

GRAPH RECOVERY FROM INCOMPLETE MOMENT INFORMATION

DIDIER HENRION^{1,2}, JEAN BERNARD LASSERRE^{1,3}

ABSTRACT. We investigate a class of moment problems, namely recovering a measure supported on the graph of a function from partial knowledge of its moments, as for instance in some problems of optimal transport or density estimation. We show that the sole knowledge of first degree moments of the function, namely linear measurements, is sufficient to obtain asymptotically all the other moments by solving a hierarchy of semidefinite relaxations (viewed as moment matrix completion problems) with a specific sparsity inducing criterion related to a weighted ℓ_1 -norm of the moment sequence of the measure. The resulting sequence of optimal solutions converges to the whole moment sequence of the measure which is shown to be the unique optimal solution of a certain infinite-dimensional linear optimization problem (LP). Then one may recover the function by a recent extraction algorithm based on the Christoffel-Darboux kernel associated with the measure. Finally, the support of such a measure supported on a graph is a meager, very thin (hence sparse) set. Therefore the LP on measures with this sparsity inducing criterion can be interpreted as an analogue for infinite-dimensional signals of the LP in super-resolution for (sparse) atomic signals.

KEYWORDS: Moment problem, Inverse problem, Sparse signals, Semidefinite programming.

1. INVERSE PROBLEM: FROM MOMENTS TO GRAPH

In data science, it is often relevant to process moments of a signal instead of the signal itself. For complex valued signals, the moments are Fourier coefficients, and many filtering operations are efficiently carried out in the sequence of moments. In numerical approximation algorithms, many operations on real valued signals are more efficiently implemented in their sequence of Chebyshev coefficients [19]. In the moment-SOS hierarchy approach, many nonlinear nonconvex problems are reformulated and solved approximately in the sequence of moments; see [13, 12] and references therein. Once moments or approximate moments have been computed, one is faced with the inverse problem of reconstructing the signal from its moments.

The recent work [15] describes an algorithm based on the Christoffel-Darboux kernel, to recover the graph of a function from knowledge of its moments. Its novelty (and distinguishing feature) is to approximate the *graph* of the function (rather than the function itself) with a *semialgebraic* function (namely a minimizer of a sum of squares of polynomials) with L_1 and pointwise convergence guarantees for an increasing number of input moments. In

¹CNRS; LAAS; Université de Toulouse, 7 avenue du colonel Roche, F-31400 Toulouse, France.

²Faculty of Electrical Engineering, Czech Technical University in Prague, Technická 2, CZ-16626 Prague, Czechia.

³Institute of Mathematics; Université de Toulouse, 118 route de Narbonne, F-31062 Toulouse, France.

Date: Draft of November 5, 2020.

Lasserre's research is partly supported by AI Interdisciplinary Institute ANITI funding, through the French "Investing for the Future PIA3" program under the Grant agreement n° ANR-19-PI3A-0004.

contrast to approximations via continuous functions (e.g. polynomials) this larger class of approximants permits to better handle nasty functions (e.g. with discontinuities). In particular, this distinguishing feature allows to sometimes avoid a typical Gibbs phenomenon (as well as oscillations) typically encountered when approximating a discontinuous function by polynomials; for more details and illustrative examples the interested reader is referred to [15].

The (algebraic) moment of degree $d := (d_x, d_y) \in \mathbb{N}^{n_x} \times \mathbb{N}^{n_y}$ of the Young measure supported on the graph of a given mapping $\mathbf{f} : X \rightarrow Y$ from a domain $X \subset \mathbb{R}^{n_x}$ to a domain $Y \subset \mathbb{R}^{n_y}$ is the real number

$$(1.1) \quad \int_X \mathbf{x}^{d_x} \mathbf{f}(\mathbf{x})^{d_y} d\lambda(\mathbf{x}), \quad (d_x, d_y) \in \mathbb{N}^{n_x} \times \mathbb{N}^{n_y},$$

where we have used the multi-index monomial notation¹. In (1.1) λ is a given reference measure, for example the uniform measure on X .

In applications treated in [15] to recover \mathbf{f} , all moments (1.1) are available; more precisely, in principle an arbitrarily large number of them can be computed. However in a number of other interesting applications this is not the case. For instance:

- In the moment-SOS approach for solving or controlling differential equations, only finitely many (approximate) moments are given, i.e. $|d| \leq r$ where $r \in \mathbb{N}$ is a given relaxation order, and $|d|$ denotes the sum of the entries of vector d ;
- In statistics for density estimation ($n_y = 1$), only finitely many moments $\int \mathbf{x}^{d_x} f(\mathbf{x}) d\mathbf{x}$, $d_x \in \mathbb{N}^{n_x}$, of the unknown density function $f \geq 0$ are available, i.e. $d_y = 1$;
- In optimal transport, only marginal moments are available, i.e. either $|d_x| = 0$ or $|d_y| = 0$.

The main feature of this paper is to consider the more general inverse problem of recovering the graph of a function \mathbf{f} given some of its moments, indexed by $d \in D$ in a given countable index set $D \subset \mathbb{N}^{n_x} \times \mathbb{N}^{n_y}$. The following *optimization / relaxation / concentration* features combined with the recovery technique described in [15], are key ingredients for a successful recovery of \mathbf{f} .

Optimization. In order to approximate the function \mathbf{f} from some of its moments (1.1), an integral functional

$$\int_X c(\mathbf{x}, \mathbf{f}(\mathbf{x})) d\lambda(\mathbf{x})$$

of the solution is typically minimized, where $c : X \times Y \rightarrow \mathbb{R}$ is a running cost. In some applications, the cost is given and there is a unique function solving the problem. Otherwise, the cost is chosen so that a specific function is recovered. The resulting optimization problem reads

$$(1.2) \quad \inf_{\mathbf{f}} \int_X c(\mathbf{x}, \mathbf{f}(\mathbf{x})) d\lambda(\mathbf{x}) \quad \text{s.t.} \quad \int_X \mathbf{x}^{d_x} \mathbf{f}(\mathbf{x})^{d_y} d\lambda(\mathbf{x}) = z_d, \quad d \in D$$

where the sequence $(z_d)_{d \in D} \subset \mathbb{R}$ and the index set D are given. Function \mathbf{f} is sought in a given functional space, for example square integrable functions, uniformly bounded functions, or measurable functions.

¹For numerical reasons, it is preferable to express moments in a basis different from the monomial basis. However, in this note we stick to monomials for notational ease.

Relaxation. Optimization problem (1.2) is nonlinear, nonconvex and quite difficult in general. Following Kantorovich’s idea proposed for transport problems [18, Introduction to optimal transport], one can instead consider the relaxed version

$$(1.3) \quad \inf_{\mu_{\mathbf{x}}} \int_{X \times Y} c(\mathbf{x}, \mathbf{y}) \mu_{\mathbf{x}}(d\mathbf{y}) \lambda(d\mathbf{x}) \quad \text{s.t.} \quad \int_{X \times Y} \mathbf{x}^{d_x} \mathbf{y}^{d_y} \mu_{\mathbf{x}}(d\mathbf{y}) \lambda(d\mathbf{x}) = z_d, \quad d \in D$$

of the non convex problem (1.2). A distinguishing feature of the relaxation (1.3) is to be *linear* and hence convex in $\mu_{\mathbf{x}}$, which is a probability measure on Y parametrized by $\mathbf{x} \in X$, also called a conditional, or Young measure. Weak star compactness arguments can be used to show that under mild assumptions on the input data, the infimum in relaxed problem (1.3) is attained. Letting $\mu_{\mathbf{x}} = \delta_{\mathbf{f}(\mathbf{x})}$ shows that the infimum in relaxed problem (1.3) is less than or equal to the infimum in problem (1.2). Techniques of calculus of variations can be used to show that under mild assumptions on the input data, the minimum in relaxed problem (1.3) is equal to the infimum in problem (1.2), i.e. that there is no relaxation gap.

Concentration. For some specific choices of c and D , it can be proved that $\mu_{\mathbf{x}} = \delta_{\mathbf{f}(\mathbf{x})}$ is the unique optimal solution of relaxed problem (1.3), and hence problem (1.2) has a unique solution \mathbf{f} . In this case we say that the Young measure $\mu_{\mathbf{x}}$ is concentrated on the graph of function \mathbf{f} .

For instance this happens in quadratic optimal transport for which the cost $c(\mathbf{x}, \mathbf{y}) := \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|^2$ is given. Brenier showed that the optimal transport function is the gradient of a convex function; see e.g. [18, Chapter 1] and references therein. If there is freedom for the cost c , it can be chosen so that concentration occurs.

It also happens in the moment-SOS approach for solving a certain class of nonlinear partial differential equations, if additional linear constraints are enforced on the measure and/or the cost if chosen appropriately; see e.g. the recent work on Burgers equation [16].

Contribution. We consider problem (1.3) in the case $|d_y| \leq 1$, i.e. the only available moments are those of the reference measure (zero-th order moments $d_y = 0$) and averages of the graph (first order moments $d_y = 1$, namely linear measurements of the function). A typical case is density estimation.

Our contribution is twofold: I. As for transport problem, we show that the measure μ to recover is the unique optimal solution of a certain infinite-dimensional LP on a measure space. II. Next, under the assumption that the input data are semialgebraic, we provide a systematic numerical scheme to approximate as closely as desired any arbitrary finite number of moments of μ . It consists of solving problem (1.3) with the moment-SOS hierarchy. Since we are given an incomplete set of moments, we optimize over the remaining set of unknown moments. In other words, each semidefinite relaxation is a *moment matrix completion problem* of increasing size.

The rationale behind the asymptotic convergence of the hierarchy is that every semidefinite relaxation is a r -truncated version of the infinite-dimensional LP where one only considers pseudo-moments of order r . It turns out that as r increases, the resulting sequence of optimal solutions converges to the vector of moments of μ .

