Tractable semi-algebraic approximation using Christoffel-Darboux kernel

Jean B. Lasserre

LAAS-CNRS and Institute of Mathematics, Toulouse, France

Joint work with D. Henrion, T. Weisser, S. Marx, E. Pauwels and M. Putinar

BIRS at BANFF, May 2019

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

European Research Council

Initial motivation

Analysis of a certain class of non-linear PDEs (e.g. Burgers equation)

$$\frac{\partial y(x,t)}{\partial t} + y(x,t)\,\frac{\partial y(x,t)}{\partial x} \,=\, 0, \quad (x,t) \in \Omega,$$
$$+ \text{ boundary conditions,}$$

• One may apply the moment-SOS approach, i.e., one solves an appropriate hierarchy of semidefinite relaxations of increasing size.

☞ Previous talk by D. Henrion

At an optimal solution z of the “step-d" semidefinite relaxation one obtains an approximation of the moments

$$z_{i,j,k} = \int_\Omega y^i\, x^j\, t^k\, d\mu(y,x,t) = \int_\Omega x^j t^k\, y(x,t)^i\, dx\, dt$$

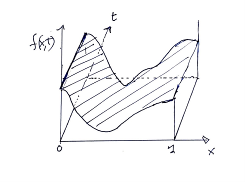
up to order $2d$ of the measure μ supported on the graph $\{(y(x,t),x,t) : (x,t) \in \Omega\}$ of the solution $y(x,t)$ of the PDE.

☞ **Problem: How to retrieve:**

the function $(x,t) \mapsto y(x,t), \quad (x,t) \in \Omega$, from the sole knowledge of $z_{i,j,k}$, for all (i,j,k) such that $i+j+k \le 2d$.

Generic univariate problem for scalar PDE

Let $(x, t) \mapsto f(x, t)$, $(x, t) \in [0, 1] \times [0, 1]$,



be an UNKNOWN bounded measurable function on $\Omega = [0, M] \times [0, 1]$, and suppose that one knows

$$z_{i,j,k} := \int_{\Omega} x^i \, t^j \, f(x, t)^k \, dx \, dt, \quad i + j + k \leq 2d.$$

☞ Approximate f as closely as desired when d increases and if possible with no Gibbs' phenomenon.

The motivation came from retrieving solutions of non-linear PDE's via the Moment-SOS hierarchy, BUT

we are concerned with the following generic situation:

Let $f : S \to \mathbb{R}$ be a bounded measurable function. Our sole knowledge on f is from the scalars

$$m_{\alpha,k} = \int_S \mathbf{x}^\alpha f(\mathbf{x})^k \, d\mathbf{x}, \quad \alpha \in \mathbb{N}^n, \, k \in \mathbb{N}.$$

and we address the generic inverse problem:

- Given $m_{\alpha,k}$, $\alpha, k \in \mathbb{N}^{n+1}_{2d}$

☞ COMPUTE an APPROXIMATION f_d of f, with CONVERGENCE GUARANTEES as d increases.

☞ ... and if possible ... with no GIBBS' phenomenon

The motivation came from retrieving solutions of non-linear PDE's via the Moment-SOS hierarchy, BUT

we are concerned with the following generic situation:

Let $f : S \to \mathbb{R}$ be a bounded measurable function. Our sole knowledge on f is from the scalars

$$m_{\alpha,k} = \int_S \mathbf{x}^\alpha f(\mathbf{x})^k \, d\mathbf{x}, \quad \alpha \in \mathbb{N}^n, \, k \in \mathbb{N}.$$

and we address the generic inverse problem:

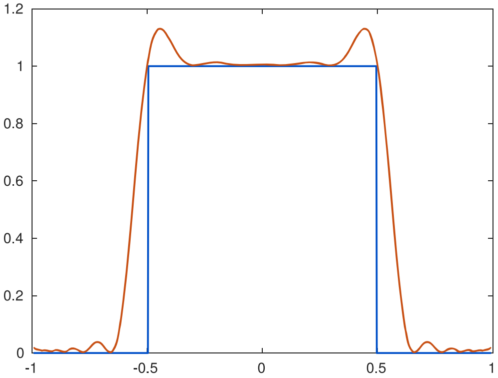
• Given $m_{\alpha,k}$, $\alpha, k \in \mathbb{N}^{n+1}_{2d}$

☞ COMPUTE an APPROXIMATION f_d of f, with CONVERGENCE GUARANTEES as d increases.

