Konrad-Zuse-Zentrum
für Informationstechnik Berlin

Takustraße 7
D-14195 Berlin-Dahlem
Germany

KURT HELMES[1] STEFAN RÖHL[2]

A Geometrical Characterization of the Multidimensional Hausdorff and Dale Polytopes with Applications to Exit Time Problems

[1]Institut für Operations Research, Humboldt-Universität zu Berlin, Spandauer Str. 1, 10178 Berlin, Germany, helmes@wiwi.hu-berlin.de

[2]Konrad-Zuse-Zentrum für Informationstechnik Berlin, roehl@zib.de

ZIB-Report 04-05 (February 2004)

A Geometrical Characterization of the Multidimensional Hausdorff and Dale Polytopes with Applications to Exit Time Problems

Kurt Helmes
Institut für Operations Research, Humboldt-Universität zu Berlin, Germany,
helmes@wiwi.hu-berlin.de

Stefan Röhl
Konrad-Zuse-Zentrum für Informationstechnik Berlin, Germany, roehl@zib.de

Abstract. We present formulae for the corner points of the multidimensional Hausdorff and Dale Polytopes and show how these results can be used to improve linear programming models for computing e. g. moments of exit distribution of diffusion processes. Specifically, we compute the mean exit time of twodimensional Brownian motion from the unit square and the unit triangle, as well as higher moments of the exit time of time space Brownian motion from a triangle.

Key Words. Linear programming, special polytopes, moment problems, Brownian motion on a triangle, numerical methods for exit time problems.

Subject Classification (AMS 2000). *Primary* 52B12. *Secondary* 90C08, 44A60, 60J65, 60J70.

1 Introduction

In articles by Cho (2000), Cho and Stockbridge (2002), Helmes (1999), (2002), Helmes et al. (2001), Helmes and Stockbridge (2000), (2001), (2003), and Röhl (2001) numerical methods for the computational analysis of exit time problems, invariant distributions of diffusions and optimal stopping and control problems have been proposed which are based on a linear programming approach to these kind of problems. The formulation of infinite dimensional linear programs for such problems is an extension of work by Manne (1960) who initiated the formulation of stochastic control problems as linear programs over a space of stationary distributions for the long term average control of finite state Markov chains, see Hernandez-Lerma et al. (1991) for details and additional references. The generalization of the LP formulation for continuous time, general state and control spaces, and different objective functions has been established by Stockbridge (1990), Kurtz and Stockbridge (1998), (1999), and Bhatt and Borkar (1996).

The basic idea of the LP approach to the analysis of controlled and uncontrolled Markov processes is to formulate such problems as linear programs over a space of stationary distributions. Specifically, the variables in these infinite dimensional linear programs are measures on the product of the state and control spaces and in the case of exit problems, each such variable is augmented by a second measure on the exterior of the state space. These variables are constrained by equations involving the generator of the Markov process and a family of test functions. Different numerical methods are determined by a judious choice of a finite set of test functions combined with a selection of a finite number of variables and/or restrictions imposed on the support of the occupation measure and the exterior measure. Such choices determine approximations of the infinite dimensional optimization problem by finite dimensional ones.

One class of approximating problems exploits the characterization of measures on bounded intervals by their moments and the identification of moment sequences by a countable family of linear inequality conditions. Hausdorff (1921) and (1923) formulated these inequalities for measures on the interval $[0,1]$. We therefore call these inequality conditions the Hausdorff Conditions, see Section 2. Hildebrandt and Schoenberg (1933) generalized these results to the multidimensional case, i. e. to measures with support in $[0,1]^d$, $d \geq 1$. For some applications of moment theory see Ang et al. (2002). Using but a finite number of these inequalities to partially describe the feasible set of finite dimensional LPs, cf. Section 4, leads us to study the geometry of what we call d-dimensional Hausdorff Polytopes. Specifically, we are interested in formulae for the vertices of these polytopes since, as will be illustrated by numerical examples, such formulae enhance the accuracy of the numerical methods to which we referred above. The geometry of moment spaces for the onedimensional case was first considered in detail in the paper by Karlin and Shapley (1953). To approximate moment sequences they introduced special simplices defined as the convex hull of specific points. In Section 2 we prove that the d-dimensional Hausdorff Polytopes are in fact extensions to higher dimensions of the simplices described by Karlin and Shapley.

Since, for dimension 1, the Hausdorff Polytopes of order n contain the first $n+1$ components of all moment sequences we work with an *outer* approximation of the projection of moment sequences onto $\mathbb{R}^{n+1}$. This approximation greatly differs from the *inner* approximation of this set by cyclic polytopes, cf. Ziegler (1995) and also Karlin and Shapley (1953). Using an outer approximation ensures that no restrictive assumptions on the support of the occupation measures of Markov processes to be analyzed need to be made, cf. Section 4. As will be seen in Sections 2 and 3 these ideas can be generalized to higher dimensions.

An alternative to the finite dimensional linear programs referred to above was recently proposed by Lasserre and Rumeau (2003). Instead of LPs they suggest using semidefinite programs, cf. also Schwerer (1996).

While Hausdorff Polytopes are associated with measures whose support is contained in a hypercube Dale Polytopes, to be introduced in Section 3, are associated with measures defined on the unit triangle $S^d = \{x \in [0,1]^d \mid \sum_{i=1}^{d} x_i \leq 1\}$. Dale (1987) gave necessary and sufficient conditions for a doubly indexed sequence to be the sequence of (joint) moments of a measure on the twodimensional triangle S^2. This result will be extended to higher dimensions and will be complemented by formulae for the vertices of the polytopes. Stockbridge (2003) recently extended Dale's result to measures defined on general polytopes and classes of more general bounded regions.

This paper is organized as follows. Section 2 presents our results on Hausdorff Polytopes while Section 3 does so for Dale Polytopes. Theorem 2.1 and 3.1 are the fundamental convergence results which reveal how distributions on $[0,1]^d$ and S^d can be recovered from their moments. They show that the reconstruction of such measures from their moments can be given in terms of a sequence of discrete distributions. These distributions incorporate iterated differences of the moments in an essential way. But requiring iterated differences (up to a finite order, cf. (8) for the Hausdorff case and (22) for the Dale case) to be nonnegative are the defining inequalities of Hausdorff and Dale Polytopes. The connection between the vertices of these polytopes and Dirac measures on $[0,1]^d$ and S^d, given in terms of iterated differences, is described by Proposition 2.6 and Proposition 3.6, while the most important results, the formulae for the corner points, can be found in Theorems 2.5 and 3.5. The basic ideas of the proofs in both sections are very much alike, and we introduce convenient notation in Section 2 which emphasizes this fact and which should facilitate reading the proofs. Section 4 includes numerical illustrations of the theory developed in Sections 2 and 3 and briefly recapitulates the formulation of the LP approach for these examples.

2 The multidimensional Hausdorff Polytope

For multiindices $j = (j_1, \ldots, j_d) \in \mathbb{Z}_+^d$, $n = (n_1, \ldots, n_d) \in \mathbb{Z}_+^d$, and a vector $x \in \mathbb{R}^d$ we use the following abbreviating notation:

$$(1)\qquad \binom{n}{j} := \prod_{i=1}^{d} \binom{n_i}{j_i}, \qquad x^j := \prod_{i=1}^{d} x_i^{j_i}, \quad \text{and} \quad \sum_{j=0}^{n} := \sum_{j_1=0}^{n_1} \cdots \sum_{j_d=0}^{n_d}.$$

For a given real valued function u on $E^d = [0,1]^d$ and multiindex $n \in \mathbb{Z}_+^d$ we call the polynomial

$$(2)\qquad B_{n,u}(x) := \sum_{j=0}^{n} u_j^{(n)} \binom{n}{j} x^j \, (\mathbf{1} - x)^{n-j}, \qquad \text{for } x \in E^d,$$

the Bernstein Polynomial of degree n corresponding to u where $u_j^{(n)} = u(\frac{j}{n}) := u(\frac{j_1}{n_1}, \ldots, \frac{j_d}{n_d})$ and $\mathbf{1} = (1, \ldots, 1)$ is the main diagonal vector in E^d (cf. Knill (1997)).

For any finite or infinite multiindexed sequence $\{x_n\}$ we define the differences

$$(\triangle_i^1 x)_{(n_1,\ldots,n_d)} := x_{(n_1,\ldots,n_i,\ldots,n_d)} - x_{(n_1,\ldots,n_i+1,\ldots,n_d)}, \quad i = 1, \ldots, d,$$

for all indices for which the right hand side is well defined. Note that we are using backward differences and therefore follow the same sign convention as Dale (1987) and

Knill (1997), in contrast to the classical notation used by Feller (1971). This convention avoids unwanted factors in some of the expressions below. Using the notation

$$(\triangle_1^0 \ldots \triangle_d^0\, x)_{(n_1,\ldots,n_d)} \;:=\; x_{(n_1,\ldots,n_d)}\,,$$

we define the iterative differences of higher order as follows:

$$(3) \qquad (\triangle_1^{j_1} \ldots \triangle_i^{j_i+1} \ldots \triangle_d^{j_d}\, x)_{(n_1,\ldots,n_d)} \;:=\; (\triangle_i^1(\triangle_1^{j_1} \ldots \triangle_i^{j_i} \ldots \triangle_d^{j_d}\, x))_{(n_1,\ldots,n_d)} \;=\; (\triangle_1^{j_1} \ldots \triangle_i^{j_i} \ldots \triangle_d^{j_d}\, x)_{(n_1,\ldots,n_i,\ldots,n_d)} \;-\; (\triangle_1^{j_1} \ldots \triangle_i^{j_i} \ldots \triangle_d^{j_d}\, x)_{(n_1,\ldots,n_i+1,\ldots,n_d)}\,.$$

For such higher order differences we use the abbreviating notation

$$(4) \qquad \triangle^j x_n \;:=\; (\triangle_1^{j_1} \ldots \triangle_d^{j_d}\, x)_{(n_1,\ldots,n_d)},$$

where $j = (j_1, \ldots, j_d)$, and $n = (n_1, \ldots, n_d)$ are multiindices.

