VOLUME OF SUB-LEVEL SETS OF HOMOGENEOUS POLYNOMIALS

JEAN B. LASSERRE

Abstract. Consider the sub level set $\mathbf{K} := \{\mathbf{x} : g(\mathbf{x}) \leq 1\}$ where g is a positive and homogeneous polynomial. We show that its Lebesgue volume can be approximated as closely as desired by solving a sequence of generalized eigenvalue problems with respect to a pair of Hankel matrices of increasing size, and whose entries are obtained in closed form.

1. Introduction

Let $\mathbf{K} \subset \mathbb{R}^n$ be the sublevel set defined by:

$$\mathbf{K} := \{\mathbf{x} \in \mathbb{R}^n : g(\mathbf{x}) \leq 1\}. \tag{1.1}$$

where $g \in \mathbb{R}[\mathbf{x}]_t$ is a nonnegative homogeneous polynomial of degree t (hence t is even). The goal of this paper is to provide en efficient numerical scheme to approximate its Lebesgue volume $\mathrm{vol}(\mathbf{K})$ as closely as desired.

Motivation. In addition of being interesting on its own, this problem has also a practical interest outside computational geometry. For instance it has a direct link with computing the integral $\int \exp(-g(\mathbf{x}))d\mathbf{x}$, called an *integral discriminant* in Dolotin and Morozov [1] and Morozov and Shakirov [9]. Indeed as proved in [9]:

$$\mathrm{vol}(\mathbf{K}) = \frac{1}{\Gamma(1+\frac{n+t}{2})} \int_{\mathbb{R}^n} \exp(-g(\mathbf{x}))\, d\mathbf{x}, \tag{1.2}$$

and to quote [9], *"averaging with exponential weights is an important operation in statistical and quantum physics"*. However, and again quoting [9], *"despite simply looking,* (1.2) *remains terra incognita"*. However, for special cases of homogeneous polynomials, the authors in [9] have been able to obtain a closed form expression for (1.2) (hence equivalently for $\mathrm{vol}(\mathbf{K})$) in terms of algebraic invariants of g.

Various consequences of formula (1.2) have been described and exploited in Lasserre [4]. For instance, $\mathrm{vol}(\mathbf{K})$ is a convex function in the coefficients of the polynomial g. In particular this strong property has been exploited for proving an extension of the Löwner-John ellipsoid theorem [5] which permits to completely characterize the sublevel set $\mathbf{K}$ (as in (1.1)) of minimum volume which contains a given set $\mathbf{\Omega} \subset \mathbb{R}^n$ (when minimizing over all positive homogeneous polynomials g of degree t). However, computing this sub-level set of minimum volume that contains $\mathbf{K}$ is a computational challenge as computing (or even approximating) the integral (1.2) is a hard problem.

We prove that $\mathrm{vol}(\mathbf{K})$ (and therefore the integral discriminant (1.2)) is the limit of a monotone sequence of generalized eigenvalue problems with respect to a pair

Research funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement ERC-ADG 666981 TAMING)..

of given real Hankel matrices of increasing size. All entries of both Hankel matrices are easy to obtain in closed-form and one Hankel matrix depends only on the degree of g. Therefore in principle the integral (1.2) can be approximated efficiently and as closely as desired by (linear algebra) eigenvalue routines. In addition, even if we do not provide a closed form expression, this new characterization of (1.2) as a limit or eigenvalue problems could bring new insights.

Methodology. Computing (and even approximating) the Lebesgue volume of a convex body is hard (let alone non-convex bodies). Often the only possibility is to use (non deterministic) Monte Carlo type methods which provide an estimate with statistical guarantees; that is, generate a sample of N points according to the uniform distribution on $[-1,1]^n$ and then the ratio $\rho_N :=$ (number of points in $\mathbf{K})/N$ provides such an estimate. However ρ_N is a random variable and is neither an upper bound or a lower bound on $\mathrm{vol}(\mathbf{K})$. For a discussion on volume computation the interested reader is referred to [2] and the many references therein.

However for basic semi-algebraic sets $\mathbf{K} \subset [-1,1]^n$, Henrion et al. [2] have provided a general methodology to approximate $\mathrm{vol}(\mathbf{K})$. it consists in solving a hierarchy $(\mathbf{Q}_d)_{d\in\mathbb{N}}$ of semidefinite programs[1] of increasing size, whose associated sequence of optimal values $(\rho_d)_{d\in\mathbb{N}}$ is monotone non increasing and converges to $\mathrm{vol}(\mathbf{K})$. An optimal solution of $\mathbf{Q}_d$ is a vector $\mathbf{y} \in \mathbb{R}^{s(2d)}$ (with $s(d) = \binom{n+d}{n}$) whose each coordinate y_α,$\alpha \in \mathbb{N}^n_{2d}$, approximates the α-moment of $\lambda_{\mathbf{K}}$, the restriction to $\mathbf{K}$ of the Lebesgue measure λ on $\mathbb{R}^n$ (and therefore y_0 approximates $\mathrm{vol}(\mathbf{K})$ from above). An optimal solution of the dual semidefinite program $\mathbf{Q}^*_d$ provides the coefficients $(p_\alpha)_{\alpha\in\mathbb{N}^n_{2d}}$ of a polynomial $p \in \mathbb{R}[\mathbf{x}]_{2d}$ which approximates on $[-1,1]$ and from above, the (indicator) function $\mathbf{x} \mapsto 1_{\mathbf{K}}(\mathbf{x}) = 1$ if $\mathbf{x} \in \mathbf{K}$ and 0 otherwise. In general the convergence $\rho_d \to \mathrm{vol}(\mathbf{K})$ is slow because of a Gibbs phenomenon[2] when one approximates the indicator function $1_{\mathbf{K}}$ by continuous functions. In [2] the authors have proposed a “trick” which accelerate drastically the convergence but at the price of loosing the monotone convergence $\rho_d \downarrow \mathrm{vol}(\mathbf{K})$. Another acceleration technique was provided in [6] which still preserves monotone convergence. It uses the fact that moments of $\lambda_{\mathbf{K}}$ satisfy linear equality constraints that follows from Stokes' theorem.