☞ ... and if possible ... with no GIBBS' phenomenon

The Gibbs phenomenon

☞ Typical when one approximates a discontinuous function (in blue) by a polynomial (in red).

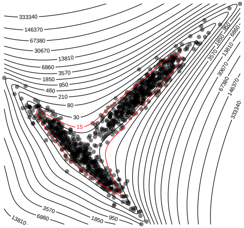


A little detour: The Christoffel function

Given a measure μ on a compact $\Omega \subset \mathbb{R}^n$, and $d \in \mathbb{N}$, one may construct a sum-of-squares (SOS) polynomial $Q_d \in \mathbb{R}[\mathbf{x}]_{2d}$ such that the levels sets

$$S_\gamma := \{ \mathbf{x} : Q_d(\mathbf{x}) \leq \gamma \}, \quad \gamma \in \mathbb{R}_+$$

capture the shape of the support Ω of μ better and better as $d \uparrow$.

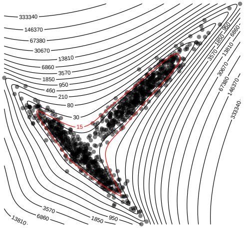


A little detour: The Christoffel function

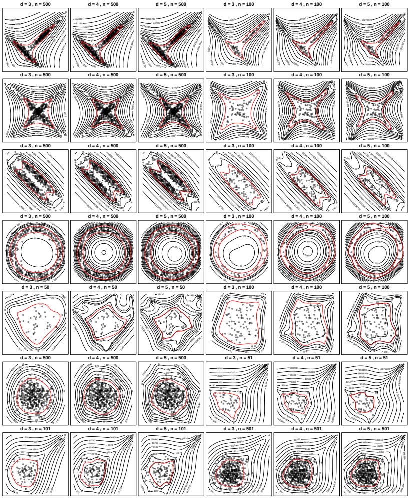
Given a measure μ on a compact $\Omega \subset \mathbb{R}^n$, and $d \in \mathbb{N}$, one may construct a sum-of-squares (SOS) polynomial $Q_d \in \mathbb{R}[\mathbf{x}]_{2d}$ such that the levels sets

$$S_\gamma \;:=\; \{\, \mathbf{x} : \; Q_d(\mathbf{x}) \leq \gamma \,\}, \quad \gamma \in \mathbb{R}_+$$

capture the shape of the support Ω of μ better and better as $d \uparrow$.



☞ Surprisingly, low degree *d* is often enough to get a pretty good idea of the shape of Ω (at least in dimension $n = 2, 3$)



The Christoffel function $C_d : \mathbb{R}^n \to \mathbb{R}_+$ is the reciprocal of the SOS polynomial Q_d and has a rich history in Approximation theory and Orthogonal Polynomials.

Theorem

Let the support Ω of μ be compact with nonempty interior and let $(P_\alpha)_{\alpha\in\mathbb{N}^n}$ be a family of orthonormal polynomials w.r.t. μ. Then for every $\xi \in \mathbb{R}^n$:

$$Q_d(\xi) = \sum_{|\alpha|\leq d} P_\alpha(\xi)^2$$

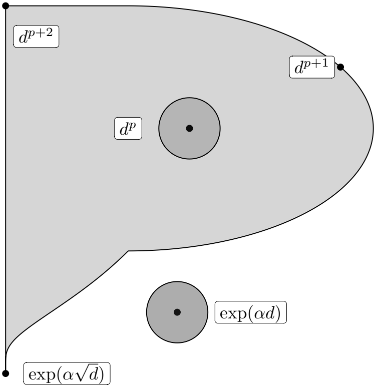
$$\frac{1}{Q_d(\xi)} = C_d(\xi) = \min_{P\in\mathbb{R}[\mathbf{x}]_d} \left\{ \int_\Omega P^2\, d\mu : \; P(\xi) = 1 \right\}$$

Theorem

Let the support Ω of μ be compact with nonempty interior. Then:

- *For all* $\mathbf{x} \in \mathrm{int}(\Omega)$: $Q_d(\mathbf{x}) = O(d^n)$.
- *For all* $\mathbf{x} \in \mathrm{int}(\mathbb{R}^n \setminus \Omega)$: $Q_d(\mathbf{x}) = \Omega(\exp(\tau d))$ *for some* $\tau > 0$.