Let T_i, $1 \le i \le d$, denotes the shift operator applied to the i-th coordinate, i. e.

$$(T_i\, x)_n = x_{n_1,\ldots,n_{i-1},n_i+1,n_{i+1},\ldots,n_d}.$$

Then we can write the operator $\triangle^j$ as a product of differences of simple commuting operators, viz.

$$\triangle^j = \prod_{i=1}^{d} (\triangle^{\mathbf{0}} - T_i)^{j_i}\,.$$

Let X be a random variable on $E^d = [0,1]^d$ distributed according to a measure μ having distribution function F and moments $\underline{\mu} = \{\mu_j\}_{j\in\mathbb{Z}_+^d}$, i.e.

$$\mu_j = \mu_{(j_1,\ldots,j_d)} = \int_{E^d} x_1^{j_1} x_2^{j_2} \cdots x_d^{j_d}\; dF(x_1, \ldots, x_d).$$

By induction over the sum of the components of k and m we see that for all $k, m \in \mathbb{Z}_+^d$ the following equation holds

$$(5) \qquad (\triangle^m \underline{\mu})_k \;=\; \int_{E^d} x^k (\mathbf{1} - x)^m \, dF(x)\,.$$

Integrating Equation (2) with respect to F we obtain the identity,

$$(6) \qquad \mathbf{E}_F\, B_{n,u} \;=\; \sum_{j=0}^{n} u_j^{(n)} \binom{n}{j} (\triangle^{n-j} \underline{\mu})_j.$$

We define

$$p_j^{(n)} \;:=\; \binom{n}{j} (\triangle^{n-j} \underline{\mu})_j \qquad \text{for} \quad j = (j_1, \ldots, j_d)\,,\; 0 \le j_i \le n_i.$$

It follows from Equation (5) that $p_j^{(n)} \geq 0$ for all $j \leq n$. Using Definition (2) with the function $u(x) \equiv 1$ we obtain $\mathbf{E}_F \, B_{n,u} = 1$, and hence

$$\sum_{j=0}^{n} p_j^{(n)} = 1 \,. \tag{7}$$

So we may interpret the vector $\{p_j^{(n)}\}_{j=0}^n$ as a discrete probability measure $\mu^{(n)}$ with distribution function $F^{(n)}$ on the set of points $\{(\frac{j}{n})\}_{j=0}^n = \{(\frac{j_1}{n_1}, \ldots, \frac{j_d}{n_d})\}_{j=0}^n$. Here the notation $\{.\}_{j=0}^n$ means $\{\, . \mid 0 \leq j_1 \leq n_1 \,, \, \ldots \,, \, 0 \leq j_d \leq n_d\}$.

This construction of discrete measures $\mu^{(n)}$ together with the following two convergence and characterization results are the motivation and justification of Defintion 2.3 below, and the numerical methods to be used in Section 4.

Theorem 2.1 *Let u be a continuous function and F be a distribution function on $E^d = [0,1]^d$. Then the following two properties hold:*

1. *For $n \to \infty$, i.e. $n_i \to \infty$ for all $i = 1, \ldots, d$, the Bernstein Polynomials $B_{n,u}$ converge uniformly to the function u.*
2. *For every $x \in E^d$ where the function F is continuous*

$$F^{(n)}(x) \;:=\; \sum_{j \leq nx} p_j^{(n)} \underset{n \to \infty}{\longrightarrow} F(x).$$

For the proofs of Theorems 2.1 and 2.2, see below, we refer the reader to Hildebrandt and Schoenberg (1933) or Knill (1997), cf. also the proofs of Theorems 3.1 and 3.2 below.

The following d-dimensional Hausdorff Conditions are straightforward generalizations of the conditions for the one- and twodimensional cases (cf. Shohat and Tamarkin (1943) p. 9 ff or the references above):

Theorem 2.2 *A multiindexed sequence $\underline{\mu} = \{\mu_j\}_{j \geq 0}$ of real numbers is a sequence of moments of a measure μ on $E^d = [0,1]^d$ iff*

$$(\triangle^m \underline{\mu})_k \;\geq\; 0 \tag{8}$$

or, equivalently,

$$\sum_{s=0}^{m} (-\mathbf{1})^s \binom{m}{s} \mu_{k+s} \;\geq\; 0 \tag{9}$$

for all multiindices $k, m \in \mathbb{Z}_+^d$.

Theorem 2.1 and Theorem 2.2 suggest the following definitions.

Definition 2.3

1. *For each element* $z \in \mathbb{R}^{\tilde{n}}$, $\tilde{n} = (n_1+1) \cdot \ldots \cdot (n_d+1)$, *we define the linear transformation* $R^{(n)} : \mathbb{R}^{\tilde{n}} \to \mathbb{R}^{\tilde{n}}$ *by*

$$\left(R^{(n)}(z)\right)_j \;:=\; \binom{n}{j}(\triangle^{n-j}z)_j \qquad \textit{for} \quad j = (j_1, \ldots, j_d)\,,\; 0 \le j_i \le n_i.$$

2. *Let* $\tilde{\mathcal{H}}_n^d \subset \mathbb{R}^{\tilde{n}}$ *be the set of all arrays* $\{z_j\}_{j=0}^n$ *fulfilling the Hausdorff Conditions up to order* n*, i.e. we require the Inequalities (8) or (9) to hold only for multiindices* k *and* m *which satisfy* $k+m \le n$. *The set* $\tilde{\mathcal{H}}_n^d$ *is called the d-dimensional Hausdorff Polygon of order* n.

3. *For* $c > 0$ *we define the set*

$$\mathcal{H}_{n,c}^d := \left\{ z \in \tilde{\mathcal{H}}_n^d \mid z_{(0,\ldots,0)} = c \right\}.$$

The set $\mathcal{H}_n^d := \mathcal{H}_{n,1}^d$ *is called the d-dimensional Hausdorff Polytope of order* n.

4. *Let* $\mathcal{K}_n^d(\partial E^d) := \left\{ \sum_\varphi \nu^\varphi \;\middle|\; \nu^\varphi \in \mathcal{H}_{n,c_\varphi}^{d-1},\ \sum_\varphi c_\varphi = 1,\ \varphi \textit{ runs over all } (n-1)\textit{-dimensional facets of } E^d \right\}$, *cf. Section 4.*

Remark: The transformation $R^{(n)}$ is an extension of the mapping $\{\mu_j\}_{j=0}^n \mapsto p_j^{(n)}$, see above, where $\{\mu_j\}_{j=0}^n$ is the truncated sequence of moments of a finite measure on $[0,1]^d$. For any finite measure μ we use the shorthand writing $R^{(n)}(\mu)$ to denote the image of the finite sequence of moments of μ.

Lemma 2.4 *For a sequence* $z = \{z_j\}_{j\ge 0}$ *the Hausdorff Conditions up to order* $n = (n_1, \ldots, n_d)$ *are equivalent to the following reduced number of conditions (10) or (11):*

$$(\triangle^{n-k}z)_k \;\ge\; 0\,, \tag{10}$$

or, equivalently,

$$\sum_{i=0}^{n-k} (-\mathbf{1})^i \binom{n-k}{i} z_{k+i} \;\ge\; 0 \tag{11}$$

for all multiindices k *with* $0 \le k \le n$ *component-wise.*

Proof. These conditions are obviously necessary. By repeatedly applying Definition (3) of the iterative differences and using Inequalities (10) it is easy to derive all missing Hausdorff Conditions (8) up to the given order. Therefore Conditions (10) are also sufficient, as are Conditions (11). $\square$

Theorem 2.5 *The Hausdorff Polytope $\mathcal{H}_n^d$ has $\tilde{n} := \prod_{i=1}^{d}(n_i+1)$ corner points. The corner point $z^{(k)}$ associated with the multiindex $k = (k_1, \ldots, k_n)$, $0 \le k_i \le n_i$, $i = 1, \ldots, d$, has the following coordinates (which are arranged according to the lexicographic order)*

$$z^{(k)} \;=\; \left(z^{(k)}_{(0,\ldots,0,0)}, \ldots, z^{(k)}_{(0,\ldots,0,n_d)}, \ldots, z^{(k)}_{(n_1,\ldots,n_{d-1},0)}, \ldots, z^{(k)}_{(n_1,\ldots,n_{d-1},n_d)} \right),$$

where

$$z_j^{(k)} \;=\; \prod_{i=1}^{d} \binom{n_i}{k_i}^{-1} \binom{n_i - j_i}{k_i - j_i} \;=\; \binom{n}{k}^{-1} \binom{n-j}{k-j}; \tag{12}$$

for $t < 0$ we define the binomial coefficient $\binom{s}{t} = 0$ for all $s \in \mathbb{Z}_+$.

