



**HAL**  
open science

# Representation of chance-constraints with strong asymptotic guarantees

Jean-Bernard Lasserre

► **To cite this version:**

Jean-Bernard Lasserre. Representation of chance-constraints with strong asymptotic guarantees. 2017. hal-01487006v1

**HAL Id: hal-01487006**

**<https://hal.science/hal-01487006v1>**

Preprint submitted on 10 Mar 2017 (v1), last revised 11 May 2017 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Representation of chance-constraints with strong asymptotic guarantees

Jean B. Lasserre<sup>1</sup>

**Abstract**—Given  $\epsilon \in (0, 1)$ , a probability measure  $\mu$  on  $\Omega \subset \mathbb{R}^p$  and a semi-algebraic set  $\mathbf{K} \subset \mathbf{X} \times \Omega$ , we consider the feasible set

$$\mathbf{X}_\epsilon^* = \{\mathbf{x} \in \mathbf{X} : \text{Prob}[(\mathbf{x}, \omega) \in \mathbf{K}] \geq 1 - \epsilon\}$$

associated with a chance-constraint. We provide a sequence of outer approximations  $\mathbf{X}_\epsilon^d = \{\mathbf{x} \in \mathbf{X} : h_d(\mathbf{x}) \geq 0\}$ ,  $d \in \mathbb{N}$ , where  $h_d$  is a polynomial of degree  $d$  whose vector of coefficients is an optimal solution of a semidefinite program. The size of the latter increases with the degree  $d$ . We also obtain the strong and highly desirable asymptotic guarantee that  $\lambda(\mathbf{X}_\epsilon^d \setminus \mathbf{X}_\epsilon^*) \rightarrow 0$  as  $d$  increases, where  $\lambda$  is the Lebesgue measure on  $\mathbf{X}$ .

## I. INTRODUCTION

We consider the following general framework for decision under uncertainty : Let  $\mathbf{x} \in \mathbf{X} \subset \mathbb{R}^n$  be a decision variable while  $\omega \in \mathbb{R}^p$  is a *disturbance* (or *noise*) parameter whose distribution  $\mu$  (with support  $\Omega \subset \mathbb{R}^p$ ) is known, i.e., its list of moments  $\mu_\beta := \int_\Omega \omega^\beta d\mu(\omega)$ ,  $\beta \in \mathbb{N}^p$ , is available in closed form or numerically.

Both  $\mathbf{x}$  and  $\omega$  are linked by constraints of the form  $(\mathbf{x}, \omega) \in \mathbf{K} \subset \mathbf{X} \times \Omega$ , where

$$\mathbf{K} = \{(\mathbf{x}, \omega) : g_j(\mathbf{x}, \omega) \geq 0, \quad j = 1, \dots, m\}, \quad (1)$$

for some polynomials  $(g_j) \subset \mathbb{R}[\mathbf{x}, \omega]$ , that is,  $\mathbf{K}$  is a basic semi-algebraic set.

Next, for each fixed  $\mathbf{x} \in \mathbf{X}$ , let  $\mathbf{K}_\mathbf{x} \subset \Omega$  be the (possibly empty) set defined by:

$$\mathbf{K}_\mathbf{x} := \{\omega \in \Omega : (\mathbf{x}, \omega) \in \mathbf{K}\}, \quad \mathbf{x} \in \mathbf{X}. \quad (2)$$

Let  $\epsilon \in (0, 1)$  be fixed. The goal of this paper is to provide tight approximations of the set

$$\begin{aligned} \mathbf{X}_\epsilon^* &:= \{\mathbf{x} \in \mathbf{X} : \mu(\mathbf{K}_\mathbf{x}) \geq 1 - \epsilon\} \\ &= \{\mathbf{x} \in \mathbf{X} : \text{Prob}((\mathbf{x}, \omega) \in \mathbf{K}) \geq 1 - \epsilon\} \end{aligned} \quad (3)$$

in the form:

$$\mathbf{X}_\epsilon^d := \{\mathbf{x} \in \mathbf{X} : h_d(\mathbf{x}) \geq 0\}, \quad d \in \mathbb{N}, \quad (4)$$

where  $h_d$  is a polynomial of degree at most  $d$ .

Such approximations are particularly useful for optimization and control problems with chance-constraints; for instance problems of the form:

$$\min \{f(\mathbf{x}) : \mathbf{x} \in C; \text{Prob}((\mathbf{x}, \omega) \in \mathbf{K}) \geq 1 - \epsilon\}. \quad (5)$$

Indeed one then replaces problem (5) with

$$\min \{f(\mathbf{x}) : \mathbf{x} \in C; h_d(\mathbf{x}) \geq 0\}, \quad (6)$$

\*This work was supported by the European Research Council (ERC) via an ERC-Advanced Grant for the # 666981 project TAMING

<sup>1</sup>Jean B Lasserre is senior researcher at LAAS-CNRS and the Institute of Mathematics, University of Toulouse, France. [lasserre@laas.fr](mailto:lasserre@laas.fr)

where the uncertain parameter  $\omega$  has disappeared. So if  $C$  is a basic semi-algebraic set then (6) is a standard polynomial optimization problem. Of course the resulting decision problem (6) may still be hard to solve because the sets  $\mathbf{X}_\epsilon^*$  and  $\mathbf{X}_\epsilon^d$  are not convex in general. But this may be the price to pay for avoiding a too conservative formulation of the problem.