Recently, Jasour et al. [3] have considered volume computation in the context of risk estimation in uncertain environments. They have provided an elegant ”trick” which reduces computing the n-dimensional volume $\mathrm{vol}(\mathbf{K})$ to computing $\phi([0,1])$ for a certain pushforward measure ϕ on the real line, whose moments are known. (With $\mathbf{K}$ as in (1.1) the pushforward measure ϕ is with respect to the mapping g.) This results in solving the hierarchy of semidefinite programs proposed in [2], but now for measures on the real line as opposed to measures on $\mathbb{R}^n$. Solving the corresponding hierarchy of dual semidefinite programs amounts to approximate the indicator of an interval on the real line by polynomials of increasing degree, and whose coefficients minimize a linear criterion.

On the one hand, it yields drastical computational savings as passing from $\mathbb{R}^n$ to $\mathbb{R}$ is indeed a big and impressive progress. But on the other hand the (monotone) convergence remains slow as one cannot one cannot apply the acceleration technique based on Stokes' theorem proposed e.g. in [6] because the density of ϕ is not known

[1] A semidefinite program ...

[2] Gibbs

explicitly. However as the problem is now one-dimensional one may then solve many more steps of the resulting hierarchy of semidefinite programs provided that one works with a nice basis of polynomials (e.g. Chebyshev polynomials) to avoid numerical problems as much as possible. Interestingly, pushforward measures were also used in Magron et al. [8] to compute the Lebesgue volume of $f(\mathbf{K})$ for a polynomial mapping $f : \mathbb{R}^n \to \mathbb{R}^m$, but in this case one has to guess the measure in $\mathbb{R}^n$ whose pushforward measure is the Lebesgue measure on $f(\mathbf{K})$, and the resulting computation is still very expensive and limited to modest dimensions.

Contribution. We provide a very efficient way to approximate $\mathrm{vol}(\mathbf{K})$ with $\mathbf{K}$ as in (1.1) and when g is positive and homogeneous. To do so we are inspired by the trick of using the pushforward measure in Jasour et al. [3]. The novelty here is that by taking into account the specific nature (homogeneity) of g in (1.1) we are able to drastically simplify computations. Indeed, the hierarchy of semidefinite programs defined in [3] can be replaced (and improved) with computing a sequence of scalars $(\tau_d)_{d\in\mathbb{N}}$ where τ_d is nothing less than the generalized maximum eigenvalue of two known Hankel matrices of size d. Therefore there is *no* optimization involved anymore. Moreover, if one uses the basis of orthonormal polynomials w.r.t. the pushforward measure, then τ_d is now the maximum eigenvalue of a single real symmetric matrix of size d.

2. Notation and definitions

Notation. Let $\mathbb{R}[\mathbf{x}]$ denote the ring of polynomials in the variables $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbb{R}[\mathbf{x}]_t \subset \mathbb{R}[\mathbf{x}]$ denote the vector space of polynomials of degree at most t, hence of dimension $s(d) = \binom{n+t}{n}$. With $\alpha \in \mathbb{N}^n$ and $\mathbf{x} \in \mathbb{R}^n$, the notation $\mathbf{x}^\alpha$ stands for $x_1^{\alpha_1} \cdots x_n^{\alpha_n}$. Also for every $\alpha \in \mathbb{N}^n$, let $|\alpha| := \sum_i \alpha_i$ and $\mathbb{N}^n_d := \{\alpha \in \mathbb{N}^n : |\alpha| \leq d\}$.

The support of a Borel measure μ on $\mathbb{R}^n$ is the smallest closed set $\mathbf{\Omega}$ such that $\mu(\mathbb{R}^n \setminus \mathbf{\Omega}) = 0$. Denote by $\mathcal{B}(\mathbf{X})$ the Borel σ-field associated with a topological space $\mathbf{X}$, and $\mathscr{M}(\mathbf{X})$ the space of finite Borel measures on $\mathbf{X}$.

Given two real symmetric matrices $\mathbf{A}, \mathbf{C} \in \mathbb{R}^{n\times n}$ denote by $\lambda_{\min}(\mathbf{A}, \mathbf{C})$ the smallest generalized eigenvalue with respect to the pair $(\mathbf{A}, \mathbf{C})$, that is, the largest scalar θ such that $\mathbf{A}\mathbf{x} = \theta\,\mathbf{C}\mathbf{x}$ for some vector $\mathbf{x} \in \mathbb{R}^n$. When $\mathbf{C}$ is the identity matrix then $\lambda_{\min}(\mathbf{A}, \mathbf{C})$ is just the largest eigenvalue of $\mathbf{A}$. Computing $\lambda_{\min}(\mathbf{A}, \mathbf{C})$ can be done via a pure and efficient linear algebra routine.

The notation $\mathbf{A} \succeq 0$ (resp. $\mathbf{A} \succ 0$) stands for $\mathbf{A}$ is positive semidefinite (resp. positive definite).

Moment matrix. Given a real sequence $\boldsymbol{\phi} = (\phi_\alpha)_{\alpha\in\mathbb{N}^n}$, let $\mathbf{M}_d(\boldsymbol{\phi})$ denote the multivariate (Hankel-type) moment matrix defined by $\mathbf{M}_d(\boldsymbol{\phi})(\alpha, \beta) = \phi_{\alpha+\beta}$ for all $\alpha, \beta \in \mathbb{N}^n_d$. For instance, in the univariate case $n = 1$, with $d = 2$, $\mathbf{M}_2$ is the Hankel matrix

$$\mathbf{M}_2(\boldsymbol{\phi}) = \begin{bmatrix} \phi_0 & \phi_1 & \phi_2 \\ \phi_1 & \phi_2 & \phi_3 \\ \phi_2 & \phi_3 & \phi_4 \end{bmatrix}.$$

If $\boldsymbol{\phi} = (\phi_j)_{j\in\mathbb{N}}$ is the moment sequence of a Borel measure ϕ on $\mathbb{R}$ then $\mathbf{M}_d(\boldsymbol{\phi}) \succeq 0$ for all $d = 0, 1, \ldots$. Conversely, if $\mathbf{M}_d(\boldsymbol{\phi}) \succeq 0$ for all $d \in \mathbb{N}$, then $\boldsymbol{\phi}$ is the moment sequence of some finite Borel measure ϕ on $\mathbb{R}$. The converse result is not true anymore in the multivariate case.

