Constructive Approximation (2021) 54:391–429
https://doi.org/10.1007/s00365-021-09535-4

CONSTRUCTIVE APPROXIMATION

Semi-algebraic Approximation Using Christoffel–Darboux Kernel

Swann Marx[1] · Edouard Pauwels[2] · Tillmann Weisser[3] · Didier Henrion[4,5] · Jean Bernard Lasserre[4,6]

Received: 20 January 2020 / Revised: 28 August 2020 / Accepted: 18 September 2020 /
Published online: 7 April 2021

Abstract
We provide a new method to approximate a (possibly discontinuous) function using Christoffel–Darboux kernels. Our knowledge about the unknown multivariate function is in terms of finitely many moments of the Young measure supported on the graph of the function. Such an input is available when approximating weak (or measure-valued) solution of optimal control problems, entropy solutions to nonlinear hyperbolic PDEs, or using numerical integration from finitely many evaluations of the function. While most of the existing methods construct a piecewise polynomial approximation, we construct a semi-algebraic approximation whose estimation and evaluation can be performed efficiently. An appealing feature of this method is that it deals with nonsmoothness implicitly so that a single scheme can be used to treat smooth or nonsmooth functions without any prior knowledge. On the theoretical side, we prove pointwise convergence almost everywhere as well as convergence in the Lebesgue one norm under broad assumptions. Using more restrictive assumptions, we obtain explicit convergence rates. We illustrate our approach on various examples from control and approximation. In particular, we observe empirically that our method does not suffer from the Gibbs phenomenon when approximating discontinuous functions.

Communicated by Remi Gribonval.

✉ Swann Marx
marx.swann@gmail.com

1 LS2N, École Centrale de Nantes and CNRS UMR 6004, 44000 Nantes, France

2 IRIT-UPS, Université de Toulouse, 118 route de Narbonne, 31400 Toulouse, France

3 Theoretical Division and Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

4 LAAS-CNRS, Université de Toulouse, 7 Avenue du Colonel Roche, 31400 Toulouse, France

5 Faculty of Electrical Engineering, Czech Technical University in Prague, Technická 4, 16206 Prague, Czechia

6 IMT-UPS, Université de Toulouse, 118 Route de Narbonne, 31400 Toulouse, France

Keywords Approximation theory · Convex optimization · Moments · Positive polynomials · Orthogonal polynomials

Mathematics Subject Classification 42C05 · 47B32 · 41A30

Contents

1 Introduction

In this paper, we address the following generic inverse problem. Let

$$\begin{aligned} f : X &\to Y \\ \mathbf{x} := (x_1,\ x_2, \ldots, x_{p-1}) &\mapsto y \end{aligned}$$

be a bounded measurable function from a given compact set $X \subset \mathbb{R}^{p-1}$ to a given compact set $Y \subset \mathbb{R}$, with $p \geq 2$. We assume that X is equal to the closure of its interior.

Given $d \in \mathbb{N}$, consider a vector of polynomials of total degree at most d

$$\mathbf{b}(\mathbf{x},\ y) :\ (\mathbf{x},\ y) \in \mathbb{R}^p \mapsto \begin{pmatrix} b_1(\mathbf{x},\ y) & b_2(\mathbf{x},\ y) & \cdots & b_{n_d}(\mathbf{x},\ y) \end{pmatrix} \in \mathbb{R}^{n_d},$$

where $n_d := \binom{p+d}{d}$ which is understood as the binomial coefficient. For example, $\mathbf{b}$ may be a vector whose entries are the polynomials of the canonical monomial basis or any orthonormal polynomial basis, e.g., Chebyshev or Legendre. Associated to $\mathbf{b}$ and f, let

$$\int_X \mathbf{b}(\mathbf{x}, f(\mathbf{x}))\mathbf{b}(\mathbf{x}, f(\mathbf{x}))^\top \mathrm{d}\mathbf{x}$$

be the moment matrix of degree $2d$, where the integral is understood entry-wise.

Problem 1 (Graph inference from moment matrix) *Given the moment matrix of degree $2d$, compute an approximation f_d of the function f, with convergence guarantees when degree d tends to infinity.*

1.1 Motivation

Inverse Problem 1 is encountered in several interesting situations. In the weak (or measure-valued) formulation of optimal control problems (OCP) [17,23,39], Markov decision processes [18], option pricing in finance [24], stochastic control and optimal stopping [37] and some nonlinear partial differential equations (PDEs) [8], nonlinear nonconvex problems are formulated as linear programming (LP) problems on occupation measures. Instead of the classical solution, the object of interest is a measure supported on the graph of the solution. Numerically, we optimize over finitely many moments of this measure.

Following the notation introduced above for stating Problem 1, and letting

$$\mathbf{z} := (\mathbf{x}, y) \in \mathbb{R}^p,$$

the moment matrix of degree $2d$ reads

$$\mathbf{M}_{\mu,d} := \int \mathbf{b}(\mathbf{z})\mathbf{b}(\mathbf{z})^\top \mathrm{d}\mu(\mathbf{z})$$

corresponding to the measure

$$\mathrm{d}\mu(\mathbf{z}) := \mathbb{I}_X(\mathbf{x})\,\mathrm{d}\mathbf{x}\,\delta_{f(\mathbf{x})}(\mathrm{d}y) \tag{1}$$

supported on the graph

$$\{(\mathbf{x}, f(\mathbf{x})) : \mathbf{x} \in X\} \subset X \times Y$$

of function f, where $\mathbb{I}_X$ denotes the indicator function of X which takes value 1 on X and 0 otherwise, and $\delta_{f(\mathbf{x})}$ denotes the Dirac measure at $f(\mathbf{x})$.

For instance, in OCPs an optimal occupation measure μ is supported on the graphs of optimal state-control trajectories. In order to recover a particular state, resp. control

trajectory, it suffices to consider the moments of the marginal of the occupation measure μ with respect to time state, resp. time control. Then, with our notation, $\mathbf{x}$ is time and y is a state, resp. control coordinate. Similarly, for the measure-valued formulation of nonlinear first-order scalar hyperbolic PDEs, an occupation measure is supported on the graph of the unique optimal entropy solution. Then, with our notation, $\mathbf{x}$ is time and space and y is the solution. The goal is to approximate the solution from the sole knowledge of moments of the occupation measure μ. The measure μ can be disintegrated into its marginal on X and its conditional on Y given $\mathbf{x}$ in X. The latter is also called a parameterized measure or a Young measure, see, e.g., [12].

This weak formulation has been used in a number of different contexts to prove existence and sometimes uniqueness of solutions. It turns out that it can also be used for effective computation as it fits perfectly the LP-based methodology described in [16] and the moment-SOS (polynomial sums of squares) methodology described in, e.g., [21–23]. In the latter methodology one may thus approximate the optimal solution μ of the measure-valued formulation by solving a hierarchy of semi-definite relaxations of the problem, whose size increases with d; see, e.g., [23] for OCPs and [26] for nonlinear PDEs. An optimal solution of each semi-definite relaxation is a finite matrix of pseudo-moments (of degree at most $2d$) which approximate those of μ. This approach allows to approximate values for the corresponding variational problems but it does not provide any information about the underlying minimizing solutions beyond moments of measures supported on these solutions. Therefore, an important and challenging practical issue consists of recovering from these moments an approximation of the trajectories of the OCP or PDEs. This is precisely an instance of Problem 1.

Another potential target application of our method is the optimal transportation problem, see, e.g., [34, Chapter 1] and references therein. In its original formulation by Monge, it is a highly nonlinear nonconvex optimization problem. Its relaxation by Kantorovich is a linear optimization problem on measures and hence on moments if the data are semi-algebraic. Under convexity assumptions, this linear problem has a unique measure solution called optimal transportation plan, supported on the graph of a function called the optimal transportation map. Our method can be used to approximate separately each coordinate of the transportation map by only considering the submatrix of moments associated with a suitable marginal, extracted from an optimal solution of the semi-definite relaxation (which considers *all* pseudo-moments up to a given degree). In view of the form of μ, one still obtains the required convergence guarantees (e.g., pointwise), under appropriate assumptions described in the paper.

More generally, moment information about the unknown function f in the format of Problem 1 is available when applying the moment-SOS hierarchy [22] to solve Generalized Moment Problems where the involved Borel measures are Young measures. The necessary moment information is also given when considering empirical measures [25,31,32] if input data points lie on the graph of an unknown function f (e.g., as is the case in interpolation). On the other hand, in some other applications like image processing or shape reconstruction, moment information is available only for the measure $f(\mathbf{x})\,d\mathbf{x}$, i.e., $y \mapsto \mathbf{b}(\mathbf{x}, y)$ is linear. Finally, our method and results would apply to measures ν supported on the graph of f, different from μ in (1). Provided

that ν and μ are mutually absolutely continuous with bounded densities, we would recover similar convergence results modulo constants.

1.2 Contribution

We address Problem 1 by providing an algorithm to approximate a (possibly discontinuous) unknown function f from the moment matrix of the measure (1) supported on its graph. The approximation converges to f (in a suitable sense described later on) when the number of moments tends to infinity.

Proposed approximation scheme

As is well known in approximation theory, the sequence of Christoffel–Darboux polynomials associated with a measure is an appropriate tool to approximate accurately the support of the measure, hence the graph of f in our case. Christoffel–Darboux kernels and functions are closely related to orthogonal polynomials [10,38] and approximation theory [6,28]. Their asymptotic behavior (i.e., when the degree goes to infinity) provides useful and even crucial information on the support and the density of the underlying measure. A quantitative analysis is provided in [15,27] for single dimensional problem and in [20] in a multivariate setting. Even more recently, in [25,31,32], Christoffel–Darboux polynomials have been used to approximate the support of Borel measures in a multivariate setting in the context of Machine Learning and Data Science.

We propose a simple scheme to approximate the graph of f based on the knowledge of the moment matrix of degree $2d$ of μ. To that end we first compute the Christoffel–Darboux polynomial using a spectral decomposition of the moment matrix. The Christoffel–Darboux polynomial is an SOS polynomial $q_d(\mathbf{x}, y)$ of degree $2d$ in p variables. For every fixed $\mathbf{x}$ we define

$$f_d(\mathbf{x}) := \operatorname{argmin}_{y \in Y} q_d(\mathbf{x}, y)$$

which is a semi-algebraic function,[1] assuming for the moment for the ease of exposition that the above argmin is uniquely defined. This class of functions is quite large. For example, all polynomials of degree at most d can be expressed using this technique: let r be a polynomial in $\mathbf{x}$ of degree d, then $q : (\mathbf{x}, y) \mapsto (r(\mathbf{x}) - y)^2$ is a degree $2d$ SOS polynomial whose partial minimization in y yields $y = r(\mathbf{x})$ for all $\mathbf{x}$. However, this class contains many more functions, including nonsmooth semi-algebraic functions such as signs or absolute values. In particular this class of functions can be used to describe efficiently discontinuous functions, a typical case encountered in, e.g., OCP problems with bang-bang controls and solutions with shocks for nonlinear PDEs.

Example 1 *(Sign function as SOS partial minimum)* Consider the polynomial

$$\begin{aligned} p_1 : \quad & \mathbb{R}^2 \mapsto \mathbb{R} \\ & (\mathbf{x}, y) \mapsto 4 - 3\mathbf{x}y - 4y^2 + \mathbf{x}y^3 + 2y^4. \end{aligned} \tag{2}$$

[1] A semi-algebraic function is a function whose graph is semi-algebraic, i.e., defined with finitely many polynomial inequalities.

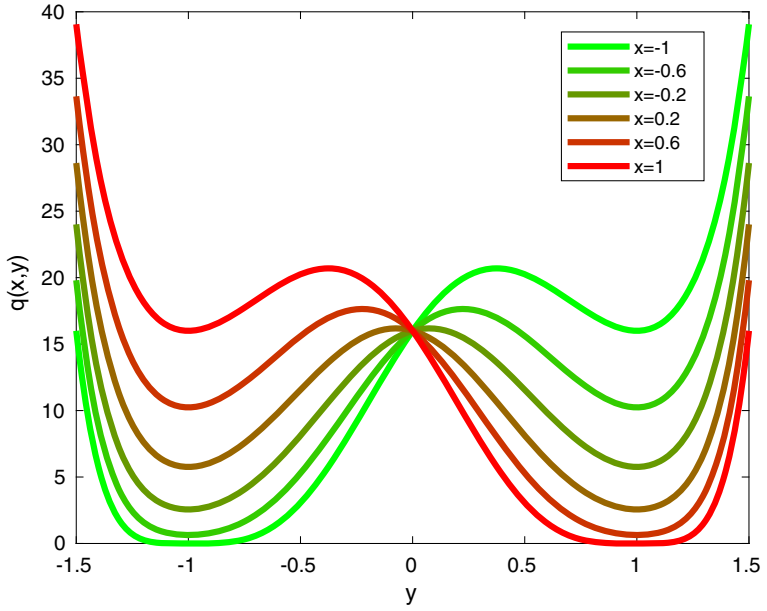

Fig. 1 SOS polynomial $q(\mathbf{x}, y)$ whose argmin in y is the sign of $\mathbf{x}$ on $[-1, 1]$

One can easily check that

$$\operatorname{argmin}_{y\in Y} p_1(\mathbf{x}, y) = \operatorname{sign}(\mathbf{x}) := \begin{cases} -1 & \text{if } \mathbf{x} < 0 \\ \{-1, 1\} & \text{if } \mathbf{x} = 0 \\ 1 & \text{if } \mathbf{x} > 0 \end{cases}$$

for any $\mathbf{x} \in X := [-1, 1]$ and $Y \subset \mathbb{R}$. Note that since $p_1(\mathbf{x}, \cdot)$ is positive for $\mathbf{x} \in [-1, 1]$, it can be squared without changing the argmin in y and hence we obtain a similar representation of the sign function in the form of partial minimization in y of the degree 8 SOS polynomial $q(\mathbf{x}, y) := p_1^2(\mathbf{x}, y)$, represented in Fig. 1.

Example 2 *(Absolute value as SOS partial minimum)* The reader may check that the argmin in y of the (square of the positive) polynomial $11-12\mathbf{x}^4y-6\mathbf{x}^2y^2+4\mathbf{x}^2y^3+3y^4$ is equal to $|\mathbf{x}|$ for all $\mathbf{x} \in [-1, 1]$.