Interestingly this result can be interpreted as an analogue for infinite-dimensional signals (densities) of *super-resolution* for sparse atomic measures [7, 6, 8]. Indeed instead of considering the measure $d\nu(\mathbf{x}) = \mathbf{f}(\mathbf{x}) d\mathbf{x}$ on X with unknown density \mathbf{f} (but with *all* its moments

available), we rather consider the graph density measure $\mu = \delta_{\mathbf{f}(\mathbf{x})}(dy)\lambda(d\mathbf{x})$ whose support $(\mathbf{x}, \mathbf{f}(\mathbf{x}))$ is a meager (very thin) degenerate set of $X \times Y$ (hence sparse), now with partial knowledge of its moments. Then as for super-resolution of sparse atomic measures, this sparse (degenerate) measure is the unique solution of an infinite LP on measures where the sparsity-inducing cost to minimize is a weighted ℓ_1 -norm of their moment vector.

Then once moments of μ up to order $2r$ have been approximated in an optimal solution at step r of the hierarchy, the function can be in turn approximated with the reconstruction technique already alluded to and described in [15]; the deeper in the hierarchy the better is the resulting approximation.

Our experiments convey the following message: In the above situations it is better to approximate the graph $G := \{(\mathbf{x}, f(\mathbf{x}) : \mathbf{x} \in X)\}$ from moments of the measure μ supported on G (a thin subset of $X \times Y$), rather than approximate the function itself (e.g. L^2 -norm approximation by polynomials) from moments of the measure $d\nu = f d\mathbf{x}$ on X with density f . Even more, this is true even if most moments of μ have to be approximated, whereas all exact moments of ν are available.

2. PROBLEM STATEMENT

Let $X := \{\mathbf{x} \in \mathbb{R}^n : g_j(\mathbf{x}) \geq 0, j = 1, \dots, m\}$ be a compact basic semi-algebraic set defined for given polynomials $(g_j) \subset \mathbb{R}[\mathbf{x}]$. With no loss of generality, we will assume that $X \subset [0, 1]^n$ (possibly after some scaling and translation).

The cone of Borel regular non-negative measures on X , topologically dual to the cone of non-negative continuous functions on X , is denoted by $\mathcal{M}(X)_+$. The Borel sigma algebra of subsets of X is denoted by $\mathcal{B}(X)$.

Let $\lambda \in \mathcal{M}(X)_+$ be a finite probability measure with moments $\boldsymbol{\lambda} = (\lambda_{d_x}) \subset \mathbb{R}$ defined by

$$\lambda_{d_x} := \int_X \mathbf{x}^{d_x} d\lambda(\mathbf{x}), \quad d_x \in \mathbb{N}^n.$$

From now on, let us restrict our attention to scalar-valued functions $f : X \rightarrow Y$. Vector-valued functions can be treated entry-wise. Let $\mu \in \mathcal{M}(X \times Y)_+$ be a finite probability measure of the form $d\mu(\mathbf{x}, y) = \delta_{f(\mathbf{x})}(dy)\lambda(d\mathbf{x})$, i.e.

$$(2.1) \quad \mu(A \times B) = \int_A 1_B(f(\mathbf{x})) d\lambda(\mathbf{x}), \quad \forall A \in \mathcal{B}(X), B \in \mathcal{B}(Y),$$

with all finite moments $\boldsymbol{\mu} = (\mu_d)_{d \in \mathbb{N}^{n+1}} \subset \mathbb{R}$ defined by

$$(2.2) \quad \mu_d := \int_X \int_Y \mathbf{x}^{d_x} y^{d_y} d\mu(\mathbf{x}, y) = \int_X \mathbf{x}^{d_x} f(\mathbf{x})^{d_y} d\lambda(\mathbf{x}), \quad d = (d_x, d_y) \in \mathbb{N}^{n+1}.$$

In particular $\mu_{d_x, 0} = \lambda_{d_x}$ for all $d_x \in \mathbb{N}^n$. We assume that $0 \leq f \in \mathcal{L}^\infty(\lambda)$ and $\|f\|_\infty \leq \gamma$ for a given $\gamma > 0$, i.e. $Y := [0, \gamma]$. (The case where f is bounded but can take negative values reduces to the previous case by taking $\tilde{f} := f + \|f\|_\infty$.) Let $\mathbb{N}_r^{n+1} := \{d \in \mathbb{N}^{n+1} : |d| := \sum_{k=1}^{n+1} d_k \leq r\}$.

Problem. Moment completion. *Let $D := \{d = (d_x, d_y) \in \mathbb{N}^{n+1} : d_y \leq 1\}$. Given moments $(\mu_d)_{d \in D}$ and given $s \in \mathbb{N}$, approximate as closely as desired all moments $(\mu_d)_{d \in \mathbb{N}_{2s}^{n+1}}$.*

Further notations. Let $g_0(\mathbf{x}) := 1$ and $g_1(\mathbf{x}) := n - \|\mathbf{x}\|^2$; as $g_1 \geq 0$ defines a redundant constraint, it can always be included in the definition of X . Let $g_{m+1}(\mathbf{x}, y) := y(\gamma - y)$. Then the quadratic module

$$(2.3) \quad Q(g) := \left\{ \sum_{j=0}^{m+1} \sigma_j g_j : \sigma_j \in \Sigma[\mathbf{x}, y] \right\} \subset \mathbb{R}[\mathbf{x}, y],$$

generated by the g_j is Archimedean.

For $j = 0, 1, \dots, m+1$, let $d_j := \lceil \deg(g_j)/2 \rceil$, and express g_j in the monomial basis as $g_j(\mathbf{x}, y) = \sum_{d=(d_x, d_y) \in \mathbb{N}^{n+1}} g_{j,d} \mathbf{x}^{d_x} y^{d_y}$.

Next, define the *localizing matrix* $\mathbf{M}_{s-d_j}(g_j, \boldsymbol{\mu})$ associated with $\boldsymbol{\mu}$ and g_j , as the real symmetric matrix of size $\binom{n+1+s-d_j}{n+1}$ whose entries indexed by $(d_r, d_c) \in \mathbb{N}_{s-d_j}^{n+1} \times \mathbb{N}_{s-d_j}^{n+1}$ are equal to $\sum_{d \in \mathbb{N}^{n+1}} g_{j,d} \mu_{d+d_r+d_c}$. In particular, for $j = 0$, $g_0 = 1$, $d_0 = 0$ and $\mathbf{M}_s(\boldsymbol{\mu})$ is called the *moment matrix* associated with $\boldsymbol{\mu}$.

A preliminary result. The following result of independent interest provides a *rationale* for the infinite-dimensional LP introduced in Section 3 as well as for the numerical scheme described in Section 4 to recover the measure μ in (2.1) supported on the graph of f .

Theorem 2.1. *Let $\mu \in \mathcal{M}(X \times Y)_+$ be as in (2.1) and let $\phi \in \mathcal{M}(X \times Y)_+$ be such that all its moments $\boldsymbol{\phi} = (\phi_d)_{d \in \mathbb{N}^{n+1}}$ are finite, with $\phi_{d_x, 0} = \lambda_{d_x}$ and $\phi_{d_x, 1} = \mu_{d_x, 1}$ for all $d_x \in \mathbb{N}^n$. Then $\phi_d \geq \mu_d$ for all $d \in \mathbb{N}^{n+1}$. In particular,*

$$(2.4) \quad \text{trace } \mathbf{M}_r(\boldsymbol{\mu}) \leq \text{trace } \mathbf{M}_r(\boldsymbol{\phi}), \quad \forall r \in \mathbb{N}.$$

The proof of Theorem 2.1 is postponed to Section 8.

Remark 1. *If $X \not\subset \mathbb{R}_+^n$ then we obtain the slightly weaker result $\phi_{2d_x, d_y} \geq \mu_{2d_x, d_y}$ for all $(2d_x, d_y) \in \mathbb{N}^{n+1}$, and therefore (2.4) still holds.*

3. AN INFINITE-DIMENSIONAL LP ON MEASURES

The moment-completion problem can be viewed as a way to embed the initial problem into that of recovering a curve (more exactly the measure μ supported on the graph of that curve), i.e., a lower dimensional object with a meager, very thin, degenerate support in a space of measures on $X \times Y$ which can have support of full-dimension $n+1$. In other words this meager support $\{(\mathbf{x}, f(\mathbf{x})) : \mathbf{x} \in X\}$ is sparse in $X \times Y$.

Therefore when viewing the problem with such glasses, Theorem 2.1 suggests a way to build up an appropriate cost criterion defined over $\mathcal{M}(X \times Y)_+$ that will select μ in (2.1) as the *unique* optimal solution of a certain infinite-dimensional LP on $\mathcal{M}(Y \times Y)_+$.

So let $\boldsymbol{\theta} = (\theta_d)_{d \in \mathbb{N}^{n+1}}$ be a nonnegative sequence of ℓ_1 , the Banach space of summable sequences $\mathbf{a} = (a_d)$ with norm $\|\mathbf{a}\|_1 = \sum_{d \in \mathbb{N}^{n+1}} |a_d|$. For instance choose

$$(3.1) \quad \theta_{d=(d_x, d_y)} = \left(c^{2d_y} \binom{n-1+|d_x|}{|d_x|} \right)^{-1} \quad \text{with } c > \max(1, \gamma).$$

Recalling $Y = [0, \gamma]$, observe that $\phi_{d_x, d_y} = \int_{X \times Y} \mathbf{x}^{d_x} y^{d_y} d\phi < \gamma^{d_y}$ for all $(d_x, d_y) \in \mathbb{N}^{n+1}$, and therefore

$$(3.2) \quad \mathcal{T}(\phi) := \sum_d \theta_d \phi_d < \sum_{|d_x|=1}^{\infty} \sum_{d_y} \frac{\gamma^{d_y}}{c^{2d_y} \binom{n-1+|d_x|}{|d_x|}} < \sum_{d_y} (\gamma/c^2)^{d_y} < \frac{1}{1 - \gamma/c^2},$$

for all $\phi \in \mathcal{M}(X \times Y)_+$. Next, with θ as in (3.1) let $T : X \times Y \rightarrow \mathbb{R}_+$,

$$(\mathbf{x}, y) \mapsto T(\mathbf{x}, y) := \sum_{d=(d_x, d_y) \in \mathbb{N}^{n+1}} \theta_d \mathbf{x}^{d_x} y^{d_y} < \infty.$$

Observe that for all $(\mathbf{x}, y) \in X \times Y$,

$$0 \leq \sum_{d=(d_x, d_y) \leq r} \theta_d \mathbf{x}^{d_x} y^{d_y} =: T_r(\mathbf{x}, y) \uparrow T(\mathbf{x}, y) \quad \text{as } r \rightarrow \infty.$$

Hence by Monotone Convergence [1, p.44],

$$(3.3) \quad \mathcal{T}(\phi) = \sum_{d \in \mathbb{N}^{n+1}} \theta_d \phi_d = \lim_{r \rightarrow \infty} \sum_{|d| \leq r} \theta_d \phi_d = \lim_{r \rightarrow \infty} \int T_r(\mathbf{x}, y) d\phi = \int T(\mathbf{x}, y) d\phi,$$

for all $\phi \in \mathcal{M}(X \times Y)_+$.