Let ϕ, ν be two finite Borel measures on $\mathbb{R}$. The notation $\phi \leq \nu$ stands for $\phi(B) \leq \nu(B)$ for all $B \in \mathcal{B}(\mathbb{R})$.

Lemma 2.1. *Let ϕ, ν be two finite Borel measures on $\mathbb{R}$ with all moments $\boldsymbol{\phi} = (\phi_j)_{j\in\mathbb{N}}$ and $\boldsymbol{\nu} = (\nu_j)_{j\in\mathbb{N}}$ finite. Then $\phi \leq \nu$ if and only if*

$$\mathbf{M}_d(\boldsymbol{\phi}) \preceq \mathbf{M}_d(\boldsymbol{\nu}), \quad \forall d = 0, 1, \ldots$$

Proof. Only if part: $\phi \leq \nu$ implies that $\nu - \phi$ with associated sequence $\boldsymbol{\nu} - \boldsymbol{\phi} = (\nu_j - \phi_j)_{j\in\mathbb{N}}$ is a finite Borel measure on $\mathbb{R}$, and therefore:

$$\mathbf{M}_d(\boldsymbol{\nu}) - \mathbf{M}_d(\boldsymbol{\phi}) = \mathbf{M}_d(\boldsymbol{\nu} - \boldsymbol{\phi}) \succeq 0, \quad d \in \mathbb{N},$$

i.e., $\mathbf{M}_d(\boldsymbol{\nu}) \succeq \mathbf{M}_d(\boldsymbol{\phi})$ for all $d \in \mathbb{N}$.

If part: If $\mathbf{M}_d(\boldsymbol{\phi}) \preceq \mathbf{M}_d(\boldsymbol{\nu})$ for all $d \in \mathbb{N}$ then the sequence $\boldsymbol{\nu} - \boldsymbol{\phi} = (\nu_j - \phi_j)_{j\in\mathbb{N}}$ satisfies $\mathbf{M}_d(\boldsymbol{\phi} - \boldsymbol{\nu}) \succeq 0$ for all $d \in \mathbb{N}$. Therefore, the moment sequence $\boldsymbol{\nu} - \boldsymbol{\phi} = (\nu_j - \phi_j)_{j\in\mathbb{N}}$ of the possibly signed measure $\nu - \phi$ is in fact the moment sequence of a finite Borel (positive) measure on $\mathbb{R}$, and therefore $\nu \geq \phi$. □

Localizing matrix. Given a real sequence $\boldsymbol{\phi} = (\phi_\alpha)_{\alpha\in\mathbb{N}^n}$ and a polynomial $\mathbf{x} \mapsto p(\mathbf{x}) := \sum_\gamma p_\gamma \mathbf{x}^\gamma$, let $\mathbf{M}_d(p\,\boldsymbol{\phi})$ denote the real symmetric matrix defined by:

$$\mathbf{M}_d(p\,\boldsymbol{\phi})(\alpha, \beta) = \sum_\gamma p_\gamma\, \phi_{\alpha+\beta+\gamma}, \quad \alpha, \beta \in \mathbb{N}^n_d.$$

For instance, with $n = 1$, $d = 2$ and $x \mapsto p(x) = x(1 - x)$:

$$\mathbf{M}_2(p\,\boldsymbol{\phi}) = \left[\begin{array}{ccc} \phi_1 - \phi_0 & \phi_2 - \phi_1 & \phi_3 - \phi_2 \\ \phi_2 - \phi_1 & \phi_3 - \phi_2 & \phi_4 - \phi_3 \\ \phi_3 - \phi_2 & \phi_4 - \phi_3 & \phi_5 - \phi_4 \end{array}\right],$$

also a Hankel matrix.

Lemma 2.2. *Let $x \mapsto p(x) = x(1 - x)$.*

(i) If a real (finite) sequence $\boldsymbol{\phi} = (\phi_j)_{j\leq 2d}$ satisfies $\mathbf{M}_d(\boldsymbol{\phi}) \succeq 0$ and $\mathbf{M}_{d-1}(p\,\boldsymbol{\phi}) \succeq 0$ then there is a measure μ on $[0, 1]$ whose moments $\boldsymbol{\mu} = (\mu_j)_{j\leq 2d}$ match $\boldsymbol{\phi}$.

(ii) If a real (infinite) sequence $\boldsymbol{\phi} = (\phi_j)_{j\in\mathbb{N}}$ satisfies $\mathbf{M}_d(\boldsymbol{\phi}) \succeq 0$ and $\mathbf{M}_d(p\,\boldsymbol{\phi}) \succeq 0$ for all d, then there is a measure μ on $[0, 1]$ whose moments $\boldsymbol{\mu} = (\mu_j)_{j\in\mathbb{N}}$ match $\boldsymbol{\phi}$.

See for instance Lasserre [7] and the many references therein.

Pushforward measure. Let $\mathbf{K} \subset \mathbb{R}^n$ be a Borel set and λ a probability measure on $\mathbf{K}$. Given a measurable mapping $f : \mathbf{K} \to \mathbb{R}^p$, the pushforward measure of λ on $\mathbb{R}^p$ w.r.t. f is denoted by $\#\lambda$ and satisfies:

$$\#\lambda(B) := \lambda(f^{-1}(B)), \qquad \forall B \in \mathcal{B}(\mathbb{R}^p).$$

In particular, its moments are given by

$$\#\lambda_\alpha := \int_{\mathbb{R}^p} \mathbf{z}^\alpha\, \#\lambda(d\mathbf{z}) = \int_{\mathbf{K}} f(\mathbf{x})^\alpha\, \lambda(d\mathbf{x}), \qquad \alpha \in \mathbb{N}^p. \tag{2.1}$$

A version of Stokes' theorem. Let $\mathbf{\Omega} \subset \mathbb{R}^n$ be an open subset with boundary $\partial\mathbf{\Omega}$ and let $\mathbf{x} \mapsto \mathbf{X}(\mathbf{x})$ be a given vector field. Then under suitable smoothness assumptions,

$$\int_{\mathbf{\Omega}} \mathrm{Div}(X) f(\mathbf{x})\, d\mathbf{x} + \int_{\mathbf{\Omega}} \langle \mathbf{X}, \nabla f(\mathbf{x})\rangle\, d\mathbf{x} \;=\; \int_{\partial\mathbf{\Omega}} \langle \vec{n}_{\mathbf{x}}, \mathbf{X}\rangle f(\mathbf{x})\, d\sigma, \tag{2.2}$$

where $\vec{n}_{\mathbf{x}}$ is the outward pointing normal to $\mathbf{\Omega}$ at $\mathbf{x} \in \partial\mathbf{\Omega}$, and σ is the $(n-1)$-dimensional Hausdorff measure on the boundary $\partial\mathbf{\Omega}$.