However, in the formulation (6) one has got rid of the disturbance parameter  $\omega$ , and so one may apply the arsenal of Non Linear Programming algorithms to get a local minimizer of (6). If  $n$  is not too large or if some sparsity is present in problem (6) one may even run a hierarchy of semidefinite relaxations to approximate its global optimal value. For the latter approach the interested reader is referred to [7] and for a discussion on approaches to various control problems with chance constraints we refer to the recent paper of Jasour et al. [4] and the references therein.

In Jasour et al. [4] the authors have considered some control problems with chance constraints. They have provided an elegant formulation and a numerical scheme for solving the related problem of computing

$$\mathbf{x}^* = \arg \max \{ \mu(\mathbf{K}_\mathbf{x}) : \mathbf{x} \in \mathbf{X} \}.$$

This problem is posed as an infinite-dimensional LP problem in an appropriate space of measures, that is, a Generalized Moment Problem (GMP) as described in Lasserre [7]. Then to obtain  $\mathbf{x}^*$  they solve a hierarchy of semidefinite relaxations, which is the *moment-SOS* approach for solving the GMP. This GMP formulation has the particular and typical feature of including a constraint of *domination*  $\phi \leq \psi$  between two measures  $\phi$  and  $\psi$ . Such domination constraints are particularly powerful and have been already used in a variety of different contexts. See for instance Henrion et al. [2] for approximating the Lebesgue volume of a compact semi-algebraic set, Lasserre [9] for computing Gaussian measures of semi-algebraic set, Lasserre [8] for “approximating” the Lebesgue decomposition of a measure with respect to another one. It has been used by Henrion and Korda [3] for approximating regions of attraction, by Korda et al. [6] for approximating maximum controlled invariant sets, and more recently in Jasour and Lagoa [5] for a unifying treatment of some control problems.

## Contribution

The approach that we propose for determining the set  $\mathbf{X}_\epsilon^d$  defined in (4) is very similar in spirit to that in [2] and [4] and can be viewed as an additional illustration of the versatility of the GMP and the moment-SOS approach in control related problems. Indeed we also define an infinite-dimensional

LP problem  $\mathbf{P}$  in an appropriate space of measures and a sequence of semidefinite relaxations  $(\mathbf{P}_d)_{d \in \mathbb{N}}$  of  $\mathbf{P}$ , whose associated monotone sequence of optimal values  $(\rho_d)_{d \in \mathbb{N}}$  converges to the optimal value  $\rho$  of  $\mathbf{P}$ . An optimal solution of the dual of  $(\mathbf{P}_d)$  allows to obtain a polynomial  $h_d$  of degree  $2d$  whose super-level set  $\{\mathbf{x} : h_d(\mathbf{x}) \geq 0\}$  is precisely the desired approximation  $\mathbf{X}_\epsilon^d$  of  $\mathbf{X}_\epsilon^*$  in (4); in fact the sets  $(\mathbf{X}_\epsilon^d)_{d \in \mathbb{N}}$  provide a sequence of *outer* approximations of  $\mathbf{X}_\epsilon^*$ . We also provide the strong asymptotic guarantee that

$$\lim_{d \rightarrow \infty} \lambda(\mathbf{X}_\epsilon^d \setminus \mathbf{X}_\epsilon^*) = 0,$$

where  $\lambda$  is the Lebesgue measure on  $\mathbf{X}$ , which to the best of our knowledge is the first result of this kind at this level of generality. (The same methodology applied to chance-constraints of the form  $\text{Prob}((\mathbf{x}, \omega) \in \mathbf{K}) < \epsilon$  would provide a sequence of *inner* approximations of the set  $\{\mathbf{x} \in \mathbf{X} : \text{Prob}((\mathbf{x}, \omega) \in \mathbf{K}) < \epsilon\}$ .)

Another contribution is to include a technique to accelerate the convergence  $\rho_d \rightarrow \rho$  which otherwise can be too slow. This technique is different from the one used in [2] for the related problem of computing the volume of a semi-algebraic set, and has the nice feature of preserving the *monotonicity* of the convergence of  $\rho_d \rightarrow \rho$ . It can be applied whenever  $d\mu$  is the Lebesgue measure  $d\omega$  on  $\Omega$  or  $d\mu = q(\omega)d\omega$  or  $d\mu = \exp(-q(\omega))d\omega$  for some homogeneous nonnegative polynomial  $q$ .

At last but not least, in principle we can also treat the case where the support  $\Omega$  of  $\mu$  and the set  $\mathbf{K}$  are not compact, which includes the important case where  $\mu$  is the normal distribution. We briefly explain what are the (technical) arguments which allow to extend the method to the non compact case.

Of course this methodology is computationally expensive and so far limited to relatively small size problems (but after all the problem is very hard). An interesting issue not discussed here is to investigate whether sparsity patterns can be exploited to handle problems with larger size.

## II. NOTATION, DEFINITIONS AND PRELIMINARY RESULTS

### A. Notation and definitions

Let  $\mathbb{R}[\mathbf{x}]$  be the ring of polynomials in the variables  $\mathbf{x} = (x_1, \dots, x_n)$  and let  $\mathbb{R}[\mathbf{x}]_d$  be the vector space of polynomials of degree at most  $d$  (whose dimension is  $s(d) := \binom{n+d}{n}$ ). For every  $d \in \mathbb{N}$ , let  $\mathbb{N}_d^n := \{\alpha \in \mathbb{N}^n : |\alpha| (= \sum_i \alpha_i) \leq d\}$ , and let  $\mathbf{v}_d(\mathbf{x}) = (\mathbf{x}^\alpha)$ ,  $\alpha \in \mathbb{N}^n$ , be the vector of monomials of the canonical basis  $(\mathbf{x}^\alpha)$  of  $\mathbb{R}[\mathbf{x}]_d$ . A polynomial  $f \in \mathbb{R}[\mathbf{x}]_d$  is written

$$\mathbf{x} \mapsto f(\mathbf{x}) = \sum_{\alpha \in \mathbb{N}^n} f_\alpha \mathbf{x}^\alpha,$$

for some vector of coefficients  $\mathbf{f} = (f_\alpha) \in \mathbb{R}^{s(d)}$ .