3.1. An LP on measures. Consider the infinite-dimensional LP:

$$(3.4) \quad \begin{aligned} \tau'_\infty &= \inf_{\phi \in \mathcal{M}(X, Y)_+} \int T d\phi \\ \text{s.t.} \quad &\int \mathbf{x}^{d_x} y^{d_y} d\phi = \mu_d, \quad d = (d_x, d_y) \in D. \end{aligned}$$

Theorem 3.1. *The measure $\mu = \delta_{f(\mathbf{x})}(dy) \lambda(d\mathbf{x}) \in \mathcal{M}(X, Y)_+$ is the unique optimal solution of the infinite-dimensional LP (3.4).*

Proof. Let $\phi \in \mathcal{M}(X, Y)_+$ be an arbitrary feasible solution of (3.4), with moments $(\phi_d)_{d \in \mathbb{N}^{n+1}}$. Then:

$$\begin{aligned} \int T(\mathbf{x}, y) d\phi &= \sum_{d=(d_x, d_y) \in \mathbb{N}^{n+1}} \theta_d \phi_d \quad [\text{by (3.3)}] \\ &\geq \sum_{d=(d_x, d_y) \in \mathbb{N}^{n+1}} \theta_d \mu_d \quad [\text{by Theorem 2.1}] \\ &= \int T(\mathbf{x}, y) d\mu, \quad [\text{by (3.3)}], \end{aligned}$$

which implies that μ is an optimal solution of LP (3.4). As measures on compact sets are moment determinate, uniqueness follows. \square

Notice that the criterion $\int T d\phi = \mathcal{T}(\phi)$ (a weighted ℓ_1 -norm of the moment vector of ϕ) is *sparsity-inducing*. Indeed it induces a sparse support for the unique optimal solution of LP (3.4). Again, by sparse we mean that the support of μ is degenerate and a meager (thin) set of $X \times Y$ (since it is the graph $(\mathbf{x}, f(\mathbf{x}))$).

Remark 2. A parallel can be drawn with super-resolution in signal processing [7, 6, 8], which consists of recovering an atomic measure with sparse finite support (a few atoms), from the sole knowledge of only a limited number of its moments. Indeed as shown in [6], the sparse atomic measure (signal) to recover is the unique optimal solution of an LP on measures with (sparsity-inducing) total-variation criterion. In addition, solving the latter LP reduces to solving a single SDP in the univariate case.

In our graph-recovering inverse problem, our knowledge is also limited to a few moments only, namely $(\mu_{d,0}, \mu_{d,1})_{d \in \mathbb{N}^n}$ out of the potentially many $(\mu_{d_x, d_y})_{(d_x, d_y) \in \mathbb{N}^{n+1}}$. (As the object to recover is infinite-dimensional one cannot expect exact recovery from finitely many moments.) Hence the spirit of the infinite-dimensional LP (3.4) is that a measure with sparse support (the graph of a function) can be recovered from a few moments only, as the unique optimal solution of an LP on measures with appropriate sparsity-inducing criterion.

3.2. A dual of (3.4). Let $\mathcal{C}(X \times Y)$ be the space of continuous functions on $X \times Y$. Observe that $T \in \mathcal{C}(X \times Y)$. Indeed let $(\mathbf{x}_n, y_n)_{n \in \mathbb{N}} \subset X \times Y$ be such that $(\mathbf{x}_n, y_n) \rightarrow (\mathbf{x}, y) \in X \times Y$, as $n \rightarrow \infty$. Let $a_d^n := \mathbf{x}_n^{d_x} y_n^{d_y}$ for all n and all $d \in \mathbb{N}^{n+1}$, so that for all d , $a_d^n \rightarrow a_d := \mathbf{x}^{d_x} y^{d_y}$ as $n \rightarrow \infty$. Moreover $|a_d^n| \leq y^{d_y}$ for all n , and $\sum_d \theta_d y^{d_y} < \infty$. Therefore by Dominated Convergence [1, p. 49]:

$$T(\mathbf{x}_n, y_n) = \sum_d \theta_d \mathbf{x}_n^{d_x} y_n^{d_y} \rightarrow \sum_d \theta_d \mathbf{x}^{d_x} y^{d_y} = T(\mathbf{x}, y) \quad \text{as } n \rightarrow \infty.$$

Next, consider the infinite-dimensional LP problem:

$$(3.5) \quad \tau_\infty^* = \sup_{h_1, h_2 \in \mathcal{C}(X \times Y)} \int_X (h_1(\mathbf{x}) + h_2(\mathbf{x}) f(\mathbf{x})) d\mathbf{x} \\ h_1(\mathbf{x}) + h_2(\mathbf{x}) y \leq T(\mathbf{x}, y), \quad \forall (\mathbf{x}, y) \in X \times Y.$$

Theorem 3.2. *There is no duality gap between (3.4) and its dual (3.5).*

Proof. The LP (3.4) reads:

$$\tau_\infty' = \inf_{\phi \in \mathcal{M}(X \times Y)_+} \{ \langle T, \phi \rangle : \mathbf{A}_1 \phi = \lambda; \mathbf{A}_2 \phi = \mathbf{A}_2 \mu \},$$

where $\mathbf{A}_i : \mathcal{M}(X \times Y) \rightarrow \mathcal{M}(X)$, $i = 1, 2$, are defined by:

$$\phi \mapsto \mathbf{A}_1 \phi(B) = \int_{B \times Y} d\phi \quad B \in \mathcal{B}(X) \\ \phi \mapsto \mathbf{A}_2 \phi(B) = \int_{B \times Y} y d\phi(\mathbf{x}, y), \quad B \in \mathcal{B}(X).$$

The adjoints $\mathbf{A}_i^* : \mathcal{C}(X) \rightarrow \mathcal{C}(X \times Y)$ are defined by:

$$h \mapsto \mathbf{A}_1^* h(\mathbf{x}, y) = h(\mathbf{x}); \quad h \mapsto \mathbf{A}_2^* h(\mathbf{x}, y) = y h(\mathbf{x}), \quad \forall (\mathbf{x}, y) \in X \times Y,$$

for all $h \in \mathcal{C}(X)$. The dual (3.5) reads:

$$\tau_\infty^* = \sup_{h_1, h_2 \in \mathcal{C}(X)} \{ \langle h_1, \lambda \rangle + \langle h_2, \mathbf{A}_2 \mu \rangle : \\ T - \mathbf{A}_1^* h_1 + \mathbf{A}_2^* h_2 \in \mathcal{C}(X \times Y)_+ \}$$

Equivalently:

$$\tau_\infty^* = \sup_{h_1, h_2 \in \mathcal{C}(X)} \left\{ \int_X (h_1(\mathbf{x}) + h_2(\mathbf{x}) f(\mathbf{x})) d\mathbf{x} : \right. \\ \left. h_1(\mathbf{x}) + y h_2(\mathbf{x}) \leq T(\mathbf{x}, y), \quad \forall (\mathbf{x}, y) \in X \times Y \right\}.$$

To prove that $\tau'_\infty = \tau_\infty^*$ it suffices to prove that the set

$$\mathcal{M} := \{(\mathbf{A}_1 \phi, \mathbf{A}_2 \phi, \langle T, \phi \rangle) : \phi \in \mathcal{M}(X \times Y)_+\}$$

is closed in the weak-star topology $\sigma(\mathcal{M}(X)^2 \times \mathbb{R}, \mathcal{C}(X)^2 \times \mathbb{R})$, see [5, Theorem IV.7.2].

So let $(\mathbf{A}_1 \phi_n, \mathbf{A}_2 \phi_n, \langle T, \phi_n \rangle) \rightarrow (a, b, c)$ as $n \rightarrow \infty$. From the definition of \mathbf{A}_1 and the weak-star convergence, one obtains $\phi_n(X \times Y) \rightarrow a(X \times Y) \geq 0$. Hence the sequence $(\phi_n)_{n \in \mathbb{N}}$ is norm-bounded. As $X \times Y$ is compact there is a measure $\psi \in \mathcal{M}(X \times Y)_+$ and a subsequence $(n_k)_{k \in \mathbb{N}}$ such that $\phi_{n_k} \rightarrow \psi$ (weak-star) as $k \rightarrow \infty$. But then

$$\langle h, a \rangle = \lim_{k \rightarrow \infty} \langle h, \mathbf{A}_1 \phi_{n_k} \rangle = \lim_{k \rightarrow \infty} \langle \underbrace{\mathbf{A}_1^* h}_{\in \mathcal{C}(X \times Y)}, \phi_{n_k} \rangle = \langle \mathbf{A}_1^* h, \psi \rangle = \langle h, \mathbf{A}_1 \psi \rangle, \quad \forall h \in \mathcal{C}(X),$$

which implies $a = \mathbf{A}_1 \psi$. Similarly,

$$\langle h, b \rangle = \lim_{k \rightarrow \infty} \langle h, \mathbf{A}_2 \phi_{n_k} \rangle = \lim_{k \rightarrow \infty} \langle \underbrace{\mathbf{A}_2^* h}_{\in \mathcal{C}(X \times Y)}, \phi_{n_k} \rangle = \langle \mathbf{A}_2^* h, \psi \rangle = \langle h, \mathbf{A}_2 \psi \rangle, \quad \forall h \in \mathcal{C}(X),$$

which implies $b = \mathbf{A}_2 \psi$. Finally, as $T \in \mathcal{C}(X \times Y)$, $c = \lim_{k \rightarrow \infty} \langle T, \phi_{n_k} \rangle = \langle T, \psi \rangle$ and therefore $(a, b, c) = (\mathbf{A}_1 \psi, \mathbf{A}_2 \psi, \langle T, \psi \rangle)$ for some $\psi \in \mathcal{M}(X \times Y)_+$, the desired result. \square

So as for super-resolution for atomic measure where the dual problem provides a certificate, the dual (3.5) indicates what would be a certificate if it has an optimal solution.