3. Main result

Let $\mathbf{B} := [-1,1]^n$ and $\mathbf{K} \subset \mathbf{B}$ be as in (1.1) with $\partial\mathbf{K} \subset \{\mathbf{x} : g(\mathbf{x}) = 1\}$. Let λ be the Lebesgue measure on $\mathbf{B}$ normalized to a probability measure so that $\mathrm{vol}(\mathbf{K}) = 2^n\lambda(\mathbf{K})$. Let g in (1.1) be a nonnegative and homogeneous polynomial of degree t. Denote by:

$$\overline{b}_g \;:=\; \max\{\, g(\mathbf{x}) : \mathbf{x} \in \mathbf{B}\,\}; \quad \underline{a}_g \;:=\; \min\{\, g(\mathbf{x}) : \mathbf{x} \in \mathbf{B}\,\}. \tag{3.1}$$

Notice that $\underline{a}_g = 0$ because $0 \in \mathbf{B}$ and g is nonnegative with $g(0) = 0$, and therefore $g(\mathbf{B}) = [0, \overline{b}_g]$. We next follow an elegant idea of Jasour et al. [3], adapted to the present context. It reduces the computation of $\mathrm{vol}(\mathbf{K})$ (in $\mathbb{R}^n$) to a certain volume computation in $\mathbb{R}$, by using a particular pushforward measure of the Lebesgue measure λ on $\mathbf{B}$.

Let $\#\lambda$ be the pushforward on the positive half line of λ, by the polynomial mapping $g : \mathbf{B} \to [0, \overline{b}_g]$. From (3.1), the support of $\#\lambda$ is the interval $I := [0, \overline{b}_g] \subset \mathbb{R}$. Then in view of (2.1):

$$\#\lambda_k \;:=\; \int_I z^k\, \#\lambda(dz) \;=\; \int_{\mathbf{B}} g(\mathbf{x})^k\, \lambda(d\mathbf{x}), \quad k = 0, 1, \ldots \tag{3.2}$$

All scalars $(\#\lambda_k)_{k\in\mathbb{N}}$ can be obtained in closed form as g is a polynomial and λ is the (normalized) Lebesgue measure on $\mathbf{B}$. Namely, writing the expansion

$$\mathbf{x} \mapsto g(\mathbf{x})^k \;=\; \sum_{\alpha\in\mathbb{N}^n_{kd}} g_{k\alpha}\, \mathbf{x}^\alpha,$$

for some coefficients $(g_{k\alpha})$, one obtains:

$$\#\lambda_k \;=\; 2^{-n} \sum_{\alpha\in\mathbb{N}^n_{kd}} g_{k\alpha} \left(\prod_{i=1}^n \frac{(1-(-1)^{\alpha_i+1})}{\alpha_i + 1} \right), \quad k = 0, 1, \ldots \tag{3.3}$$

Next observe that $2^{-n}\mathrm{vol}(\mathbf{K}) = \#\lambda(g(\mathbf{K}))$ and note that $g(\mathbf{K}) = [0,1]$. Therefore following the recipe introduced in Henrion et al. [2], and with $S := [0,1] \subset [0, \overline{b}_g]$:

$$\#\lambda(S) \;=\; \max_{\phi\in\mathscr{M}(S)} \{\, \phi(S) : \; \phi \;\leq\; \#\lambda\,\}. \tag{3.4}$$

Denote by ϕ^* the mesure on the real line which is the restriction to $S \subset I$ of the pushforward measure $\#\lambda$. That is,

$$\phi^*(B) \;:=\; \#\lambda(B \cap S), \qquad \forall B \in \mathcal{B}(\mathbb{R}). \tag{3.5}$$

Then ϕ^* is the unique optimal solution of (3.4); see e.g. Henrion et al. [2]. Then to approximate ϕ^*_0 from above one possibility is to solve the hierarchy of semidefinite relaxations:

$$\rho_d \,=\, \max_{\boldsymbol{\phi}} \{\, \phi_0 :\, 0 \preceq \mathbf{M}_d(\boldsymbol{\phi}) \,\preceq\, \mathbf{M}_d(\#\lambda); \quad \mathbf{M}_{d-1}(x(1-x)\,\boldsymbol{\phi}) \succeq 0 \,\} \tag{3.6}$$

where $\boldsymbol{\phi} = (\phi_j)_{j\leq 2d}$, $\mathbf{M}_d(\#\lambda)$ is the (Hankel) moment matrix (with moments up to order $2d$) associated with the pushforward measure $\#\lambda$, and $\mathbf{M}_d(\boldsymbol{\phi})$ (resp. $\mathbf{M}_{d-1}(z(1-z)\,\boldsymbol{\phi})$) is the Hankel moment (resp. localizing) matrix (with moments up to order $2d$) associated with the sequence $\boldsymbol{\phi}$ and the polynomial $x \mapsto p(x) = x(1-x)$; see §2. Indeed (3.6) is a relaxation of (3.4) and the sequence $(\rho_d)_{d\in\mathbb{N}}$ is monotone non increasing and converges to $\#\lambda(S)$ from above; see e.g. [2].

The dual of (3.6) is the semidefinite program

$$\rho_d^* \,=\, \max_{p\in\mathbb{R}[x]_{2d}} \{ \int p\, d\#\lambda :\, p \geq 1 \text{ on } [0,1];\, p \geq 0 \text{ on } [0, \overline{b}_g] \,\}, \tag{3.7}$$

and if $\mathbf{K}$ has nonempty interior then $\rho_d^* = \rho_d$.