In particular as $d \to \infty$, $d^n\, C_d(\mathbf{x}) \to 0$ *very fast whenever* $\mathbf{x} \notin \Omega$


d^{p+2}
d^{p+1}
d^p
$\exp(\alpha d)$
$\exp(\alpha\sqrt{d})$

The Christoffel function can be used in several important applications of Machine Learning (e.g. outlier detection, density estimation). In this case the measure μ is the empirical probability measure associated with a cloud of points $\mathcal{C} \subset \mathbb{R}^n$ (the data of interest).

For instance one may decide that points $\xi \in \mathcal{C}$ such that $Q_d(\xi) > \binom{n+d}{d}$ can be classified as outliers. Such a strategy (even with relatively low degree d) is as efficient as more elaborated techniques, and with no optimization involved.

☞ Lass & Pauwels Sorting out typicality via the inverse moment matrix SOS polynomial, NIPS 2016.
Lass & Pauwels The empirical Christoffel function with applications in data analysis, Adv. Comp. Math. 2019
Pauwels, Putinar & Lass Data analysis from empirical moments and the Christoffel function, arXiv:1810.08480

The Christoffel function can be used in several important applications of Machine Learning (e.g. outlier detection, density estimation). In this case the measure μ is the empirical probability measure associated with a cloud of points $\mathcal{C} \subset \mathbb{R}^n$ (the data of interest).

For instance one may decide that points $\xi \in \mathcal{C}$ such that $Q_d(\xi) > \binom{n+d}{d}$ can be classified as outliers. Such a strategy (even with relatively low degree d) is as efficient as more elaborated techniques, and with no optimization involved.

☞ Lass & Pauwels Sorting out typicality via the inverse moment matrix SOS polynomial, NIPS 2016.
Lass & Pauwels The empirical Christoffel function with applications in data analysis, Adv. Comp. Math. 2019
Pauwels, Putinar & Lass Data analysis from empirical moments and the Christoffel function, arXiv:1810.08480

The Christoffel function can be used in several important applications of Machine Learning (e.g. outlier detection, density estimation). In this case the measure μ is the empirical probability measure associated with a cloud of points $\mathcal{C} \subset \mathbb{R}^n$ (the data of interest).

For instance one may decide that points $\xi \in \mathcal{C}$ such that $Q_d(\xi) > \binom{n+d}{d}$ can be classified as outliers. Such a strategy (even with relatively low degree d) is as efficient as more elaborated techniques, and with no optimization involved.

☞ Lass & Pauwels Sorting out typicality via the inverse moment matrix SOS polynomial, NIPS 2016.
Lass & Pauwels The empirical Christoffel function with applications in data analysis, Adv. Comp. Math. 2019
Pauwels, Putinar & Lass Data analysis from empirical moments and the Christoffel function, arXiv:1810.08480

Back to our recovery problem

Take home message

In our problem, the support Ω of μ on $\mathbb{R}^{n+1}$ IS the graph $\{(\mathbf{x}, f(\mathbf{x})) : \mathbf{x} \in S \subset \mathbb{R}^n\}$ of an unknown function $f : S \to \mathbb{R}$.

☞ Hence the Christoffel function is an appropriate tool for getting information on f from moments of μ.

An illustrative example: Let $f : [0,1] \to [0,1]$ be the step function:

$$f(t) := \begin{cases} 0 & t \in [0,1/2] \\ 1 & t \in (1/2,1] \end{cases}$$

and let μ be a measure on $[0,1]^2$ supported on the graph $\Omega = \{(t, f(t)) : t \in [0,1]\}$ of f.