Proof. A vector $z = (z_{(0,\ldots,0)}, \ldots, z_{(n_1,\ldots,n_d)})$ is an element of $\mathcal{H}_n^d$ if and only if (i) $z_{(0,\ldots,0)} = 1$ and (ii), cf. Lemma 2.4, the $\tilde{n} = \prod_{i=1}^{d}(n_i+1)$ inequalities

$$\sum_{j=0}^{n-l} (-\mathbf{1})^j \binom{n-l}{j} z_{l+j} \;\ge\; 0 \tag{13}$$

are satisfied for all multiindices l with $0 \le l \le n$. In order to show that the set of vectors $z^{(k)}$, see (12), is the set of all corner points of $\mathcal{H}_n^d$ we need to prove that, first of all, $z^{(k)}_{(0,\ldots,0)} = 1$. Moreover, since the vectors $z^{(k)}$ are linearly independent, we need to show that for all vectors $z^{(k)} = (1, z^{(k)}_{(0,\ldots,0,1)}, \ldots, z^{(k)}_{(n_1,\ldots,n_d)}) \in \{1\} \times \mathbb{R}^{\tilde{n}-1}$, $\tilde{n}-1$ of the Inequalities (13) become equations while the remaining inequality is a strict inequality.

Let $k \le n$ be given, then the equation $z_0^{(k)} = 1$ trivially holds. Moreover, we know that $z_l^{(k)} = 0$ for all multiindices l for which $l_i > k_i$ for at least one $i \in \{1, \ldots, d\}$. For such indices l Inequality (13) is actually an equation, because on the left hand side of (13) we only sum over multiindices which satisfy this index condition. For the case $l = k$ the left hand side of (13) reduces to

$$\sum_{j=0}^{n-k} (-\mathbf{1})^j \binom{n-k}{j} z^{(k)}_{k+j} \;=\; (-\mathbf{1})^0 \binom{n-k}{0} z^{(k)}_k \;=\; \binom{n}{k}^{-1},$$

which is obviously a positive number.

It remains to be shown that the numbers $\{z_j^{(k)}\}_{j \le k}$ satisfy

$$\sum_{j=0}^{k-l} (-\mathbf{1})^j \binom{n-l}{j} z^{(k)}_{l+j} \;=\; 0 \qquad \text{for} \quad l \le k, \;\; l \ne k\,.$$

By definition,

$$\sum_{j=0}^{k-l} (-\mathbf{1})^j \binom{n-l}{j} z^{(k)}_{l+j} \;=\; \binom{n}{k}^{-1} \sum_{j=0}^{k-l} (-\mathbf{1})^j \binom{n-l}{j} \binom{n-l-j}{k-l-j}.$$

For onedimensional binomial coefficients we know that

$$\binom{r}{s}\binom{r-s}{t-s} = \binom{t}{s}\binom{r}{t};$$

this identity easily extends to the multidimensional case by forming products on both sides. Hence, since $l \neq k$, we get

$$\begin{aligned}\sum_{j=0}^{k-l}(-\mathbf{1})^j\binom{n-l}{j} z_{l+j}^{(k)} &= \binom{n}{k}^{-1}\binom{n-l}{k-l}\sum_{j=0}^{k-l}(-\mathbf{1})^j\binom{k-l}{j} \\ &= \binom{n}{k}^{-1}\binom{n-l}{k-l}(\mathbf{1}-\mathbf{1})^{k-l} \;=\; 0.\end{aligned}$$

□

Proposition 2.6 *Let n and k be given multiindices satisfying $k \leq n$ component-wise. The transformation $R^{(n)}$ applied to the corner points $z^{(k)}$ of the Hausdorff Polytope $\mathcal{H}_n^d$ yields the set of unit vectors, i. e. $j = (j_1, \ldots, j_d) \leq n$,*

$$R^{(n)}(z^{(k)})_j = \begin{cases} 1 & \textit{for} \quad j = k \\ 0 & \textit{otherwise.} \end{cases}$$

Proof. The Hausdorff Polytope $\mathcal{H}_n^d$ is characterized by $\tilde{n}$ inequality conditions. For each corner point $z^{(k)}$ exactly $\tilde{n}-1$ of these conditions are active. We know, cf. proof of Theorem 2.5, that $\sum_{j=0}^{n} R^{(n)}(z^{(k)})_j = 1$ and that for $\tilde{n}-1$ different multiindices $j \leq n$

$$(\triangle^{n-j} z^{(k)})_j = 0.$$

Thus

$$R^{(n)}(z^{(k)})_j = \binom{n}{j}(\triangle^{n-j} z^{(k)})_j = 0$$

for all but one multiindex $m \leq n$. For this particular index m the equation

$$R^{(n)}(z^{(k)})_m = 1$$

holds. But

$$\begin{aligned}R^{(n)}(z^{(k)})_k &= \binom{n}{k}(\triangle^{n-k} z^{(k)})_k \\ &= \binom{n}{k}\sum_{s=0}^{n-k}(-\mathbf{1})^s\binom{n-k}{s} z_{k+s}^{(k)} \\ &= \binom{n}{k}\binom{n-k}{0}\binom{n}{k}^{-1}\binom{n-k}{0} \\ &= 1,\end{aligned}$$

since, by definition of $z^{(k)}$, all other terms of this sum disappear. Hence, $m = k$ and $R^{(n)}(z^{(k)})$ is a unit vector. $\square$

So we can think of the image of $R^{(n)}$ of the corner point $z^{(k)}$ as being the Dirac measure at the point $(\frac{k_1}{n_1}, \ldots, \frac{k_d}{n_d})$, and we can think of the range $R^{(n)}(\mathcal{H}_n^d)$ as the set of all discrete probability measures on the set of points $\{(\frac{j}{n})\}_{j=0}^n$.

3 The multidimensional Dale Polytope

Let $S^d \subset \mathbb{R}^d$, $d \in \mathbb{N}$, denote the d-dimensional triangle, i. e.

$$S^d = \left\{ x \in [0,1]^d \;\middle|\; \Sigma(x) \leq 1 \right\},$$

and let $\mathcal{I}_N^d = \{j = (j_1, \ldots, j_d) \in \mathbb{Z}_+^d \mid \Sigma(j) \leq N\}$ for $d, N \in \mathbb{N}$. For vectors $x \in \mathbb{R}^d$, multiindices $j \in \mathbb{Z}_+^d$ resp., we use the notation $\Sigma(x) := \sum_{i=1}^d x_i$, $\Sigma(j) := \sum_{i=1}^d j_i$ resp.

For a real valued function $\tilde{u}$ on S^d and single index N we call the polynomial

$$\tilde{B}_{N,\tilde{u}}(x) \;:=\; \sum_{j \in \mathcal{I}_N^d} \tilde{u}_j^{(N)} \begin{bmatrix} N \\ j \end{bmatrix} x^j \; (1 - \Sigma(x))^{N - \Sigma(j)}\,, \qquad x \in S^d, \tag{14}$$

the modified Bernstein Polynomial of degree N associated with $\tilde{u}$. We define, cf. Section 2, $\tilde{u}_j^{(N)} := \tilde{u}(\frac{j_1}{N}, \ldots, \frac{j_d}{N})$ and

$$\begin{bmatrix} N \\ j \end{bmatrix} := \begin{bmatrix} N \\ j_1, \ldots, j_d \end{bmatrix} := \frac{N!}{j_1!\; j_2! \cdots j_d!\, (N - \Sigma(j))!}\,, \quad j \in \mathcal{I}_N^d;$$

we call $\begin{bmatrix} N \\ j \end{bmatrix}$ a pseudo multinomial coefficient of dimension d and order N (cf. Dale (1987)). Note that

$$\begin{bmatrix} N \\ j \end{bmatrix} = (\Sigma(j);\, j_1, \ldots, j_d) \binom{N}{\Sigma(j)},$$

where $(\Sigma(j);\, j_1, \ldots, j_d)$ is the multinomial coefficient of j, cf. Abramowitz and Stegun (1965), Section 24.1.2.

For any finite or infinite multiindexed sequence $x = \{x_r\}$ we denote by $\tilde{\triangle}$ the linear transformation $x \mapsto \tilde{\triangle}x$, where

$$(\tilde{\triangle}x)_j \;:=\; x_j - \sum_{i=1}^d x_{j_1,\ldots,j_{i-1},j_i+1,j_{i+1},\ldots,j_d}$$

for all multiindices j for which the right hand side is well defined.

Let $\tilde{\triangle}^0 x :\equiv x$, then the N-th iterate of $\tilde{\triangle}$ is defined by

$$\tilde{\triangle}^N x \;:=\; \tilde{\triangle}(\tilde{\triangle}^{N-1} x) \qquad \text{for } N \in \mathbb{N}.$$

Using the shift operators T_i the operator $\tilde{\triangle}^N$ can be written as the N-th power of a sum of simple commuting operators:

$$\tilde{\triangle}^N \;=\; (\tilde{\triangle}^0 - T_1 - \cdots - T_d)^N\,.$$

By the multinomial theorem we obtain

$$\begin{aligned}
\tilde{\triangle}^N \;&=\; \sum_{\substack{(j_1,\ldots,j_d):\\ \Sigma(j)\leq N}} (-1)^{\Sigma(j)} \begin{bmatrix} N \\ j_1,\ldots,j_d \end{bmatrix} T_1^{j_1}\,\cdots\,T_d^{j_d} \\
&=\; \sum_{j\in\mathcal{I}_N^d} (-\mathbf{1})^j \begin{bmatrix} N \\ j \end{bmatrix} T_1^{j_1}\,\cdots\,T_d^{j_d}\,.
\end{aligned} \tag{15}$$

Let $\tilde{X}$ be a random variable on S^d distributed according to the measure ν with distribution function $\tilde{F}$ and moments $\underline{\nu} = \{\nu_j\}_{j\geq 0}$, i.e.

$$\nu_j \;=\; \int\limits_{S^d} x_1^{j_1} x_2^{j_2}\cdots x_d^{j_d}\; d\tilde{F}(x_1,\ldots,x_d).$$