Indeed assume that $(h_1^*, h_2^*) \in \mathcal{C}(X \times Y)$ is an optimal solution of (3.5). Then

$$\int_X (T(\mathbf{x}, f(\mathbf{x})) - h_1^*(\mathbf{x}) - h_2^*(\mathbf{x}) f(\mathbf{x})) d\lambda(\mathbf{x}) = 0.$$

Equivalently:

$$\begin{aligned} T(\mathbf{x}, f(\mathbf{x})) &= h_1^*(\mathbf{x}) + h_2^*(\mathbf{x}) f(\mathbf{x}), \quad \text{a.e. in } X. \\ T(\mathbf{x}, y) &\geq h_1^*(\mathbf{x}) + h_2^*(\mathbf{x}) y, \quad \forall (\mathbf{x}, y) \in X \times Y. \end{aligned}$$

Example 1. $X = [0, 1]$; $Y = [0, 1]$ and $f(\mathbf{x}) = 1$. We can take

$$T(\mathbf{x}, y) = \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \frac{y^j \mathbf{x}^k}{j! k!} = \exp(\mathbf{x}) \exp(y), \quad (\mathbf{x}, y) \in [0, 1] \times [0, 1].$$

Then with $h_1^* = 0$, $h_2^* := e \cdot \exp(\mathbf{x})$ we obtain

$$T(\mathbf{x}, f(\mathbf{x})) = e \cdot \exp(\mathbf{x}) = f(\mathbf{x}) \cdot h_2^*(\mathbf{x}); \quad T(\mathbf{x}, y) = \exp(\mathbf{x}) \cdot \exp(y) \geq e \cdot \exp(\mathbf{x}) y$$

as $\exp(y) \geq e y$ for all $y \in [0, 1]$.

Unfortunately solving the infinite-dimensional LP (3.4) is out of reach. However Theorem 2.1 still suggests a practical numerical scheme which in principle allow to approximate as closely as desired moments, any finite number of moments of μ in (2.1).

4. ASYMPTOTIC RECOVERY

In this section we provide a systematic numerical scheme to recover asymptotically either an arbitrary (but fixed) number of moments of μ , or all moments of μ .

4.1. Recovery of finitely many moments. Given an integer $s \in \mathbb{N}$ fixed, and a sequence $\phi = (\phi_d)_{d \in \mathbb{N}^{n+1}}$, let

$$(4.1) \quad \mathcal{I}_s(\phi) := \sum_{d \in \mathbb{N}_s^{n+1}} \phi_{2d} = \text{trace}(\mathbf{M}_s(\phi)),$$

and consider the following hierarchy of semidefinite programs:

$$(4.2) \quad \begin{aligned} \tau_r = \inf_{\phi} \quad & \mathcal{I}_s(\phi) \\ \text{s.t.} \quad & \mathbf{M}_r(\phi) \succeq 0, \quad \mathbf{M}_{r-d_j}(g_j \phi) \succeq 0, \quad j = 1, \dots, m+1, \\ & \phi_d = \mu_d, \quad d \in D, \end{aligned}$$

indexed by the relaxation order $r \geq d_0 = \max(s, \max_j d_j)$, and where $\phi = (\phi_d)_{d \in \mathbb{N}_{2r}^{n+1}}$.

Theorem 4.1. *The semidefinite program (4.2) has an optimal solution $\phi^r = (\phi_d^r)_{d \in \mathbb{N}_{2r}^{n+1}}$, and $\tau_r = \mathcal{I}_s(\phi^r) \leq \mathcal{I}_s(\mu)$ for all $r \geq d_0$. In addition, $\lim_{r \rightarrow \infty} \mathcal{I}_s(\phi^r) = \mathcal{I}_s(\mu)$ and*

$$(4.3) \quad \lim_{r \rightarrow \infty} \phi_d^r = \mu_d \quad \forall d \in \mathbb{N}_{2s}^{n+1}.$$

Moreover if ν is a measure on $X \times Y$ such that $\nu_d = \mu_d$ for all $d \in D$ then $\mathcal{I}_s(\mu) \leq \mathcal{I}_s(\nu)$.

In other words, with a suitable choice of an objective function, the moment completion problem is solved asymptotically via the hierarchy of semidefinite relaxations (4.2). That is, arbitrary finitely many moments of a graph can be recovered from linear measurements, i.e. from the sole knowledge of zero-th and first order moments.

The proof of Theorem 4.1 is relegated to Section 8.

4.2. Recovery of all moments. Theorem 4.1 states that finitely many moments of the graph can be recovered. This result can be extended to recover (asymptotically) infinitely many moments. Recall the weighted criterion:

$$\mathcal{T}(\phi) := \sum_{d \in \mathbb{N}^{n+1}} \theta_d \phi_d$$

already considered in (3.2). When $\phi = (\phi_d)_{d \in \mathbb{N}_{2r}^{n+1}}$ then $\mathcal{T}(\phi)$ is the obvious finite truncation

$$\mathcal{T}(\phi) = \sum_{d \in \mathbb{N}_{2r}^{n+1}} \theta_d \phi_d.$$

Next, consider the following hierarchy of semidefinite programs:

$$(4.4) \quad \begin{aligned} \tau'_r = \inf_{\phi} \quad & \mathcal{T}(\phi) : \\ \text{s.t.} \quad & \mathbf{M}_r(\phi) \succeq 0, \quad \mathbf{M}_{r-d_j}(g_j \phi) \succeq 0, \quad j = 1, \dots, m+1, \\ & \phi_d = \mu_d, \quad d \in D, \end{aligned}$$

indexed by $r \geq d_0 := \max_j d_j$. Notice that in contrast to (4.2), the cost function in (4.4) changes with the index r as more and more terms of ϕ are taken into account. Of course as μ is a feasible solution, we also have $\tau'_d \leq \mathcal{T}(\mu)$ for every r .

Theorem 4.2. *The semidefinite program (4.4) has an optimal solution $\phi^r = (\phi_d^r)_{d \in \mathbb{N}_{2r}^{n+1}}$, and $\tau_r' = \mathcal{T}(\phi^r) \leq \mathcal{T}(\mu)$ for all $r \geq d_0$. In addition, $\lim_{r \rightarrow \infty} \mathcal{T}(\phi^r) = \mathcal{T}(\mu)$ and*

$$(4.5) \quad \lim_{r \rightarrow \infty} \phi_d^r = \mu_d \quad \forall d \in \mathbb{N}^{n+1}.$$

The proof of Theorem 4.2 is relegated to Section 8.

5. APPLICATION TO THE L-MOMENT PROBLEM WITH UNKNOWN SUPPORT

In the generalized L-moment problem with $\mathbf{K} \subset \mathbb{R}^n$ known, the goal is to retrieve a measure ϕ on \mathbf{K} with density $0 \leq h \leq L$ in $\mathcal{L}^\infty(\mathbf{K})_+$, whose moments $\phi = (\phi_{d_x})_{d_x \in D} = (\int_{\mathbf{K}} \mathbf{x}^{d_x} d\phi)_{d_x \in D}$, are given, and where $D \subset \mathbb{N}^n$ is also given. The classical univariate L-moment problem on \mathbb{R} was solved in [2].

So when \mathbf{K} is known and moments of Lebesgue measure on \mathbf{K} are known as well, then we have seen how to apply the moment-completion approach to recover the unknown density in $h \in \mathcal{L}(\mathbf{K})_+$; see Section 4 with $X = \mathbf{K}$.

When \mathbf{K} is unknown this is just the \mathbf{K} -moment problem with the additional information that ϕ has a bounded density with respect to Lebesgue measure on the unknown \mathbf{K} . We next show how to formulate this problem as a completion problem investigated in previous sections.

Assuming \mathbf{K} is compact, let X be a box such that $\mathbf{K} \subset X$. Then

$$(5.1) \quad \mu_{d_x} = \int_{\mathbf{K}} \mathbf{x}^{d_x} d\phi = \int_X \mathbf{x}^{d_x} \underbrace{1_{\mathbf{K}}(\mathbf{x}) h(\mathbf{x})}_{f(\mathbf{x})} d\mathbf{x}, \quad d_x \in D.$$

Formulation (5.1) for the full L-moment problem (i.e., where $D = \mathbb{N}^n$) is exactly the framework developed in previous sections where $\mathbf{x} \mapsto f(\mathbf{x}) := 1_{\mathbf{K}}(\mathbf{x}) h(\mathbf{x}) \in \mathcal{L}^\infty(X)_+$, is the function whose graph is to be recovered. Indeed, letting $d\mu(\mathbf{x}, y) = \delta_{1_{\mathbf{K}}(\mathbf{x}) h(\mathbf{x})}(dy) d\mathbf{x}$,

$$(5.2) \quad \mu_{d_x, 0} = \int_X \mathbf{x}^{d_x} d\mathbf{x}, \quad d_x \in \mathbb{N}^n \quad \text{known in closed form}$$

$$(5.3) \quad \mu_{d_x, 1} = \int_{X \times Y} \mathbf{x}^{d_x} y d\mu := \int_X \mathbf{x}^{d_x} f(\mathbf{x}) d\mathbf{x} = \int_{\mathbf{K}} \mathbf{x}^{d_x} h(\mathbf{x}) d\mathbf{x} = \phi_{d_x}, \quad d_x \in \mathbb{N}^n.$$

So applying the moment completion approach developed earlier, one may approximate as closely as desired any finite number of moments of μ , and in a second-step to approximate the function $\mathbf{x} \mapsto 1_{\mathbf{K}}(\mathbf{x}) h(\mathbf{x})$.

It follows that the moment completion approach allows to both approximate the unknown support and density of the measure ϕ on \mathbf{K} .