This is the approach advocated by Jasour et al. [3] and indeed this reduction of the initial (Lebesgue) volume computation in $\mathbb{R}^n$

$$\mathrm{vol}(\mathbf{K}) \,=\, \max_{\phi\in\mathscr{M}(\mathbf{K})} \{\, \phi(\mathbf{K}) :\, \phi \,\leq\, \lambda \,\} \tag{3.8}$$

to instead compute $\#\lambda([0,1])$ (in $\mathbb{R}$) by solving (3.4) is quite interesting as it yields drastic computational savings; in fact solving the multivariate analogues for (3.8) of the univariate semidefinite relaxations (3.6) for (3.4), becomes rapidly impossible even for moderate d, except for problems of modest dimension (say e.g. $n \leq 4$).

However it is important to notice that in general the convergence $\rho_d \downarrow \#\lambda(S)$ is very slow and numerical problems are expected for large values of d. To partially remedy this problem the authors of [3] suggest to express moment and localizing matrices in (3.6) in the Chebyshev basis (as opposed to the standard monomial basis). This allows to solve a larger number of relaxations but it does not change the typical slow convergence. The trick based on Stokes' theorem used in [6] cannot be used here because the dominating (or reference) measure $\#\lambda$ in (3.4) is *not* the Lebesgue measure λ anymore (as in (3.8)). On the other hand the trick to accelerate convergence used in [2] can still be used, that is, in (3.6) one now maximizes $L_{\boldsymbol{\phi}}(x(1-x)) = \phi_1 - \phi_2$ instead of ϕ_0. If $\boldsymbol{\phi}^d = (\phi_j^d)_{j\leq 2d}$ is an optimal solution of (3.6) then $\phi_0^d \to \#\lambda(S)$ as d increases but one looses the monotone convergence from above.

In the sequel we show that in the particular case where g is positive and homogeneous then one can avoid solving the hierarchy (3.6) and instead solve a hierarchy of simple generalized eigenvalue problems with *no* optimization involved and with a much faster convergence.

3.1. Exploiting homogeneity.

A crucial observation. Recall that $2^{-n}\mathrm{vol}(\mathbf{K}) = \#\lambda(S)$. So let ϕ^* be as in (3.5), and let $\boldsymbol{\phi}^* = (\phi_j^*)_{j\in\mathbb{N}}$ be its associated sequence of moments. Consider the vector field $\mathbf{x} \mapsto \mathbf{X}(\mathbf{x}) := \mathbf{x}$. Then $\mathrm{Div}(\mathbf{X}) = n$. In addition, as g is homogeneous of degree t then by Euler's identity, $\langle \mathbf{x}, \nabla g(\mathbf{x})\rangle = t\, g(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^n$. Recall that $\mathbf{K} \subset \mathrm{int}(\mathbf{B})$ and therefore $g(\mathbf{x}) = 1$ for every $\mathbf{x} \in \partial\mathbf{K}$. Next, for every $j \in \mathbb{N}$, as

$g(\mathbf{x})^j = 1$ on $\partial\mathbf{K}$ for all $j \in \mathbb{N}$, Stokes' Theorem (2.2) yields:

$$\begin{aligned}
0 &= \int_{\partial\mathbf{K}} \langle \vec{n}_{\mathbf{x}}, \mathbf{x}\rangle\,(1-g(\mathbf{x})^j)\,d\sigma \quad [\text{as } g(\mathbf{x})=1 \text{ on } \partial\mathbf{K}]\\
&= n\int_{\mathbf{K}} (1-g(\mathbf{x})^j)\,\lambda(d\mathbf{x}) + \int_{\mathbf{K}} \langle \mathbf{x}, \nabla(1-g(\mathbf{x})^j)\rangle\,\lambda(d\mathbf{x}) \quad [\text{by Stokes}]\\
&= n\,\lambda(\mathbf{K}) - (n+jt)\int_{\mathbf{K}} g(\mathbf{x})^j\,\lambda(d\mathbf{x})\\
&= n\,\#\lambda(S) - (n+jt)\int_{g(\mathbf{K})} z^j\,\#\lambda(dz) \;=\; n\,\phi_0^* - (n+jt)\,\phi_j^*,
\end{aligned}$$

so that we have proved:

Lemma 3.1. *Let ϕ^* be the Borel measure on $\mathbb{R}$ which is the restriction to $S = [0,1]$ of the pushforward measure $\#\lambda$ on I. Then its moments $\boldsymbol{\phi}^* = (\phi_j^*)_{j\in\mathbb{N}}$ satisfy :*

$$\phi_0^* \;=\; 2^{-n}\,\mathrm{vol}(\mathbf{K}); \quad \phi_j^* \;:=\; \frac{n}{n+jt}\,\phi_0^*, \quad j = 1,2,\ldots \tag{3.9}$$

Remarkably, Lemma 3.1 relates all moments of ϕ^* to the mass $\phi_0^* = 2^{-n}\mathrm{vol}(\mathbf{K})$ in very simple manner! However it now remains to compute ϕ_0^*.

Computing ϕ_0^*. Define $\mathbf{M}_d^*$ to be the Hankel (moment) matrix with entries:

$$\mathbf{M}_d^*(k,\ell) \;:=\; \frac{n}{n+(k+\ell-2)\,t}, \quad k,\ell \;=\; 1,2,\ldots, \tag{3.10}$$

so that $\phi_0^*\,\mathbf{M}_d^* = \mathbf{M}_d(\boldsymbol{\phi}^*)$ for all $d \in \mathbb{N}$, where $\mathbf{M}_d(\boldsymbol{\phi}^*)$ is the Hankel moment matrix associated with the sequence $\boldsymbol{\phi}^*$.