To the best of our knowledge, this work is the first contribution where this class of semi-algebraic functions is used for graph approximations. The present work shows how these functions may be used to approximate discontinuous functions accurately.
Comparison to existing approximation approaches: A classical alternative to our approach would be to use $\mathscr{L}^2$-norm Legendre or Chebyshev approximations of the function f which are also based on moment information. However, these approaches only use moment information on the measure $f(\mathbf{x})\,d\mathbf{x}$, i.e., $y \mapsto \mathbf{b}(\mathbf{x}, y)$ is linear. Moreover, the support of this measure is not the graph of f.

We claim that the full moment information provides useful additional data on the graph of f which we can access using Christoffel–Darboux kernels associated with

the measure μ in (1). Note that in interpolation applications, we have the possibility to estimate the higher-order moments of f from finitely many evaluations of f through Riemann integral approximations or Monte Carlo approximations, for example. However, in situations where we have neither access to higher moments nor pointwise evaluation of f, our method cannot be applied; signal processing applications are a typical case of the latter situations.

In general, approximating a discontinuous function f is a challenge. Indeed, most well-known techniques suffer from the Gibbs phenomenon, i.e., the approximation produces oscillations at each point of discontinuity of f, see, e.g., [14] for a good survey on this topic. The main tools usually rely on properties of orthogonal polynomials [38] and the resulting approximations are based on a finite number of Fourier coefficients of the latter functions, i.e., typically first degree moment information on f. Projecting a discontinuous function on a class of infinitely smooth functions is the typical mechanism producing Gibbs phenomenon. In order to get rid of such a curse, additional techniques and prior information is needed. A possible approach is reported in [11] in the univariate case ($p = 2$ in our notations), where a good approximation of locations of discontinuity points and jump magnitude is obtained by solving an appropriate (univariate) polynomial equation. Recent developments have effectively shown that such approaches may tame the Gibbs phenomenon [3,4]. Iterative numerical methods can also be used to identify the points of discontinuities of f (and of its derivatives) so as to construct accurate approximations locally in each identified interval, see, e.g., [30] in the case of Chebyshev polynomials. However, such ad hoc techniques are very specific to the univariate setting.

On the contrary, our approach is not based on projection on a subspace of smooth functions, or identification of points of discontinuities of the function to be approximated. It is based on geometric approximation of the support of a singular measure using semi-algebraic techniques. An important feature of this approach is that the resulting approximant functions are not necessarily smooth, and furthermore, discontinuities (if any) can be treated implicitly only based on the whole moment information. As a result, we obtain a single approximation scheme, which (i) may adapt to the smoothness features of the target function f without requiring prior knowledge of it, and (ii) can be used for multivariate f, both points being important challenges regarding numerical approximation.

Description of our contribution

1. We first provide a numerical scheme which allows to approximate the compact support of a measure which is singular with respect to the Lebesgue measure. We need to adapt the strategy of [25] which covered the absolutely continuous case. This result may be considered of independent interest and will be instrumental to providing convergence guarantees for our approach.
2. Next, given a degree $d \in \mathbb{N}$, we provide an approximation f_d of the function f and prove that the sequence $(f_d)_{d\in\mathbb{N}}$ converges pointwise to f almost everywhere as well as in $\mathscr{L}^1$-norm as d goes to infinity (under broad assumptions on f). Furthermore, if we assume more regularity on f, then we also provide estimates for the rate of convergence. More precisely, we obtain $O(d^{-1/2})$ for multivariate Lipschitz functions and $O(d^{-1/4})$ for univariate functions of bounded variation.

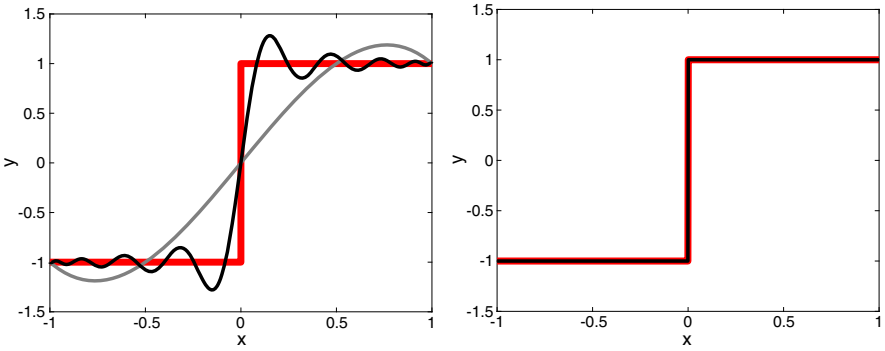

Fig. 2 On the left, Chebyshev interpolant of degrees 4 (gray) and 20 (black) of the step function (red), featuring the typical Gibbs phenomenon. On the right, the proposed semi-algebraic approximation of degree 4 (black) of the same step function (red). The approximation cannot be distinguished from the step function (Color figure online)

3. Finally, we provide some numerical examples to illustrate the efficiency of the method. We first use our algorithm to approximate discontinuous or nonsmooth solutions of OCP or PDE problems based on the moment-SOS hierarchy. These experiments empirically demonstrate the absence of Gibbs phenomena. We also provide an example where only samples of the measure under consideration are given and show that our algorithm also works well, even for moderate size samples, showing that moment input data could in principle be approximated using numerical integration methods.

Example 3 *(Sign function)* To give a flavor of what can be obtained numerically, consider the measure (1) supported on the graph $\{(\mathbf{x}, f(\mathbf{x})) : \mathbf{x} \in X\} \subset X \times Y$ of the function $\mathbf{x} \mapsto f(\mathbf{x}) := \mathrm{sign}(\mathbf{x})$, with $X := Y := [-1, 1]$. In Fig. 2 (right) the resulting approximation f_2 with a moment matrix of size 6 and degree 4 (i.e., 15 moments) cannot be distinguished from f. On the other hand, on the left, their Chebsyhev interpolants of degrees 4 and 20 (obtained with `chebfun` [9]) illustrate the typical Gibbs phenomenon, namely oscillations near the discontinuity points that cannot be attenuated by increasing the degree. This phenomenon can be reduced or suppressed by identifying the discontinuity points and splitting X into intervals (as in, e.g., [30] and also implemented in `chebfun`), but this strategy works only in the univariate case. In contrast, our algorithm does not attempt to approximate directly a univariate function with one or several univariate polynomials of increasing degree, but with the argmin of a bivariate polynomial of increasing degree. Moreover, our algorithm works also for multivariate functions, as shown by numerical examples later on.

1.3 Organization of the Paper

Section 2 introduces the Christoffel–Darboux polynomial, its regularized version and our semi-algebraic approximant. In Sect. 3 the main results of the paper are collected, while their proofs are provided in Sect. 4. More precisely, we first give some quanti-

tative estimates on the support of the measure μ and then prove the $\mathscr{L}^1$ convergence of our approximant. Section 5 discusses computational issues, and with the help of a simple MATLAB prototype, it illustrates the efficiency of our method on some examples. Finally, Sect. 6 collects some concluding remarks together with further research lines to be followed.

Notation

The Euclidean space of real-valued symmetric matrices of size n is denoted by $\mathbb{S}^n$. Given a set X in Euclidean space, its diameter is denoted by $\mathrm{diam}(X)$ and its volume, or Lebesgue measure, is denoted by $\mathrm{vol}(X)$. For $k \geq 1$, the Lebesgue space $\mathscr{L}^k(X)$ consists of functions on X whose k-norms are bounded. Given a positive Borel measure μ, we denote by $\mathrm{supp}(\mu)$ its support, defined as the smallest closed subset whose complement has measure zero.

Throughout the paper, p denotes the dimension of the ambient space. Consistently with the notations introduced in Sect. 1 for stating Problem 1, we let $\mathbf{z} = (\mathbf{x}\ \ y) \in \mathbb{R}^p$. We denote by $\mathbb{R}[\mathbf{z}]$ the algebra of multivariate polynomials of $\mathbf{z} \in \mathbb{R}^p$ with real coefficients. For a given degree $d \in \mathbb{N}$, the dimension of the vector space of polynomials of degree less than or equal to d is denoted by $n_d := \binom{p+d}{d}$.

2 Christoffel–Darboux Approximation

This section describes our main approximant based on the Christoffel–Darboux kernel. We first introduce notations and definitions, describe our regularization scheme for the Christoffel–Darboux kernel and then describe our functional approximant based on moments.

2.1 Polynomials and Moments

Following the notations introduced for stating Problem 1, any polynomial $q \in \mathbb{R}[\mathbf{z}]$ of degree d can be expressed in the polynomial basis $\mathbf{b}(\mathbf{z})$ as $q : \mathbf{z} \mapsto \mathbf{q}^\top \mathbf{b}(\mathbf{z})$ with $\mathbf{q} \in \mathbb{R}^{n_d}$ denoting its vector of coefficients. Recall that the moment matrix of degree $2d$ of the measure μ reads

$$\mathbf{M}_{\mu,d} = \int \mathbf{b}(\mathbf{z})\mathbf{b}(\mathbf{z})^\top \mathrm{d}\mu(\mathbf{z}) \in \mathbb{S}^{n_d}.$$

Since $\mathbf{M}_{\mu,d}$ is positive semi-definite, it has a spectral decomposition

$$\mathbf{M}_{\mu,d} = \mathbf{P}\mathbf{E}\mathbf{P}^\top, \tag{3}$$

where $\mathbf{P} \in \mathbb{R}^{n_d \times n_d}$ is an orthonormal matrix whose columns are denoted $\mathbf{p}_i$, $i = 1, 2, \ldots, n_d$ and satisfy $\mathbf{p}_i^\top \mathbf{p}_i = 1$ and $\mathbf{p}_i^\top \mathbf{p}_j = 0$ if $i \neq j$ and $\mathbf{E} \in \mathbb{S}^{n(d)}$ is a diagonal matrix whose diagonal entries are eigenvalues $e_{i+1} \geq e_i \geq 0$ of the moment matrix. Each column $\mathbf{p}_i \in \mathbb{R}^{n_d}$ is the vector of coefficients of a polynomial $p_i \in \mathbb{R}[\mathbf{z}]$, $i = 1, \ldots n_d$, so that

$$\begin{aligned} \mathbf{p}_i^\top \mathbf{M}_{\mu,d}\, \mathbf{p}_i &= e_i = \int p_i^2(\mathbf{z}) \mathrm{d}\mu(\mathbf{z}), \\ \mathbf{p}_i^\top \mathbf{M}_{\mu,d}\, \mathbf{p}_j &= 0 = \int p_i(\mathbf{z}) p_j(\mathbf{z}) \mathrm{d}\mu(\mathbf{z}), \quad i \neq j. \end{aligned} \tag{4}$$

2.2 Approximating the Support from Moments

Let us assume for now that the support of the measure μ has nonempty interior, then $\mathbf{M}_{\mu,d}$ is positive definite for any $d \in \mathbb{N}$, i.e., $e_i > 0$, $i = 1, \ldots, n_d$. In this case, one can define the Christoffel–Darboux polynomial

$$q_{\mu,d}(\mathbf{z}) := \sum_{i=1}^{n_d} \frac{p_i^2(\mathbf{z})}{e_i} = \mathbf{b}(\mathbf{z})^\top \mathbf{M}_{\mu,d}^{-1} \mathbf{b}(\mathbf{z}). \tag{5}$$

It is known that sublevel sets of $q_{\mu,d}$ can be used to recover the support of μ for large d, see, for example, [25] for an overview.

The goal of this work is to approximate the function f. From a set theoretic perspective, this amounts to approximating the graph of f whose closure is actually the support of the measure μ in (1). Hence, the sublevel sets of $q_{\mu,d}$ are interesting candidates for this goal. However, in the case of the graph of a function, the construction given in (5) is not valid anymore since this graph is a singular set so that $\mathbf{M}_{\mu,d}$ may not be positive definite and invertible. In this singular setting, one should ideally consider the following extended value Christoffel–Darboux polynomial:

$$q^e_{\mu,d}(\mathbf{z}) := \begin{cases} +\infty & \text{if } \exists i,\ e_i = 0,\ p_i(\mathbf{z}) \neq 0 \\ \sum_{e_i > 0} \frac{p_i^2(\mathbf{z})}{e_i} = \mathbf{b}(\mathbf{z})^\top \mathbf{M}_{\mu,d}^\dagger \mathbf{b}(\mathbf{z}) & \text{otherwise,} \end{cases} \tag{6}$$

where † denotes the Moore–Penrose pseudo-inverse. This is a natural extension of the Christoffel–Darboux polynomial to the singular case [32] and amounts to working in the Zariski closure of the graph of f.

2.3 Regularization Scheme

Spectral filtering: Computing an object such as in (6) can be numerically sensitive since it essentially relies on pseudo-inverse which requires an eigenvalue thresholding scheme. This means that small perturbations of the moment matrix may lead to large changes of the output. Furthermore, the candidate function takes finite values only on an algebraic set, and this situation is difficult to handle in finite precision arithmetic. One may rewrite the extended value polynomial (6) in the following form

$$q^e_{\mu,d}(\mathbf{z}) = \sum_{i=1}^{n_d} g(e_i) p_i^2(\mathbf{z})$$

where $g\colon [0, +\infty) \mapsto [0, +\infty]$ with $g(s) = \frac{1}{s}$ for any $s > 0$ and $g(0) = +\infty$. One approach to restore stability is to use regularization techniques which replace the pseudo-inversion expressed through the function g, by spectral filtering expressed

through a different spectral function (see, for example, [7] for an illustration in the context of support estimation). Since the function g is not regular, instead of studying the above defined extended value polynomial, one rather looks at the following polynomial

$$\sum_{i=1}^{n_d} g_\beta(e_i)\, p_i^2(\mathbf{z}) \tag{7}$$

where g_β is a parameterized family of spectral filtering regular functions satisfying, for any $\beta > 0$, $g_\beta : [0, +\infty) \mapsto [0, +\infty]$. Common examples include

$$\begin{aligned}
\text{Tikhonov regularization:} \quad & g_\beta : s \mapsto \frac{1}{\beta + s}, \\
\text{Spectral cut-off:} \quad & g_\beta : s \mapsto \frac{1}{\beta}\mathbb{I}_{[0,\beta]}(s) + \frac{1}{s}\mathbb{I}_{(\beta,+\infty)}(s), \\
\text{Ideal low-pass:} \quad & g_\beta : s \mapsto \frac{1}{\beta}\mathbb{I}_{[0,\beta]}(s).
\end{aligned}$$

We choose to work with the Tikhonov regularization as it has an intuitive measure space interpretation. We believe that our results can be generalized to different spectral filters.