Given a closed set  $\mathcal{X} \subset \mathbb{R}^n$ , denote by  $M(\mathcal{X})$  the space of finite Borel measures on  $\mathcal{X}$  and by  $\mathcal{P}(\mathcal{X})$  the convex cone of polynomials that are nonnegative on  $\mathcal{X}$ .

**Moment matrix.** Given a sequence  $\mathbf{y} = (y_\alpha)_{\alpha \in \mathbb{N}^n}$ , let  $L_{\mathbf{y}} : \mathbb{R}[\mathbf{x}] \rightarrow \mathbb{R}$  be the linear (Riesz) functional

$$f (= \sum_{\alpha} f_\alpha \mathbf{x}^\alpha) \mapsto L_{\mathbf{y}}(f) := \sum_{\alpha} f_\alpha y_\alpha.$$

Given  $\mathbf{y}$  and  $d \in \mathbb{N}$ , the *moment* matrix associated with  $\mathbf{y}$ , is the real symmetric  $s(d) \times s(d)$  matrix  $\mathbf{M}_d(\mathbf{y})$  with rows and columns indexed in  $\mathbb{N}_d^n$  and with entries

$$M_d(\mathbf{y})(\alpha, \beta) := L_{\mathbf{y}}(\mathbf{x}^{\alpha+\beta}) = y_{\alpha+\beta}, \quad \alpha, \beta \in \mathbb{N}_d^n.$$

**Localizing matrix.** Given a sequence  $\mathbf{y} = (y_\alpha)_{\alpha \in \mathbb{N}^n}$ , and a polynomial  $g \in \mathbb{R}[\mathbf{x}]$ , the *localizing* moment matrix associated with  $\mathbf{y}$  and  $g$ , is the real symmetric  $s(d) \times s(d)$  matrix  $\mathbf{M}_d(g\mathbf{y})$  with rows and columns indexed in  $\mathbb{N}_d^n$  and with entries

$$\begin{aligned} M_d(g\mathbf{y})(\alpha, \beta) &:= L_{\mathbf{y}}(g(\mathbf{x}) \mathbf{x}^{\alpha+\beta}) \\ &= \sum_{\gamma} g_\gamma y_{\alpha+\beta+\gamma}, \quad \alpha, \beta \in \mathbb{N}_d^n. \end{aligned}$$

### B. The volume of a compact semi-algebraic set

In this section we recall how to approximate as closely as desired the Lebesgue volume of a compact semi-algebraic set  $\mathbf{K} \subset \mathbb{R}^n$ . It will be the building block of the methodology to approximate the set  $\mathbf{X}_\epsilon^*$  in (3).

Let  $\mathbf{X} \subset \mathbb{R}^n$  be a box and let  $\lambda \in M(\mathbf{X})$  be the Lebesgue measure on  $\mathbf{X}$ . Let  $\mathbf{K} := \{\mathbf{x} : g_j(\mathbf{x}) \geq 0, j = 1, \dots, m\}$ , assumed to be compact. For convenience and with no loss of generality we may and will assume that  $g_1(\mathbf{x}) = M - \|\mathbf{x}\|^2$  for some  $M > 0$ .

*Theorem 2.1* ([2], [4]): Let  $\mathbf{K} \subset \mathbf{X}$  and with nonempty interior. Then

$$\lambda(\mathbf{K}) = \sup_{\phi \in M(\mathbf{K})} \{\phi(\mathbf{K}) : \phi \leq \lambda\}, \quad (7)$$

and  $d\phi^* = 1_{\mathbf{K}}(\mathbf{x})d\lambda$  is the unique optimal solution.

Problem (7) is an infinite-dimensional LP with dual

$$\rho = \inf_{p \in \mathbb{R}[\mathbf{x}]} \left\{ \int_{\mathbf{X}} p d\lambda : p \geq 0 \text{ on } \mathbf{X}; p \geq 1 \text{ on } \mathbf{K} \right\}. \quad (8)$$

and there is no duality gap, i.e.,  $\rho = \lambda(\mathbf{K})$ .

Let  $d_j = \lceil \deg(g_j)/2 \rceil$ ,  $j = 1, \dots, m$ . To approximate  $\lambda(\mathbf{K})$  one solves the hierarchy of semidefinite programs, indexed by  $d \in \mathbb{N}$ :

$$\begin{aligned} \rho_d = \sup_{\mathbf{y}, \mathbf{z}} \{ & L_{\mathbf{y}}(1) : y_\alpha + z_\alpha = \lambda_\alpha, \quad \forall \alpha \in \mathbb{N}_{2d}^n \\ & \mathbf{M}_d(\mathbf{y}), \mathbf{M}_d(\mathbf{z}) \succeq 0 \\ & \mathbf{M}_{d-d_j}(g_j \mathbf{y}) \succeq 0, \quad j = 1, \dots, m \}. \end{aligned} \quad (9)$$