Similarly, define $\mathbf{M}_{d,x(1-x)}^*$ to be the Hankel matrix with entries:

$$\mathbf{M}_{d,x(1-x)}^*(k,\ell) \;:=\; \frac{n}{n+(k+\ell-1)\,t} - \frac{n}{n+(k+\ell-2)\,t}, \quad k,\ell \;=\; 1,2,\ldots, \tag{3.11}$$

so that $\phi_0^*\,\mathbf{M}_{d,x(1-x)}^* = \mathbf{M}_d(x(1-x)\,\boldsymbol{\phi}^*)$ is the localizing matrix associated with $\boldsymbol{\phi}^*$ and the polynomial $x \mapsto x(1-x)$, for all $d \in \mathbb{N}$. As ϕ^* is supported on $[0,1]$ then $\phi_0^*\,\mathbf{M}_{d,x(1-x)}^* \succeq 0$ for all d, which in turn implies

$$\mathbf{M}_{d,x(1-x)}^* \;\succeq\; 0, \qquad \forall d \in \mathbb{N}, \tag{3.12}$$

because $\phi_0^* > 0$.

Theorem 3.2. *For each $d \in \mathbb{N}$, let $\mathbf{M}_d^*$ be as in* (3.10) *and let $\mathbf{M}_d(\#\lambda)$ be the Hankel moment matrix associated with $\#\lambda$ (hence with sequence of moments as in* (3.3)*). Then :*

$$\phi_0^* \;=\; \lim_{d\to\infty} \lambda_{\min}(\mathbf{M}_d(\#\lambda), \mathbf{M}_d^*), \tag{3.13}$$

i.e., ϕ_0^* is the limit of a sequence of maximum generalized eigenvalues associated with the pair $(\mathbf{M}_d(\#\lambda), \mathbf{M}_d^*))$, $d \in \mathbb{N}$.

Proof. For every $d \in \mathbb{N}$, let

$$\tau_d := \lambda_{\min}(\mathbf{M}_d(\#\lambda), \mathbf{M}_d^*) \;=\; \max\{\,\tau :\; \tau\,\mathbf{M}_d^* \;\preceq\; \mathbf{M}_d(\#\lambda)\,\}, \tag{3.14}$$

which implies $\tau_d\,\mathbf{M}_d^* \preceq \mathbf{M}_d(\#\lambda)$. In addition, $\tau_d\,\mathbf{M}_{d,x(1-x)}^* \succeq 0$ follows from (3.12). On the other hand, as $\phi^* \leq \#\lambda$, we also have $\phi_0^*\mathbf{M}_d^* = \mathbf{M}_d(\boldsymbol{\phi}^*) \preceq \mathbf{M}_d(\#\lambda)$ for all $d \in \mathbb{N}$. Hence, $\phi_0^* \leq \tau_d$ for all $d \in \mathbb{N}$, and the sequence $(\tau_d)_{d\in\mathbb{N}}$ is monotone non increasing.

We next show that $\tau_d \downarrow \phi_0^*$ as $d \to \infty$. As $\tau_d \geq \phi_0^*$ for all d, $\lim_{d\to\infty} \tau_d = \tau^* \geq \phi_0^*$. Consider the sequence $\boldsymbol{\mu} = (\mu_j)_{j\in\mathbb{N}}$ defined by:

$$\mu_j \,=\, \tau^* \,\frac{n}{n+jt}, \quad j \in \mathbb{N}.$$

Then from $\tau_d \mathbf{M}_d^* \succeq 0$ for all d, and the convergence $\tau_d \to \tau^*$, we obtain

$$\tau^* \,\mathbf{M}_d^* \,=\, \mathbf{M}_d(\boldsymbol{\mu}) \,\succeq\, 0, \quad \forall d \in \mathbb{N}.$$

For similar reasons

$$\tau^* \,\mathbf{M}_{d,x(1-x)}^* \,=\, \mathbf{M}_d(x(1-x)\,\boldsymbol{\mu}) \,\succeq\, 0, \quad \forall d \in \mathbb{N}.$$

By Lemma 2.2, $\boldsymbol{\mu}$ is the moment sequence of a measure μ supported on $[0,1]$ with mass $\mu_0 = \mu([0,1]) = \tau^*$, and by construction we also have $\mu \leq \#\lambda$. Therefore $\mu \in \mathscr{M}(S)$ is a feasible solution of (3.4) which implies $\mu([0,1]) \leq \phi_0^*$. But on the other hand,

$$\phi_0^* \,\leq\, \tau^* \,=\, \mu([0,1]) \,\leq\, \phi_0^*,$$

which yields the desired result $\tau^* = \phi_0^*$. □

Therefore to approximate $\mathrm{vol}(\mathbf{K})$ from above, one proceeds as follows. Start with $d = 1$ and then

- Compute all moments of $\#\lambda$ up to order $2d$ by (3.3).
- Compute $\tau_d := \lambda_{\min}(\mathbf{M}_d(\#\lambda), \mathbf{M}_d^*)$
- set $d = d + 1$ and repeat.

This produces the required monotone sequence of upper bounds $(\tau_d)_{d\in\mathbb{N}}$ on ϕ_0^*, which converges to $\phi_0^* = 2^{-n}\mathrm{vol}(\mathbf{K})$ as d increases. Finally, the following result shows that $\tau_d \leq \rho_d$.

Proposition 3.3. *For each $d \in \mathbb{N}$, let ρ_d (resp. τ_d) be as in* (3.6) *(resp.* (3.14)*). Then $\rho_d \geq \tau_d$.*

Proof. Consider the sequence $\boldsymbol{\mu} = (\mu_j)_{j\leq 2d}$ defined by:

$$\mu_j \,=\, \tau_d \,\frac{n}{n+jt}, \quad j \leq 2d.$$

Then from (3.14), $\tau_d \mathbf{M}_d^* = \mathbf{M}_d(\boldsymbol{\mu})$ and therefore, $0 \preceq \mathbf{M}_d(\boldsymbol{\mu}) \preceq \mathbf{M}_d(\#\lambda)$. Similarly $\tau_d \,\mathbf{M}_{d-1,x(1-x)} = \mathbf{M}_{d-1}(x(1-x)\,\boldsymbol{\mu}) \succeq 0$. In other words, the sequence $\boldsymbol{\mu}$ is a feasible solution of (3.6), which implies $\mu_0\,(=\tau_d)\, \leq \rho_d$. □

Hence the above eigenvalue procedure (with no optimization involved) provides a monotone sequence of upper bounds on ϕ_0^* that are better than the sequence of upper bounds $(\rho_d)_{d\in\mathbb{N}}$ obtained by solving the hierarchy of semidefinite relaxations (3.6). Notice also that the matrix $\mathbf{M}_d^*$ depends only on the degree of g and not on g itself.