Tikhonov regularization and measures: Applying the Tikhonov spectral filter to (7) yields the following polynomial

$$\mathbf{z} \mapsto \sum_{i=1}^{n_d} \frac{p_i^2(\mathbf{z})}{e_i + \beta} = \mathbf{b}(\mathbf{z})^\top (\mathbf{M}_{\mu,d} + \beta \mathbf{I}_{n_d})^{-1} \mathbf{b}(\mathbf{z}) \tag{8}$$

where $\mathbf{I}_{n_d}$ denotes the identity matrix of size n_d. In order to use analytic tools, we need to provide an interpretation of the addition of diagonal elements in terms of measures. One therefore has to choose a polynomial basis for which the diagonal matrix is the moment matrix of a reference Borel measure on $\mathbb{R}^p$ that we will denote μ_0. We make the following assumption which will be standing throughout the paper.

Assumption 1 (Reference measure and polynomial basis)

- The reference measure μ_0 is absolutely continuous with respect to the Lebesgue measure and it has compact support.
- The polynomial basis $\mathbf{b}$ is orthonormal with respect to the bilinear form induced by μ_0, that is $\int b_i(\mathbf{z}) b_j(\mathbf{z}) \mathrm{d}\mu_0(\mathbf{z}) = 1$ if $i = j$ and 0 otherwise.

The first part of Assumption 1 ensures that the moment matrix of μ_0 is always positive definite. The second part of Assumption 1 provides the following relation:

$$\mathbf{M}_{\mu,d} + \beta \mathbf{I}_{n_d} = \mathbf{M}_{\mu+\beta\mu_0,d}. \tag{9}$$

An easy example of such a measure should be the following: considering a function f whose domain of definition is contained in the unit cube of dimension $p - 1$ and

which takes values in $[-1, 1]$, one might pick the uniform measure on the unit cube of dimension p, for which moments are easy to compute.

Most importantly, using the notation in (5), this allows to express the polynomial of interest (8) as follows.

Definition 1 *(Regularized Christoffel–Darboux polynomial)* The regularized Christoffel–Darboux polynomial is the SOS

$$q_{\mu+\beta\mu_0,d}(\mathbf{z}) := \sum_{i=1}^{n_d} \frac{p_i^2(\mathbf{z})}{e_i + \beta}. \tag{10}$$

This provides a geometric interpretation of the regularization parameter as a combination of two measures: μ which is supported on the graph of the function of interest and μ_0 which is a reference measure, used for regularization purposes. The supported boundedness hypothesis in Assumption 1 will allow to provide quantitative estimates in further sections and it could in principle be replaced by a fast decreasing tail condition. An important example for measures satisfying Assumption 1 is the restriction of Lebesgue measure to the unit hypercube together with Legendre polynomials which form an orthonormal basis.

Making Assumption 1 is a slight restriction for which a few comments are in order. Firstly, this could be relaxed in various ways to remove the restriction on the polynomial basis, for example:

- Replace the identity matrix by the moment matrix of μ_0;
- Add assumption on the spectrum of the moment matrix μ_0.

These would lead to a lot of technical complications and we find our results clearer and easier to state under Assumption 1. Secondly, working numerically with polynomials in the standard monomial basis is problematic in many situations. Better conditioned polynomial bases are often those enjoying orthonormality properties with respect to a certain reference measure, such as, e.g., Chebyshev or Legendre polynomials. We would like to argue here that the restrictions induced by Assumption 1 are quite benign since it is already common in practice to work in such polynomial bases for numerical reasons.

2.4 Semi-algebraic Approximant

Definition 2 (*Semi-algebraic approximant*) The regularized Christoffel–Darboux semi-algebraic approximant $f_{\beta,d}$ is defined as follows:

$$\mathbf{x} \in X \mapsto f_{\beta,d}(\mathbf{x}) := \min\{\operatorname{argmin}_{y\in Y} \; q_{\mu+\beta\mu_0,d}(\mathbf{x}, y)\}. \tag{11}$$

Remark 1 If X and Y are semi-algebraic, then the set-valued map which associates to each $\mathbf{x} \in X$ the set

$$\operatorname{argmin}_{y\in Y} \; q_{\mu+\beta\mu_0,d}(\mathbf{x}, y)$$

is semi-algebraic. Recall that a map is semi-algebraic if its graph is semi-algebraic. By the Tarski–Seidenberg Theorem (see, for example, [5, Theorem 2.6]), any first-order formula involving semi-algebraic sets describes a semi-algebraic set. Since minima are described by first-order formulas, the argmin of a polynomial on the compact semi-algebraic set Y is a compact semi-algebraic subset of Y, which is itself a subset of the real line. Hence, the argmin set has a minimal element and the function in (11) is well defined.

Remark 2 For clarity of exposition we describe our main results by considering that the partial minimization in y over Y is exact. As will be seen from arguments in the proof, especially in the proof of Lemma 7, approximate minimization up to a factor of the order d^{p+2} enjoys similar approximation properties. Indeed, in Remark 5, we mention the precision $\frac{\gamma_d}{2}$, where γ_d is chosen later on with more justifications in (14) to be of order d^{p+2}.

The two parameters d and β control the behavior of the approximant $f_{\beta,d}$. In latter sections, we describe an explicit dependency between d and β which allows to construct a sequence of regularization parameters $(\beta_d)_{d\in\mathbb{N}}$, and we investigate the asymptotic properties of the sequence of approximants $\left(f_{\beta_d,d}\right)_{d\in\mathbb{N}}$.

3 Main Results

Our main theoretical contribution is an investigation of the relations between the function f to be approximated and its regularized Christoffel–Darboux approximant $f_{\beta,d}$ under Assumption 1. In particular, we are interested in building an explicit sequence $(\beta_d)_{d\in\mathbb{R}}$ and investigating the convergence $f_{\beta_d,d}(\mathbf{x}) \to f(\mathbf{x})$ for $\mathbf{x} \in X$, fixed, as well as the convergence $\|f - f_{\beta_d,d}\|_{\mathscr{L}^1(X)} \to 0$, when $d \to \infty$.

3.1 Convergence Under Continuity Assumptions

The following section describes our main result regarding convergence of the approximant $f_{\beta_d,d}$ in (11) under different regularity assumptions for the function f to be approximated. Let us define

$$\delta_0 := \operatorname{diam}(\operatorname{supp}(\mu + \mu_0)), \quad m := \mu(\mathbb{R}^p), \quad m_0 := \mu_0(\mathbb{R}^p).$$

Theorem 1 *Under Assumption 1 and with the choice $\beta_d = 2^{3-\sqrt{d}}$ in Definition 2, it holds:*

(i) *If the set $S \subset X$ of continuity points of f is such that $X \setminus S$ has Lebesgue measure zero, then*

$$f_{\beta_d,d}(\mathbf{x}) \underset{d\to\infty}{\to} f(\mathbf{x})$$

for almost all $\mathbf{x} \in X$, *and*

$$\|f - f_{\beta_d,d}\|_{\mathscr{L}^1(X)} \underset{d\to\infty}{\to} 0.$$

(ii) *If* f *is* L*-Lipschitz on* X *for some* $L > 0$, *then for any* $d > 1$ *and any* $r > p$,

$$\|f - f_{\beta_d,d}\|_{\mathscr{L}^1(X)} \leq \mathrm{vol}(X)\frac{\delta_0}{\sqrt{d}-1}(1+L) + \mathrm{diam}(Y)\frac{8(m+m_0)(3r)^{2r}e^{\frac{p^2}{d}}}{p^p e^{2r-p} d^{r-p}}.$$

Remark 3 • As proved recently in [35], the solutions to scalar conservation laws are continuous almost everywhere, i.e., the Lebesgue measure of the set of discontinuity points reduces to 0. It is then clear that Item (i) of Theorem 1 can be applied directly to the case of scalar conservation laws.

- Thanks to Egorov's Theorem, pointwise convergence almost everywhere implies almost uniform convergence, that is uniform convergence up to a subset of X whose measure can be taken arbitrarily small. Since we manipulate bounded functions, this in turn implies convergence in $\mathscr{L}^1$.
- For Lipschitz continuous functions, we obtain an $O(d^{-1/2})$ convergence rate in $\mathscr{L}^1$ norm by setting $r = p + 1/2$.
- The convergence rate for Lipschitz functions is slow and we observe in practice a much faster convergence. We conjecture that faster rates can be obtained for special classes of functions.

Theorem 1 is a special case of a more general result described in Theorem 3 and proven later on.

3.2 Convergence for Univariate Functions of Bounded Variation

Spaces of functions of bounded variations are of interest because many PDE problems are formulated on such spaces, see [1] for an introduction. Modern construction of such spaces is done by duality through measure theoretic arguments. Our main proof mechanisms rely on pointwise properties of the function f, which are not completely captured by the measure theoretic construction.

We prove $\mathscr{L}^1$ convergence for univariate bounded variation functions. The reason we are limited to the univariate setting is that we can use the classical definition of total variation which is directly connected to pointwise properties of the function of interest. We conjecture that the proposed approximation scheme is also convergent for multivariate functions of bounded variation, but we leave this question for future work.

Definition 3 Let $f : \mathbb{R} \mapsto \mathbb{R}$ be a measurable function. The (Jordan) total variation norm of f is given by

$$V(f) = \sup_{n\in\mathbb{N}} \sup_{x_0<x_1<x_2<\ldots<x_n} \sum_{i=1}^{n} |f(x_i) - f(x_{i-1})|.$$

Theorem 2 *Under Assumption 1, let X and Y be intervals of the real line, and assume that $V(f) < +\infty$. With the choice $\beta_d = 2^{3-\sqrt{d}}$ in Definition 2, for any $r > 2$ and for any $d > 1$, it holds*

$$\begin{aligned}
&\|f - f_{\beta_d,d}\|_{\mathscr{L}^1(X)} \\
&\quad \leq \operatorname{vol}(X)\left(\frac{2\delta_0}{\sqrt{d}-1} + d^{-\frac{1}{4}}\right) \\
&\qquad + \operatorname{diam}(Y)\left(\frac{8(m+m_0)(3r)^{2r}e^{\frac{4}{d}}}{4e^{2r-2}d^{r-2}} + \frac{4d^{\frac{1}{4}}V(f)\delta_0}{\sqrt{d}-1}\right).
\end{aligned}$$

We remark that we obtain a convergence rate in $O(d^{-1/4})$. This result is a special case of a more general result described in Theorem 4 and proven later on.

3.3 Robustness to Small Perturbations

In many situations, one only has access to an approximation of the regularized Christoffel–Darboux polynomial. This is, for example, the case when applying the moment-SOS hierarchy. At the end, one indeed obtains pseudo-moments of the measure under consideration, i.e., a vector of real numbers close to the actual moments of the measure. The moment matrix is then not $\mathbf{M}_{\mu+\beta_d\mu_0}$ as in (9) but a matrix $\mathbf{M}$ close to it. The effect of this perturbation on the Christoffel–Darboux polynomial is captured by the following result.

Lemma 1 *Assume that the approximate moment matrix $\mathbf{M} \in \mathbb{S}^{n_d}$ is positive definite and let*

$$\alpha := \|\mathbf{I}_{n_d} - \mathbf{M}^{\frac{1}{2}}_{\mu+\beta_d\mu_0}\mathbf{M}^{-1}\mathbf{M}^{\frac{1}{2}}_{\mu+\beta_d\mu_0}\|$$

where we used the matrix operator norm. Then, the polynomial $q_d^\alpha : \mathbf{z} \mapsto \mathbf{b}(\mathbf{z})^\top\mathbf{M}^{-1}\mathbf{b}(\mathbf{z})$ satisfy

$$\sup_{\mathbf{z}\in\mathbb{R}^p}\left|1 - \frac{q_d^\alpha(\mathbf{z})}{q_{\mu+\beta_d\mu_0,d}(\mathbf{z})}\right| \leq \alpha.$$

Proof For any $\mathbf{z} \in \mathbb{R}^p$, we have

$$\begin{aligned}
&\left| \frac{q_d^\alpha(\mathbf{z})}{q_{\mu+\beta_d\mu_0,d}(\mathbf{z})} - 1 \right| \\
&= \frac{1}{q_{\mu+\beta_d\mu_0,d}(\mathbf{z})} \left| \mathbf{b}(\mathbf{z})^\top (\mathbf{M}^{-1} - \mathbf{M}_{\mu+\beta_d\mu_0}^{-1}) \mathbf{b}(\mathbf{z}) \right| \\
&= \frac{1}{q_{\mu+\beta_d\mu_0,d}(\mathbf{z})} \left| (\mathbf{M}_{\mu+\beta_d\mu_0}^{-\frac{1}{2}} \mathbf{b}(\mathbf{z}))^\top (\mathbf{I}_{n_d} - \mathbf{M}_{\mu+\beta_d\mu_0}^{\frac{1}{2}} \mathbf{M}^{-1} \mathbf{M}_{\mu+\beta_d\mu_0}^{\frac{1}{2}}) \mathbf{M}_{\mu+\beta_d\mu_0}^{-\frac{1}{2}} \mathbf{b}(\mathbf{z}) \right| \\
&\leq \left\| \mathbf{I}_{n_d} - \mathbf{M}_{\mu+\beta_d\mu_0}^{\frac{1}{2}} \mathbf{M}^{-1} \mathbf{M}_{\mu+\beta_d\mu_0}^{\frac{1}{2}} \right\| = \alpha.
\end{aligned}$$

□

Remark 4 *(Accuracy of the moments)* In this paper, we assume the existence of a positive number α which makes the link between an approximated moment matrix and the real one, which exists in general, but estimates for α are not known in general. Such an analysis has been performed for the moments themselves in [29], but not for moment matrices. Indeed, the bounds linking moment matrices and their corresponding depend nonlinearly on too many variables to obtain easily bounds on the approximated moment matrix and the real one. This is a topic of further investigation.

More generally, we can consider a robust Christoffel–Darboux function satisfying the following inequality.