☞ The support $\Omega \subset \mathbb{R}^2$ of μ has an empty interior as $d\mu(x,t)$ is singular w.r.t. Lebesgue measure on $\mathbb{R}^2$.

Therefore we instead use $\mu + \varepsilon\lambda$ where λ is the Lebesgue measure on $[0,1]^2$ and $\varepsilon > 0$ is very small.

An illustrative example: Let $f : [0,1] \to [0,1]$ be the step function:

$$f(t) := \begin{cases} 0 & t \in [0, 1/2] \\ 1 & t \in (1/2, 1] \end{cases}$$

and let μ be a measure on $[0,1]^2$ supported on the graph $\Omega = \{(t, f(t)) : t \in [0,1]\}$ of f.

☞ The support $\Omega \subset \mathbb{R}^2$ of μ has an empty interior as $d\mu(x,t)$ is singular w.r.t. Lebesgue measure on $\mathbb{R}^2$.

Therefore we instead use $\mu + \varepsilon\lambda$ where λ is the Lebesgue measure on $[0,1]^2$ and $\varepsilon > 0$ is very small.

Suppose that we only know the moments $(z_{i,j})_{i,j\leq 2d}$, up to order $2d$, of μ.

From the moments:

- $z = (z_{i,j})_{i+j\leq 2d}$, up to order $2d$, and
- $\boldsymbol{\lambda} = (\lambda_{i,j})_{i+j\leq 2d}$ up to order $2d$ of the Lebesgue measure on $[0,1]^2$, and for $\varepsilon > 0$ small (and fixed),

☞ form the moment matrix $\mathbf{M}_d(z + \varepsilon\,\boldsymbol{\lambda})$.

☞ Compute the Christoffel polynomial $Q_d(x,t)$.

For arbitrary $t \in [0,1]$, let:

$$h_d(t) \;:=\; x^* = \arg\min_{x\in[0,1]} Q_d(x,t).$$

☞ As $x \mapsto Q_d(x,t)$ is a UNIVARIATE polynomial, x^* can be obtained efficiently.

Suppose that we only know the moments $(z_{i,j})_{i,j\leq 2d}$, up to order $2d$, of μ.

From the moments:

- $z = (z_{i,j})_{i+j\leq 2d}$, up to order $2d$, and
- $\boldsymbol{\lambda} = (\lambda_{i,j})_{i+j\leq 2d}$ up to order $2d$ of the Lebesgue measure on $[0,1]^2$, and for $\varepsilon > 0$ small (and fixed),

- ☞ form the moment matrix $\mathbf{M}_d(z + \varepsilon\,\boldsymbol{\lambda})$.
- ☞ Compute the Christoffel polynomial $Q_d(x,t)$.

For arbitrary $t \in [0,1]$, let:

$$h_d(t) \;:=\; x^* = \arg\min_{x\in[0,1]} Q_d(x,t).$$

☞ As $x \mapsto Q_d(x,t)$ is a UNIVARIATE polynomial, x^* can be obtained efficiently.

Suppose that we only know the moments $(z_{i,j})_{i,j\leq 2d}$, up to order $2d$, of μ.

From the moments:

- $z = (z_{i,j})_{i+j\leq 2d}$, up to order $2d$, and
- $\boldsymbol{\lambda} = (\lambda_{i,j})_{i+j\leq 2d}$ up to order $2d$ of the Lebesgue measure on $[0,1]^2$, and for $\varepsilon > 0$ small (and fixed),

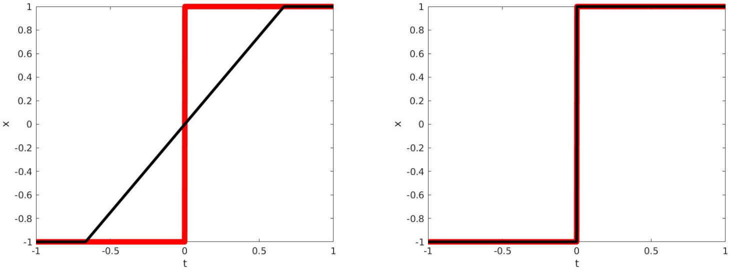
☞ form the moment matrix $\mathbf{M}_d(z + \varepsilon\,\boldsymbol{\lambda})$.