The sequence  $(\rho_d)_{d \in \mathbb{N}}$  is monotone non increasing and  $\rho_d \rightarrow \lambda(\mathbf{K})$  as  $d \rightarrow \infty$ . However the convergence is rather slow and in [2] the authors have proposed to replace the criterion  $L_{\mathbf{y}}(1)$  by  $L_{\mathbf{y}}(h)$  where  $h \in \mathbb{R}[\mathbf{x}]$  is a polynomial that is nonnegative on  $\mathbf{K}$  and vanishes on the boundary of  $\mathbf{K}$ . If one denotes by  $\mathbf{y}^d$  an optimal solution of (9) then  $\rho_d \rightarrow \int_{\mathbf{K}} h d\lambda$  and  $y_0^d \rightarrow \lambda(\mathbf{K})$  as  $d \rightarrow \infty$ . The convergence  $y_0^d \rightarrow \lambda(\mathbf{K})$  is much faster but is not monotone anymore, which can be annoying because we do not obtain a decreasing sequence of upper bounds on  $\lambda(\mathbf{K})$  as was the case with (9). For more details the interested reader is referred to [2].

### C. Stokes can help

This is why we propose another technique to accelerate the convergence  $\rho_d \rightarrow \lambda(\mathbf{K})$  in (9) while maintaining its monotonicity. So let  $h \in \mathbb{R}[\mathbf{x}]$  be such that  $h(\mathbf{x}) = 0$  for all  $\mathbf{x} \in \partial\mathbf{K}$  (but  $h$  is not required to be nonnegative on  $\mathbf{K}$ ). Then by Stokes' theorem, for each  $\alpha \in \mathbb{N}^n$ :

$$\int_{\mathbf{K}} \underbrace{((n + |\alpha|)h + \langle \mathbf{x}, \nabla h \rangle)}_{\mathbf{x} \mapsto \theta_\alpha(\mathbf{x})} \mathbf{x}^\alpha d\lambda(\mathbf{x}) = 0,$$

and so the optimal solution  $\phi^*$  of Theorem 2.1 must satisfy

$$\int_{\mathbf{K}} \theta_\alpha(\mathbf{x}) d\phi^*(\mathbf{x}) = 0, \quad \forall \alpha \in \mathbb{N}^n.$$

Therefore in (9) we may impose the additional moment constraints

$$L_{\mathbf{y}}(\theta_\alpha) = 0, \quad \forall \alpha \in \mathbb{N}_{2d-\deg(h)}^n. \quad (10)$$

To appreciate the impact of such additional constraints on the convergence  $\rho_d \rightarrow \lambda(\mathbf{K})$ , consider first the toy problem where  $\mathbf{X} = [-1, 1]$  and  $\mathbf{K} = [-1/2, 1/2]$  with  $\lambda(\mathbf{K}) = 1$ . Let  $\rho_d$  (resp.  $\rho'_d$ ) be the optimal value of (9) without (resp. with) the Stokes constraints. Results are displayed in Table I. Similarly, with  $n = 2$  and  $\mathbf{X} = [-a, a]^2$ , let  $\mathbf{K} := \{\mathbf{x} :$

TABLE I  
THE EFFECT OF STOKES CONSTRAINTS;  $n = 1$

|           | d=4   | d=6   | d=8   | d=10  | d=12  |
|-----------|-------|-------|-------|-------|-------|
| $\rho_d$  | 1.689 | 1.463 | 1.423 | 1.382 | 1.305 |
| $\rho'_d$ | 1.156 | 1.069 | 1.025 | 1.010 | 1.003 |

$\|\mathbf{x}\|^2 \leq 1\}$  so that  $\lambda(\mathbf{K}) = \pi$ . For different values of  $a$  and  $d = 3, 4$ , results are displayed in Table II.

TABLE II  
THE EFFECT OF STOKES CONSTRAINTS.  $n = 2$

|       | $\rho_3$ | $\rho'_3$ | $\rho_4$ | $\rho'_4$ |
|-------|----------|-----------|----------|-----------|
| a=1.4 | 5.71     | 3.55      | 5.38     | 3.27      |
| a=1.3 | 5.38     | 3.41      | 5.04     | 3.21      |
| a=1.2 | 5.02     | 3.31      | 4.70     | 3.17      |
| a=1.1 | 4.56     | 3.36      | 4.32     | 3.15      |
| a=1.2 | 3.91     | 3.20      | 3.85     | 3.144     |

*Remark 2.2:* Theorem 2.1 is valid for any measure  $\mu \in M(\mathbf{X})$  and not only the Lebesgue measure  $\lambda$ . On the other hand, the additional Stokes constraints (10) are valid provided that  $d\mu = f d\lambda$  or  $d\mu = \exp(f) d\lambda$  for some homogeneous polynomial  $f \in \mathbb{R}[\mathbf{x}]$ . Then with  $d_f = \deg(f)$ ,

$$\theta_\alpha(\mathbf{x}) = \begin{cases} (n + d_f)\mathbf{x}^\alpha h + \langle \mathbf{x}, \nabla(\mathbf{x}^\alpha h) \rangle & \text{for } f d\lambda \\ \mathbf{x}^\alpha h(n + d_f f) + \langle \mathbf{x}, \nabla(\mathbf{x}^\alpha h) \rangle & \text{for } \exp(f) d\lambda \end{cases}$$