The particular case when $h = 1$. When the density $h \in \mathcal{L}(\mathbf{K})_+$ is the indicator $1_{\mathbf{K}}$ then the above procedure described in Section 4 allows to approximate the indicator function $1_{\mathbf{K}}$ from knowledge of moments of the Lebesgue measure on \mathbf{K} (e.g. obtained from measurements). Notice that in this case (5.3) yields:

$$\mu_{d_x, d_y} = \int_X \mathbf{x}^{d_x} y^{d_y} d\mu(\mathbf{x}, y) = \int_X \mathbf{x}^{d_x} 1_{\mathbf{K}}(\mathbf{x})^{d_y} d\mathbf{x} = \mu_{d_x, 1}, \quad \forall (d_x, d_y) \in \mathbb{N}^{n+1}.$$

Therefore the matrix completion problem is trivial. There is no optimization involved. Only the second step of recovery via the Christoffel-Darboux kernel is needed. In particular the

moment matrix $\mathbf{M}_r(\mu)$ is highly singular since all columns indexed by (d_x, d_y) with $d_y > 1$ are identical to the column indexed by $(d_x, 1)$. So we may replace $\mathbf{M}_r(\mu)$ with the smaller size matrix $\widehat{\mathbf{M}}_r(\mu)$ with rows and columns indexed by monomials $(\mathbf{x}^{d_x})_{d_x \in \mathbb{N}_r^n}$ and $(\mathbf{x}^{d_x} y)_{d_x \in \mathbb{N}_{r-1}^n}$.

6. GRAPH APPROXIMATION WITH THE CHRISTOFFEL-DARBOUX KERNEL

Let us briefly recall one of the main results of [15] where the Christoffel-Darboux kernel is used to approximate the support of a measure concentrated on a graph given its moments.

Let $\mathbf{b}(\mathbf{x}, y)$ be a polynomial vector of degree at most d whose elements are orthonormal with respect to the bilinear form induced by a measure which is absolutely continuous with respect to the uniform measure on $X \times Y$. Let

$$\mathbf{M}_r := \int_X \mathbf{b}(\mathbf{x}, f(\mathbf{x})) \mathbf{b}(\mathbf{x}, f(\mathbf{x}))^\top d\mathbf{x} = \int_{X \times Y} \mathbf{b}(\mathbf{x}, y) \mathbf{b}(\mathbf{x}, y)^\top d\mu(\mathbf{x}, y)$$

be the moment matrix of order d of the measure $d\mu(\mathbf{x}, y) := \mathbb{I}_X(\mathbf{x}) d\mathbf{x} \delta_{f(\mathbf{x})}(dy)$ concentrated on the graph $\{(\mathbf{x}, f(\mathbf{x})) : \mathbf{x} \in X\} \subset X \times Y$. Given \mathbf{M}_r , we want to compute an approximation f_r of function f , with convergence guarantees when $r \rightarrow \infty$.

Given a regularization parameter $\beta_r := 2^{3-\sqrt{r}}$, define the Christoffel-Darboux (CD) polynomial

$$q_r(\mathbf{x}, y) := \mathbf{b}(\mathbf{x}, y)^\top (\mathbf{M}_r + \beta_r \mathbf{I})^{-1} \mathbf{b}(\mathbf{x}, y).$$

Polynomial q_r can be computed numerically by a spectral decomposition of the positive semi-definite matrix \mathbf{M}_r . Instead of trying to approximate the (possibly discontinuous) function f with polynomials, we approximate it with a class of semi-algebraic functions. We define the semi-algebraic approximant

$$f_r(\mathbf{x}) := \min\{\operatorname{argmin}_{y \in Y} q_r(\mathbf{x}, y)\}$$

as the minimum of the argument of the minimum of the CD polynomial, always well defined since this polynomial is a sum of squares (SOS).

Theorem 6.1. [15, Theorem 1] *If the set $S \subset X$ of continuity points of f is such that $X \setminus S$ has Lebesgue measure zero, then $\lim_{r \rightarrow \infty} f_r(\mathbf{x}) = f(\mathbf{x})$ for almost all $\mathbf{x} \in X$ and $\lim_{r \rightarrow \infty} \|f - f_r\|_{\mathcal{L}^1(X)} = 0$.*

This graph approximation method is implemented in the `momgraph` function described in [15].

7. EXAMPLES

For our numerical examples we modeled our moment problems with GloptiPoly [10] and solved them with MOSEK [3] on Matlab. For reconstructing graphs from approximate moments we used the `momgraph` function described in [15].

7.1. Convex optimal transport. Let us first illustrate the CD kernel graph approximation method of Theorem 6.1 with a standard quadratic optimal transport problem, a particular case of optimization problem (1.2). Let $\mathbf{x} = x \in X := [-1, 1]$ and $y \in Y := [-1, 1]$ be scalar, and $c(x, y) = (x - y)^2/2$. Only marginal moments are available: $D = \{(d_x, 0) : d_x \in \mathbb{N}\} \cup \{(0, d_y) : d_y \in \mathbb{N}\}$ and we want to transport the density $d\lambda(x) = (1 - x)dx/2$ on X onto the density $(1 + y)dy/2$ on Y .

In this case there exists a unique transport map

$$f(x) = -1 + \sqrt{(1+x)(3-x)}$$

yielding the optimal cost

$$\int_X c(x, f(x)) d\lambda(x) = \frac{\pi}{2} - \frac{4}{3} \approx 0.23746.$$

relaxation order r	1	2	3	4	5	6	7
lower bound	0.22222	0.23494	0.23725	0.23742	0.23743	0.23745	0.23746

TABLE 1. Lower bounds on optimal cost at various relaxation orders r .

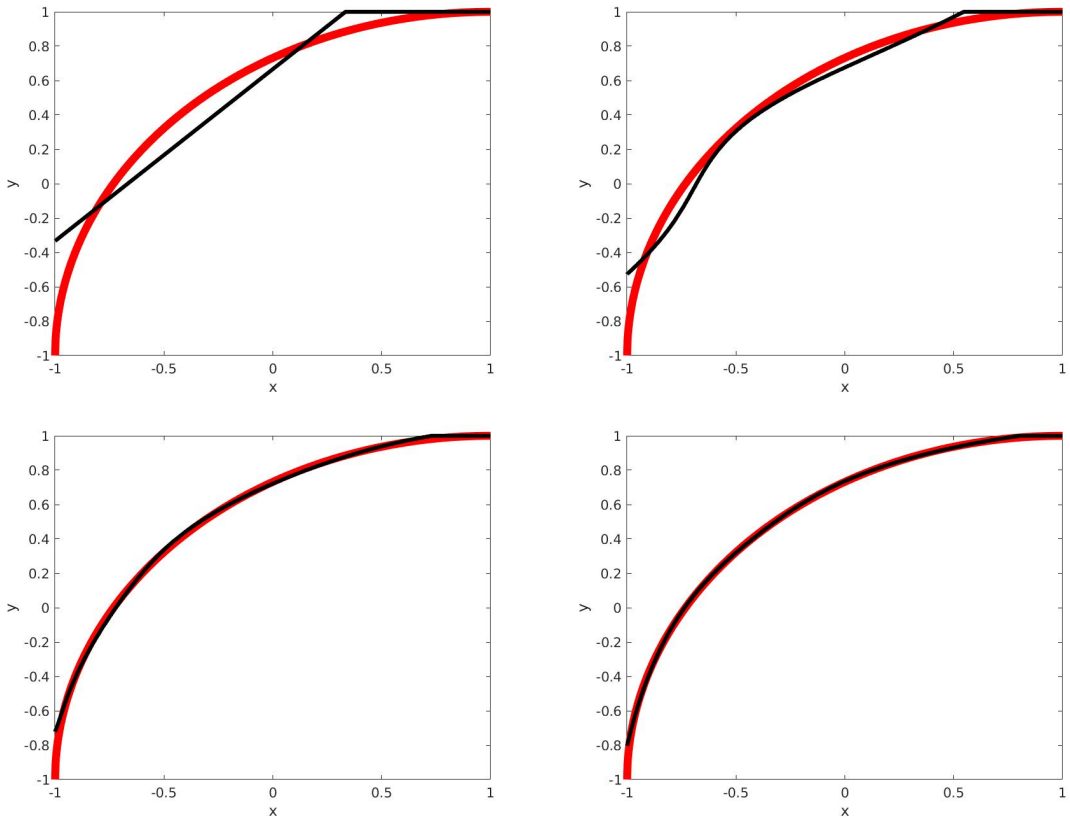


FIGURE 1. CD approximations (thin black) of optimal transport map (thick gray) for relaxation orders $r = 1$ (top left), $r = 2$ (top right), $r = 3$ (bottom left), $r = 4$ (bottom right).

In Table 1 we report the lower bounds of the moment relaxations for various relaxation orders r . In Figure 1 we represent the approximate graphs obtained from the CD kernel method described in Section 6.