Assumption 2 For a given $\alpha \in [0, 1)$, let $(\beta_d)_{d\in\mathbb{N}}$ be a sequence of positive numbers and $(q_d^\alpha)_{d\in\mathbb{N}}$ be a sequence of continuous functions over $\mathbb{R}^p$ such that for any $d \in \mathbb{N}$ and any $\mathbf{z} \in \mathbb{R}^p$, we have

$$\left| 1 - \frac{q_d^\alpha(\mathbf{z})}{q_{\mu+\beta_d\mu_0,d}(\mathbf{z})} \right| \leq \alpha.$$

Note that Assumption 2 ensures that

$$(1-\alpha) q_{\mu+\beta_d\mu_0,d}(\mathbf{z}) \leq q_d^\alpha(\mathbf{z}) \leq (1+\alpha) q_{\mu+\beta_d\mu_0,d}(\mathbf{z}). \tag{12}$$

Furthermore, one can always choose $q_d^\alpha(\mathbf{z}) = q_{\mu+\beta_d\mu_0,d}(\mathbf{z})$ and $\alpha = 0$ which corresponds to the nominal case. The robust approximant is then defined similarly as in Definition 2.

Definition 4 *(Robust semi-algebraic approximant)* Given a degree $d \in \mathbb{N}$, a regularizing parameter $\beta > 0$, and a scalar $\alpha \in [0, 1)$, our robust approximant $f_{\beta_d,d}^\alpha$ is defined as follows:

$$\mathbf{x} \in X \mapsto f_{\beta_d,d}^\alpha(\mathbf{x}) := \min\{\operatorname{argmin}_{y\in Y} q_d^\alpha(\mathbf{x}, y)\}. \tag{13}$$

The main technical result of this paper is the following from which Theorem 1 directly follows.

Theorem 3 *Under Assumptions 1 and 2, and with the choice* $\alpha \in [0,1)$ *and* $\beta_d = 2^{3-\sqrt{d}}$ *in Definition 4, it holds:*

(i) *If the set* $S \subset X$ *of continuity points of* f *is such that* $X \setminus S$ *has Lebesgue measure zero, then*

$$f^{\alpha}_{\beta_d,d}(\mathbf{x}) \underset{d\to\infty}{\to} f(\mathbf{x})$$

for almost all $\mathbf{x} \in X$*, and*

$$\|f - f^{\alpha}_{\beta_d,d}\|_{\mathscr{L}^1(X)} \underset{d\to\infty}{\to} 0.$$

(ii) *If* f *is L-Lipschitz on X for some* $L > 0$*, then for any* $d > 1$ *and any* $r > p$*,*

$$\begin{aligned}&\|f - f^{\alpha}_{\beta_d,d}\|_{\mathscr{L}^1(X)}\\ &\quad\le \operatorname{vol}(X)\frac{\delta_0}{\sqrt{d}-1}(1+L) + \operatorname{diam}(Y)\frac{1+\alpha}{1-\alpha}\frac{8(m+m_0)(3r)^{2r}e^{\frac{p^2}{d}}}{p^p e^{2r-p} d^{r-p}}.\end{aligned}$$

Furthermore, we have the following robust convergence result for univariate functions of bounded variation, from which Theorem 2 directly follows.

Theorem 4 *Under Assumptions 1 and 2, let X and Y be intervals of the real line and assume that* $V(f) < +\infty$*. With the choice* $\alpha \in [0,1)$ *and* $\beta_d = 2^{3-\sqrt{d}}$ *in Definition 4, for any* $r > 2$ *and for any* $d > 1$*, it holds*

$$\begin{aligned}&\|f - f^{\alpha}_{\beta_d,d}\|_{\mathscr{L}^1(X)}\\ &\quad\le \operatorname{vol}(X)\left(\frac{2\delta_0}{\sqrt{d}-1} + d^{-\frac{1}{4}}\right)\\ &\qquad + \operatorname{diam}(Y)\left(\frac{1+\alpha}{1-\alpha}\frac{8(m+m_0)(3r)^{2r}e^{\frac{4}{d}}}{4e^{2r-2}d^{r-2}} + \frac{4d^{\frac{1}{4}}V(f)\delta_0}{\sqrt{d}-1}\right).\end{aligned}$$

The next section is dedicated to the proof of these theorems.

4 Proofs

This section is divided into several subsections. Section 4.1 gives some quantitative results on the estimation of the support of μ which does not depend on the nature of μ and could be of independent interest. More precisely, we show that the regularized Christoffel–Darboux polynomial takes large values outside the support of μ and

smaller values inside. This is expressed by describing properties of certain sublevel sets of the polynomial being close to the support of μ. In Sect. 4.2, we translate these geometric results in functional terms. In Sect. 4.3, we prove our main results: the argument of the minimum of the regularized Christoffel–Darboux polynomial is close to the graph of the function f.

4.1 Estimation of the Support

In this section we build a polynomial sublevel set that will be instrumental for our proofs of convergence in functional terms. Note that in practice this sublevel set is not computed: we just focus on the argmin of the regularized Christoffel–Darboux polynomial. The contents of this section may be considered of independent interest.

For any $d \in \mathbb{N}$ and $r \in \mathbb{N}$ such that $r > p$ and for any $\alpha \in [0, 1)$, define

$$\gamma_d := \frac{1-\alpha}{8(m+m_0)} \frac{e^{2r} d^r}{(3r)^{2r}} \tag{14}$$

and

$$S_d := \{\mathbf{z} \in \mathbb{R}^p : q_d^\alpha(\mathbf{z}) < \gamma_d\}. \tag{15}$$

We aim at proving that the sublevel set S_d is approaching the support of μ as d goes to infinity, with a given convergence rate. It is precisely quantified with the following result which is illustrated in Fig. 3.

Theorem 5 *For $d > 1$ it holds*

(i)

$$\mu(\{\mathbf{z} \in \mathbb{R}^p : \mathbf{z} \notin S_d\}) \leq \frac{1+\alpha}{1-\alpha} \frac{8(m+m_0)(3r)^{2r} e^{\frac{p^2}{d}}}{p^p e^{2r-p} d^{r-p}}.$$

(ii) For any $\mathbf{z} \in S_d$,

$$\operatorname{dist}(\mathbf{z}, \operatorname{supp}(\mu)) \leq \frac{\delta_0}{\sqrt{d}-1}.$$

The results and techniques that we will use to prove this theorem are adapted from [25] which considers the absolutely continuous setting without regularization.

Proof of Theorem 5 (i) Using (10) and (4), we obtain

$$\int_{\mathbb{R}^p} q_{\mu+\beta_d\mu_0,d}(\mathbf{z}) \mathrm{d}\mu(\mathbf{z}) = \sum_{i=1}^{n_d} \frac{e_i}{e_i+\beta_d} \leq n_d \leq d^p \left(\frac{e}{p}\right)^p e^{\frac{p^2}{d}} \tag{16}$$

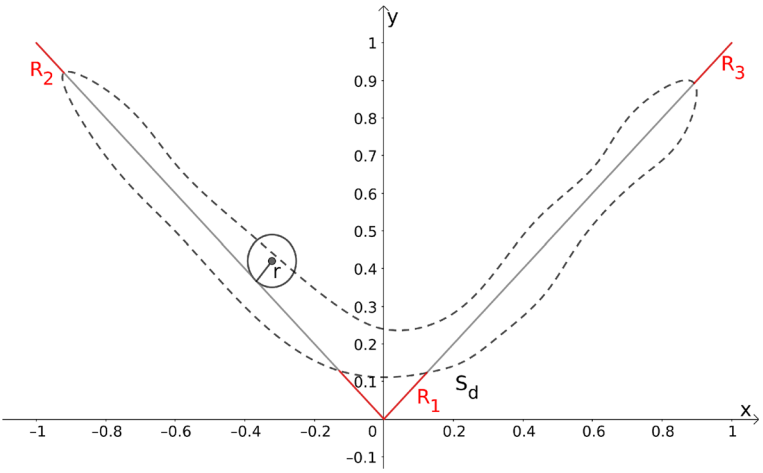

Fig. 3 Illustration of the result of Theorem 5, the dotted curve represents the boundary of the set S_d and the considered function f is the absolute value. The theorem states that (i) most of the points of the graph of f will be in S_d and that (ii) all points in S_d will be close to the graph of f. More precisely, (i) states that the measure of $R_1 \cup R_2 \cup R_3$ will vanish and (ii) the distance r of any point in S_d to the graph of f will go to zero for $d \to \infty$, respectively

where the last inequality is given in [25, Lemma 6.5]. Using Markov's inequality [36, Page 91] and (16) yields

$$\begin{aligned}
&\mu(\{\mathbf{z} \in \mathbb{R}^p : q_{\mu+\beta_d\mu_0,d}(\mathbf{z}) \geq \frac{\gamma_d}{1+\alpha}\}) \\
&\leq \frac{\int_{\mathbb{R}^p}(1+\alpha)q_{\mu+\beta_d\mu_0,d}(\mathbf{z})\mathrm{d}\mu(\mathbf{z})}{\gamma_d} \\
&\leq (1+\alpha)\frac{d^p\left(\frac{e}{p}\right)^p e^{\frac{p^2}{d}}}{\gamma_d}.
\end{aligned} \tag{17}$$

Now using (12) and (15), we have the following implications

$$\mathbf{z} \notin S_d \quad \Leftrightarrow \quad q_d^\alpha(\mathbf{z}) \geq \gamma_d \quad \Rightarrow \quad q_{\mu+\beta_d\mu_0,d}(\mathbf{z}) \geq \frac{\gamma_d}{1+\alpha}.$$

Hence, $\mu(\{\mathbf{z} \in \mathbb{R}^p : \mathbf{z} \notin S_d\}) \leq \mu(\{\mathbf{z} \in \mathbb{R}^p : q_{\mu+\beta_d\mu_0,d}(\mathbf{z}) \geq \frac{\gamma_d}{1+\alpha}\})$. Using the expression of γ_d in (14) and the inequality (17), one has:

$$\mu(\{\mathbf{z} \in \mathbb{R}^p : \mathbf{z} \notin S_d\}) \leq \frac{1+\alpha}{1-\alpha}\frac{8(m+m_0)(3r)^{2r}e^{\frac{p^2}{d}}}{p^p e^{2r-p}d^{r-p}}$$

which concludes the proof of item (i) of Theorem 5. □

To carry out the proof of Theorem 5 *(ii)*, we begin with a few lemmas. The following result is classical, see, e.g., [25, Remark 3.6.] and [20, Equation (1.1.)].

Lemma 2 *Let $d \in \mathbb{N}$, $\mathbf{z} \in \mathbb{R}^p$, $\beta > 0$, and q be a polynomial of degree at most d. Then,*

$$\frac{q^2(\mathbf{z})}{\int_{\mathbb{R}^p} q^2(\mathbf{z}) d(\mu + \beta\mu_0)(\mathbf{z})} \leq q_{\mu+\beta\mu_0,d}(\mathbf{z}).$$

The following lemma defines the needle polynomial, introduced first in [20], and gives a quantitative result crucial for our analysis.

Lemma 3 (Existence of a needle polynomial) *Let B_δ denote the Euclidean ball of radius δ. Then, for all $\delta \in (0, 1)$ and $d \in \mathbb{N}$, $d > 0$, there exists a polynomial q of degree $2d$ such that*

$$\begin{aligned} &q(0) = 1,\ q(\mathbf{z}) \in [-1, 1] \text{ for all } \mathbf{z} \in B_1, \text{ and} \\ &|q(\mathbf{z})| \leq 2^{1-\delta d} \text{ for all } \mathbf{z} \in B_1 \setminus B_\delta. \end{aligned} \tag{18}$$

A detailed proof is provided in [25, Lemma 6.3]. Thanks to the latter lemma, we can characterize the behavior of the regularized Christoffel–Darboux polynomial $q_{\mu+\beta_d\mu_0,d}$ outside the support of μ:

Lemma 4 *Let $d \in \mathbb{N}$, $d > 1$ and $\mathbf{z} \in \mathbb{R}^p$. Recall that $\beta_d := 2^{3-\sqrt{d}}$. Assume that* $\operatorname{dist}(\mathbf{z}, \operatorname{supp}(\mu)) \geq \frac{\delta_0}{\sqrt{d}-1}$. *Then,*

$$\frac{2^{\sqrt{d}-3}}{m + m_0} \leq q_{\mu+\beta_d\mu_0,d}(\mathbf{z}). \tag{19}$$

Proof of Lemma 4 Let $d > 1$, $\mathbf{z} \in \mathbb{R}^p$, $\delta = \operatorname{dist}(\mathbf{z}, \operatorname{supp}(\mu))$. Let $d' \in \mathbb{N}$ and $t > 0$, arbitrary for the moment. Consider the affine map $T : \mathbf{w} \mapsto \frac{\mathbf{w}-\mathbf{z}}{\delta+\delta_0}$. Let q be the degree $2d'$ polynomial given as in Lemma 3 such that

$$\begin{aligned} &q(0) = 1,\ q(\mathbf{w}) \in [-1, 1] \text{ for all } \mathbf{w} \in B_1, \text{ and} \\ &|q(\mathbf{w})| \leq 2^{1-\frac{\delta d'}{\delta+\delta_0}} \text{ for all } \mathbf{w} \in B_1 \setminus B_{\frac{\delta}{\delta+\delta_0}}. \end{aligned} \tag{20}$$

Let $r = q \circ T$. The polynomial r satisfies

$$\begin{aligned} |r(\mathbf{z}')| &\leq 1, \qquad \forall \mathbf{z}' \in \operatorname{supp}(\mu + \mu_0), \\ r(\mathbf{z}') &\leq 2^{1-\frac{\delta d'}{\delta+\delta_0}}, \qquad \forall \mathbf{z}' \in \operatorname{supp}(\mu), \\ r(\mathbf{z}) &= 1. \end{aligned} \tag{21}$$

Using Lemma 2 and the fact that $r(\mathbf{z}) = 1$, we obtain

$$\left(\int_{\mathbb{R}^p} r^2(\mathbf{w}) d(\mu + \beta\mu_0)(\mathbf{w}) \right)^{-1} \leq q_{\mu+\beta\mu_0,2d'}(\mathbf{z}) \tag{22}$$

and

$$\left(\int_{\mathbb{R}^p} r^2(\mathbf{w}) d(\mu+\beta\mu_0)(\mathbf{w})\right)^{-1} \leq q_{\mu+\beta\mu_0,2d'+1}(\mathbf{z}) \tag{23}$$

From (21), we deduce

$$\int_{\mathbb{R}^p} r^2(\mathbf{w}) d(\mu+\beta\mu_0)(\mathbf{w}) \leq 2^{2-\frac{\delta(2d')}{\delta+\delta_0}} m + \beta m_0 \tag{24}$$

$$\int_{\mathbb{R}^p} r^2(\mathbf{w}) d(\mu+\beta\mu_0)(\mathbf{w}) \leq 2^{3-\frac{\delta(2d'+1)}{\delta+\delta_0}} m + \beta m_0 \tag{25}$$

Combining (22), (23) , (24) and (25), we obtain the following bounds

$$\begin{aligned} q_{\mu+\beta\mu_0,2d'}(\mathbf{z}) &\geq \left(2^{3-\frac{\delta(2d')}{\delta+\delta_0}} m + \beta m_0\right)^{-1}, \\ q_{\mu+\beta\mu_0,2d'+1}(\mathbf{z}) &\geq \left(2^{3-\frac{\delta(2d'+1)}{\delta+\delta_0}} m + \beta m_0\right)^{-1}. \end{aligned} \tag{26}$$

Recall that d' and β were arbitrary. Now we can choose $d' = \lfloor d/2 \rfloor$, $\beta = \beta_d$ in one of the identities in (26) (depending on the parity of d) to obtain

$$q_{\mu+\beta_d\mu_0,d}(\mathbf{z}) \geq \left(2^{3-\frac{\delta d}{\delta+\delta_0}} m + \beta_d m_0\right)^{-1} \geq \frac{2^{\sqrt{d}-3}}{m+m_0}, \tag{27}$$

where the last inequality follows because the right-hand side is strictly increasing as a function of δ and $\delta \geq \frac{\delta_0}{\sqrt{d}-1}$. This proves the desired result. □

Let us give two additional simple technical lemmas.