### III. MAIN RESULT

After the preliminary results of Section §II, we are now in position to state our main result. Let  $\mu$  be the distribution of the noise parameter  $\omega \in \Omega$ , and let  $\lambda$  be the Lebesgue

measure on  $\mathbf{X}$ . The notation  $\lambda \otimes \mu$  denotes the product measure on  $\mathbf{X} \times \Omega$ , that is,

$$\lambda \otimes \mu(A \times B) = \lambda(A)\mu(B), \quad \forall A \in \mathcal{B}(\mathbf{X}), B \in \mathcal{B}(\Omega).$$

With  $\mathbf{K} \subset \mathbf{X} \times \Omega$  as in (1), and for every  $\mathbf{x} \in \mathbf{X}$ , let  $\mathbf{K}_{\mathbf{x}}$  be as in (2) (possibly empty). Consider the infinite dimensional LP:

$$\rho = \sup_{\phi \in M(\mathbf{K})} \{ \phi(\mathbf{K}) : \phi \leq \lambda \otimes \mu \}. \quad (11)$$

*Theorem 3.1:* The unique optimal solution of (11) is

$$d\phi^*((\mathbf{x}, \omega)) = 1_{\mathbf{K}}((\mathbf{x}, \omega)) d\lambda \otimes \mu((\mathbf{x}, \omega)),$$

and the optimal value  $\rho$  of (11) satisfies

$$\begin{aligned} \rho &= \int_{\mathbf{X} \times \Omega} 1_{\mathbf{K}}((\mathbf{x}, \omega)) \lambda \otimes \mu(d(\mathbf{x}, \omega)) \\ &= \int_{\mathbf{X}} \mu(\mathbf{K}_{\mathbf{x}}) \lambda(d\mathbf{x}). \end{aligned} \quad (12)$$

*Proof:* That  $\rho = \int_{\mathbf{X} \times \Omega} 1_{\mathbf{K}}((\mathbf{x}, \omega)) \lambda \otimes \mu(d(\mathbf{x}, \omega))$  follows from Theorem 2.1 (with  $\lambda \otimes \mu$  instead of  $\lambda$  in Theorem 2.1). By Tonelli's Theorem

$$\begin{aligned} &\int_{\mathbf{X} \times \Omega} 1_{\mathbf{K}}((\mathbf{x}, \omega)) \lambda \otimes \mu(d(\mathbf{x}, \omega)) = \\ &\int_{\mathbf{X}} \underbrace{\left( \int_{\Omega} 1_{\mathbf{K}}((\mathbf{x}, \omega)) \mu(d\omega) \right)}_{\mu(\mathbf{K}_{\mathbf{x}})} \lambda(d\mathbf{x}) = \int_{\mathbf{X}} \mu(\mathbf{K}_{\mathbf{x}}) \lambda(d\mathbf{x}) \end{aligned}$$

### Semidefinite relaxations

Let  $d_j = \lceil \deg(g_j)/2 \rceil$  for all  $j$ . As we did for (9) in §II, let  $\mathbf{y} = (y_{\alpha,\beta})$  and  $\mathbf{z} = (z_{\alpha,\beta})$ ,  $(\alpha, \beta) \in \mathbb{N}^{n+p}$ , and relax (11) to the following hierarchy of semidefinite programs, indexed by  $d \in \mathbb{N}$ :

$$\begin{aligned} \rho_d &= \sup_{\mathbf{y}, \mathbf{z}} \{ y_0 : \\ &\text{s.t. } y_{\alpha,\beta} + z_{\alpha,\beta} = \lambda_\alpha \cdot \mu_\beta, \quad (\alpha, \beta) \in \mathbb{N}_{2d}^{n+p} \\ &\quad \mathbf{M}_d(\mathbf{y}), \mathbf{M}_d(\mathbf{z}) \succeq 0 \\ &\quad \mathbf{M}_{d-d_j}(g_j \mathbf{y}) \succeq 0, \quad j = 1, \dots, m \}, \end{aligned} \quad (13)$$

and of course  $\rho_d \geq \rho_{d+1} \geq \rho$  for all  $d$ . The dual of (13) is the semidefinite program:

$$\begin{aligned} \rho_d^* &= \inf_{p \in \mathbb{R}[\mathbf{x}, \omega]_{2d}} \left\{ \int_{\mathbf{X} \times \Omega} p(\mathbf{x}, \omega) \lambda \otimes \mu(d(\mathbf{x}, \omega)) \right. \\ &\text{s.t. } p(\mathbf{x}, \omega) \geq 1, \quad \forall (\mathbf{x}, \omega) \in \mathbf{K} \\ &\quad p \text{ is SOS} \left. \right\}. \end{aligned} \quad (14)$$

Again as  $\mathbf{K}$  is compact, for technical reasons (but with no loss of generality) we may and will assume that in the definition (1) of  $\mathbf{K}$ ,  $g_1(\mathbf{x}) = M - \|\mathbf{x}\|^2$  for some  $M > 0$ .