Then, for all $N \in \mathbb{Z}_+$ and $k \in \mathcal{I}_N^d$, the following equation holds:

$$(\tilde{\triangle}^N \underline{\nu})_k \;=\; \int\limits_{S^d} x^k (1-\Sigma(x))^N\, d\tilde{F}(x)\,. \tag{16}$$

Integrating Equation (14) with respect to $\tilde{F}$ we obtain

$$\mathbf{E}_{\tilde{F}}\,\tilde{B}_{N,\tilde{u}} \;=\; \sum_{j\in\mathcal{I}_N^d} \tilde{u}_j^{(N)} \begin{bmatrix} N \\ j \end{bmatrix} (\tilde{\triangle}^{N-\Sigma(j)}\underline{\nu})_j. \tag{17}$$

Next, we define

$$\tilde{p}_j^{(N)} \;:=\; \begin{bmatrix} N \\ j \end{bmatrix} (\tilde{\triangle}^{N-\Sigma(j)}\underline{\nu})_j \qquad \text{for} \quad j\in\mathcal{I}_N^d$$

and we define the linear transformation $\tilde{R}^{(N)}$ as $\nu \mapsto \tilde{R}^{(N)}(\nu) := \left\{\tilde{p}_j^{(N)}\right\}_{j\in\mathcal{I}_N^d}$. Equation (16) implies that $\tilde{p}_j^{(N)} \geq 0$ for all $j\in\mathcal{I}_N^d$. Choosing the function $\tilde{u}(x)\equiv 1$ Equation (14) implies $\mathbf{E}_{\tilde{F}}\,\tilde{B}_{N,\tilde{u}} = 1$; therefore,

$$\sum_{j\in\mathcal{I}_N^d} \tilde{p}_j^{(N)} = 1\,. \tag{18}$$

So we may interpret $\tilde{R}^{(N)}(\nu)$ as a discrete probability measure $\nu^{(N)}$ with distribution function $\tilde{F}^{(N)}$ on the set of points $\{(\frac{j}{N})\}_{j\in\mathcal{I}_N^d}$.

For modified Bernstein Polynomials and the distribution functions $\tilde{F}^{(N)}$ we obtain convergence results similar to Theorem 2.1 (cf. Dale (1987)):

Theorem 3.1 *Let $\tilde{u}$ be a continuous function on S^d, and $\tilde{F}$ a distribution function on S^d. Then the following two assertions hold:*

1. *For $N \to \infty$ the modified Bernstein Polynomials $\tilde{B}_{N,\tilde{u}}$ converge uniformly to the function $\tilde{u}$.*
2. *For all points $x \in S^d$ where the function $\tilde{F}$ is continuous*

$$\tilde{F}^{(N)}(x) \;:=\; \sum_{j\leq Nx} \tilde{p}_j^{(N)} \underset{N\to\infty}{\longrightarrow} \tilde{F}(x).$$

Proof. 1. Let $\{e_k\}_{1\leq k\leq d}$ denote the set of unit vectors in $\mathbb{R}^d$. For fixed $x \in S^d$ we consider a sequence $\{X_i\}_{i\geq 1}$ of i. i. d. discrete random vectors with distribution $P(X_1 = e_k) = x_k$ for $k = 1,\ldots,d$ and $P(X_1 = 0) = 1 - \Sigma(x)$. From the weak law of large numbers it follows that $\frac{1}{N}\sum_{i=1}^{N} X_i \underset{N\to\infty}{\longrightarrow} x$ in distribution. Since $\tilde{u}$ is a continuous function it follows that $\tilde{u}\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right) \underset{N\to\infty}{\longrightarrow} \tilde{u}(x)$ in distribution too. This fact implies the convergence of the expected values, i. e.

$$\mathbf{E}\left[\tilde{u}\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right)\right] \underset{N\to\infty}{\longrightarrow} \tilde{u}(x)\,. \tag{19}$$

By the definition of the random vectors X_i we get:

$$\begin{aligned}\mathbf{E}\left[\tilde{u}\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right)\right] &= \sum_{j\in\mathcal{I}_N^d} \tilde{u}\left(\frac{j}{N}\right)\begin{bmatrix} N \\ j \end{bmatrix} x^j(1-\Sigma(x))^{N-j} \\ &= \tilde{B}_{N,\tilde{u}}(x).\end{aligned} \tag{20}$$

Equation (20) together with Equation (19) show the pointwise convergence of the Bernstein polynomials $\tilde{B}_{N,\tilde{u}}$.

Since $\tilde{u}$ is a continuous function on S^d we know that $\tilde{u}$ is bounded and uniformly continuous, i. e. $|\tilde{u}(x)| \leq \gamma$ for all $x \in S^d$, and for any $\varepsilon > 0$ there exists a $\delta > 0$ such that $|\tilde{u}(x) - \tilde{u}(y)| < \varepsilon$ for all $x, y \in S^d$ whenever $\|x - y\| < \delta$. Let A_x, $x \in S^d$, be the following event:

$$A_x = \left\{\omega : \left\|\frac{1}{N}\sum_{i=1}^{N} X_i(\omega) - x\right\| < \delta\right\};$$

we denote the complement by $\bar{A}_x$. By Tchebychev's Inequality we have for $x \in \mathcal{S}^d$

$$\begin{aligned}
P(\bar{A}_x) &= P\left(\left\|\frac{1}{N}\sum_{i=1}^{N} X_i - x\right\| \geq \delta\right) \\
&\leq \frac{1}{\delta^2}\,\mathbf{E}\left[\left\|\frac{1}{N}\sum_{i=1}^{N} X_i - x\right\|^2\right] \\
&= \frac{1}{\delta^2}\,\sum_{k=1}^{d}\mathbf{E}\left[\frac{1}{N}\sum_{i=1}^{N} X_i^{(k)} - x_k\right]^2 \\
&= \frac{1}{\delta^2}\,\sum_{k=1}^{d}\frac{x_k(1-x_k)}{N} \\
&\leq \frac{d}{4\delta^2 N}\,.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\left|\mathbf{E}\left[\tilde{u}\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right)\right] - \tilde{u}(x)\right| &\leq \mathbf{E}\left[\mathbb{1}_{A_x}(\omega)\cdot\left|\tilde{u}\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right) - \tilde{u}(x)\right|\right] \\
&\quad + \mathbf{E}\left[\mathbb{1}_{\bar{A}_x}(\omega)\cdot\left|\tilde{u}\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right) - \tilde{u}(x)\right|\right] \\
&\leq \varepsilon\,+\,2\gamma P(\bar{A}_x) \\
&\leq \varepsilon\,+\,\frac{\gamma d}{2\delta^2 N}\,.
\end{aligned}$$

The right hand side of the last inequality does not depend on x, and the second term of the sum tends to 0 for $N \to \infty$. Thus, modified Bernstein Polynomials $\tilde{B}_{N,\tilde{u}}$ converge uniformly to $\tilde{u}$.

2. The convergence of $\tilde{F}^{(N)}(x)$ to $\tilde{F}(x)$, whenever F is continuous at x, follows from part 1 and the compactness of S^d. Let k be any multiindex and set

$$\tilde{u}(x) = x^k = x_1^{k_1} x_2^{k_2} \cdot \ldots \cdot x_d^{k_d}.$$

Then, cf. (17),

$$\int_{S^d} x^k d\tilde{F}^{(N)}(x) \longrightarrow m_k \qquad \text{as } N \to \infty, \tag{21}$$

where m_k is the k-th moment of the distribution function $\tilde{F}$. Next, take any convergent subsequence $\{\tilde{F}^{(N_r)}\}$ with limit distribution G. Part 1 again implies that the moments of $\tilde{F}^{(N_r)}$ converge to the moments of G which, by (21), are equal to the moments of F. Since distribution on S^d are uniquely determined by their moments, $G \equiv F$. Hence, the sequence $\{F^{(N)}\}_{N\geq 0}$ converges to F, which implies part 2. □

Theorem 3.2 *A multiindexed sequence* $\underline{\nu} = \{\nu_n\}_{n\geq 0}$ *of real numbers is a sequence of moments of a measure* ν *on* S^d *iff*

$$(\tilde{\triangle}^N \underline{\nu})_k \;\geq\; 0 \tag{22}$$

or, equivalently,

$$\sum_{j\in\mathcal{I}_N^d} (-\mathbf{1})^j \begin{bmatrix} N \\ j \end{bmatrix} \nu_{k+j} \;\geq\; 0 \tag{23}$$

for all $N \in \mathbb{Z}_+$ *and multiindices* $k \in \mathbb{Z}_+^d$.

For a proof of Theorem 3.2 and generalizations of such characterization theorems see Stockbridge (2003).

Definition 3.3

1. *We denote by* $\tilde{\mathcal{D}}_N^d$ *the set of all arrays* $\tilde{z} = \{\tilde{z}_n\}_{n\in\mathcal{I}_N^d} \in \mathbb{R}^K$, $K = \binom{N+d}{d}$, *which have the following property:*

   $$(\tilde{\triangle}^M \tilde{z})_n \geq 0 \text{ for all } M \geq 0 \text{ and } n \in \mathcal{I}_N^d \text{ such that } M + \Sigma(n) \leq N.$$

   The set $\tilde{\mathcal{D}}_N^d$ *is called the d-dimensional Dale Polygon of order* N.

2. *For* $c > 0$ *we define the set*

   $$\mathcal{D}_{N,c}^d := \left\{ \tilde{z} \in \tilde{\mathcal{D}}_N^d \mid \tilde{z}_{(0,\ldots,0)} = c \right\},$$

   and call $\mathcal{D}_N^d := \mathcal{D}_{N,1}^d$ *the d-dimensional Dale Polytope of order* N.

Similar to the case of the Hausdorff Polytope we are able to reduce the number of constraints to characterize the Dale Polytope $\mathcal{D}_N^d$.