7.2. Non-convex optimal transport. Let us now illustrate the CD kernel graph approximation method of Theorem 6.1 with a non-convex optimal transport problem. Let $\mathbf{x} = x \in X := [0, 1]$ and $y \in Y := [0, 1]$ be scalar, and $c(x, y) := (4x - y)xy$. Only marginal

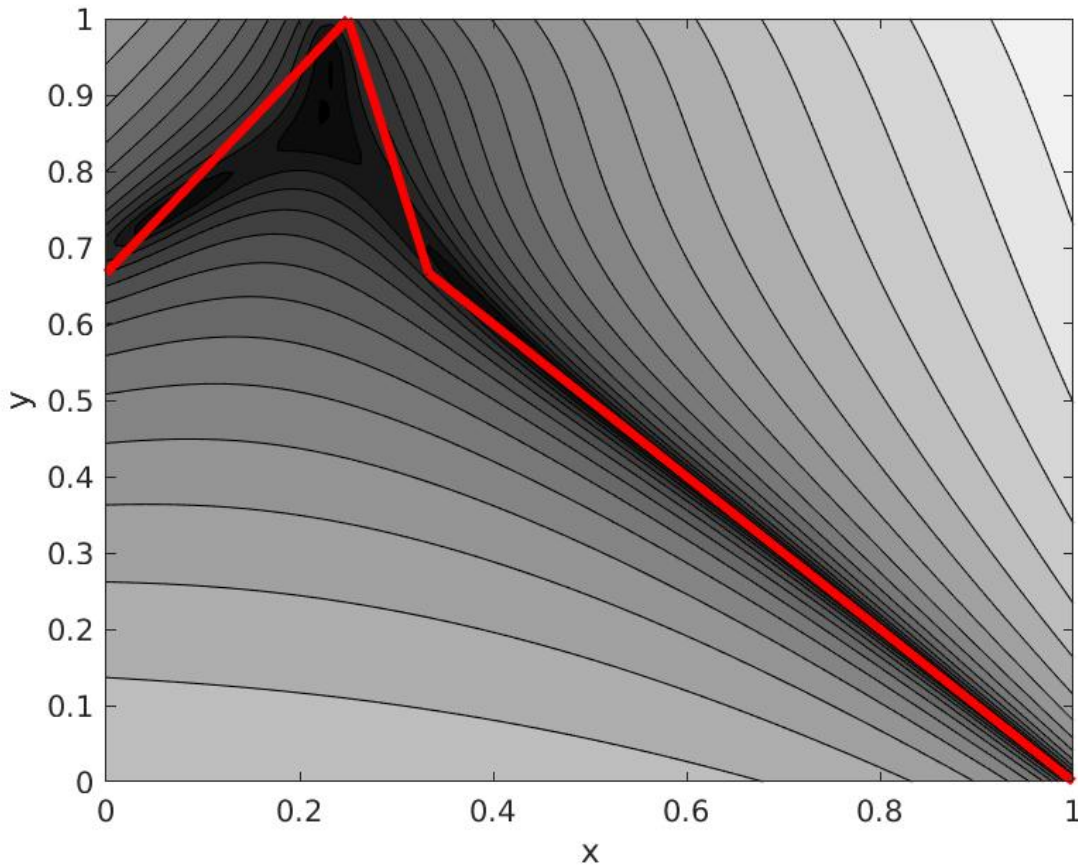


FIGURE 2. Logarithmic level sets (gray, smaller values darker) of the CD polynomial of degree 12 approximating the graph (thick red) of the optimal transport map.

moments are available: $D = \{(d_x, 0) : d_x \in \mathbb{N}\} \cup \{(0, d_y) : d_y \in \mathbb{N}\}$ and we want to transport the Lebesgue measure on X onto the Lebesgue measure on Y . An analytic optimal transport map is known [4, Section 5.3] for this example:

$$f(x) = \begin{cases} \frac{2}{3} + \frac{4}{3}x & \text{on } [0, \frac{1}{4}] \\ 2 - 4x & \text{on } [\frac{1}{4}, \frac{1}{3}] \\ 1 - x & \text{on } [\frac{1}{3}, 1] \end{cases}$$

yielding the optimal cost

$$\int_X c(x, f(x))dx = \frac{107}{432} \approx 0.24769.$$

Formulating problem (1.2) as a generalized problem of moments (1.3) with $\lambda(dx) = dx$, we can use the moment-SOS hierarchy to approximate numerically all the moments of the measure $dx\delta_{f(x)}(dy)$.

We solved the moment relaxation of order $r = 6$, corresponding to approximate moments of degree up to $2r = 12$ and a positive semidefinite moment matrix of size 28. The lower bound on the optimal cost obtained with the moment relaxation is 0.24741, matching 3 digits of the exact cost. The maximum absolute error between the approximate moments and the

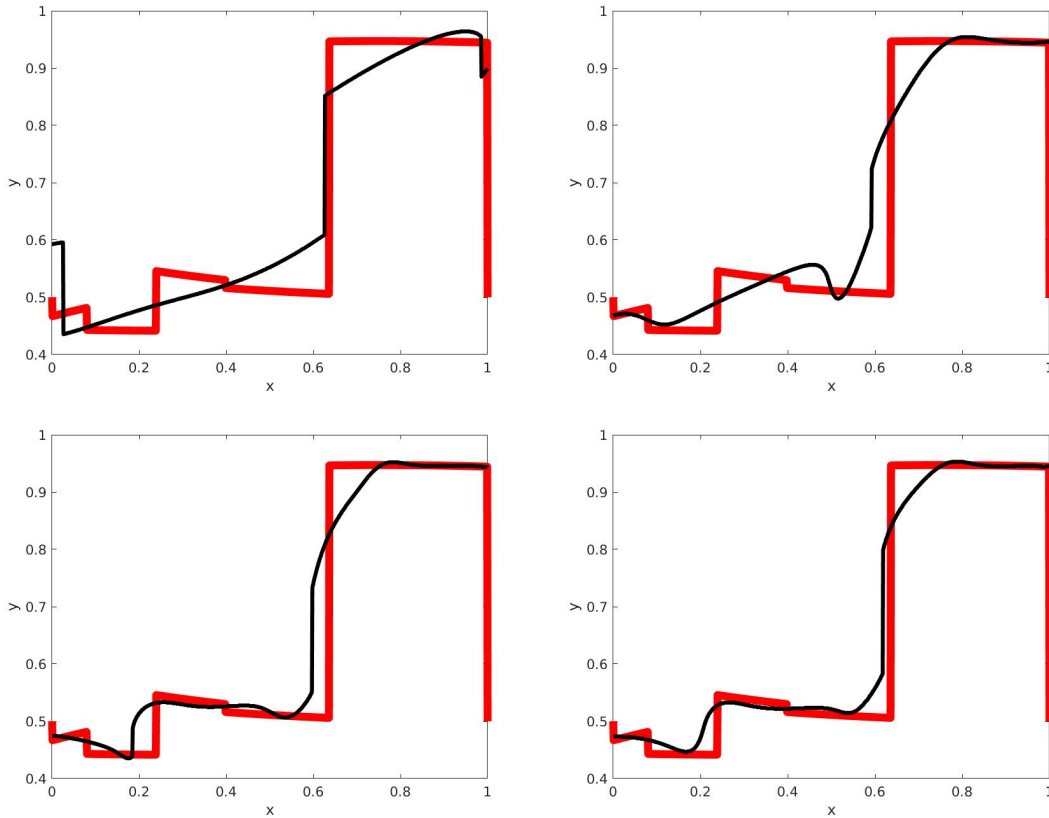


FIGURE 3. Discontinuous benchmark function (red) and its approximations (black) obtained from approximate moments recovered at relaxation order $r = 4$ (top left), $r = 8$ (top right), $r = 12$ (bottom left), $r = 16$ (bottom right).

exact moments is $9.6 \cdot 10^{-4}$. On Figure 2 we represent sublevel sets of the CD polynomial q_6 of degree 12. We observe that they closely match with the graph of the transport map.

7.3. Graph reconstruction from linear measurements. Example 67 of [9] was used in [15, Section 5.4] to illustrate the CD kernel graph approximation algorithm when all the moments are available. Let us revisit the same example to illustrate the reconstruction algorithm of Theorem 4.2, when θ is the truncated sequence of all ones.

Approximate graphs for various relaxation orders are reported in Figure 3.

7.4. Comparison with L_2 estimate. Consider the problem of reconstructing a step function

$$f(x) = \begin{cases} 0.1 & \text{if } 0 \leq x < 0.5 \\ 0.9 & \text{if } 0.5 \leq x \leq 1. \end{cases}$$

The only input data are the moments $\int_0^1 x^{d_x} dx$ and $\int_0^1 x^{d_x} f(x) dx$ for d_x less than a given degree $r \in \mathbb{N}$.

The reconstruction algorithm of Theorem 4.2, when θ is the truncated sequence of all ones, and $r = 12$ gives the approximate graph of the right of Figure 4. It is not perfect, especially at the discontinuity point, yet it compares favorably with the L_2 density estimate of degree

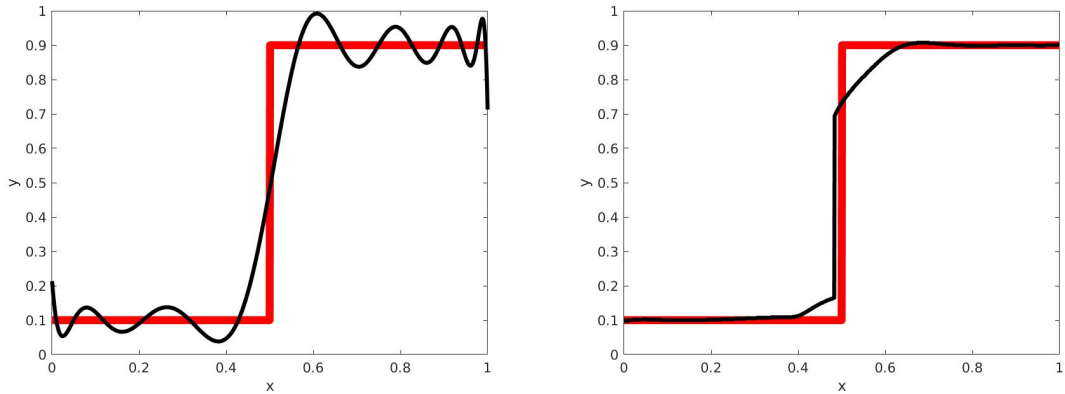


FIGURE 4. Approximations of the graph (thick red) of a step function. Left: L_2 estimate of degree 12 (black). Right: CD polynomial of degree 12 obtained from first order moments (black).