*Theorem 3.2:* Let  $\mathbf{K}$  and  $(\mathbf{X} \times \Omega) \setminus \mathbf{K}$  be with nonempty interior. There is no duality gap between (13) and its dual (14), i.e.,  $\rho_d = \rho_d^*$  for all  $d$ . In addition (14) has an optimal solution  $p_d^* \in \mathbb{R}[\mathbf{x}, \omega]_{2d}$  such that

$$\rho_d = \rho_d^* = \int_{\mathbf{X} \times \Omega} p_d^*(\mathbf{x}, \omega) \lambda \otimes \mu(d(\mathbf{x}, \omega)).$$

Define  $h_d^* \in \mathbb{R}[\mathbf{x}]_{2d}$  to be:

$$\mathbf{x} \mapsto h_d^*(\mathbf{x}) := \int_{\Omega} p_d^*(\mathbf{x}, \omega) \mu(d\omega), \quad \mathbf{x} \in \mathbb{R}^n.$$

Then  $h_d^*(\mathbf{x}) \geq \mu(\mathbf{K}_{\mathbf{x}})$  for all  $\mathbf{x} \in \mathbf{X}$  and

$$\rho_d = \int_{\mathbf{X}} h_d^*(\mathbf{x}) \lambda(d\mathbf{x}) \rightarrow \rho = \int_{\mathbf{X}} \mu(\mathbf{K}_{\mathbf{x}}) \lambda(d\mathbf{x})$$

as  $d \rightarrow \infty$ .

*Proof:* That  $\rho_d = \rho_d^*$  is because Slater's condition holds for (13). Indeed let  $\mathbf{y}^*$  be the moments of  $\phi^*$  in Theorem 3.1 and  $\mathbf{z}^*$  be the moments of  $\lambda \otimes \mu - \phi^*$  (on  $(\mathbf{X} \times \Omega) \setminus \mathbf{K}$ ). Then as  $\mathbf{K}$  has nonempty interior,  $\mathbf{M}_d(\mathbf{y}^*) \succ 0$  and  $\mathbf{M}_d(g_j \mathbf{y}^*) \succ 0$  for all  $d$ . Similarly as  $(\mathbf{X} \times \Omega) \setminus \mathbf{K}$  has nonempty interior,  $\mathbf{M}_d(\mathbf{z}^*) \succ 0$ . Moreover since the optimal value  $\rho_d$  is finite for all  $d$  this implies that (14) has an optimal solution  $p_d^* \in \mathbb{R}[\mathbf{x}, \omega]_{2d}$ . Therefore:

$$\begin{aligned} \rho_d &= \int_{\mathbf{X} \times \Omega} p_d^*(\mathbf{x}, \omega) \lambda \otimes \mu(d(\mathbf{x}, \omega)) \\ &= \int_{\mathbf{X}} \underbrace{\left( \int_{\Omega} p_d^*(\mathbf{x}, \omega) \mu(d\omega) \right)}_{h_d^*(\mathbf{x}) \geq \mu(\mathbf{K}_{\mathbf{x}})} \lambda(d\mathbf{x}) = \int_{\mathbf{X}} h_d^*(\mathbf{x}) \lambda(d\mathbf{x}) \end{aligned}$$

where  $h_d^*(\mathbf{x}) \geq \mu(\mathbf{K}_{\mathbf{x}})$  follows from  $p_d^* \geq 1$  on  $\mathbf{K}$ . Finally the convergence  $\lim_{d \rightarrow \infty} \rho_d = \rho$  follows from Theorem 2.1.  $\blacksquare$

Then as  $h^*(\mathbf{x}) \geq \mu(\mathbf{K}_{\mathbf{x}})$  on  $\mathbf{X}$ , the sets  $\mathbf{X}_{\epsilon}^d = \{\mathbf{x} \in \mathbf{X} : h_d^*(\mathbf{x}) > 1 - \epsilon\}$ ,  $d \in \mathbb{N}$ , form a sequence of outer approximations of the set  $\mathbf{X}_{\epsilon}^*$ . In fact more can be said.

*Corollary 3.3:* Let  $h_d^* \in \mathbb{R}[\mathbf{x}, \omega]_{2d}$  be as in Theorem 3.2. Then the function  $\mathbf{x} \mapsto \psi_d^*(\mathbf{x}) := h_d^*(\mathbf{x}) - \mu(\mathbf{K}_{\mathbf{x}})$  is nonnegative on  $\mathbf{X}$  and converges to 0 in  $L_1(\mathbf{X}, \lambda)$ . In particular  $\psi_d^* \rightarrow 0$  in  $\lambda$ -measure, and  $\lambda$ -almost uniformly for some subsequence  $(\psi_{d_k}^*)_{k \in \mathbb{N}}$ .

*Proof:* As  $\rho_d \rightarrow \rho$  as  $d \rightarrow \infty$ ,

$$\lim_{d \rightarrow \infty} \int_{\mathbf{X}} \underbrace{(h_d^*(\mathbf{x}) - \mu(\mathbf{K}_{\mathbf{x}}))}_{\geq 0} \lambda(d\mathbf{x}) = 0,$$

whence the convergence to 0 in  $L_1(\mathbf{X}, \lambda)$ . Then convergence  $\psi_d^* \rightarrow 0$  in  $\lambda$ -measure, and  $\lambda$ -almost sure convergence for a subsequence follow from standard results from Real Analysis. See e.g. Ash [1, Theorem 2.5.1].  $\blacksquare$

As we next see, the convergence  $h_d^*(\mathbf{x}) \rightarrow \mu(\mathbf{K}_{\mathbf{x}})$  in  $\lambda$ -measure established in Corollary 3.3 will be useful to obtain strong asymptotic guarantees.

#### A. Strong asymptotic guarantees

We here investigate asymptotic properties of the sequence of sets  $(\mathbf{X}_{\epsilon}^d)_{d \in \mathbb{N}}$ , as  $d \rightarrow \infty$ .