Lemma 5 *For any $r > 0$,*

$$\begin{aligned} \min_{x>0} \{\log(2)x - (2r)\log(x)\} = \; & (2r)\left(1 - \log\left(\frac{2r}{\log(2)}\right)\right) \\ \geq \; & (2r)\left(1 - \log(3r)\right). \end{aligned}$$

Proof of Lemma 5 A simple analysis shows that the minimum is attained at $x = \frac{2r}{\log(2)}$. The lower bound follows because $\frac{2}{\log(2)} \leq 3$. □

Lemma 6 *For any $d \in \mathbb{N}$, we have*

$$\frac{2^{\sqrt{d}-3}}{m+m_0} \geq \frac{\gamma_d}{1-\alpha}.$$

Proof of Lemma 6 Using the definition of γ_d in (14), we have

$$\log\left(\gamma_d \frac{8(m+m_0)}{1-\alpha}\right) = 2r(1-\log(3r)+\log(\sqrt{d})) \leq \log(2)\sqrt{d}, \tag{28}$$

where the inequality follows from Lemma 5 with $x = \sqrt{d}$. This proves the desired result. □

Proof of Theorem 5 (ii) We prove the result by contraposition. We have the following chain of implications for $\mathbf{z} \in \mathbb{R}^p$,

$$\begin{aligned} &\operatorname{dist}(\mathbf{z}, \operatorname{supp}(\mu)) \geq \frac{\delta_0}{\sqrt{d}-1} \\ &\Rightarrow \frac{2^{\sqrt{d}-3}}{m+m_0} \leq q_{\mu+\beta_d\mu_0,d}(\mathbf{z}) \\ &\Rightarrow \gamma_d/(1-\alpha) \leq q_{\mu+\beta_d\mu_0,d}(\mathbf{z}) \\ &\Rightarrow \gamma_d \leq q_d^{\alpha}(\mathbf{z}) \\ &\Rightarrow \mathbf{z} \notin S_d, \end{aligned} \tag{29}$$

where the first implication is from Lemma 4, the second implication is due to Lemma 6, the third implication is deduced from (12) and the last implication is from the definition of S_d in (15). □

4.2 Estimation of Functions

We now translate Theorem 5 in functional terms. Considering that $\mathbf{z}$ can be written as follows $\mathbf{z} = (\mathbf{x}, y)$, let us introduce a specific set which will be of interest throughout the proof:

$$I_d := \{\mathbf{x} \in X : \inf_{y \in Y} q_d^{\alpha}(\mathbf{x}, y) \geq \gamma_d\}. \tag{30}$$

Lemma 7 *Suppose that $d > 1$. Then, we have*

$$\int_{I_d} d\mathbf{x} \leq \frac{1+\alpha}{1-\alpha} \frac{8(m+m_0)(3r)^{2r} e^{\frac{p^2}{d}}}{p^p e^{2r-p} d^{r-p}}. \tag{31}$$

Proof of Lemma 7 For all $\mathbf{x} \in X$ and all $A \subset \mathbb{R}^p$ measurable, one has

$$\int_Y \mathbb{I}_A(\mathbf{x}, y)\delta_{f(\mathbf{x})}(dy) = \mathbb{I}_A(\mathbf{x}, f(\mathbf{x})) \tag{32}$$

and hence

$$\mu(A) = \int_Y \mathbb{I}_A(\mathbf{x}, y)\mathrm{d}\mu(\mathbf{x}, y) = \int_X \mathbb{I}_A(\mathbf{x}, f(\mathbf{x}))\mathrm{d}\mathbf{x} = \int_{I_A} \mathrm{d}\mathbf{x}, \tag{33}$$

where $I_A := \{\mathbf{x} \in X : (\mathbf{x}, f(\mathbf{x})) \in A\}$. One can see from (30) that $\mathbf{x} \in I_d$ implies that $(\mathbf{x}, f(\mathbf{x})) \notin S_d$ and hence $I_d \subset I_{S_d^c}$ where S_d^c denotes the complement of S_d given in (15). We deduce that

$$\int_{I_d} \mathrm{d}\mathbf{x} \leq \int_{I_{S_d^c}} \mathrm{d}\mathbf{x} = \mu(S_d^c) \tag{34}$$

and the result follows from Item (i) of Theorem 5. □

Remark 5 Letting $\tilde{I}_d := \{\mathbf{x} \in X : \inf_{y\in Y} q_d^\alpha(\mathbf{x}, y) \geq \gamma_d/2\}$ it can be seen using the exact same arguments that a bound on $\int_{\tilde{I}_d} \mathrm{d}\mathbf{x}$ holds similarly as in Lemma 7 with a multiplicative factor of 2. This can be used to handle the situation where the argmin in (11) is computed up to a precision of the order $\gamma_d/2$. See also Remark 2.

Thanks to Lemma 7, it is sufficient to prove the convergence of the approximated function in the set $I_d^c := X \setminus I_d$.

Proposition 1 *Let $d \in \mathbb{N}$, $d > 1$. Let $\alpha \in [0, 1)$ be as in Assumption 2 with $\beta_d = 2^{3-\sqrt{d}}$. Consider $f_{\beta_d,d}^\alpha$ as in (13) and let I_d be defined by (30). Then, for any $r > p$, we have*

$$\int_X |f(\mathbf{x}) - f_{\beta_d,d}^\alpha(\mathbf{x})|d\mathbf{x} \leq \int_{I_d^c} |f(\mathbf{x}) - f_{\beta_d,d}^\alpha(\mathbf{x})|d\mathbf{x} + \mathrm{diam}(Y)\frac{1+\alpha}{1-\alpha}\frac{8(m+m_0)(3r)^{2r}e^{\frac{p^2}{d}}}{p^p e^{2r-p} d^{r-p}}.$$

Proof of Proposition 1 Since y takes values in the compact set Y, it is clear that $f_{\beta_d,d}^\alpha \in \mathscr{L}^\infty(X)$. We have

$$\begin{aligned}
&\int_X |f(\mathbf{x}) - f_{\beta_d,d}^\alpha(\mathbf{x})|\mathrm{d}\mathbf{x} \\
&= \int_{X\setminus I_d} |f(\mathbf{x}) - f_{\beta_d,d}^\alpha(\mathbf{x})|\mathrm{d}\mathbf{x} + \int_{I_d} |f(\mathbf{x}) - f_{\beta_d,d}^\alpha(\mathbf{x})|\mathrm{d}\mathbf{x} \\
&\leq \int_{I_d^c} |f(\mathbf{x}) - f_{\beta_d,d}^\alpha(\mathbf{x})|\mathrm{d}\mathbf{x} + \|f - f_{\beta_d,d}^\alpha\|_{\mathscr{L}^\infty(I_d)} \int_{I_d} \mathrm{d}\mathbf{x} \\
&\leq \int_{I_d^c} |f(\mathbf{x}) - f_{\beta_d,d}^\alpha(\mathbf{x})|\mathrm{d}\mathbf{x} \\
&\quad + \mathrm{diam}(Y)\frac{1+\alpha}{1-\alpha}\frac{8(m+m_0)(3r)^{2r}e^{\frac{p^2}{d}}}{p^p e^{2r-p} d^{r-p}}
\end{aligned} \tag{35}$$

where we have used Lemma 7 for the last inequality. □

4.3 Proofs of the Main Theorems

We are now in position of proving Theorem 3. We start with the Lipschitz case, which is the simplest and conveys most of the ideas.

Proof of Theorem 3 (ii) Using Proposition 1 It remains to bound the term

$$\int_{I_d^c} |f(\mathbf{x}) - f^{\alpha}_{\beta_d,d}(\mathbf{x})| \mathrm{d}\mathbf{x}. \tag{36}$$

For any $\mathbf{x} \in X$ define

$$\mathbf{u}_d(\mathbf{x}) \in \mathrm{argmin}_{\mathbf{u}\in X} \left\| \left(\mathbf{x},\ f^{\alpha}_{\beta_d,d}(\mathbf{x})\right) - (\mathbf{u},\ f(\mathbf{u})) \right\|, \tag{37}$$

with an arbitrary choice in the case where the argmin is not unique. Note that by continuity, the graph of f is closed so that the minimum is attained. Using the definition of I_d in (30), the fact that $\mathbf{x} \in I_d^c$ implies that

$$\left(\mathbf{x},\ f^{\alpha}_{\beta_d,d}(\mathbf{x})\right) \in S_d. \tag{38}$$

Moreover, Theorem 5 implies that

$$\begin{aligned} |f^{\alpha}_{\beta_d,d}(\mathbf{x}) - f(\mathbf{u}_d(\mathbf{x}))| &\leq \frac{\delta_0}{\sqrt{d}-1}, \\ \|\mathbf{x} - \mathbf{u}_d(\mathbf{x})\| &\leq \frac{\delta_0}{\sqrt{d}-1}, \end{aligned} \tag{39}$$

where $\|\cdot\|$ denotes the usual Euclidean distance in $\mathbb{R}^{p-1}$. Therefore, using Lipschitz continuity of f, we have, for any $\mathbf{x} \in I_d^c$,

$$\begin{aligned} |f^{\alpha}_{\beta_d,d}(\mathbf{x}) - f(\mathbf{x})| &\leq |f^{\alpha}_{\beta_d,d}(\mathbf{x}) - f(\mathbf{u}_d(\mathbf{x}))| + |f(\mathbf{x}) - f(u_d(\mathbf{x}))| \\ &\leq \frac{\delta_0}{\sqrt{d}-1}(1+L). \end{aligned} \tag{40}$$

We deduce that

$$\int_{I_d^c} |f(\mathbf{x}) - f^{\alpha}_{\beta_d,d}(\mathbf{x})| \mathrm{d}\mathbf{x} \leq \mathrm{vol}(X) \frac{\delta_0}{\sqrt{d}-1}(1+L) \tag{41}$$

which concludes the proof. □

We now turn to case (i), starting with the pointwise convergence.

Proof of Theorem 3 (i) We rely on a slightly different use of Lemma 7. Choose $r = p+2$ and let

$$I := \{\mathbf{x} \in X : \forall d_0 \in \mathbb{N},\ \exists d \in \mathbb{N},\ d \geq d_0,\ \mathbf{x} \in I_d\} = \cap_{d_0 \in \mathbb{N}} \cup_{d \geq d_0} I_d. \tag{42}$$

Lemma 7 ensures that $\mathrm{vol}(I_d) = O(1/d^2)$ so that

$$\mathrm{vol}\left(\cup_{d \geq d_0} I_d\right) \leq \sum_{d \geq d_0} \mathrm{vol}(I_d) \quad \underset{d_0 \to \infty}{\to} 0.$$

We have $\mathrm{vol}(I) = \lim_{d_0 \to \infty} \left(\cup_{d \geq d_0} I_d\right) = 0$. This means that we have the two following properties, for almost every $\mathbf{x} \in X$:

- f is continuous at $\mathbf{x}$ (by assumption),
- $\exists d_0 \in \mathbb{N}, \forall d \in \mathbb{N}, d \geq d_0, \mathbf{x} \notin I_d$ (because $\mathrm{vol}(I) = 0$).

Fix any such $\mathbf{x}$ and for any $d \geq d_0$ consider

$$(\mathbf{u}_d, v_d) \in \mathrm{argmin}_{\mathbf{u} \in X, v \in Y} \left\| \left(\mathbf{x},\ f^{\alpha}_{\beta_d,d}(\mathbf{x})\right) - (\mathbf{u},\ v) \right\|, \qquad \text{s.t.} \qquad (\mathbf{u}, v) \in \mathrm{supp}(\mu) \tag{43}$$

with an arbitrary choice when the argmin is not unique. Note that the support of μ is actually the closure of the graph of f so that the minimum is attained. Using the definition of I_d in (30), we have that $\mathbf{x} \in I_d^c$ for all $d \geq d_0$ which implies that

$$\left(\mathbf{x},\ f^{\alpha}_{\beta_d,d}(\mathbf{x})\right) \in S_d, \tag{44}$$

and Theorem 5 implies that

$$\begin{aligned} |f^{\alpha}_{\beta_d,d}(\mathbf{x}) - v_d| &\leq \frac{\delta_0}{\sqrt{d}-1}, \\ \|\mathbf{x} - \mathbf{u}_d\| &\leq \frac{\delta_0}{\sqrt{d}-1}. \end{aligned}$$

Since $(\mathbf{u}_d, v_d) \in \mathrm{supp}(\mu)$ and $\mathrm{supp}(\mu)$ is the closure of the graph of f, there exists $\mathbf{h}_d \in X$, such that

$$\begin{aligned} |f^{\alpha}_{\beta_d,d}(\mathbf{x}) - f(\mathbf{h}_d)| &\leq \frac{2\delta_0}{\sqrt{d}-1}, \\ \|\mathbf{x} - \mathbf{h}_d\| &\leq \frac{2\delta_0}{\sqrt{d}-1}. \end{aligned} \tag{45}$$

This concludes the proof of pointwise convergence since

$$|f^{\alpha}_{\beta_d,d}(\mathbf{x}) - f(\mathbf{x})| \leq |f^{\alpha}_{\beta_d,d}(\mathbf{x}) - f(\mathbf{h}_d)| + |f(\mathbf{x}) - f(\mathbf{h}_d)|$$

and both terms tend to 0 as $d \to \infty$, using (45) and continuity of f at $\mathbf{x}$.