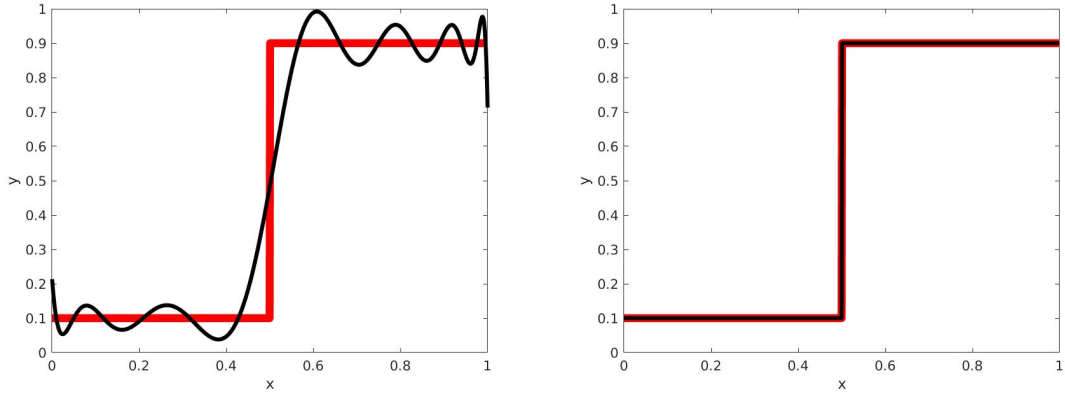


FIGURE 5. Approximations of the graph (thick red) of a step function. Left: L_2 estimate of degree 12 (black). Right: CD polynomial obtained from all moments of degree up to 4 (black).

12 of [11, Proposition 1] which oscillates around the graph to be recovered, as seen at the left of Figure 4.

For comparison, we report on the right of Figure 5 the graph obtained with the CD polynomial constructed from moments up to degree 4, as already studied in [15, Example 3, Figure 2]. We observe that in this case the graph approximation is almost perfect, thanks to the additional information provided by the higher degree moments.

7.5. Disconnected L-moment problem. This example illustrates our solution in Section 5 to the L-moment problem when $h = 1$ and hence the graph to be reconstructed is the indicator function of an unknown set \mathbf{K} . To compute the moments we choose a disconnected set $\mathbf{K} := [1/5, 2/5] \cup [3/5, 4/5] \subset X := [0, 1]$. Solving the moment relaxation of order $r = 10$ we obtain the approximate graph represented on Figure 6. We observe that it matches closely the indicator function of \mathbf{K} .

The level sets of the CD polynomial (obtained with the `momgraph` function with regularization parameter 10^{-6}) are represented on the left of Figure 7. We see that the CD polynomial

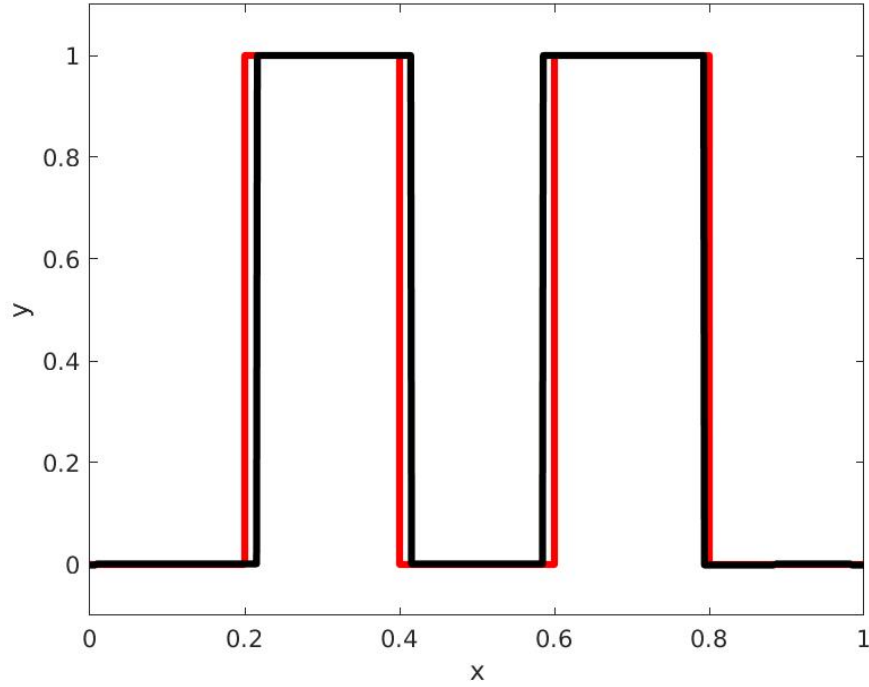


FIGURE 6. Indicator function (red) and its approximation (black) obtained from approximate moments recovered at relaxation order $r = 10$.

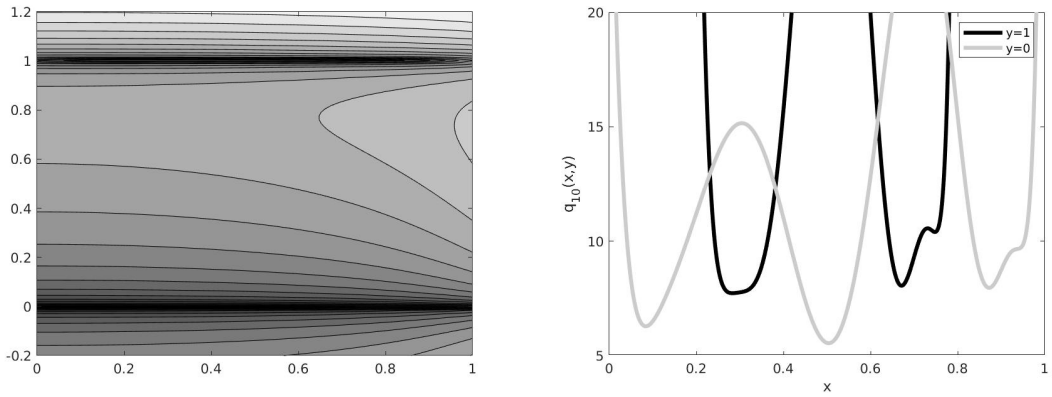


FIGURE 7. Logarithmic level sets (gray, smaller values darker) of the CD polynomial $q_{10}(x, y)$ of degree 20 (left), and its values at $y = 0$ (light gray) and $y = 1$ (black).

$q_{10}(x, y)$ is small around $y = 0$ and $y = 1$, as expected. For each x , the value of y for which $q_{10}(x, y)$ is minimal is the approximation $f_{10}(x)$ of the indicator function of \mathbf{K} . On the right of Figure 7 we represent $q_{10}(x, y)$ for $y = 0$ and $y = 1$. For each x , the value $f_{10}(x)$ is the minimum of $q_{10}(x, 0)$ and $q_{10}(x, 1)$.

8. PROOFS

Proof of Theorem 2.1. As $\phi_{d_x,0} = \lambda_{d_x}$ for all $d_x \in \mathbb{N}^n$ and $X \subset [0,1]^n$, the marginal of ϕ on X is the measure λ . Next, disintegrate ϕ with respect to its marginal λ on X and its conditional $\varphi(\cdot|\mathbf{x})$ on Y given $\mathbf{x} \in X$, i.e.:

$$\phi(A \times B) = \int_A \varphi(B|\mathbf{x}) \lambda(d\mathbf{x}), \quad A \in \mathcal{B}(X), B \in \mathcal{B}(Y).$$

In particular,

$$\phi_{d_x,1} = \int_{X \times Y} \mathbf{x}^{d_x} y d\mu(\mathbf{x}, y) = \int_X \mathbf{x}^{d_x} \left(\underbrace{\int_Y y \varphi(dy|\mathbf{x})}_{h_1(\mathbf{x})} \right) d\lambda(\mathbf{x}) = \int_X \mathbf{x}^{d_x} h_1(\mathbf{x}) d\lambda(\mathbf{x})$$

for some measurable function h_1 on X . Therefore

$$\int_X \mathbf{x}^{d_x} f(\mathbf{x}) d\lambda(\mathbf{x}) = \mu_{d_x,1} = \phi_{d_x} = \int_X \mathbf{x}^{d_x} h_1(\mathbf{x}) d\lambda(\mathbf{x}), \quad \forall d_x \in \mathbb{N}^n.$$

As signed measures on compact sets are moment-determinate, one deduces that $h_1 = f$ for λ -almost-all \mathbf{x} in X . Next, fix $(d_x, d_y) \in \mathbb{N}^{n+1}$, arbitrary and recall that $X \subset [0,1]^n \subset \mathbb{R}_+^n$. Then

$$\begin{aligned} \phi_{d_x, d_y} &= \int \mathbf{x}^{d_x} y^{d_y} d\phi(\mathbf{x}, y) = \int_X \mathbf{x}^{d_x} \left(\int_Y y^{d_y} \varphi(dy|\mathbf{x}) \right) d\lambda(\mathbf{x}) \\ &\geq \int_X \mathbf{x}^{d_x} \left(\int_Y y \varphi(dy|\mathbf{x}) \right)^{d_y} d\lambda(\mathbf{x}) \quad [\text{by Jensen's inequality}] \\ (8.1) \quad &\geq \int_X \mathbf{x}^{d_x} h_1(\mathbf{x})^{d_y} d\lambda(\mathbf{x}) = \int_X \mathbf{x}^{d_x} f(\mathbf{x})^{d_y} d\lambda(\mathbf{x}) = \mu_{d_x, d_y}. \end{aligned}$$

(We can apply Jensen's inequality because y^{d_y} is convex on $Y \subset \mathbb{R}_+$ and $X \subset \mathbb{R}_+^n$.) Therefore with $r \in \mathbb{N}$ fixed, arbitrary, summing up over all $(2d_x, 2d_y) \in \mathbb{N}_{2r}^{n+1}$ yields

$$\text{trace } \mathbf{M}_r(\phi) = \sum_{(d_x, d_y) \in \mathbb{N}_r^{n+1}} \phi_{2d_x, 2d_y} \geq \sum_{(d_x, d_y) \in \mathbb{N}_r^{n+1}} \mu_{2d_x, 2d_y} = \text{trace } \mathbf{M}_r(\mu),$$

the desired result.

(Concerning Remark 1. If $X \not\subset \mathbb{R}_+^n$ the above arguments still hold whenever d_x is even (hence $\mathbf{x}^{d_x} \geq 0$) so that Jensen's inequality applies.) \square

Proof of Theorem 4.1. (i) Let $\phi = (\phi_{d_x, d_y})_{(d_x, d_y) \in \mathbb{N}_{2r}^{n+1}}$ be an arbitrary feasible solution. From $\mathbf{M}_r(g_j \phi) \succeq 0$, we may deduce

$$(8.2) \quad L_\phi(y^{2r}) \leq \gamma^{2r}; \quad L_\phi(x_i^{2r}) \leq 1 \quad i = 1, \dots, n.$$

Hence by [14, Proposition 2.38, p. 41], $|\phi_{d_x, d_y}| \leq \max[1, \gamma^{2r}] =: \rho_r$ for all $(d_x, d_y) \in \mathbb{N}_{2r}^{n+1}$, which implies that that feasible set is compact and (4.2) has an optimal solution ϕ^r with $\mathcal{T}_s(\phi^r) = \tau_r$.