☞ Compute the Christoffel polynomial $Q_d(x,t)$.

For arbitrary $t \in [0,1]$, let:

$$h_d(t) \;:=\; x^* = \arg\min_{x\in[0,1]} Q_d(x,t).$$

☞ As $x \mapsto Q_d(x,t)$ is a UNIVARIATE polynomial, x^* can be obtained efficiently.



In black (left) the approximation with moments of order 2 and in black (right) the approximation with moments of order 4.

☞ Observe the absence of any Gibbs phenomenon ...

For the Burgers equation

Let μ be our unknown measure supported on the graph $\Omega = \{(f(x,t), x, t) : (x,t) \in S\}$ of the entropy solution of the Burgers equation. Suppose that we only know the moments $(z_{i,j,k})_{i,j,k \leq 2d}$, up to order $2d$, of μ.

- ☞ Recall that in practice, $(z_{i,j,k})$ is an optimal solution of the step-d semidefinite relaxation associated with the Burgers equation.
- ☞ $\Omega \subset \mathbb{R}^3$ has an empty interior as $d\mu(y,x,t)$ is singular w.r.t. Lebesgue measure on $\mathbb{R}^3$.

Therefore we instead use $\mu + \varepsilon\lambda$ where λ is the Lebesgue measure on $[0,R] \times \underbrace{[0,M] \times [0,1]}_{S}$ and $\varepsilon > 0$ is very small. (For the Burgers equation, R and M are determined from bounds on the boundary condition $y_0(x,0)$.)

For the Burgers equation

Let μ be our unknown measure supported on the graph $\Omega = \{(f(x,t), x, t) : (x,t) \in S\}$ of the entropy solution of the Burgers equation. Suppose that we only know the moments $(z_{i,j,k})_{i,j,k \leq 2d}$, up to order $2d$, of μ.

☞ Recall that in practice, $(z_{i,j,k})$ is an optimal solution of the step-d semidefinite relaxation associated with the Burgers equation.

☞ $\Omega \subset \mathbb{R}^3$ has an empty interior as $d\mu(y,x,t)$ is singular w.r.t. Lebesgue measure on $\mathbb{R}^3$.

Therefore we instead use $\mu + \varepsilon\lambda$ where λ is the Lebesgue measure on $[0,R] \times \underbrace{[0,M] \times [0,1]}_{S}$ and $\varepsilon > 0$ is very small. (For the Burgers equation, R and M are determined from bounds on the boundary condition $y_0(x,0)$.)

Let μ be our unknown measure supported on the graph $\Omega = \{(f(x,t), x, t) : (x,t) \in S\}$ of the entropy solution of the Burgers equation. Suppose that we only know the moments $(z_{i,j,k})_{i,j,k \leq 2d}$, up to order $2d$, of μ.

☞ Recall that in practice, $(z_{i,j,k})$ is an optimal solution of the step-d semidefinite relaxation associated with the Burgers equation.

☞ $\Omega \subset \mathbb{R}^3$ has an empty interior as $d\mu(y,x,t)$ is singular w.r.t. Lebesgue measure on $\mathbb{R}^3$.

Therefore we instead use $\mu + \varepsilon\lambda$ where λ is the Lebesgue measure on $[0,R] \times \underbrace{[0,M] \times [0,1]}_{S}$ and $\varepsilon > 0$ is very small. (For the Burgers equation, R and M are determined from bounds on the boundary condition $y_0(x,0)$.)

Recovery strategy

From the moments:

- $z = (z_{i,j,k})_{i+j+k \leq 2d}$, up to order $2d$, and
- $\boldsymbol{\lambda} = (\lambda_{i,j,k})_{i+j+k \leq 2d}$ up to order $2d$ of the Lebesgue measure on $[0, R] \times [0, M] \times [0, 1]$,

and for $\varepsilon > 0$ small (and fixed),

- ☞ form the moment matrix $\mathbf{M}_d(z + \varepsilon\,\boldsymbol{\lambda})$.
- ☞ Compute the Christoffel polynomial $Q_d(y, x, t)$.