Convergence in $\mathscr{L}^1$ follows from Egorov's Theorem (see, e.g., [33, chapter 18]): for any $\epsilon > 0$, there exists $S_\epsilon \subset X$, measurable, of Lebesgue measure smaller than ϵ such that $f^\alpha_{\beta_d,d} \to f$ uniformly on $X \setminus S_\epsilon$. We have

$$\begin{aligned}
\|f^\alpha_{\beta_d,d} - f\|_{\mathscr{L}^1(X)} &= \int_X |f^\alpha_{\beta_d,d}(\mathbf{x}) - f(\mathbf{x})| \mathrm{d}\mathbf{x} \\
&= \int_{S_\epsilon} |f^\alpha_{\beta_d,d}(\mathbf{x}) - f(\mathbf{x}) \mathrm{d}\mathbf{x} + \int_{X\setminus S_\epsilon} |f^\alpha_{\beta_d,d}(\mathbf{x}) - f(\mathbf{x})| \mathrm{d}\mathbf{x} \\
&\leq \mathrm{vol}(S_\epsilon)\,\mathrm{diam}(Y) + \mathrm{vol}(X)\|f^\alpha_{\beta_d,d} - f(\mathbf{x})\|_{\mathscr{L}^\infty(X\setminus S_\epsilon)} \\
&\leq \epsilon\,\mathrm{diam}(Y) + \mathrm{vol}(X)\|f^\alpha_{\beta_d,d} - f(\mathbf{x})\|_{\mathscr{L}^\infty(X\setminus S_\epsilon)}.
\end{aligned}$$

By uniform convergence the second term goes to 0 as $d \to \infty$, this shows that

$$\lim_{d\to\infty} \sup \|f^\alpha_{\beta_d,d} - f\|_{\mathscr{L}^1(X)} \leq \epsilon\,\mathrm{diam}(Y).$$

Moreover, since $\epsilon > 0$ was arbitrary, the limit is 0. □

We now turn to the proof of convergence in $\mathscr{L}^1$ norm for univariate functions of bounded variation.

Lemma 8 *Let $f : \mathbb{R} \mapsto \mathbb{R}$ be such that $V(f)$ is finite and f vanishes outside a segment I. Let $a, b > 0$ be positive constants. Let*

$$J = \{t \in I : \exists u, |t-u| \leq b, u \in I, |f(u) - f(t)| > a\}.$$

Then,

$$\int_J dx \leq 2V(f)\frac{b}{a}.$$

Proof of Lemma 8 This is a packing argument. Let $J_0 = J$ and follow the recursive process, for $k \in \mathbb{N}, k \geq 1$,

$$\begin{cases}
\text{if} & J_{k-1} \cap J \neq \emptyset \\
& \text{let } t_k \in J_{k-1} \cap J,\ u_k \in J,\ |t_k - u_k| \leq b \\
& \text{let } J_k = J_{k-1} \setminus [t_k - b, t_k + b] \\
\text{otherwise} & \text{stop.}
\end{cases}$$

This process must stop after a finite number of iterations. Indeed, the set $I_k = I \setminus \cup_{i=1}^k [t_k - b, t_k + b]$ consists of a finite union of intervals. At iteration k, either one of these intervals is a subset of $[t_{k+1} - b, t_{k+1} + b]$ and then it is removed entirely from I_k, or otherwise t_{k+1} is contained in an interval which contains either $[t_{k+1} - b, t_{k+1}]$

or $[t_{k+1}, t_{k+1}+b]$ and the Lebesgue measure of I_{k+1} is reduced by at least b compared to I_k, possibly creating a new interval.

Let K be the last iteration, so that $J_K \cap J = \emptyset$. By the iterative process, at each step, the measure of J_k is reduced by at most $2b$ and we have

$$0 = \int_{J_K \cap J} dx \geq \int_{J_{K-1} \cap J} dx - 2b \geq \ldots \geq \int_{J_0 \cap J} dx - 2Kb$$

so that

$$\int_J dx \leq 2Kb.$$

Finally, since the intervals $[t_k, u_k]$, $k = 1, \ldots, K$ are disjoint, we have

$$V(f) \geq \sum_{i=1}^{K} |f(t_k) - f(u_k)| \geq Ka \geq \frac{a}{2b} \int_J dx,$$

which proves the desired result. □

Proof of Theorem 4 For any $d > 1$ consider

$$J_d = \left\{ x \in X : \exists u \in X, |x-u| \leq 2\delta_0/(\sqrt{d}-1), |f(u) - f(t)| > d^{-1/4} \right\}.$$

By Lemma 8, for any $d > 1$ we have

$$\mathrm{vol}(J_d) \leq \frac{4\delta_0 d^{1/4} V(f)}{\sqrt{d}-1}. \tag{46}$$

Choose any $d > 1$ and any x such that $x \notin I_d$ and $x \notin J_d$. Consider

$$(u_d, v_d) \in \mathrm{argmin}_{u \in X, v \in Y} \left\| \left(x,\ f^{\alpha}_{\beta_d, d}(x)\right) - (u,\ v) \right\| \quad \text{s.t.} \quad (u,\ v) \in \mathrm{supp}(\mu) \tag{47}$$

with an arbitrary choice when the argmin is not unique.

Note that the support of μ is actually the closure of the graph of f so that the minimum is attained. Using the definition of I_d in (30), $x \in I_d^c$ implies that

$$\left(x,\ f^{\alpha}_{\beta_d, d}(x)\right) \in S_d \tag{48}$$

and Theorem 5 implies that

$$|f^{\alpha}_{\beta_d, d}(x) - v_d| \leq \frac{\delta_0}{\sqrt{d}-1},$$

$$|x - u_d| \leq \frac{\delta_0}{\sqrt{d}-1}.$$

Since $(u_d, v_d) \in \mathrm{supp}(\mu)$ and $\mathrm{supp}(\mu)$ is the closure of the graph of f, there exists $h_d \in X$, such that

$$\begin{aligned} |f^{\alpha}_{\beta_d,d}(x) - f(h_d)| &\leq \frac{2\delta_0}{\sqrt{d}-1}, \\ |x - h_d| &\leq \frac{2\delta_0}{\sqrt{d}-1}. \end{aligned} \tag{49}$$

Now since $x \notin J_d$ and $|x - h_d| \leq \frac{2\delta_0}{\sqrt{d}-1}$, we have $|f(x) - f(h_d)| \leq d^{-1/4}$. This entails

$$\begin{aligned} |f^{\alpha}_{\beta_d,d}(x) - f(x)| &\leq |f^{\alpha}_{\beta_d,d}(x) - f(h_d)| + |f(x) - f(h_d)| \\ &\leq \frac{2\delta_0}{\sqrt{d}-1} + d^{-1/4}. \end{aligned}$$

The latter expression does not depend on x which was arbitrarily chosen outside of I_d and J_d. We deduce that

$$\begin{aligned} &\|f - f^{\alpha}_{\beta_d,d}\|_{\mathscr{L}^1(X)} \\ &\quad \leq: \mathrm{vol}(X)\left(\frac{2\delta_0}{\sqrt{d}-1} + d^{-1/4}\right) + \mathrm{diam}(Y)\,(\mathrm{vol}(I_d) + \mathrm{vol}(J_d)) \end{aligned}$$

and the result follows by invoking Lemma 7 and using Inequality (46). □

5 Numerical Examples

5.1 Computational Tractability

Working with a large class of approximation functions may pose computational difficulties. An advantage of our Christoffel–Darboux semi-algebraic approximant is that it can be computed efficiently.

If the input moments are exactly known and given in rational form, the Christoffel–Darboux polynomial to be partially minimized is obtained through formal inversion of the moment matrix. This operation has efficient implementations, namely polynomial time algorithms over rational entries, an example is given in [2].

In most of the applications, the moments are, however, known only approximately, and the Christoffel–Darboux polynomial is constructed via the numerical eigenvalue decomposition of the approximate moment matrix. Since the moment matrix is symmetric, its eigenvalue decomposition can be computed efficiently with numerically stable algorithms in floating point arithmetic [13].

In addition, the computational overhead of evaluating the semi-algebraic approximant at a given point $\mathbf{x}$ is that of minimizing a univariate polynomial over the segment $[-1, 1]$. The Lipschitz constant of a univariate polynomial with coefficients $\mathbf{p} = (p_0, \ldots, p_{2d})$ over $[-1, 1]$ is at most $\|\mathbf{p}\|_1$, and hence, grid search finds an ϵ-accurate solution to this problem using $\frac{2\|\mathbf{p}\|_1}{\epsilon}$ evaluations. In our case the entries of $\mathbf{p}$ are polynomials in $\mathbf{x}$ which are deduced from moment data so that for a fixed d estimating and evaluating our semi-algebraic approximant up to a fixed arbitrary precision with rational inputs (moment matrix and $\mathbf{x}$) can be done in polynomial time. Note also that our analysis shows that a level of precision of the order d^{-p-2} is sufficient so that the cost of the overall procedure has a complexity which is polynomial in the bit size of the moment matrix, the target evaluation point $\mathbf{x}$, as well as in d, the degree bound.

5.2 Prototype Code

We provide a simple MATLAB prototype to validate our algorithm. All the examples described below are reproducible, and the MATLAB scripts can be found at

`homepages.laas.fr/henrion/software/momgraph`

The calling syntax of the main routine is

```
[Y,P] = momgraph(M,X)
```

It takes as an input an approximate moment matrix

$$\mathbf{M} = \int_X \mathbf{b}(\mathbf{x}, f(\mathbf{x}))\mathbf{b}(\mathbf{x}, f(\mathbf{x}))^\top \mathrm{d}\mathbf{x}$$

(in MATLAB floating point format) for $\mathbf{b}$ the monomial basis vector (in grevlex ordering), and a collection of points

$$\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\} \subset X \subset [-1, 1]^{p-1}$$

(in MATLAB floating point format) with $p > 1$. It outputs an approximation

$$\{y_1, y_2, \ldots, y_N\} \subset Y := [-1, 1]$$

(in MATLAB floating point format) of the values $\{f(\mathbf{x}_1), f(\mathbf{x}_2), \ldots, f(\mathbf{x}_N)\}$, as well as a matrix $\mathbf{P}$ of coefficients (in MATLAB floating point format) of the vector of polynomials $\mathbf{p}$ whose sum of squares yields the Christoffel–Darboux polynomial.

Our implementation is straightforward, not optimized for efficiency. The regularization parameter β is set to the default value of 10^{-8}, and the Christoffel–Darboux polynomial is computed from the eigenvalue decomposition (MATLAB's command `eig`) of the approximate moment matrix $\mathbf{M} + \beta\mathbf{I}$.

5.3 Sign Function

Consider the measure supported on the graph of the sign function $f(\mathbf{x}) = \mathrm{sign}(\mathbf{x})$ whose moments in the monomial basis on $X := [-1, 1]$ are

$$\int_{-1}^{1} \mathbf{x}^{a_1} f(\mathbf{x})^{a_2} \mathrm{d}\mathbf{x} = (-1)^{a_2} \int_{-1}^{0} \mathbf{x}^{a_1} \mathrm{d}\mathbf{x} + (1)^{a_2} \int_{0}^{1} \mathbf{x}^{a_1} \mathrm{d}\mathbf{x}$$
$$= \frac{(-1)^{a_2}(0^{a_1+1} - (-1)^{a_1+1}) + 1 - 0^{a_1+1}}{a_1 + 1}$$

for $(a_1,\ a_2) \in \mathbb{N}^2$. Here is an example of the use of `momgraph` to recover the sign function from the (floating point approximations) of the (exact) moments:

```
>> M

M =

    2.0000         0         0    0.6667    1.0000    2.0000
         0    0.6667    1.0000         0         0         0
         0    1.0000    2.0000         0         0         0
    0.6667         0         0    0.4000    0.5000    0.6667
    1.0000         0         0    0.5000    0.6667    1.0000
    2.0000         0         0    0.6667    1.0000    2.0000

>> X = linspace(-1,1,1e3)';
>> [Y,P] = momgraph(M,X);
>> plot(X,graph(X),'-r','linewidth',6); hold on;
>> plot(X,Y,'-k','linewidth',3); xlabel('x'); ylabel('y');
```

This code corresponds to a degree 4 approximation from a moment matrix of size 6 with 15 moments. A degree 2 approximation can be obtained from its 3 by 3 submatrix

```
>> M(1:3,1:3)

ans =

    2.0000         0         0
         0    0.6667    1.0000
         0    1.0000    2.0000
```

In Fig. 4 we see that the resulting degree 4 semi-algebraic approximation cannot be distinguished from the graph of the sign function. Our semi-algebraic approximant is $\mathbf{x} \mapsto \mathrm{argmin}_{y \in [-1,1]} q(\mathbf{x}, y)$ with q the Christoffel–Darboux polynomial constructed as the sum of squares of the polynomials returned by the `momgraph` function:

```
>> mpol x y; b = mmon([x y],2);
>> P*b

6-by-1 polynomial vector
```

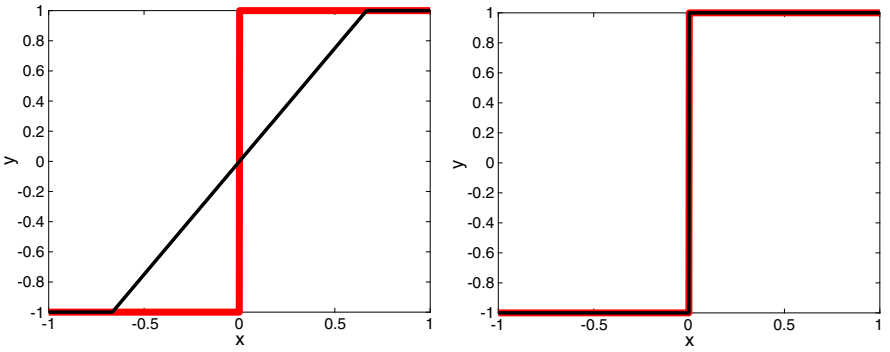

Fig. 4 Graph of the sign function (red) and its degree 2 (left) and degree 4 (right) semi-algebraic approximations (black) (Color figure online)

```
(1,1):7071.0678-7071.0678y^2
(2,1):0.86713+9.4x^2-9.7305xy+0.86713y^2 (3,1):2.4315x-1.3013y
(4,1):-0.53517+1.2757x^2+1.137xy-0.53517y^2 (5,1):0.29635x+0.55374y
(6,1):0.29443+0.10761x^2+0.15643xy+0.29443y^2
```

We see in particular that the first polynomial is $(\mathbf{x}, y) \mapsto 1 - y^2$ with a large scaling factor. This polynomial vanishes on the graphs of the functions $y \mapsto -1$ and $y \mapsto 1$. The other polynomials are instrumental to determining which one of the two graphs corresponds to a given value of $\mathbf{x}$.