Lemma 3.4 *A sequence* $\tilde{z} = \{\tilde{z}_n\}_{n\in\mathcal{I}_N^d} \in \mathbb{R}^K$, $K = \binom{N+d}{d}$, *is an element of the Dale Polytope* $\mathcal{D}_N^d$ *iff* $\tilde{z}_{(0,\ldots,0)} = 1$ *and, for all* $n \in \mathcal{I}_N^d$, $M = N - \Sigma(n)$, *the inequalities*

$$(\tilde{\triangle}^M \tilde{z})_n \;=\; \sum_{j\in\mathcal{I}_M^d} (-\mathbf{1})^j \begin{bmatrix} M \\ j \end{bmatrix} \tilde{z}_{n+j} \;\geq\; 0 \tag{24}$$

are satisfied.

Proof. Condition (24) is of course necessary, cf. (15). To see that it is also sufficient, we shall recursively repeat the arguments spelled out in detail in Step 1, cf. also Step M, below. These arguments are based on the following identity, cf. the definition of $\tilde{\triangle}^M$,

$$\tilde{\triangle}^M = \tilde{\triangle}(\tilde{\triangle}^{M-1}) = \tilde{\triangle}^{M-1} - \sum_{i=1}^{d} T_i \circ \tilde{\triangle}^{M-1},$$

i. e.

$$\tilde{\triangle}^{M-1} = \tilde{\triangle}^M + \sum_{i=1}^{d} T_i \circ \tilde{\triangle}^{M-1}. \tag{25}$$

Step 0: The following inequalities hold by assumption:

- $\tilde{z}_n \geq 0$, if $\Sigma(n) = N$;
- $(\tilde{\triangle}\tilde{z})_n \geq 0$, if $\Sigma(n) = N - 1$;
  $\vdots$
- $(\tilde{\triangle}^M \tilde{z})_n \geq 0$, if $\Sigma(n) = N - M$;
  $\vdots$
- $(\tilde{\triangle}^N \tilde{z})_n \geq 0$, if $\Sigma(n) = N - N$.

Step 1: Now it follows:

- $\tilde{z}_n \geq 0$, if $\Sigma(n) = N - 1$, by the first and second inequality of Step 0 and Equation (25) for $M = 1$;
- $(\tilde{\triangle}\tilde{z})_n \geq 0$, if $\Sigma(n) = N - 2$, by the second and third inequality of Step 0 and Equation (25) for $M = 2$;

  $\vdots$

- $(\tilde{\triangle}^{M-1} \tilde{z})_n \geq 0$, if $\Sigma(n) = N - M$, by the M-th and $(M + 1)$-th inequality of Step 0 and Equation (25) for M;

  $\vdots$

- $(\tilde{\triangle}^{N-1} \tilde{z})_n \geq 0$, if $\Sigma(n) = N - N$, by the N-th and $(N+1)$-th inequality of Step 0 and Equation (25) for $M = N$.

Step M: The M-th step only comprises $(N - M + 1)$ inequalities:

- $\tilde{z}_n \geq 0$, if $\Sigma(n) = N - M$, by the first and second inequality of Step $M-1$ and Equation (25);

  $\vdots$

- $(\tilde{\triangle}^{N-M}\tilde{z})_n \geq 0$, if $\Sigma(n) = N - N$, by the next to last and the last inequality of Step $M-1$ and Equation (25);

and the final step is but the inequality $\tilde{z}_{(0,\ldots,0)} \geq 0$ which holds by Assumption (24). $\square$

Theorem 3.5 *The Dale Polytope $\mathcal{D}_N^d$ has $K = \binom{N+d}{d}$ corner points which can be indexed by the elements of $\mathcal{I}_N^d$. The k-th vertex, $k \in \mathcal{I}_N^d$, is characterized as follows:*

$$\tilde{z}^{(k)} = \begin{bmatrix} N \\ k \end{bmatrix}^{-1} \left(\begin{bmatrix} N-\Sigma(j) \\ k-j \end{bmatrix}_{j \leq k}, 0, \ldots, 0 \right) \in \mathbb{R}^K, \tag{26}$$

where $j \leq k$ is understood component-wise and the coordinates of the corner points are arranged according to the lexicographic order on $\mathcal{I}_N^d$.

Proof. To verify that each $\tilde{z}^{(k)}$ is a corner point of $\mathcal{D}_N^d$ it is sufficient to show that each $\tilde{z}^{(k)}$ satisfies $(K-1)$ of the Inequalities (24) as equations and one as a strict inequality. For any vector $\tilde{z}^{(k)}$ defined by (26) the following properties hold:

(i) $\tilde{z}_0^{(k)} = 1$.

(ii) For $n = k$ Inequality (24) in Lemma 3.4 is a strict inequality. This follows from

$$\tilde{z}_k^{(k)} = \begin{bmatrix} N \\ k \end{bmatrix}^{-1} > 0.$$

(iii) For all Inequalities (24) indexed by k such that $n \leq k$ component-wise, $n \neq k$, we have

$$\sum_{j:\, j+n \in \mathcal{I}_N^d} (-\mathbf{1})^j \begin{bmatrix} N-\Sigma(n) \\ j \end{bmatrix} \tilde{z}_{n+j}^{(k)}$$
$$= \begin{bmatrix} N \\ k \end{bmatrix}^{-1} \sum_{j_1=0}^{k_1-n_1} \cdots \sum_{\substack{j_d=0 \\ \Sigma(j) \leq N-\Sigma(n)}}^{k_d-n_d} (-1)^{\Sigma(j)} \begin{bmatrix} N-\Sigma(n) \\ j \end{bmatrix} \begin{bmatrix} N-\Sigma(n+j) \\ k-n-j \end{bmatrix};$$

since $n \leq k$, $k \in \mathcal{I}_N^d$, and the summation indices j satisfy $j \leq k-n$ component-wise we obtain $\Sigma(j+n) \leq N$; thus the restriction $\Sigma(j) \leq N - \Sigma(n)$ is redundant. Henceforth, it will be dropped.

Next, we set the multiindex $l = k - n$ and $L = N - \Sigma(n)$; so the previous expression becomes

$$\begin{bmatrix} N \\ k \end{bmatrix}^{-1} \sum_{j_1=0}^{l_1} (-1)^{j_1} \cdots \sum_{j_d=0}^{l_d} (-1)^{j_i} \begin{bmatrix} L \\ j \end{bmatrix} \begin{bmatrix} L-\Sigma(j) \\ l-j \end{bmatrix}.$$

Recombining the (pseudo) multinomial coefficients, i. e.

$$
\begin{aligned}
\begin{bmatrix} L \\ j \end{bmatrix} \begin{bmatrix} L - \Sigma(j) \\ l - j \end{bmatrix} &= \frac{L!}{\left(\prod_{i=1}^{d} j_i!\right)(L-\Sigma(j))!} \cdot \frac{(L-\Sigma(j))!}{\left(\prod_{i=1}^{d}(l_i - j_i)!\right)(L-\Sigma(l))!} \\
&= \frac{L!}{(L-\Sigma(l))! \prod_{i=1}^{d} l_i!} \cdot \frac{\prod_{i=1}^{d} l_i!}{\left(\prod_{i=1}^{d} j_i!\right)\left(\prod_{i=1}^{d}(l_i - j_i)!\right)},
\end{aligned}
$$

we finally see that the left hand side of the n-th inequality equals zero, since

$$
\begin{bmatrix} N \\ k \end{bmatrix}^{-1} \begin{bmatrix} N - \Sigma(n) \\ k - n \end{bmatrix} \sum_{j_0=0}^{l_1} (-1)^{j_1} \binom{l_1}{j_1} \cdots \sum_{j_d=0}^{l_d} (-1)^{j_d} \binom{l_d}{j_d} \;=\; 0\,.
$$

(iv) For Inequalities (24) such that $n_i > k_i$ for at least one $i \in \{1, \ldots, d\}$ all coordinates $\tilde{z}^{(k)}_{n+j}$ in the sum on the left hand side are zero by definition.

Hence (i) – (iv) show that except for one strict inequality all other inequalities are actually equations. □

Proposition 3.6 *Let $N \in \mathbb{N}$ and $k \in \mathcal{I}^d_N$. The transformation $\tilde{R}^{(N)}$ applied to the corner points $\tilde{z}^{(k)}$ of the Dale Polytope $\mathcal{D}^d_N$ yields the set of unit vectors, i. e.*

$$
\tilde{R}^{(N)}(\tilde{z}^{(k)})_j \;=\; \begin{cases} 1 & \text{for} \quad j = k \\ 0 & \text{otherwise.} \end{cases}
$$

PROOF. The Dale Polytope $\mathcal{D}^d_N$ is characterized by $K = \binom{N+d}{d}$ inequality conditions. At each corner point $\tilde{z}^{(k)}$ exactly K–1 of these conditions are active. We know, because of the proof of Theorem 3.5, that $\sum_{j \in \mathcal{I}^d_N} \tilde{R}^{(N)}(\tilde{z}^{(k)})_j = 1$ and that for $K-1$ different multiindices $j \in \mathcal{I}^d_N$:

$$
(\tilde{\triangle}^{N-\Sigma(j)} \tilde{z}^{(k)})_j \;=\; 0,
$$

i. e.