For arbitrary $(x, t) \in [0, M] \times [0, 1]$, let:

$$h_d(x, t) \;:=\; y^* = \arg \min_{y \in [0,R]} Q_d(y, x, t).$$

☞ As $y \mapsto Q_d(y, x, t)$ is a UNIVARIATE polynomial, y^* is obtained exactly by solving a single SDP.

Recovery strategy

From the moments:

- $z = (z_{i,j,k})_{i+j+k \leq 2d}$, up to order $2d$, and
- $\boldsymbol{\lambda} = (\lambda_{i,j,k})_{i+j+k \leq 2d}$ up to order $2d$ of the Lebesgue measure on $[0, R] \times [0, M] \times [0, 1]$,

and for $\varepsilon > 0$ small (and fixed),

☞ form the moment matrix $\mathbf{M}_d(z + \varepsilon \, \boldsymbol{\lambda})$.

☞ Compute the Christoffel polynomial $Q_d(y, x, t)$.

For arbitrary $(x, t) \in [0, M] \times [0, 1]$, let:

$$h_d(x, t) \; := \; y^* = \arg \min_{y \in [0, R]} Q_d(y, x, t).$$

☞ As $y \mapsto Q_d(y, x, t)$ is a UNIVARIATE polynomial, y^* is obtained exactly by solving a single SDP.

Convergence guarantees

Under some relatively weak conditions on $(x,t) \mapsto f(x,t)$:

☞ $h_d \to f$ in $L_1([0,M] \times [0,1])$.

☞ $h_d(x,t) \to f(x,t)$ for almost all $(x,t) \in [0,M] \times [0,1]$.

Importantly:

☞ the APPROXIMANT f_d belongs to the class of semi-algebraic functions, as opposed to standard approximation schemes where f_d is a polynomial.

☞ provides a RATIONALE why the GIBBS' phenomenon disappears in our numerical experiments.

Convergence guarantees

Under some relatively weak conditions on $(x, t) \mapsto f(x, t)$:

☞ $h_d \to f$ in $L_1([0, M] \times [0, 1])$.

☞ $h_d(x, t) \to f(x, t)$ for almost all $(x, t) \in [0, M] \times [0, 1]$.

Importantly:

☞ the APPROXIMANT f_d belongs to the class of semi-algebraic functions, as opposed to standard approximation schemes where f_d is a polynomial.

☞ provides a RATIONALE why the GIBBS' phenomenon disappears in our numerical experiments.

Convergence guarantees

Under some relatively weak conditions on $(x, t) \mapsto f(x, t)$:

☞ $h_d \to f$ in $L_1([0, M] \times [0, 1])$.

☞ $h_d(x, t) \to f(x, t)$ for almost all $(x, t) \in [0, M] \times [0, 1]$.

Importantly:

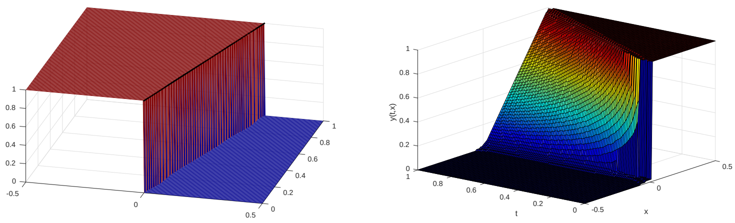
☞ the APPROXIMANT f_d belongs to the class of semi-algebraic functions, as opposed to standard approximation schemes where f_d is a polynomial.

☞ provides a RATIONALE why the GIBBS' phenomenon disappears in our numerical experiments.

Ex: The Burgers equation

We consider two initial conditions: One yields a solution $f(x,t)$ with a discontinuity (shock) and the other yields a continuous solution (rarefaction).

