynomial number of convex sets.
- (c) The semidefinite property plays an important role in the proof of Theorem 11. Not only we check whether the matrix $\hat{\mathbf{A}}$ is semidefinite as a first step of both Algorithms A and B, but we also solve convex quadratic optimization problems over polyhedral spaces (assuming that the sets D_k are polyhedral, which is typically the case in applications), which can be formulated as semidefinite programming problems.

4.2 Bounds on option prices given moment information

In this section, we derive an explicit upper bound for the case of a single stock, and a European call option with strike k , $\phi(x) = \max(0, x - k)$. The solution is due to Lo [22]. Our proof follows Bertsimas and Popescu [8]. The proof illustrates the use of duality.

Theorem 12 (Tight upper bound on option prices) *The tight upper bound on the price of an option with strike k , on a stock whose price at maturity has a known mean μ and variance σ^2 , is computed by:*

$$\max_{X \sim (\mu, \sigma^2)^+} E[\max(0, X - k)] = \begin{cases} \frac{1}{2} \left[(\mu - k) + \sqrt{\sigma^2 + (\mu - k)^2} \right], & \text{if } k \geq \frac{\mu^2 + \sigma^2}{2\mu}, \\ \mu - k + k \frac{\sigma^2}{\mu^2 + \sigma^2}, & \text{if } k < \frac{\mu^2 + \sigma^2}{2\mu}. \end{cases}$$

Proof:

The dual in this case can be formulated by associating dual variables y_0, y_1, y_2 with the probability-mass, mean and respectively, variance constraints. We obtain the following dual formulation:

$$Z_D = \text{minimize}_y \quad (\mu^2 + \sigma^2) y_2 + \mu y_1 + y_0$$

$$\text{subject to} \quad g(x) = y_2 x^2 + y_1 x + y_0 \geq \max(0, x - k), \quad \forall x \geq 0.$$

A dual feasible function $g(\cdot)$ is any quadratic function that, on the positive orthant, is nonnegative and lies above the line $(x - k)$. In an optimal solution, such a quadratic should be tangent to the line $(x - k)$, so we can write $g(x) - (x - k) = a(x - b)^2$, for some $a \geq 0$. The nonnegativity constraint on $g(\cdot)$ can be expressed as $a(x - b)^2 + x - k \geq 0, \forall x \geq 0$. Let $x_0 = b - \frac{1}{2a}$ be the point of minimum of this quadratic. Depending whether x_0 is nonnegative or not, either the inequality at $x = x_0$ or at $x = 0$ is binding in an optimal solution. We have two cases:

(a) If $b \geq \frac{1}{2a}$, then $-\frac{1}{4a} + b - k = 0$ (binding constraint at x_0);

Substituting $a = \frac{1}{4(b - k)}$ in the objective, we obtain:

$$Z_D = \min_b \frac{((\mu - k) + (b - k))^2 + \sigma^2}{4(b - k)} = \frac{1}{2} \left[(\mu - k) + \sqrt{\sigma^2 + (\mu - k)^2} \right],$$

achieved at $b_0 = \frac{\mu^2 + \sigma^2}{\mu}$. This bound is valid whenever $b_0 \geq \frac{1}{2a_0} = 2(b_0 - k)$, that is

$$\frac{\mu^2 + \sigma^2}{2\mu} \leq k.$$

(b) If $b < \frac{1}{2a}$, then $ab^2 - k = 0$ (binding constraint at $x = 0$).

Substituting $a = \frac{k}{b^2}$ in the objective, we obtain:

$$Z_D = \min_b \frac{k}{b^2} (\mu^2 + \sigma^2) - 2\frac{k}{b} \mu + \mu = \mu - k \frac{\mu^2}{\mu^2 + \sigma^2},$$

achieved at $b_0 = \frac{\mu^2 + \sigma^2}{\mu}$.