$$
\tilde{R}^{(N)}(\tilde{z}^{(k)})_j \;=\; \begin{bmatrix} N \\ j \end{bmatrix} (\tilde{\triangle}^{N-\Sigma(j)} \tilde{z}^{(k)})_j \;=\; 0
$$

for all but one multiindex $m \in \mathcal{I}^d_N$. For this particular index m the equation

$$
\tilde{R}^{(N)}(\tilde{z}^{(k)})_m \;=\; 1
$$

holds. But

$$
\begin{aligned}
\tilde{R}^{(N)}(\tilde{z}^{(k)})_k &= \begin{bmatrix} N \\ k \end{bmatrix} (\tilde{\triangle}^{N-\Sigma(k)} \tilde{z}^{(k)})_k \\
&= \begin{bmatrix} N \\ k \end{bmatrix} \sum_{j \in \mathcal{I}^d_{N-\Sigma(k)}} \begin{bmatrix} N-\Sigma(k) \\ j \end{bmatrix} (-\mathbf{1})^j \tilde{z}^{(k)}_{k+j} \\
&= \begin{bmatrix} N \\ k \end{bmatrix} \begin{bmatrix} N-\Sigma(k) \\ 0 \end{bmatrix} \tilde{z}^{(k)}_k \\
&= \begin{bmatrix} N \\ k \end{bmatrix} \begin{bmatrix} N-\Sigma(k) \\ 0 \end{bmatrix} \begin{bmatrix} N \\ k \end{bmatrix}^{-1} \begin{bmatrix} N-\Sigma(k) \\ 0 \end{bmatrix} \quad = \quad 1.
\end{aligned}
$$

since, by definition of $z^{(k)}$, all other terms of this sum disappear. □

Similar to the case of the Hausdorff Polytope we may interpret the image of $\tilde{R}^{(N)}$ of the corner point $\tilde{z}^{(k)}$ as the Dirac measure at the point $(\frac{k_1}{N}, \ldots, \frac{k_d}{N})$ and the range $\tilde{R}^{(N)}(\mathcal{D}^d_N)$ as the set of all discrete probability measures on the set of points $\{(\frac{j}{N})\}_{j \in \mathcal{I}^d_N}$.

4 Numerical examples

The examples to be considered in this section are uncontrolled diffusion processes. To make the exposition self-contained, we briefly recapitulate the formulation of the LP approach to exit time problems of Markov processes; see Kurtz and Stockbridge (1998) for an exposition of the general case of controlled processes.

Let $X = \{X_t\}_{t \geq 0}$ be a Markov process on $\mathbb{R}^d$ with $X_0 = x_0$ and generator A; let $\mathcal{D}(A)$ denote the domain of the generator. We shall partition the states of the process, Ω, into two disjoint regions, a bounded open region Ω_0 and a set Ω_1. For simplicity of the exposition we assume that $\Omega_0 \subset [0,1]^d$; the general case can be reduced to the special one by a change of variables. Let τ denote the first time X hits Ω_1. So we may consider the set Ω_1 as the stopping region for the process X. For the case of a diffusion process it is convenient to choose Ω_1 as the boundary of Ω_0. Since we are interested in the distribution of τ we shall explain how to compute upper and lower bounds for the moments of the random variable τ. Since the formulation of the LP approach for the mean exit time is simpler than the formulation for higher moments we begin with the simpler case; afterwards we extend the formulation to higher moments.

The basic fact which underlies the LP approach is that for each $f \in \mathcal{D}(A)$,

$$
f(X_t) - f(X_0) - \int_0^t Af(X_s)\, ds
$$

is a martingale and, if τ has finite expectation, the martingale property implies the equation

$$\mathbf{E}[f(X_\tau)] - \mathbf{E}[f(X_0)] - \mathbf{E}\left[\int_0^\tau Af(X_s)ds\right] = 0. \tag{27}$$

Define the occupation measure μ_0 and the exit distribution μ_1 by

$$\mu_0(B) = \mathbf{E}\left[\int_0^\tau I_B(X_s)\,ds\right] \qquad \text{and} \qquad \mu_1(B) = P(Y_\tau \in B)$$

for Borel sets B. It then follows, since $X_0 = x_0$, that Equation (27) can be written as

$$\langle f, \mu_1\rangle := \int_{\Omega_1} f(x)\,\mu_1(dx) = f(x_0) + \int_{\Omega_0} Af(x)\,\mu_0(dx) =: f(x_0) + \langle Af, \mu_0\rangle. \tag{28}$$

We formally define the adjoint operator A^*, $\mu \mapsto A^*\mu$, by

$$\langle f, A^*\mu\rangle := \langle Af, \mu\rangle,$$

for all $f \in \mathcal{D}(A)$, and use the shorthand writing

$$\mu_1 - \delta_{x_0} - A^*\mu_0 = 0, \tag{29}$$

where δ_{x_0} denotes the Dirac measure at x_0, to express Equation (28). The results in Kurtz and Stockbridge (1998) imply that for each μ_0 and μ_1 which satisfy Equation (29) – to be understood in the sense of (28) – there is a process X and an exit time τ for which Equation (27) holds. Thus Equation (29) characterizes the occupation measure μ_0 and the exit distribution μ_1 of a Markov process starting at x_0 and having generator A. Assuming that the constant function $\mathbb{1}(x) \equiv 1$ is an element of the range of A, Equation (29) implies that the mean exit time can be described as the value of two infinite dimensional linear programs which have but *one* feasible solution (μ_0, μ_1), viz.

$$\begin{aligned}\mathbf{E}_{x_0}[\tau] &= \inf_{\mu_0,\mu_1\geq 0}\left\{\int_{\Omega_0}\mu_0(dx)\;\middle|\;\mu_1(\Omega_1) = 1 \text{ and Equation (28) holds for all } f \in \mathcal{D}(A)\right\}\\ &= \sup_{\mu_0,\mu_1\geq 0}\left\{\int_{\Omega_0}\mu_0(dx)\;\middle|\;\mu_1(\Omega_1) = 1 \text{ and Equation (28) holds for all } f \in \mathcal{D}(A)\right\}.\end{aligned}$$

More generally, if R, l resp., are bounded measurable functions on Ω_1, Ω_0 resp., using the shorthand writing (29), we may write

$$\begin{aligned}\mathbf{E}_{x_0}\left[R(X_\tau) + \int_0^\tau l(X_s)\,ds\right] &= \inf_{\mu_0,\mu_1\geq 0}\{\langle R,\mu_1\rangle + \langle l,\mu_0\rangle \mid \langle \mathbb{1},\mu_1\rangle = 1,\ \mu_1 - A^*\mu_0 = \delta_{x_0}\}\\ &= \sup_{\mu_0,\mu_1\geq 0}\{\langle R,\mu_1\rangle + \langle l,\mu_0\rangle \mid \langle \mathbb{1},\mu_1\rangle = 1,\ \mu_1 - A^*\mu_0 = \delta_{x_0}\}. \end{aligned}\tag{30}$$

It needs to be stressed that for exit time problems the linear programs (30) are actually an artifact since the constraints uniquely determine the feasible pair (μ_0, μ_1). While this is true for the infinite dimensional problems (30) this is no longer the case for the finite dimensional approximating LPs, see below.

Next, if R and l are polynomials, i.e.

$$R(x) \;=\; \sum_{j=0}^{n_R} r_j x^j\,, \qquad l(y) \;=\; \sum_{j=0}^{n_l} \lambda_j y^j$$

for $x \in \Omega_1,\ y \in \Omega_0$ and multiindices n_R and n_l, and A maps monomials onto polynomials, e.g. for $f(x) = x^k$, k a multiindex, there exist n_k and coefficients c_{k_j}, $0 \le j \le n_k$, such that

$$Af(x) \;=\; \sum_{i=0}^{n_k} c_{k_j} x^j,$$

if Ω_1 is a bounded set and moment sequences on Ω_0 and Ω_1 can be characterized by linear inequalities then the infinite dimensional LPs (30) are equivalent to infinite dimensional LPs whose variables are moment sequences.

Using but finite sequences and a finite number of the characterizing inequalities as constraints will determine finite dimensional linear programs whose values provide upper and lower bounds on the expected value which is to be computed. For instance, assuming R and l to be polynomial functions defined on $[0,1]^d$, A the generator of a diffusion with polynomial coefficients, $\Omega_0 = (0,1)^d$, $\Omega_1 = [0,1]^d \backslash (0,1)^d$ and choosing the Hausdorff Conditions up to a given order n the following finite dimensional LPs provide bounds on the quantity of interest, i.e.

$$\underline{\varphi}_n \;\le\; \mathbf{E}_{x_0}\left[R(X_\tau) + \int_0^\tau l(X_s)\, ds \right] \;\le\; \bar{\varphi}_n, \tag{31}$$

where

$$\bar{\varphi}_n \;:=\; \max_{z^{(0)},\, z^{(1)}} \left\{ \sum_{j=0}^{n_R} r_j z_j^{(1)} + \sum_{j=0}^{n_l} \lambda_j z_j^{(0)} \,\middle|\, z^{(0)} \in \tilde{\mathcal{H}}_n^d,\ z^{(1)} \in \mathcal{K}_n^d, \right.$$
$$\left. \text{and for all } k \text{ such that } n_k \le n : z_k^{(1)} - \sum_{j=0}^{n_k} c_{k_j} z_j^{(0)} = x_0^k \right\}$$

and

$$\underline{\varphi}_n \;:=\; \min_{z^{(0)},\, z^{(1)}} \left\{ \sum_{j=0}^{n_R} r_j z_j^{(1)} + \sum_{j=0}^{n_l} \lambda_j z_j^{(0)} \,\middle|\, z^{(0)} \in \tilde{\mathcal{H}}_n^d,\ z^{(1)} \in \mathcal{K}_n^d, \right.$$
$$\left. \text{and for all } k \text{ such that } n_k \le n : z_k^{(1)} - \sum_{j=0}^{n_k} c_{k_j} z_j^{(0)} = x_0^k \right\}.$$

Note, each vector $z^{(1)} \in \mathcal{K}_n^d$ is the convex combination of scaled elements of Hausdorff Polytopes associated with $(d-1)$-dimensional hypercubes. Since for exit time problems the total mass of the corresponding occupation measure is typically different from one, the vectors $z^{(0)}$ are also scaled versions of elements – this time – of d-dimensional Hausdorff Polytopes. For the 2-dimensional example below, see Example 4.1, our LP-code incorporates the corner point formulas for $\mathcal{H}_n^2$ only once, while we use the 1-dimensional formulas four times (for each side of the unit square).