*Corollary 3.4:* With  $\mathbf{X}_{\epsilon}^*$  as in (3), let  $\mathbf{X}_{\epsilon}^d := \{\mathbf{x} \in \mathbf{X} : h_d^*(\mathbf{x}) \geq 1 - \epsilon\}$  where  $h_d^*$  is as in Theorem 3.2,  $d \in \mathbb{N}$ . Then:

$$\lim_{d \rightarrow \infty} \lambda(\mathbf{X}_{\epsilon}^d) = \lambda(\mathbf{X}_{\epsilon}^*). \quad (15)$$

*Proof:* Observe that

$$\mathbf{X} \setminus \mathbf{X}_{\epsilon}^* = \bigcup_{\ell=1}^{\infty} \{\mathbf{x} \in \mathbf{X} : \mu(\mathbf{K}_{\mathbf{x}}) < 1 - \epsilon - 1/\ell\},$$

and therefore

$$\lambda(\mathbf{X} \setminus \mathbf{X}_{\epsilon}^*) = \lim_{\ell \rightarrow \infty} \lambda(\underbrace{\{\mathbf{x} \in \mathbf{X} : \mu(\mathbf{K}_{\mathbf{x}}) < 1 - \epsilon - 1/\ell\}}_{R_{\ell}}).$$

Next, for each  $\ell = 1, \dots$ , write

$$\lambda(R_{\ell}) = \lambda(R_{\ell} \cap \{\mathbf{x} \in \mathbf{X} : h_d^*(\mathbf{x}) < 1 - \epsilon\}) + \lambda(R_{\ell} \cap \mathbf{X}_{\epsilon}^d).$$

By the convergence  $h_d^* \rightarrow \mu(\mathbf{K}_{\mathbf{x}})$  in  $\lambda$ -measure as  $d \rightarrow \infty$ ,  $\lim_{d \rightarrow \infty} \lambda(R_{\ell} \cap \mathbf{X}_{\epsilon}^d) = 0$  and so

$$\begin{aligned} \lambda(R_{\ell}) &= \lim_{d \rightarrow \infty} \lambda(R_{\ell} \cap \{\mathbf{x} \in \mathbf{X} : h_d^*(\mathbf{x}) < 1 - \epsilon\}) \\ &\leq \lim_{d \rightarrow \infty} \lambda(\{\mathbf{x} \in \mathbf{X} : h_d^*(\mathbf{x}) < 1 - \epsilon\}) \\ &\leq \lambda(\mathbf{X} \setminus \mathbf{X}_{\epsilon}^*). \end{aligned}$$

This implies

$$\lim_{d \rightarrow \infty} \lambda(\{\mathbf{x} \in \mathbf{X} : h_d^*(\mathbf{x}) < 1 - \epsilon\}) = \lambda(\mathbf{X} \setminus \mathbf{X}_{\epsilon}^*),$$

which in turn yields the desired result (15).  $\blacksquare$

#### Inner Approximations

In the previous section we have provided a converging sequence  $(\mathbf{X}_{\epsilon}^d)_{d \in \mathbb{N}}$  of outer approximations of  $\mathbf{X}_{\epsilon}^*$ . Clearly, the same methodology now applied to a chance constraint of the form

$$\text{Prob}((\mathbf{x}, \omega) \in \mathbf{K}) < \epsilon$$

would provide a converging sequence of *inner* approximations of the set  $\tilde{\mathbf{X}}_{\epsilon}^* := \{\mathbf{x} \in \mathbf{X} : \text{Prob}((\mathbf{x}, \omega) \in \mathbf{K}) < \epsilon\}$ .

#### B. Accelerating convergence

As we already have seen in Section II-C for the semidefinite program (9), as  $d \rightarrow \infty$  the convergence  $\rho_d \rightarrow \rho$  of the optimal value of (13) can also be slow due to the Gibb's effect that appears in the dual (14) when approximating the indicator function  $\mathbf{x} \mapsto 1_{\mathbf{K}}(\mathbf{x})$  by a polynomial.

So assume that  $\mu$  is the Lebesgue measure on  $\Omega$  scaled to be a probability measure (but the same idea works if  $d\mu = h(\omega)d\omega$ , or if  $d\mu = \exp(h(\omega))d\omega$  for some homogeneous polynomial  $h$ ). Then again we propose to include additional constraints on the moments  $\mathbf{y}$  in (13) coming from additional properties of the optimal solution  $\phi^*$  of (11). Again these additional properties are coming from Stokes' formula but now for integrals on  $\mathbf{K}_{\mathbf{x}}$ , then integrated over  $\mathbf{X}$ .

Let  $f \in \mathbb{R}[\mathbf{x}, \omega]$  be the polynomial  $(\mathbf{x}, \omega) \mapsto f(\mathbf{x}, \omega) = \prod_j g_j(\mathbf{x}, \omega)$ . For each fixed  $\mathbf{x} \in \mathbf{K}_{\mathbf{x}}$ , the polynomial  $\omega \mapsto f(\mathbf{x}, \omega)$  vanishes on the boundary  $\partial \mathbf{K}_{\mathbf{x}}$  of  $\mathbf{K}_{\mathbf{x}}$ . Therefore for each  $\beta \in \mathbb{N}^p$ , Stokes' Theorem states that

$$\int_{\mathbf{K}_{\mathbf{x}}} \omega^{\beta} [(p + |\beta|) f(\mathbf{x}, \omega) + \langle \omega, \nabla_{\omega} f(\mathbf{x}, \omega) \rangle] d\mu(\omega) = 0.$$

So define the polynomial  $\theta \in \mathbb{R}[\mathbf{x}, \omega]$  of degree  $d_{\theta}$ , by:

$$(\mathbf{x}, \omega) \mapsto \theta(\mathbf{x}, \omega) := (p + |\beta|) f(\mathbf{x}, \omega) + \langle \omega, \nabla_{\omega} f(\mathbf{x}, \omega) \rangle.$$