This bound is valid whenever $b_0 < \frac{1}{2a_0} = \frac{b_0^2}{2k}$, that is $\frac{\mu^2 + \sigma^2}{2\mu} > k$. ■

5 Moment Problems in Discrete Optimization

In this section, we explore the connection of moment problems and discrete optimization. We consider the maximum $s - t$ cut problem. Goemans and Williamson [13] showed that a natural semidefinite relaxation is within 0.878 of the value of the maximum $s - t$ cut. Bertsimas and Ye [10] provide an alternative interpretation of their method that makes the connection of moment problems and discrete optimization explicit. We review this development in this section.

Given an undirected graph on n nodes, and weights c_{ij} on the edges we would like to find an $s - t$ cut of maximum weight. We formulate the problem as follows:

$$\begin{aligned} Z_{IP} = \text{maximize} \quad & \frac{1}{2} \sum_{i,j} c_{ij} (1 - x_i x_j) \\ \text{subject to} \quad & x_s + x_t = 0, \\ & x_j^2 = 1, \quad j = 1, \dots, n, \end{aligned} \tag{37}$$

and consider the semidefinite relaxation

$$\begin{aligned} Z_{SD} = \text{maximize} \quad & \frac{1}{2} \sum_{i,j} c_{ij} (1 - y_{ij}) \\ \text{subject to} \quad & y_{st} = 0, \\ & y_{jj} = 1, \quad j = 1, \dots, n, \\ & \mathbf{Y} \succeq \mathbf{0}. \end{aligned} \tag{38}$$

We solve the semidefinite relaxation (38) and obtain the semidefinite matrix \mathbf{Y} . In order to understand the closeness of Z_{SD} and Z_{IP} , we create a feasible solution to the $s - t$ maximum cut problem by interpreting the matrix \mathbf{Y} as a covariance matrix.

Randomized heuristic H :

1. We generate a vector $\bar{\mathbf{x}}$ from a **multivariate normal distribution** with 0 mean and covariance matrix \mathbf{Y} , that is,

$$\bar{\mathbf{x}} \sim N(0, \mathbf{Y}).$$

2. We create a vector $\hat{\mathbf{x}}$ with components equal to 1 or -1:

$$\hat{\mathbf{x}} = \text{sign}(\bar{\mathbf{x}}), \tag{39}$$

i.e., $\hat{x}_j = 1$ if $\bar{x}_j > 0$, and $\hat{x}_j = -1$ if $\bar{x}_j \leq 0$.

Notice that instead of using a multivariate normal distribution for $\bar{\mathbf{x}}$ we can use any distribution that has covariance $Cov(\bar{\mathbf{x}}) = \mathbf{Y}$. What is interesting is that we show the degree of closeness of Z_{SD} and Z_{IP} by considering results regarding the normal distributed that were known in 1900.

Proposition 1

$$\begin{aligned} E[\hat{x}_j] &= 0, \quad E[\hat{x}_j^2] = 1, \quad j = 1, 2, \dots, n, \\ E[\hat{x}_i \hat{x}_j] &= \frac{2}{\pi} \arcsin(y_{ij}), \quad i, j = 1, 2, \dots, n. \end{aligned}$$

Proof:

The marginal distribution of \bar{x}_i is $N(0, 1)$, and thus $P(\hat{x}_i = 1) = P(\hat{x}_i = -1) = 1/2$. Thus, $E[\hat{x}_i] = 0$ and $E[\hat{x}_i^2] = 1$. Furthermore,

$$\begin{aligned} E[\hat{x}_i \hat{x}_j] &= P(\hat{x}_i = 1, \hat{x}_j = 1) + P(\hat{x}_i = -1, \hat{x}_j = -1) - P(\hat{x}_i = 1, \hat{x}_j = -1) - P(\hat{x}_i = -1, \hat{x}_j = 1) \\ &= P(x_i \geq 0, x_j \geq 0) + P(x_i < 0, x_j < 0) - P(x_i \geq 0, x_j < 0) - P(x_i < 0, x_j \geq 0). \end{aligned}$$

The tail probabilities of a multivariate normal distribution is a problem that has been studied in the last 100 years. Sheppard [28] shows (see Johnson and Kotz [18], p. 95) that

$$\begin{aligned} P(x_i \geq 0, x_j \geq 0) &= P(x_i < 0, x_j < 0) = \frac{1}{4} + \frac{1}{2\pi} \arcsin(y_{ij}) \\ P(x_i \geq 0, x_j < 0) &= P(x_i < 0, x_j \geq 0) = \frac{1}{4} - \frac{1}{2\pi} \arcsin(y_{ij}). \end{aligned}$$