The numerical examples below will demonstrate that for reasonably large values of n the difference between $\bar{\varphi}_n$ and $\underline{\varphi}_n$ will typically be small. In earlier publications, e.g. Helmes, Röhl, and Stockbridge (2001), we have implemented the requirements $z^{(0)} \in \tilde{\mathcal{H}}_n^d$ and $z^{(1)} \in \mathcal{K}_n^d$ using Inequalities (8) or (9) instead of the corner point formulas; actually, to increase numerical stability we usually used the recursive definition of iterated differences, i.e. (3), and (8). Such implementations result in large programs, whose size and run times restrict the values of n. For typical twodimensional problems, e. g. the computation of the mean exit time of a twodimensional Brownian motion from a square, cf. Example 4.1 below, we took $n = (M, M)$, $M \leq 14$. But larger values of M provide better bounds on $\mathbf{E}[\tau]$. Larger values of M become a possibility when the corner point formulas, cf. Sections 2 and 3, are used. The following examples will show that by using the corner point formulas and larger values of M the resulting LPs not only provide better bounds than the old programs do, but also require fewer iterations.

Finally, to compute bounds for higher moments of the exit times of such processes we need to augment the time coordinate to the state and consider the time-space generator $\tilde{A}$,

$$\tilde{A}f(t,x) \;=\; \frac{\partial f}{\partial t}(t,x) + Af(t,x),$$

acting on functions f depending on t and x. The measures μ_0 and μ_1 are now defined by

$$\mu_0(\Gamma) \;=\; \mathbf{E}_{x_0}\left[\int_0^\tau I_\Gamma(s, X_s)ds\right] \qquad \text{and} \qquad \mu_1(\Gamma) \;=\; P_{x_0}\left((\tau, X_\tau) \in \Gamma\right)$$

for subsets Γ of $[0,\infty) \times \Omega$. The extension of the fundamental Equation (28) takes the form

$$\text{(32)} \qquad \int_{\mathbb{R}^+ \times \Omega_1} f(s,x)\mu_0(ds \times dx) - f(t_0, x_0) - \int_{\mathbb{R}^+ \times \Omega_0} \left[\frac{\partial f}{\partial t}(s, X_s) + Af(s, X_s)\right] \;=\; 0$$

for all functions f such that $f(t,\cdot) \in \mathcal{D}(A)$, and $f(\cdot, x)$ is continuously differentiable in t and vanishes at ∞. Note that the n-th moment of the exit time is given by

$$\mathbf{E}_{x_0}[\tau^n] \;=\; n \int_{\mathbb{R}^+ \times \Omega_0} s^{n-1}\mu_0(ds \times dx).$$

It follows from the general theory (see Kurtz and Stockbridge (1998)) that, as in the case of the mean exit time, measures μ_0 and μ_1 which satisfy (32) characterize processes having generator A up to the exit time τ. But while the process $\{X_t\}_{t\geq 0}$ evolves on a bounded set the process $\{t, X_t\}_{t\geq 0}$ does not. To mimick the LP approach for X, i.e. exploit the characterization of μ_0 and μ_1 by their moments, we truncate the unbounded time domain $\mathbb{R}^+$ to a finite interval $[0,T]$. This introduces an additional approximation into the method in all those cases where $\tau \neq \tau \wedge T$. Rather than the exit time τ, the analysis will evaluate the distribution of $\tau \wedge T$. But, by the assumption imposed on τ, for large values of T the difference between the exit time and the truncated exit time will be zero with high probability. Moreover, there are many problems, cf. Example 4.3 below, for which there is a natural bound T.

Example 4.1 *Twodimensional Brownian motion on the unit square:*

Let $(X_s, Y_s) = \left(x_0 + W_s^{(1)}, y_0 + W_s^{(2)}\right)$, where $x_0, y_0 \in (0,1)$ and $W = \left(W_s^{(1)}, W_s^{(2)}\right)$ is a twodimensional standard Brownian motion process. The generator of the process (X_s, Y_s) is given for $f \in \mathcal{D}(A) = \{f \mid f \text{ twice continuously differentiable on } \mathbb{R}^2\}$ by

$$Af(x,y) \;=\; \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(x,y) + \frac{1}{2}\frac{\partial^2 f}{\partial y^2}(x,y).$$

In this example we consider the exit time of the process from the bounded domain $\Omega_0 = (0,1)\times(0,1)$; the boundary consists of four parts $\Gamma_{(b)} = [0,1]\times\{0\}$, $\Gamma_{(t)} = [0,1]\times\{1\}$, $\Gamma_{(\ell)} = \{0\}\times(0,1)$ and $\Gamma_{(r)} = \{1\}\times(0,1)$. Therefore, we need to work with five measures: the occupation measure μ_0 of the process in the open unit square, and measures on the four boundary parts which together comprise the exit distribution μ_1 of the Brownian motion process from the unit square. We associate with each boundary measure elements of the onedimensional Hausdorff Polytope of order M, while the occupation measure is associated with elements of the twodimensional Hausdorff Polygon, i.e. $\tilde{\mathcal{H}}^2_{(M,M)}$.

Using the corner point formula (12) the constraints of the associated finite dimensional LP problems can be expressed as follows. Let $z = \{z_{ij}\}_{i,j=0}^M \in \tilde{\mathcal{H}}^2_{(M,M)}$ be associated with the occupation measure μ_0 on Ω_0 and let $z^{(b)} = \{z_i^{(b)}\}_{i=0}^M$, $z^{(t)}$, $z^{(\ell)}$, $z^{(r)}$, being associated with the exit distribution on $\Gamma_{(b)}$, $\Gamma_{(t)}$, $\Gamma_{(\ell)}$ and $\Gamma_{(r)}$, be elements of $\mathcal{H}^1_M$. Then for $0 \leq m,n \leq M$,

$$\begin{aligned} 0 \;&=\; \frac{m(m-1)}{2}\, z_{m-2,n} + \frac{n(n-1)}{2}\, z_{m,n-2} \\ &\quad + x_0^m y_0^n - 0^n z_m^{(b)} - 1^n z_m^{(t)} - 0^m z_n^{(\ell)} - 1^m z_n^{(r)}, \end{aligned}$$

and all the z-vectors are convex combinations of the corner points described in Section 2.

Table 1 reports our results on the mean exit time for a sample of initial values. Because of the symmetry of the problem we only report the values for initial values in the first quarter of the unit square. The numbers in parantheses are those which we reported in Helmes, Röhl, and Stockbridge (2001). There we used the Hausdorff Conditions expressed in terms of the Inequalities (9). This restricted our choice of M; typically, M was chosen between 11 and 14 resulting in LPs which required approximately 13000 simplex iterations and $\sim$ 25 minutes on an Ultra Sun 10/300 using AMPL/CPLEX 6.5. The corner point formula allows us to choose $M \sim 30$ which increases the accuracy of the values while actually requiring fewer iterations, $\sim$ 10000 simplex iterations, and less time on the same machine. The numbers in the column "exact value" are based on the formula for the mean exit time of Brownian motion from a square, cf. Helmes, Röhl, and Stockbridge (2001).

y_0	Lower Bound	Upper Bound	exact value
0.5	0.14728996	0.14745482	0.147340
	(0.14693254)	(0.14801697)	
0.4	0.14224363	0.14240372	0.142310
	(0.14189854)	(0.14299119)	
0.3	0.12669597	0.12688587	0.126760
	(0.12635959)	(0.12741458)	
0.2	0.09934259	0.09950494	0.099396
	(0.09904923)	(0.10002026)	
0.1	0.05805426	0.05814112	0.058084
	(0.05784276)	(0.05844282)	

Table 1: Bounds on the mean exit time of twodimensional Brownian motion from the unit square as a function of y_0 using corner point formulae; $x_0 = 0.5$, $M = 30$.

We have observed that the average of the bounds usually give good estimates for the quantity of interest. In Table 2 we report the bounds and compare the estimates with the exact values for a different set of initial values. Note that the numerical values nicely reflect the symmetry of the problem.

y_0	Lower Bound	Upper Bound	Estimate	exact value
0.9	0.04171744	0.04183004	0.041774	0.041761
0.8	0.06924839	0.06936745	0.069308	0.069294
0.7	0.08664556	0.08674721	0.086696	0.086682
0.6	0.09627489	0.09637400	0.096324	0.096340
0.5	0.09936082	0.09946310	0.099412	0.099396
0.4	0.09626428	0.09640394	0.096334	0.096311
0.3	0.08663224	0.08678911	0.086711	0.086682
0.2	0.06921902	0.06943718	0.069328	0.069294
0.1	0.04168128	0.04189773	0.041790	0.041761

Table 2: Bounds and estimates of the mean exit time of twodimensional Brownian motion from the unit square as a function of the initial position using corner point formulae; $x_0 = 0.8$, $M = 30$.