5.4 Discontinuous Functions

Let us revisit the discontinuous univariate examples of [11]. Since in this case we do not have the analytic moments of the measure supported on the graph of the function f to input to our algorithm, we use the empirical moment matrix computed by uniform sampling, i.e.,

$$M = \frac{1}{N} \sum_{k=1}^{N} \mathbf{b}(\mathbf{x}_k, f(\mathbf{x}_k)) \mathbf{b}(\mathbf{x}_k, f(\mathbf{x}_k))^\top \tag{50}$$

for N sufficiently large, i.e., 10^3, and $\mathbf{b}$ the monomial basis vector. Degree 10 semi-algebraic approximations are reported in Fig. 5 for three benchmarks [11, Examples 65, 66, 67] of discontinuous functions f, appropriately scaled in $X = Y = [-1, 1]$.

We observe that the second rightmost discontinuity in the middle example is not detected. Increasing the degree of the approximations does not fix the issue, and we believe that it is due to the poor resolution of the monomial basis. It would be more appropriate to use here a complex exponential basis (i.e., Fourier coefficients) or an orthogonal basis (e.g., Chebyshev or Legendre polynomials).

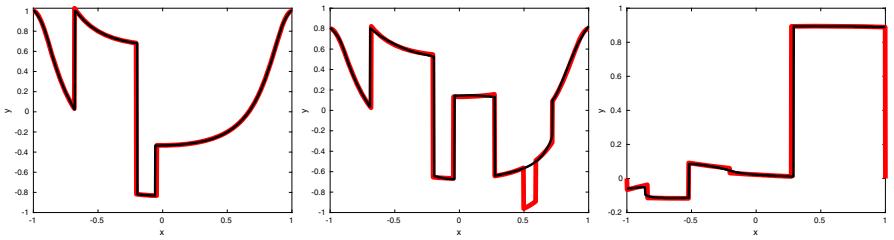

Fig. 5 Degree 10 semi-algebraic approximations (black) for the discontinuous univariate functions (red) of Examples 65 (left), 66 (middle) and 67 (right) of reference [11] (Color figure online)

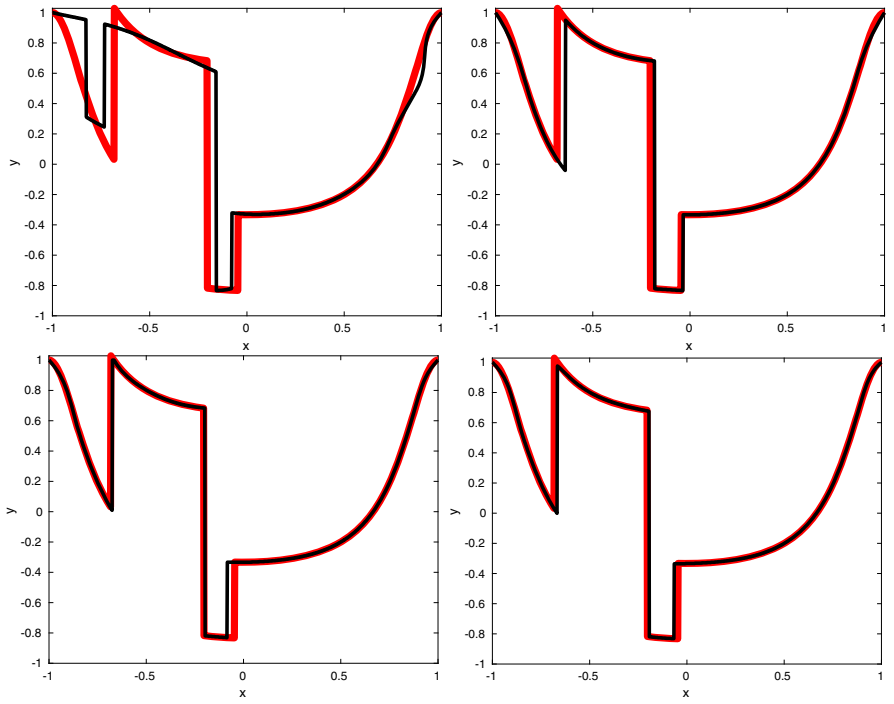

Fig. 6 Degree 10 semi-algebraic approximations (black) of a discontinuous function (red) computed from empirical moments evaluated at 10 (upper left), 20 (upper right), 30 (lower left) and 40 (lower right) uniformly distributed samples (Color figure online)

5.5 Interpolation

Suppose now that we have access only to the values $\{f(\mathbf{x}_k)\}_{k=1,\ldots,N}$ of the function to be approximated at given sampling points $\{\mathbf{x}_k\}_{k=1,\ldots,N}$, for N small. Our algorithm takes as input the empirical moment matrix (50). In Fig. 6, we revisit [11, Example 65] to study the effect of the number of samples N on the quality of the approximation, for a uniform distribution of samples. We see that with 20 samples the function is already well approximated.

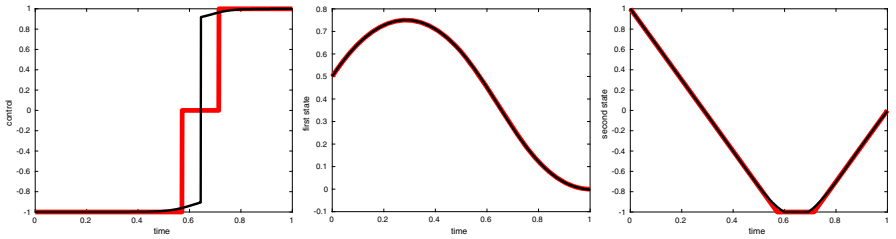

Fig. 7 Minimum time double integrator with state constraint: control (left), first state (middle) and second state (right) trajectories (red) and their degree 8 semi-algebraic approximations (black) constructed from the pseudo-moments of the occupation measure (Color figure online)

5.6 Recovering Trajectories for Optimal Control

In [23], the moment-SOS hierarchy is applied to solve numerically nonlinear optimal control of ODEs with polynomial data and semi-algebraic state and control constraints. Nonlinear optimal control is formulated as a linear problem on moments of occupation measures supported on optimal trajectories. Let us show how numerical approximations of these moments obtained by semi-definite programming can be input to our algorithm to approximate optimal state and control trajectories.

Let us revisit the state-constrained double integrator problem of [23, Section 5.1] to approximate the time optimal trajectories. After a scaling of time, state and control, we use the MATLAB interface GloptiPoly 3 and the conic solver MOSEK to compute the pseudo-moments of the occupation measure of degree up to 8. This can be achieved in less than 2 seconds on a standard desktop computer. From this output, we construct the 45-by-45 moment matrices of the control and state marginals, conditioned w.r.t. time. Using our notations, the independent variable $\mathbf{x}$ is time, while the dependent variable y is, respectively, the control, the first state and the second state. For this example, the analytic trajectories are available for comparison.

We see in Fig. 7 that the state trajectory approximations are tight, whereas the control trajectory approximation misses partly the central region corresponding to the saturation of the second state. Indeed, since it is obtained by solving numerically the degree 8 semi-definite relaxation of the moment-SOS hierarchy, the approximated moment matrix differs from the exact moment matrix, and this has an impact on the quality of the Christoffel–Darboux approximation.

For this example, we can construct analytically the exact moment matrix of the control trajectory and observe that indeed its Christoffel–Darboux semi-algebraic approximation of degree 8 identifies well the optimal control trajectory switching times, see Fig. 8.

5.7 Bivariate Examples

Consider the indicator function

$$f(\mathbf{x}) := \mathbb{I}_{\{\mathbf{x}\in\mathbb{R}^2 : \mathbf{x}_1^2+\mathbf{x}_2^2\le 1/4\}}(\mathbf{x})$$

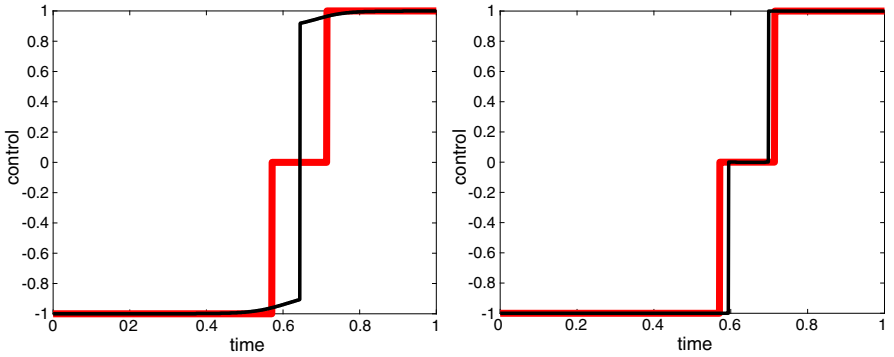

Fig. 8 Minimum time double integrator with state constraint: control trajectories (red) and their degree 8 semi-algebraic approximations (black) constructed from the pseudo-moments (left, same as left of Fig. 7) and from the analytic moments (right) of the occupation measure (Color figure online)

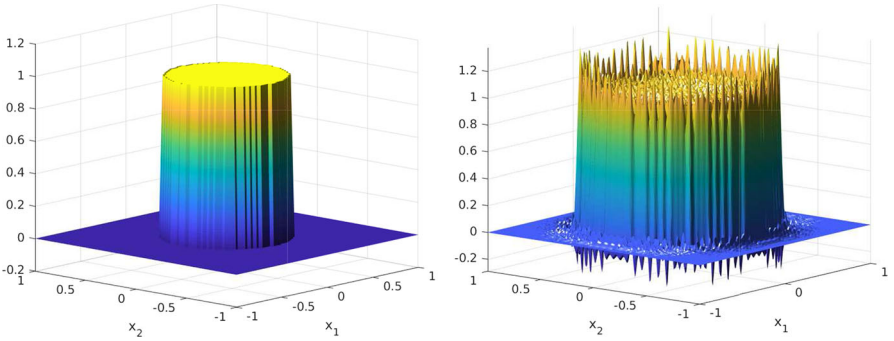

Fig. 9 Degree 4 (left) semi-algebraic approximation, and Chebyshev polynomial approximation (right) of the indicator function of a disk

of a centered disk of radius $1/2$. We compute the empirical moments obtained by sampling 100^2 points on a uniform grid of $X := [-1, 1]^2$. With this input, our algorithm computes the degree 8 semi-algebraic approximation reported in Fig. 9, to be compared with the Chebyshev polynomial approximation obtained from 100^2 points by the `chebfun2` command, showing the typical Gibbs phenomenon.

We perform the same computations for the piecewise constant function

$$f(\mathbf{x}) := \mathbb{I}_{\{\mathbf{x}\in\mathbb{R}^2:\mathbf{x}_1^2+\mathbf{x}_2^2\le 1/4\}}(\mathbf{x}) - \frac{1}{2}\mathbb{I}_{\{\mathbf{x}\in\mathbb{R}^2:(\mathbf{x}_1+\frac{1}{2})^2+(\mathbf{x}_2+\frac{1}{2})^2\le 1/4\}}(\mathbf{x})$$

obtained as a superposition of signed indicator functions of two disks. Its degree 8 and 16 semi-algebraic approximations are reported in Fig. 10. The absolute pointwise error between the approximations and the original function is displayed in Fig. 11.

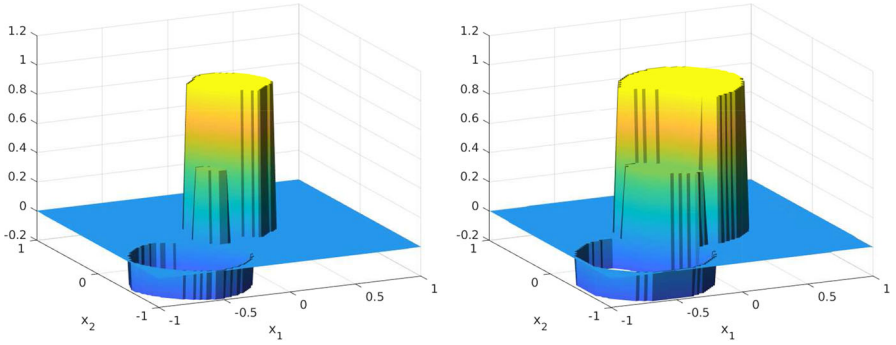

Fig. 10 Degree 8 (left) and degree 16 (right) semi-algebraic approximations of the superposition of signed indicator functions of two disks

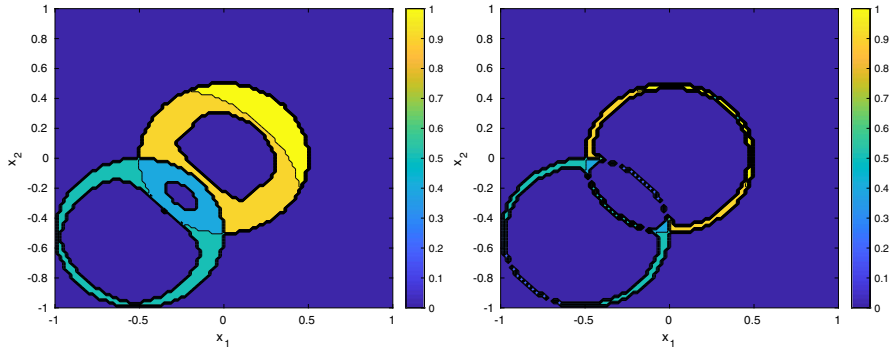

Fig. 11 Contour plots of the absolute error between the two disk indicator function and its degree 8 (left) and degree 16 (right) semi-algebraic approximations of Fig. 10

5.8 Discontinuous Solutions of Nonlinear PDEs

In [26], the moment-SOS hierarchy is applied to solve numerically a class of nonlinear PDEs for which we known that classical (i.e., differentiable) solutions do not exist. The advantage of optimizing over occupation measures is that they can be supported on graphs of weak (i.e., possibly discontinuous) solutions. Let us show how approximate moments of these measures computed by semi-definite programming can be processed by our algorithm so as to recover these discontinuous solutions.