This leads to

$$E[\hat{x}_i \hat{x}_j] = \frac{2}{\pi} \arcsin(y_{ij}). \tag{40}$$

■

Theorem 13 *Heuristic H provides a feasible solution for the s – t maximum cut problem with objective value Z_H :*

$$E[Z_H] \geq 0.878Z_{SD}.$$

Proof

First, we notice that $E[\hat{x}_s + \hat{x}_t] = 0$, and

$$E[(\hat{x}_s + \hat{x}_t)^2] = 2 + 2E[\hat{x}_s \hat{x}_t] = 2 + \frac{4}{\pi} \arcsin(-1) = 2 - 2 = 0,$$

i.e., $\hat{x}_s + \hat{x}_t = 0$ with probability 1, i.e., the solution is feasible.

Moreover, the value of the heuristic solution is

$$\begin{aligned} \mathbb{E}[Z_H] &= \sum_{ij} c_{ij} \left(P(\hat{x}_i = 1, \hat{x}_j = -1) + P(\hat{x}_i = -1, \hat{x}_j = 1) \right) \\ &= \sum_{ij} c_{ij} \left(\frac{1}{2} - \frac{1}{\pi} \arcsin(y_{ij}) \right) \\ &\geq 0.878 \frac{1}{2} \sum_{ij} c_{ij} (1 - y_{ij}) \\ &= 0.878 Z_{SD}, \end{aligned}$$

where we used the inequality $\frac{1}{2} - \frac{1}{\pi} \arcsin(y_{ij}) \geq (0.878) \frac{1}{2} (1 - y_{ij})$ from Goemans and Williamson [13]. ■

Bertsimas and Ye [10] show that the interpretation of the matrix \mathbf{Y} as a covariance matrix leads to interesting bounds for several problems, such as the graph bisection problem, the s, t, u maximum cut problem, and constrained quadratic maximization. If the distribution used is the multivariate normal distribution the method is equivalent to the Goemans and Williamson [13] method. However, it would be interesting to explore the generation of random variables $\bar{\mathbf{x}}$ using a distribution other than the multivariate normal distribution. For the max cut problem, the bound 0.878 is not known to be tight. Generating a random vector using a distribution other than the normal, might lead to a sharper bound.

6 Concluding Remarks

We would like to leave the reader with the following closing thoughts:

- (a) Convex optimization has a central role to play in moment problems arising in probability and finance. Optimization not only offers a natural way to formulate and study such problems, it leads to unexpected improvements of classical results.
- (b) Semidefinite optimization represents a promising direction for further research in the area of stochastic optimization. The two major areas in which semidefinite programming has had a very positive impact are discrete optimization and control theory. Paralleling the development in discrete optimization, we believe that semidefinite programming can play an important role in stochastic optimization by increasing our ability to find better lower bounds.

References

- [1] N. I. Akhiezer. *The Classical Moment Problem*. Hafner, New York, NY, 1965.
- [2] F. Avram, D. Bertsimas, and M. Ricard. Optimization of multiclass queueing networks: a linear control approach. In *Stochastic Networks, Proceedings of the IMA*, pages 199–234. F. Kelly and R. Williams eds., 1995.
- [3] D. Bertsimas. The achievable region method in the optimal control of queueing systems; formulations, bounds and policies. *Queueing Systems*, 21:337–389, 1995.
- [4] D. Bertsimas and J. Niño-Mora. Conservation laws, extended polymatroids and multi-armed bandit problems; a polyhedral approach to indexable systems. *Math. Oper. Res.*, 21:257–306, 1996.
- [5] D. Bertsimas and J. Niño-Mora. Optimization of multiclass queueing networks with changeover times via the achievable region approach: Part I, the single-station case. *Math. Oper. Res.*, to appear.
- [6] D. Bertsimas and J. Niño-Mora. Optimization of multiclass queueing networks with changeover times via the achievable region approach: Part II, the multi-station case. *Math. Oper. Res.*, to appear.
- [7] D. Bertsimas, I. Paschalidis, and J. Tsitsiklis. Optimization of multiclass queueing networks: polyhedral and nonlinear characterizations of achievable performance. *Ann. Appl. Probab.*, 4:43–75, 1994.
- [8] D. Bertsimas and I. Popescu. On the relation between option and stock prices: an optimization approach. Submitted for publication, 1998.
- [9] D. Bertsimas and I. Popescu. Optimal inequalities in probability: a convex programming approach. Submitted for publication, 1998.
- [10] D. Bertsimas and Y. Ye. Semidefinite relaxations, multivariate normal distributions, and order statistics. In *Handbook of Combinatorial Optimization (Vol. 3)*, volume 3, pages 1–19. Kluwer Academic Publishers, D.-Z. Du and P.M. Pardalos (Eds.), 1998.