In fact there is a simple interpretation of this improvement. In Problem (3.4) and in its associated semidefinite relaxations (3.6), one may include the additional constraints

$$\phi_j \,=\, n\,\phi_0/(n+jt), \quad \forall j \leq 2d, \tag{3.15}$$

coming from Stokes' theorem applied to ϕ^*; see Lemma 3.1. Indeed we are allowed to do that because ϕ^* (which is the unique optimal solution of (3.4)) satisfies these additional constraints. If it it does not change the optimal value of (3.4) it changes that of (3.6) as it makes the corresponding relaxation stronger and therefore $\tau_d \leq \rho_d$ for all d.

Remark 3.4. *If one uses the basis of orthonormal polynomials with respect to the pushforward measure $\#\lambda$ then the new moment matrix $\tilde{\mathbf{M}}_d(\#\lambda)$ (expressed in this basis) is the identity matrix and therefore τ_d is the maximum eigenvalue of the corresponding matrix $\tilde{M}_d^*$ (also expressed in that basis). This basis of orthonormal polynomials can be obtained from the decomposition $\mathbf{M}_d(\#\lambda) = \mathbf{D}\mathbf{D}^T$ for triangular matrices $\mathbf{D}$ and $\mathbf{D}^T$.*

Example 1. *The following elementary example (the unit disk in $\mathbb{R}^2$) is to show the first two steps and to confirm the faster convergence. Let $n = 2$ and $g = \|\mathbf{x}\|^2 = x_1^2 + x_2^2$, and $\mathbf{B} = [-1, 1]^2$, so that $\mathrm{vol}(\mathbf{K}) = \pi$.*

$$\mathbf{M}_1^* = \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{bmatrix}; \quad \mathbf{M}_1(\#\lambda) = \begin{bmatrix} 1 & 2/3 \\ 2/3 & 28/45 \end{bmatrix}$$

This yields $4 \cdot \tau_1 \approx 3.48$ which is already a good upper bound on π whereas $4 \cdot \rho_1 = 4$.

$$\mathbf{M}_2^* = \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{bmatrix}; \quad \mathbf{M}_2(\#\lambda) = \begin{bmatrix} 1 & 2/3 & 28/45 \\ 2/3 & 28/45 & 24/35 \\ 28/45 & 24/35 & 2/9 + 8/21 + 6/25 \end{bmatrix}$$

This yields $4 \cdot \tau_2 \approx 3.1440$ while $4 \cdot \rho_2 = 3.8928$. Hence τ_2 provides a very good upper bound on π with only moments of order 4.

In dimension n, $\mathrm{vol}(\mathbf{K}) = \frac{1}{\Gamma(1+n/2)}\,\pi^{n/2}$.

For $n = 3$, $\mathrm{vol}(\mathbf{K}) = 4.1888$ and we obtain $2^n\tau_1 \approx 4.6881$ whereas $2^n\rho_1 = 8$. $2^n\tau_2 \approx 4.2517$ whereas $2^n\rho_2 = 6.626$. $2^n\tau_3 \approx 4.1955$ whereas $2^n\rho_2 = 6.398$. $2^n\tau_4 \approx 4.1894$ whereas $2^n\rho_2 = 6.086$.

For $n = 5$, $\mathrm{vol}(\mathbf{K}) \approx 5.2638$ and we obtain $2^n\tau_1 \approx 10.2892$, $2^n\tau_2 \approx 6.5248$, $2^n\tau_3 \approx 5.5755$, whereas $2^n\rho_1 = 16$, $2^n\rho_2 = 15.63$, and $2^n\rho_=13.36$.

4. Extensions

An immediate extension is when $\mathbf{K} = \{\mathbf{x} : g_j(\mathbf{x}) \leq 1,\ j = 1, \ldots, m\} \subset (-1, 1)^n$ for a family $(g_j)_{j=1}^m$ of positive homogeneous polynomials, not necessarily of same degree, say $\deg(g_j) = t_j$. In this case one may proceed again as suggested in Jasour et al. [3]. Now $\#\lambda$ is the pushforward on $\mathbb{R}^m$ of λ on $\mathbf{B}$, by the mapping:

$$g : \mathbf{B} \to \mathbb{R}^m, \quad g(\mathbf{x}) = \begin{bmatrix} g_1(\mathbf{x}) \\ \cdots \\ g_m(\mathbf{x}) \end{bmatrix}.$$

In particular $\#\lambda$ has its moments defined by:

$$\#\lambda_\alpha = \int_{\mathbf{B}} g_1(\mathbf{x})^{\alpha_1} \cdots g_m(\mathbf{x})^{\alpha_m}\, \lambda(d\mathbf{x}) = \int_{g(\mathbf{B})} \mathbf{z}^\alpha\, \#\lambda(d\mathbf{z}), \quad \forall \alpha \in \mathbb{N}^m.$$

Again all moments $\#\lambda_\alpha$ can be computed in closed form, and again with $S = [0, 1]^m$ $2^{-n}\mathrm{vol}(\mathbf{K}) = \#\lambda(S)$.