We focus on the Burgers equation and choose the initial data (a function of one space coordinate, at time zero) in a way that at a given time a shock appears, i.e., the solution becomes a discontinuous function of the space coordinate. Once the shock appeared, it propagates through, i.e., the discontinuity remains but its location varies. In Figs. 12, 13 and 14, we show the graphs obtained from the moment relaxations proposed in [26]. In all cases we use the 969 triviate moments of degree 16 of the occupation measure (supported on time, space and solution) to recover the graph of the approximated solution. For comparison we also sketch the analytic solution with red lines.

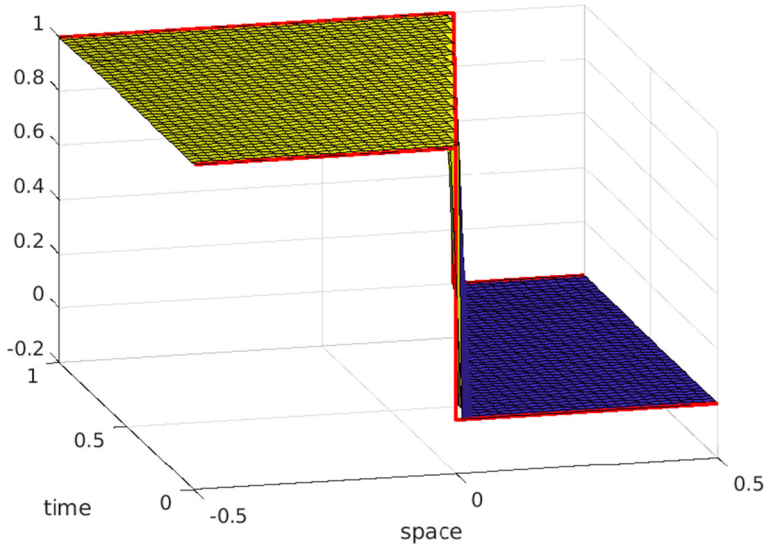

Fig. 12 Graph of the solution (a function of time and space) recovered from approximate moments for the Burgers PDE: Discontinuous initial data. The shock propagates linearly with time

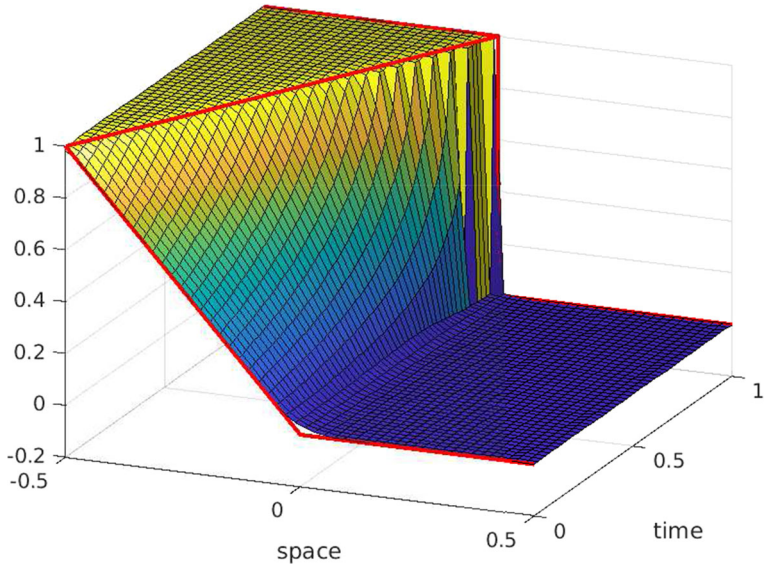

Fig. 13 Graph of the solution (a function of time and space) recovered from approximate moments for the Burgers PDE: Initial condition chosen to produce a shock at final time

For the graphs in Figs. 12 and 13, the approximated moments match the analytic moments up to an error of the order of 10^{-8}. Our semi-algebraic approximations are almost identical to the analytic solution.

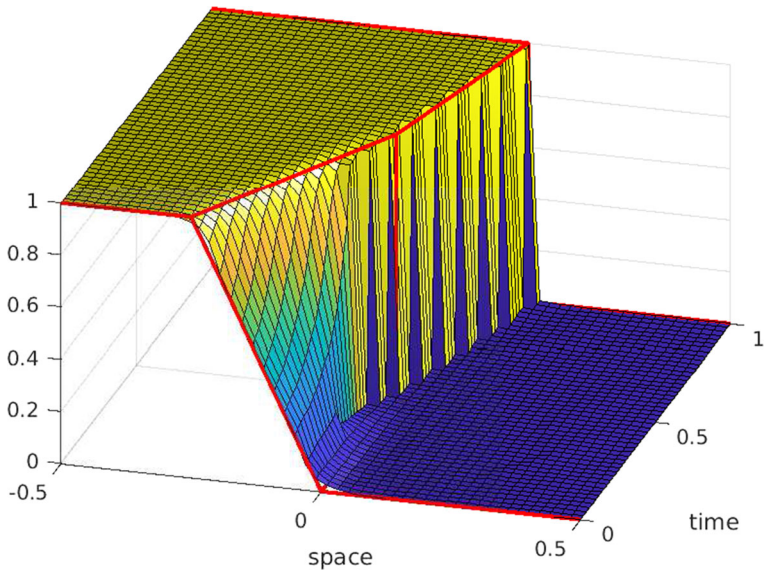

Fig. 14 Graph of the solution (a function of time and space) recovered from approximate moments for the Burgers PDE: Initial function chosen such that the shock occurs at $t = \frac{1}{2}$

For the graph in Fig. 14, the approximated moments are noticeably incorrect, i.e., the error is of order 10^{-4}. Nevertheless, our semi-algebraic approximation is able to reproduce the graph of the solution quite accurately. In particular the propagation of the shock is retrieved from the moment data. However, the approximation is erroneous when the solution passes over from its continuous to its discontinuous part.

6 Conclusion

In this paper, we describe a new technique to estimate discontinuous functions from moment data, based on Christoffel–Darboux kernels. Instead of using polynomial or piecewise polynomial approximants, we use a class of semi-algebraic approximants, namely arguments of minima of polynomials. This is another occurrence of a lifting technique: instead of using only moments depending linearly on the function so as to recover directly the function, we use also moments depending nonlinearly on the function so as to approximate the support of a measure concentrated on the graph of the function. We provide functional analytic and geometric convergence proofs. Finally, some numerical examples illustrate the efficiency of our algorithm.

We believe that this work opens the way to many other further research lines:

- When applying the moment-SOS hierarchy, the moments are numerical approximations of the real ones. It would be interesting to provide a sensibility analysis of the application of our algorithm for the real moments and the approximated ones. We believe that such an analysis can be performed, since promising results were

achieved recently in [19] for the case of zero dimensional manifolds, i.e., unions of finitely many points.

- It could also be interesting to investigate in a more quantitative way why the Gibbs phenomenon might be avoided or at least attenuated with the technique we provide. We believe that it is mainly due to the semi-algebraic point of view we are following.
- We could also check whether our algorithm works as well when considering only the knowledge of Fourier coefficients, namely moments depending linearly on the function that we want to approximate. In many problems, this is the only measurement that we might have. This is therefore a partial moment information, and we may want to complement it with estimates of higher degree moments. This makes the problem challenging.

Acknowledgements We are grateful to Quentin Vila for his technical input on discontinuous solutions of PDEs and to Milan Korda, Victor Magron and Matteo Tacchi for interesting discussions. This work was partly funded by the ERC Advanced Grant Taming and was also conducted in the framework of the regional program "Atlanstic 2020, Research, Education and Innovation in Pays de la Loire", supported by the French Region Pays de la Loire and the European Regional Development Fund. E. Pauwels and J.B. Lasserre are also partially supported by the AI Interdisciplinary Institute ANITI funding through the French program "Investing for the Future PIA3", under the Grant agreement number ANR-19-PI3A-0004.

References

1. Ambrosio, L., Fusco, N., Pallara, D.: Functions of bounded variation and free discontinuity problems. Clarendon Press, Oxford (2000)
2. Bareiss, E.H.: Sylvester's identity and multistep integer-preserving Gaussian elimination. Math. Comput. **22**(103), 565–578 (1968)
3. Batenkov, D., Yomdin, Y.: Algebraic Fourier reconstruction of piecewise smooth functions. Math. Comput. **81**(277), 277–318 (2012)
4. Batenkov, D.: Complete algebraic reconstruction of piecewise-smooth functions from Fourier data. Math. Comput. **84**(295), 2329–2350 (2015)
5. Coste, M.: An introduction to semialgebraic geometry. RAAG Netw. Sch. **145**, 30 (2002)
6. De Marchi, S., Sommariva, A., Vianello, M.: Multivariate Christoffel functions and hyperinterpolation. Dolomites Res. Notes Approx. **7**, 26–33 (2014)
7. De Vito, E., Rosasco, L., Toigo, A.: Learning sets with separating kernels. Appl. Comput. Harmon. Anal. **37**(2), 185–217 (2014)
8. DiPerna, R.J.: Measure-valued solutions to conservation laws. Arch. Ration. Mech. Anal. **88**(3), 223–270 (1985)
9. Driscoll, T.A., Hale, N., Trefethen, L.N. (eds.): Chebfun Guide. Pafnuty Publications, Oxford (2014)
10. Dunkl, C.F., Xu, Y.: Orthogonal Polynomials of Several Variables. Cambridge University Press, Cambridge (2014)
11. Eckhoff, K.S.: Accurate and efficient reconstruction of discontinuous functions from truncated series expansions. Math. Comput. **61**(204), 745–763 (1993)
12. Fattorini, H.O.: Infinite Dimensional Optimization and Control Theory. Cambridge University Press, Cambridge (1999)
13. Golub, G.H., Van Loan, C.F.: Matrix Computations, 3rd edn. The Johns Hopkins University Press, London (1996)
14. Gottlieb, D., Shu, C.-W.: On the Gibbs phenomenon and its resolution. SIAM Rev. **39**(4), 644–668 (1997)
15. Gustafsson, B., Putinar, M., Saff, E.B., Stylianopoulos, N.: Bergman polynomials on an archipelago: estimates, zeros and shape reconstruction. Adv. Math. **222**(4), 1405–1460 (2009)

16. Helmes, K., Röhl, S., Stockbridge, R.H.: Computing moments of the exit time distribution for Markov processes by linear programming. Oper. Res. **49**, 516–530 (2001)
17. Hernández-Hernández, D., Hernández-Lerma, O., Taksar, M.: The linear programming approach to deterministic optimal control problems. Appl. Math. **24**, 17–33 (1996)
18. Hernández-Lerma, O., Lasserre, J.B.: The Linear Programming Approach. Handbook of Markov Decision Processes, pp. 377–407. Springer, Boston (2002)
19. Klep, I., Povh, J., Volčič, J.: Minimizer extraction in polynomial optimization is robust. SIAM J. Optim. **28**(4), 3177–3207 (2018)
20. Kroó, A., Lubinsky, D.S.: Christoffel functions and universality in the bulk for multivariate orthogonal polynomials. Can. J. Math. **65**(3), 600–620 (2012)
21. Lasserre, J.B.: Moments, Positive Polynomials and Their Applications. World Scientific, Singapore (2010)
22. Lasserre, J.B.: The moment-SOS hierarchy. Proc. Int. Congress Math. (ICM) **4**, 3773–3794 (2018)
23. Lasserre, J.B., Henrion, D., Prieur, C., Trélat, E.: Nonlinear optimal control via occupation measures and LMI relaxations. SIAM J. Control Optim. **47**(4), 1643–1666 (2008)
24. Lasserre, J.B., Prieto-Rumeau, T., Zervos, M.: Pricing a class of exotic options via moments and SDP relaxations. Math. Finance **16**, 469–494 (2006)
25. Lasserre, J.B., Pauwels, E.: The empirical Christoffel function with applications in data analysis. Adv. Comput. Math. **45**, 1439–1468 (2019)
26. Marx, S., Weisser, T., Henrion, D., Lasserre, J.B.: A moment approach for entropy solutions to nonlinear hyperbolic PDEs. Math. Control. Related Fields **10**, 113–140 (2019)
27. Máté, A., Nevai, P.G.: Bernstein's inequality in L^p for $0<p<1$ and $(C,1)$ bounds for orthogonal polynomials. Ann. Math. **111**, 145–154 (1980)
28. Nevai, P.G., Freud, Géza: orthogonal polynomials and Christoffel functions. A case study. J. Approx. Theory **48**(1), 3–167 (1986)
29. Nie, J., Schweighofer, M.: On the complexity of Putinar's Positivstellensatz. J. Complex. **23**(1), 135–150 (2007)
30. Pachón, R., Platte, R.B., Trefethen, L.N.: Piecewise-smooth chebfuns. IMA J. Numer. Anal. **30**(4), 898–916 (2010)
31. Lasserre, J.B., Pauwels, E.: Sorting out typicality with the inverse moment matrix sos polynomial. In: Advances in Neural Information Processing Systems, pp. 190–198 (2016)
32. Pauwels, E., Putinar, M., Lasserre, J.B.: Data analysis from empirical moments and the Christoffel function. Found. Comput. Math. (**to appear**) (2020)
33. Royden, H., Fitzpatrick, P.: Real Analysis. 4th Edition, Pearson Education, (2010)
34. Santambrogio, F.: Optimal Transport for Applied Mathematicians-Calculus of Variations, PDEs, and Modeling. Birkhäuser, Basel (2015)
35. Silvestre, L.: Oscillation properties of scalar conservation laws. Commun. Pure Appl. Math. **72**(6), 1321–1348 (2019)
36. Stein, E.M., Shakarchi, R.: Real Analysis: Measure Theory, Integration, and Hilbert Spaces. Princeton University Press, Princeton (2009)
37. Stockbridge, R.H.: Discussion of dynamic programming and linear programming approaches to stochastic control and optimal stopping in continuous time. Metrika **77**(1), 137–162 (2014)
38. Szegö, G.: Orthogonal Polynomials, vol. 23. American Mathematical Society, Providence (1939)
39. Vinter, R.: Convex duality and nonlinear optimal control. SIAM J. Control. Optim. **31**(2), 518–538 (1993)