- [11] J. Cox and S. Ross. The valuation of options for alternative stochastic processes. *Journal of Financial Economics*, 3:145–166, 1976.
- [12] K. Fujisawa, M. Kojima, and K. Nakata. SDPA (Semidefinite Programming Algorithm User’s Manual, Version 4.10. Research Report on Mathematical and Computing Sciences, Tokyo Institute of Technology, Tokyo, Japan, 1998.
- [13] M.X. Goemans and D.P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of ACM*, 42:1115–1145, 1995.
- [14] M. Grötschel, L. Lovász, and A. Schrijver. The ellipsoid method and its consequences in combinatorial optimization. *Combinatorica*, 1:169–197, 1981.
- [15] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric Algorithms and Combinatorial Optimization*. Algorithms and Combinatorics; 2. Springer, Berlin-Heidelberg, 1988.
- [16] M. Harrison and D. Kreps. Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory*, 20:381–408, 1979.
- [17] K. Isii. On sharpness of chebyshev-type inequalities. *Ann. Inst. Stat. Math.*, 14:185–197, 1963.
- [18] N. Johnson and S. Kotz. *Distributions in Statistics: Continuous Multivariate Distributions*. John Wiley & Sons, New York, NY, 1972.
- [19] S. Karlin and L.S. Shapley. Geometry of moment spaces. *Memoirs Amer. Math. Soc.*, 12, 1953.
- [20] J. H. B. Kempermann. On the role of duality in the theory of moments. In *Lecture Notes in Econom. and Math. Systems*, volume 215, pages 63–92, Berlin-New York, 1983. Springer.
- [21] S. Kumar and P.R. Kumar. Performance bounds for queueing networks and scheduling policies. *IEEE Trans. Autom. Control*, 39:1600–1611, 1994.
- [22] A. Lo. Semiparametric upper bounds for option prices and expected payoffs. *Journal of Financial Economics*, 19:373–388, 1987.

- [23] L. Lovász and A. Schrijver. Cones of matrices and set-functions and 0–1 optimization. *SIAM J. Optim.*, 1:166–190, 1991.
- [24] X. Luo and D. Bertsimas. A new algorithm for state constrained separated continuous linear programs. *SIAM J. Contr. Optim.*, 1:177–210, 1998.
- [25] K.G. Murty and S.N. Kabadi. Some np-complete problems in quadratic and nonlinear programming. *Math. Progr.*, 39:117–129, 1987.
- [26] C.H. Papadimitriou and J.N. Tsitsiklis. The complexity of optimal queueing network control. *Math. Oper. Res.*, 1, 1998.
- [27] S. Sengupta, W. Siu, and N.C. Lind. A linear programming formulation of the problem of moments. *Z. Angew. Math. Mecg.*, 10:533–537, 1979.
- [28] W.F. Sheppard. On the calculation of the double integral expressing normal correlation. *Trans. Cambr. Phil. Soc.*, 19:23–66, 1900.
- [29] J. Smith. Generalized chebyshev inequalities: theory and applications in decision analysis. *Oper. Res.*, 43:807–825, 1995.
- [30] G. Weiss. On the optimal draining of re-entrant fluid lines. In *Stochastic Networks, Proceedings of the IMA*, pages 91–104. F. Kelly and R. Williams eds., 1995.