Then for each  $(\alpha, \beta) \in \mathbb{N}^{n+p}$

$$\int_{\mathbf{X}} \int_{\mathbf{K}_{\mathbf{x}}} \mathbf{x}^{\alpha} \omega^{\beta} \theta(\mathbf{x}, \omega) d\mu(\omega) d\lambda(\mathbf{x}) = 0.$$

Equivalently, in view of what is  $\phi^*$  in Theorem 3.1,

$$\int_{\mathbf{K}} \mathbf{x}^\alpha \omega^\beta \theta(\mathbf{x}, \omega) d\phi^*((\mathbf{x}, \omega)) = 0.$$

Therefore in (13) we may include the additional moments constraints  $L_{\mathbf{y}}(\mathbf{x}^\alpha \omega^\beta \theta(\mathbf{x}, \omega)) = 0$ , for all  $(\alpha, \beta) \in \mathbb{N}^{n+p}$  such that  $|\alpha + \beta| \leq 2d - d_\theta$ .

### C. The non-compact case

In some applications the noise  $\omega$  is assumed to follow a normal distribution  $\mu$  on  $\Omega = \mathbb{R}^p$ . Therefore  $\Omega$  is not compact anymore and the machinery used in [2] cannot be applied directly. However the normal distribution satisfies the important Carleman's property. That is, let  $L_{\mathbf{y}}$  be the Riesz functional associated with  $\mu$ , i.e.,  $L_{\mathbf{y}}(f) = \int f d\mu$  for all  $f \in \mathbb{R}[\omega]$ . Then

$$\sum_{k=1}^{\infty} L_{\mathbf{y}}(\omega_i^{2k})^{-1/2k} = +\infty, \quad i = 1, \dots, p. \quad (16)$$

In particular  $\mu$  is *moment determinate*, that is,  $\mu$  is completely determined by its moments. These two properties have been used extensively in e.g. Lasserre [8] and also in [9], precisely to show that with  $\mathbf{K} \subset \Omega$  not necessarily compact, one may still approximate its Gaussian measure  $\mu(\mathbf{K})$  as closely as desired. Again one solves (7) via the same hierarchy of semidefinite relaxations (9) (but now with  $\mu$  instead of  $\lambda$ ). For more details the interested reader is referred to Lasserre [8], [9].

In view of the above (technical) remarks, one may then extend the machinery described in §III for the case where  $\Omega = \mathbb{R}^p$ ,  $\mu$  is the Gaussian measure, and  $\mathbf{K}_{\mathbf{x}}$  ( $\mathbf{x} \in \mathbf{X}$ ) is not necessarily compact. A version of Stokes' Theorem for non compact sets is even described in [9] to accelerate the convergence of the semidefinite relaxations (9) (with  $\mu$  instead of  $\lambda$ ). It can be used to accelerate the convergence of the semidefinite relaxations (13), exactly as we do in §III-B for the compact case.

## IV. CONCLUSION

We have presented a systematic numerical scheme to provide an sequence of outer approximations ( $\mathbf{X}_\epsilon^d$ ) of the feasible set  $\mathbf{X}_\epsilon^*$  associated with chance constraints of the form  $\text{Prob}((\mathbf{x}, \omega) \in \mathbf{K}) > 1 - \epsilon$ . Each outer approximation  $\mathbf{X}_\epsilon^d$  is the 0-super level set of some polynomial whose coefficients are computed via solving a certain semidefinite program. As  $d$  increases  $\lambda(\mathbf{X}_\epsilon^d \setminus \mathbf{X}_\epsilon^*) \rightarrow 0$ , a nice and highly desirable asymptotic property. Of course this methodology is computationally expensive and limited to problems of small size. But we hope it can pave the way to define more efficient heuristics (perhaps at the price of loosing the nice asymptotic properties). Also checking whether this methodology can accommodate potential sparsity patterns present in larger size problems, is a topic of further investigation.

## REFERENCES

- [1] R. Ash. Real Analysis and Probability, Academic Press Inc., Boston, USA, 1972
- [2] D. Henrion, J. B. Lasserre, C. Savorgnan, Approximate volume and integration for basic semialgebraic sets, SIAM Review 51, No. 4, pp. 722-743, 2009.
- [3] D. Henrion, M. Korda, Convex computation of the region of attraction of polynomial control systems. IEEE Trans. Aut. Control 59, No 2, pp. 297-312, 2014.
- [4] A. M. Jasour, N.S. Aybat, C. M. Lagoa, Semidefinite programming for chance constrained optimization over semi-algebraic sets, SIAM J. Optim. 25, No 3, pp. 1411-1440, 2015
- [5] A. Jasour, C. Lagoa, Convex constrained semialgebraic volume optimization: Application in systems and control, 2017, arXiv:1701.08910
- [6] M. Korda, D. Henrion, C. N. Jones, Convex computation of the maximum controlled invariant set for polynomial control systems, SIAM J. Control Optim. 52(5):2944-2969, 2014.
- [7] J. B. Lasserre, Moments, positive polynomials and their applications. Imperial College Press, London, 2010.
- [8] J.B. Lasserre, Lebesgue decomposition in action via semidefinite relaxations, Adv. Comput. Math. 42, pp. 1129-1148, 2016.
- [9] J.B. Lasserre, Computing gaussian and exponential measures of semi-algebraic sets, arXiv:1508.06132. Submitted.