Example 4.2 *Twodimensional Brownian motion on the unit triangle:*

Let (X_s, Y_s) be again twodimensional Brownian motion but, in contrast to Example 4.1, restricted to S^2. This time the variables of the LP model comprise the occupation measure on the interior of S^2 and three boundary measures, $\mu^{(\ell)}$, $\mu^{(b)}$ and $\mu^{(d)}$ resp., on the left, bottom and diagonal boundary resp. The dynamics of the process and the geometry of the state space is reflected in the following equality constraints, where $\tilde{z} \in \tilde{\mathcal{D}}^2_M$ and $z^{(\ell)}$, $z^{(b)}$ and $z^{(d)} \in \tilde{\mathcal{H}}^1_M$ are again expressed as multiples of convex combinations of the extreme points of $\mathcal{D}^2_M$ and $\mathcal{H}^1_M$:

$$\begin{aligned} 0 &= \frac{m(m-1)}{2}\,\tilde{z}_{m-2,n} + \frac{n(n-1)}{2}\,\tilde{z}_{m,n-2} \\ &\quad + x_0^m y_0^n - 0^n z^{(b)}_m - 0^m z^{(\ell)}_n - \sum_{k=0}^{n} \binom{n}{k} (-1)^k z^{(d)}_{k+m}, \end{aligned}$$

for $m, n \geq 0;\quad m + n \leq M$.

Table 3 displays the bounds on the mean exit time of Brownian motion from the unit triangle for a sample of initial values. We compare the bounds and the estimates based on the LP-values with the numerical results, labeled "NE", obtained by using the Matlab PDE-Toolbox. The initial values we have choosen in Table 3 are part of the grid points generated by Matlab's PDE-solver.

x_0	y_0	lower bound	upper bound	average value	NE
0.187418	0.576648	0.035876	0.035899	0.035888	0.035934
0.539775	0.079868	0.024828	0.024845	0.024837	0.024837
0.563342	0.320957	0.028271	0.028297	0.028284	0.028277
0.369505	0.540196	0.024245	0.024266	0.024255	0.024275
0.768800	0.153916	0.011122	0.011125	0.011124	0.011108
0.102203	0.853922	0.004365	0.004366	0.004366	0.004363
0.433926	0.356116	0.046865	0.046911	0.046888	0.046884
0.298395	0.651147	0.012590	0.012599	0.012595	0.012594

Table 3: Bounds on the mean exit time of twodimensional Brownian motion from the unit triangle as a function of (x_0, y_0) using corner point formulae; $M = 30$.

Note that for the case of the unit square the Matlab results are also only accurate up to 4 digits if the mesh which is automatically generated by the PDE-solver is refined but twice. Of course, precision does increase if a finer mesh is used.

Example 4.3 *Onedimensional Brownian motion with a moving boundary:*

Let $Z = \{t, X_t\}_{t\geq 0}$, where $\{X_t\}_{t\geq 0}$ denotes onedimensional Brownian motion starting at $x_0 = 0.5$. We consider the process X on $[0,1]$ assuming that the right boundary point $\{1\}$ is moving to the left at constant speed 1. Let τ denote the first time when $X_\tau = 0$ or $X_\tau = 1 - \tau$. This description of the situation is equivalent to Z starting at $(0, x_0)$ and evolving on the unit triangle until Z hits either the diagonal $\{(t,x) \mid t + x \leq 1\}$ or the bottom part of the boundary $[0,1] \times \{0\}$. The generator A of Z is given by

$$Af(t,x) \;=\; \frac{\partial f}{\partial t}(t,x) \,+\, \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(t,x)\,.$$

The dynamics of the process Z (up to the time of stopping) and the geometry of the state space determine the following equality constraints, where again $\tilde{z} \in \tilde{\mathcal{D}}^2_M$ and $z^{(b)}$ and $z^{(d)} \in \tilde{\mathcal{H}}^1_M$:

$$0 \;=\; m\tilde{z}_{m-1,n} + \frac{1}{2}n(n-1)\tilde{z}_{m,n-2} + 0^m x_0^n - 0^n z_m^{(b)} - \sum_{k=0}^{n}\binom{n}{k}(-1)^k z_{k+m}^{(d)},$$

for $m, n \geq 0$; $m + n \leq M$.

Table 4 illustrates the dependence of the accuracy of the bounds on higher moments of τ on the parameter M. We report bounds for moments up to order 4 for $M = 10$ and $M = 20$.

moments	M	lower bound	upper bound	mean
0th	10	0.17163196	0.17163199	0.17163198
	20	0.17163198	0.17163198	0.17163198
1st	10	0.02013079	0.02013089	0.02013084
	20	0.02013083	0.02013083	0.02013083
2nd	10	0.00390564	0.00390591	0.00390578
	20	0.02013083	0.02013083	0.02013083
3rd	10	0.00098930	0.00098997	0.00098963
	20	0.00098961	0.00098961	0.00098961
4th	10	0.00029713	0.00029874	0.00029793
	20	0.00029786	0.00029786	0.00029786

Table 4: Higher moments of the exit time of time-space Brownian motion from the unit triangle for $M = 10$ and $M = 20$; $(t_0,\, x_0) = (0,\, 0.5)$.

References

Abramowitz, M., I. Stegun. 1965. *Handbook of Mathematical Functions.* Dover, New York.

Ang, D.D., R. Gorenflo, V.K. Le, D.D. Trong. 2002. *Moment theory and some inverse problems in potential theory and heat conduction*. Volume 1792 of *Lecture Notes in Math.* Berlin: Springer–Verlag.

Bhatt, A., V. Borkar. 1996. Occupation measures for controlled Markov processes: Characterisation and optimality. *Ann. Probab.* **24**(3), 1531–1562.

Cho, M. 2000. *Linear programming formulation for optimal stopping*. Ph. D. thesis, The Graduate School, University of Kentucky, Lexington.

Cho, M., R. Stockbridge. 2002. Linear programming formulation for optimal stopping problems. *SIAM J. Control Optim.* **40**, 1965–1982.

Dale, A. 1987. Two-dimensional moment problems. *Math. Sci.* **12**, 21–29.

Feller, W. 1971. *An introduction to probability theory and its applications* (2nd ed.), Volume 2. New York: John Wiley & Sons.

Hausdorff, F. 1921. Summationsmethoden und Momentenfolgen. I. & II. *Math. Z.* **9**, 74–109, 280–299.

Hausdorff, F. 1923. Momentenprobleme für ein endliches Intervall. *Math. Z.* **16**, 220–248.

Helmes, K. 1999. Computing moments of the exit distribution for diffusion processes using linear programming. In P. Kall and J.-H. Lüthi (Eds.), *Oper. Res. Proc. 1998*, Berlin, pp. 231–240. Springer–Verlag.

Helmes, K. 2002. Numerical methods for optimal stopping using linear and non-linear programming. In B. Pasik-Duncan (Ed.), *Proceedings of a Workshop "Stochastic Theory and Control"*, Lecture Notes in Control and Information Sciences, Berlin, pp. 185–202. Springer–Verlag.

Helmes, K., S. Röhl, R. Stockbridge. 2001. Computing moments of the exit time distribution for Markov processes by linear programming. *Oper. Res.* **49**, 516–530.

Helmes, K., R. Stockbridge. 2000. Numerical comparison of controls and verification of optimality for stochastic control problems. *J. Optim. Theory Appl.* **106**, 107–127.

Helmes, K., R. Stockbridge. 2001. Numerical evaluation of resolvents and Laplace transforms of Markov processes. *Math. Methods Oper. Res.* **53**, 309–331.

Helmes, K., R. Stockbridge. 2003. Extension of dale's moment conditions with application to the Wright-Fisher model. *Stoch. Models* **19**(2), 255–267.

Hernandez-Lerma, O., J. Hennet, J. Lasserre. 1991. Average cost Markov decision processes: Optimality conditions. *J. Math. Anal. Appl.* **158**, 396–406.

Hildebrandt, T., I. Schoenberg. 1933. On linear functional operations and the moment problem for a finite interval in one or several dimensions. *Ann. of Math.* **34**, 317–328.

Karlin, S., L. Shapley. 1953. Geometry of moment spaces. *Mem. Amer. Math. Soc.* **12**.

Knill, O. 1997. On Hausdorff's moment problem in higher dimensions. Preprint, `http://abel.math.harvard.edu/~knill/preprints/stability.ps`.

Kurtz, T., R. Stockbridge. 1998. Existence of Markov controls and characterization of optimal Markov controls. *SIAM J. Control Optim.* **36**, 609–653.

Kurtz, T., R. Stockbridge. 1999. Martingale problems and linear programs for singular control. In *Thirty-Seventh Annual Allerton Conference on Communication, Control, and Computing (Monticello, Ill.)*, Urbana-Champaign, Ill., pp. 11–20. Univ. Illinois.

Lasserre, J., T. P. Rumeau. 2003. SDP vs. LP relaxations for the moment approach in some performance evaluation problems. *Rapport LAAS No. 03125*.

Manne, A. 1960. Linear programming and sequential decisions. *Management Sci.* **6**, 259–267.

Röhl, S. 2001. *Ein linearer Programmierungsansatz zur Lösung von Stopp- und Steuerungsproblemen*. Ph. D. thesis, Humboldt-Universität zu Berlin, Berlin, Germany.

Schwerer, E. 1996. *A linear programming approach to the steady-state analysis of reflected Brownian motion*. Ph. D. thesis, Stanford University.

Shohat, J., J. Tamarkin. 1943. *The problem of moments* (1st ed.). Providence, Rhode Island: American Mathematical Society.

Stockbridge, R. 1990. Time-average control of martingale problems: A linear programming formulation. *Ann. Probab.* **18**, 206–217.

Stockbridge, R. 2003. The problem of moments on a polytope and other bounded regions. *J. Math. Anal. Appl.* **285**, 356–375.

Ziegler, G. 1995. *Lectures on polytopes*, Volume 152 of *Graduade texts in mathematics*. New York: Springer–Verlag.