Let us describe how the generalization works for the case $m = 2$. Again denote by ϕ^* on $\mathbb{R}^2$ the restriction of $\#\lambda$ to S and let $\boldsymbol{\phi}^* = (\phi_{ij}^*)_{i,j\in\mathbb{N}}$ with:

$$\phi_{ij}^* := \int_S z_1^i z_2^j\, \phi^*(d\mathbf{z}), \quad \forall i, j = 0, 1, \ldots.$$

So the bivariate analogues of the semidefinite relaxations (3.6) read:

$$\rho_d = \max_{\boldsymbol{\phi}} \{ \phi_0 : \quad 0 \preceq \mathbf{M}_d(\boldsymbol{\phi}) \preceq \mathbf{M}_d(\#\lambda) \qquad \mathbf{M}_d(x_j(1-x_j)\,\boldsymbol{\phi}) \succeq 0, \quad j=1,2\,\}, \tag{4.1}$$

where $\boldsymbol{\phi} = (\phi_{ij})_{i+j\leq 2d}$, and $\mathbf{M}_d(\boldsymbol{\phi})$ (resp. $\mathbf{M}_{d-1}(x_j(1-x_j)\,\boldsymbol{\phi})$, $j=1,2$) is the moment (resp. localizing) matrix associated with $\boldsymbol{\phi}$ (resp. with $\boldsymbol{\phi}$ and $\mathbf{x}\mapsto x_j(1-x_j)$, $j=1,2$). Then $\rho_d\downarrow\#\lambda(S)$ as $d\to\infty$. Again the semidefinite relaxations (4.1) are a lot cheaper to solve than those associated with the n-variate problem (3.8).

As we did for the univariate case we can improve the above convergence by adding additional constraints that must be satisfied at the optimal solution ϕ^* of (3.8). Again $\phi_0^* = 2^{-n}\mathrm{vol}(\mathbf{K})$. Let $(i,j,k,\ell)\in\mathbb{N}^4$ with $k,\ell\geq 1$. Then with $X(\mathbf{x})=\mathbf{x}$, Stokes's Theorem yields

$$\begin{aligned}
0 &= n\int_{\mathbf{K}} g_1^i g_2^j\,(1-g_1)^k(1-g_2)^\ell\lambda(d\mathbf{x})\\
&\quad+\int_{\mathbf{K}}\langle\mathbf{x},\nabla[g_1^ig_2^j(1-g_1)^k(1-g_2)^\ell]\rangle\,\lambda(d\mathbf{x})\\
&= n\int_S z_1^iz_2^j\,(1-z_1)^k(1-z_2)^\ell\,\#\lambda(d\mathbf{z})\\
&\quad+it_1\int_S z_1^iz_2^j\,(1-z_1)^k(1-z_2)^\ell\,\#\lambda(d\mathbf{z})\\
&\quad+jt_2\int_S z_1^iz_2^j\,(1-z_1)^k(1-z_2)^\ell\,\#\lambda(d\mathbf{z})\\
&\quad-kt_1\int_S z_1^{i+1}z_2^j\,(1-z_1)^{k-1}(1-z_2)^\ell\,\#\lambda(d\mathbf{z})\\
&\quad-\ell t_2\int_S z_1^iz_2^{j+1}(1-z_1)^k(1-z_2)^{\ell-1}\,\#\lambda(d\mathbf{z})
\end{aligned}$$

That is, for each $(i,j,k,\ell)\in\mathbb{N}^4$ with $k,\ell\geq 1$ one obtains a linear constraint that links the moments of ϕ^*, that we denote by $L_{ijk\ell}(\boldsymbol{\phi}^*)=0$. For instance,

$$0 = L_{0011}(\phi^*) = n(\phi_0^*-\phi_{10}^*-\phi_{01}^*+\phi_{11}^*)-t_1(\phi_0^*-\phi_{01}^*)-t_2(\phi_0^*-\phi_{10}^*)$$

$$0 = L_{1111}(\phi^*) = (n+t_1+t_2)\,(\phi_{11}^*+\phi_{22}^*-\phi_{21}^*-\phi_{12}^*)-t_1(\phi_{21}^*-\phi_{22}^*)-t_2(\phi_{12}^*-\phi_{22}^*),$$

etc. So we can add this additional constraints to (4.1) and solve

$$\tau_d = \max_{\boldsymbol{\phi}}\{\phi_0 : \quad 0\preceq\mathbf{M}_d(\boldsymbol{\phi})\preceq\mathbf{M}_d(\#\lambda) \qquad \mathbf{M}_d(x_i(1-x_j)\,\boldsymbol{\phi})\succeq 0,\quad j=1,2 \qquad L_{i,j,k,\ell}(\boldsymbol{\phi})=0,\quad k,\ell\geq 1;\ i+j+k+\ell\leq 2d\,\} \tag{4.2}$$

Of course and again $\tau_d\leq\rho_d$ for all d and therefore $\tau_d\downarrow\#\lambda(S)$ as d increases. The difference with the univariate case is that now computing τ_d still requires to solve a semidefinite program, namely (4.2). However it is of same dimension as (4.1) and the convergence $\tau_d\downarrow\#\lambda(S)$ is expected to be much faster than $\rho_d\downarrow\#\lambda(S)$.

5. Conclusion

References

[1] Dolotin V., Morozov. A. Introduction to non-linear algebra, hep-th/0609022 [SPIRES].

[2] Henrion D., Lasserre J.B., Savorgnan C. Approximate volume and integration for basic semi-algebraic sets *SIAM Review* **51**, pp. 722–743, 2009

[3] Jasour A., Hofmann A., Williams, B.C. Moment-Sum-Of-Squares Approach For Fast Risk Estimation In Uncertain Environments, `arXiv:1810.01577`, 2018.

[4] Lasserre J.B. Level sets and non Gaussian integrals of positively homogeneous functions *Int. Game Theory Review* **17**, No 1, 28 pp., 2015. 2015

[5] Lasserre J.B. A generalization of Löwner-John's ellipsoid theorem *Math. Program.* **152**, No 1 pp. 559–591, 2015

[6] Lasserre J.B. Computing Gaussian & exponential measures of semi-algebraic sets, *Adv. Appl. Math.* **91**, pp. 137–163, 2017.

[7] Lasserre J.B. *Moments, Positive Polynomials and Their Applications*, Imperial College Press, 2010.

[8] Magron V., Henrion D., Lasserre J.B. Semidefinite approximations of projections and polynomial images of semi-algebraic sets, *SIAM J. Optim.* **25**, pp. 2143–2164, 2015.

[9] Morozov, A., Shakirov, S. Introduction to integral discriminants, J. High Energy Phys. 12, No. 12, 2009.

LAAS-CNRS and Institute of Mathematics, University of Toulouse, LAAS, 7 avenue du Colonel Roche, 31077 Toulouse Cédex 4, France, Tel: +33561336415

E-mail address: `lasserre@laas.fr`