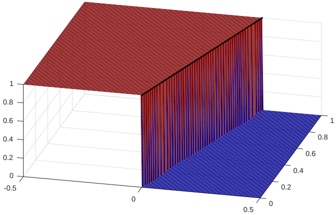
☞ With moments up to order $2d = 12$, the moments z match those of the measure μ supported on the graph of f (with at least 4 digits of precision). Then after discretizing $[0, M] \times [0, 1]$ and computing $h_d(x,t)$ on this grid, one obtains the two approximations (with almost no Gibbs phenomenon):

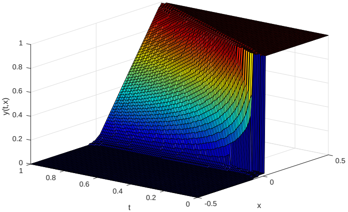


Ex: The Burgers equation

We consider two initial conditions: One yields a solution $f(x, t)$ with a discontinuity (shock) and the other yields a continuous solution (rarefaction).

☞ With moments up to order $2d = 12$, the moments z match those of the measure μ supported on the graph of f (with at least 4 digits of precision). Then after discretizing $[0, M] \times [0, 1]$ and computing $h_d(x, t)$ on this grid, one obtains the two approximations (with almost no Gibbs phenomenon):

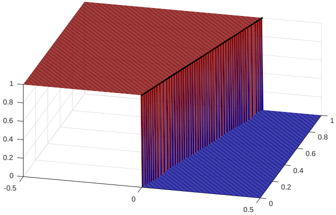


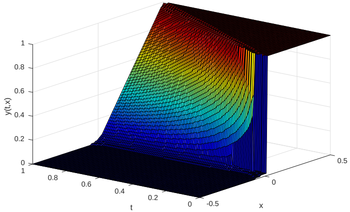


Ex: The Burgers equation

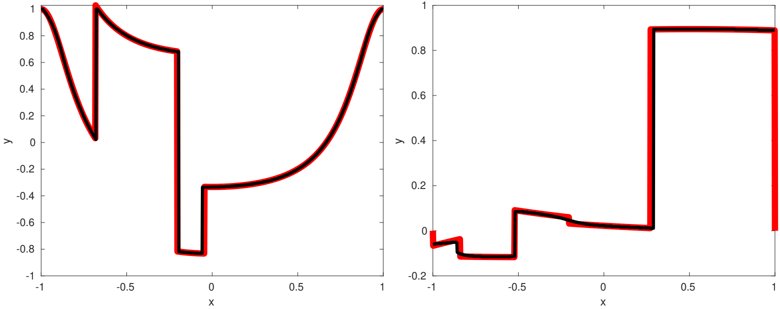
We consider two initial conditions: One yields a solution $f(x,t)$ with a discontinuity (shock) and the other yields a continuous solution (rarefaction).

☞ With moments up to order $2d = 12$, the moments z match those of the measure μ supported on the graph of f (with at least 4 digits of precision). Then after discretizing $[0, M] \times [0, 1]$ and computing $h_d(x,t)$ on this grid, one obtains the two approximations (with almost no Gibbs phenomenon):

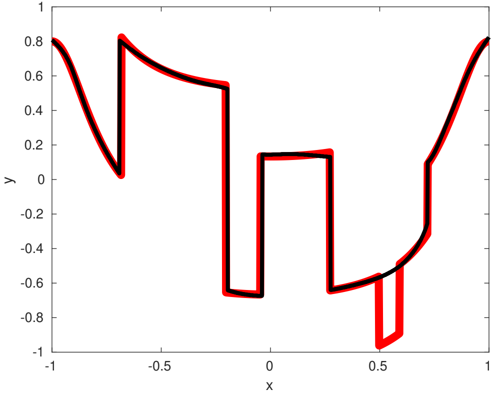




Examples from Eckhoff



Examples from Eckhoff continued



More details in :

S. Marx, T. Weisser, D. Henrion and J.B. Lass (2018). A moment approach for entropy solutions to nonlinear hyperbolic PDEs. Math. Control & Related Fields, 2019

S. Marx, E. Pauwels, T. Weisser, D. Henrion and J.B. Lass (2018). Tractable semi-algebraic approximation using Christoffel-Darboux kernel. arXiv:1807.02306

THANK YOU !