(ii) Let $(\phi^r)_{r \in \mathbb{N}}$ be a sequence of arbitrary optimal solutions of (4.2) as r increases. With each r fixed, recalling the definition of ρ_r in the proof of (i), one has $\phi_{0,0}^r = 1$ and:

$$|\phi_d^r| \leq \rho_t, \quad \forall 2t-1 \leq |d| \leq 2t; \quad t = 1, \dots, r.$$

So for each $r \in \mathbb{N}$ let

$$|\psi_d^r| := \frac{|\phi_d^r|}{\rho_t}, \quad \text{for all } 2t-1 \leq |d| \leq 2t; t = 1, \dots, r.$$

By construction $|\psi_d^r| < 1$ for all $d \in \mathbb{N}_{2r}^{n+1}$. Hence by completing the finite sequence $\boldsymbol{\psi}^r$ with zeros to make it an infinite sequence indexed by \mathbb{N}^{n+1} , the infinite sequence $\boldsymbol{\psi}^r = (\psi_d^r)_{d \in \mathbb{N}^{n+1}}$ is an element of the unit ball \mathbf{B}_1 of the Banach space ℓ_∞ of uniformly bounded sequence. As this unit ball is weak-star sequentially compact, there exists an element $\boldsymbol{\psi}^* \in \mathbf{B}_1$ and a subsequence $(r_j)_{j \in \mathbb{N}}$ such that $\boldsymbol{\psi}^{r_j} \xrightarrow{\text{weak}^*} \boldsymbol{\psi}^*$ as $j \rightarrow \infty$. In particular,

$$\lim_{j \rightarrow \infty} \psi_d^{r_j} = \psi_d^*, \quad \forall d \in \mathbb{N}^{n+1},$$

and equivalently:

$$(8.3) \quad \lim_{j \rightarrow \infty} \phi_{k,\alpha}^{r_j} = \rho_t \psi_d^* =: \phi_d^*, \quad \forall 2t-1 \leq |d| \leq 2t; t = 1, 2, \dots$$

In addition, the latter convergence (8.3) implies $\phi_d^* \geq 0$ for all $d \in \mathbb{N}^{n+1}$, and

$$\mathbf{M}_r(g_j \boldsymbol{\phi}^*) \succeq 0, \quad j = 0, 1, \dots, m+1, \forall r \in \mathbb{N}.$$

Next, recall that the quadratic module $Q(g)$ in (2.3) is Archimedean and therefore by Putinar's Positivstellensatz [17], $\boldsymbol{\phi}^*$ has a representing measure ϕ^* on $X \times Y$. Moreover, by (8.3) again,

$$\phi_{d_x,0}^* = \lambda_{d_x} \quad \text{and} \quad \phi_{d_x,1}^* = \mu_{d_x,1}, \quad \forall d_x \in \mathbb{N}^n,$$

and $\lim_{j \rightarrow \infty} \tau_{r_j} = \lim_{j \rightarrow \infty} \mathcal{I}_s(\boldsymbol{\phi}^{r_j}) = \mathcal{I}_s(\boldsymbol{\phi}^*)$. Therefore since $\tau_{r_j} \leq \mathcal{I}_s(\boldsymbol{\mu})$ for all j , we obtain $\mathcal{I}_s(\boldsymbol{\phi}^*) \leq \mathcal{I}_s(\boldsymbol{\mu})$. But by Lemma 2.1, one concludes that $\mathcal{I}_s(\boldsymbol{\phi}^*) = \mathcal{I}_s(\boldsymbol{\mu})$, and $\phi_d^* = \mu_d$ for all $d \in \mathbb{N}_{2s}^{n+1}$, i.e.,

$$\lim_{j \rightarrow \infty} \phi_d^{r_j} = \mu_d = \int_{\Omega} \mathbf{x}^{d_x} f(\mathbf{x})^{d_y} d\lambda(\mathbf{x}), \quad \forall d = (d_x, d_y) \in \mathbb{N}_{2s}^{n+1}.$$

As this limit does not depend on the particular converging subsequence $(\boldsymbol{\phi}^{r_j})_{j \in \mathbb{N}}$, we also obtain the full convergence (4.3). \square

Proof of Theorem 4.2. The proof of (i) is almost a verbatim copy of that of Theorem 4.1(i).

(ii) Again as in the proof of Theorem 4.1(ii) there is a subsequence $(r_j)_{j \in \mathbb{N}}$ and a sequence $\boldsymbol{\phi}^*$ such that (8.3) holds and $\boldsymbol{\phi}^*$ has a representing measure on $X \times Y$. In addition, recall that (8.2) holds for $\boldsymbol{\phi}^{r_j}$, and so $\phi_d^{r_j} \leq \gamma^{2\lceil d_y/2 \rceil}$ for all $d = (d_x, d_y) \in \mathbb{N}^{n+1}$.

Hence letting $\boldsymbol{\psi} := (\psi_d)_{d \in \mathbb{N}^{n+1}}$ with $\psi_d = \gamma^{2\lceil d_y/2 \rceil}$, $\boldsymbol{\phi}^{r_j} \leq \boldsymbol{\psi}$ for all j . Moreover notice that $\mathcal{I}(\boldsymbol{\psi}) < \infty$. Therefore by the Dominated Convergence Theorem [1, p. 49], $\mathcal{I}(\boldsymbol{\phi}^{r_j}) \rightarrow \mathcal{I}(\boldsymbol{\phi}^*)$, as $j \rightarrow \infty$. But recall that $\mathcal{I}(\boldsymbol{\phi}^r) \leq \mathcal{I}(\boldsymbol{\mu})$ for all r , and therefore $\mathcal{I}(\boldsymbol{\phi}^*) \leq \mathcal{I}(\boldsymbol{\mu})$.

On the other hand, by Theorem 2.1, the reverse inequality $\mathcal{I}(\boldsymbol{\phi}^*) \geq \mathcal{I}(\boldsymbol{\mu})$ holds (since $\phi_d^* \geq \mu_d$ for all d), which yields $\mathcal{I}(\boldsymbol{\phi}^*) = \mathcal{I}(\boldsymbol{\mu})$. But from $\mathcal{I}(\boldsymbol{\phi}^* - \boldsymbol{\mu}) = 0$ we obtain

$$0 = \sum_{d \in \mathbb{N}^{n+1}} \theta_d \underbrace{(\phi_d^* - \mu_d)}_{\geq 0} \Rightarrow \phi_d^* = \mu_d, \quad \forall d \in \mathbb{N}^{n+1}.$$

Finally, as the limit $\phi_d^* = \mu_d$ does not depend on the particular converging subsequence $(r_j)_{j \in \mathbb{N}}$, we conclude that the whole sequence converges, i.e., the desired result (4.5) holds. \square

9. CONCLUSION

This paper deals with the problem of recovering the graph of a function when only partial moment information is available, here only zero-th and first-order moments, namely linear measurements of the function. We show how to recover the graph asymptotically by solving a hierarchy of semidefinite programs, with an appropriately chosen objective function that promotes sparsity.

As shown in non trivial numerical experiments, our results suggest a perhaps counter-intuitive message. Namely we observe experimentally that it is better

- to approximate the graph $G = \{(\mathbf{x}, f(\mathbf{x})) : \mathbf{x} \in X\} \subset X \times Y$ from moments of the measure supported on G (a thin subset of $X \times Y$) with the recovery technique of [15],

rather than

- to approximate the function f itself (e.g. via L^2 -norm approximation by polynomials) from moments of the measure with density f on X .

Moreover, this is true even if *all* exact moments of the measure with density f are available, whereas only a few exact moments of the measure supported on the graph G are available (and the other must be appropriately approximated).

ACKNOWLEDGMENTS

The authors wish to thank Swann Marx and Edouard Pauwels for fruitful and helpful discussions on the topic.

REFERENCES

- [1] R. Ash. Real Analysis and Probability, Academic Press Inc., San Diego, 1972
- [2] A.I. Akhiezer and M.G. Krein. Some Questions in the Theory of Moments, Kharkov, 1938, Amer. Math. Soc., Providence, R.I. 1962.
- [3] E. D. Andersen, K. D. Andersen. The MOSEK interior point optimizer for linear programming: an implementation of the homogeneous algorithm. In H. Frenk, K. Roos, T. Terlaky, S. Zhang (Editors). High Performance Optimization. Kluwer Academic Publishers, 2000.
- [4] E. J. Anderson, P. Nash. Linear programming in infinite-dimensional spaces: theory and applications. Wiley, New York, NY, 1987.
- [5] A. Barvinok. A Course in Convexity. American Math. Soc., Providence, R.I., 2002.
- [6] E. J. Candès, C. Fernández-Granda. Towards a mathematical theory of superresolution. Communications on Pure and Applied Mathematics, 67(6):906-956, 2014.
- [7] Y. de Castro, F. Gamboa. Exact reconstruction using Beurling minimal extrapolation. Journal of Mathematical Analysis and Applications, 395(1):336-354, 2012.
- [8] Y. de Castro, F. Gamboa, D. Henrion, J. B. Lasserre. Exact solutions to Super Resolution on semi-algebraic domains in higher dimensions. IEEE Transactions on Information Theory, 63(1):621-630, 2017.
- [9] K. S. Eckhoff. Accurate and efficient reconstruction of discontinuous functions from truncated series expansions. Math. Comput. 61(204):745-763, 1993.
- [10] D. Henrion, J. B. Lasserre, J. Löfberg. GloptiPoly 3: moments, optimization and semidefinite programming. Optimization Methods and Software 24(4-5):761-779, 2009.
- [11] D. Henrion, J. B. Lasserre, M. Mevissen. Mean squared error minimization for inverse moment problems. Applied Mathematics and Optimization, 70(1):83-110, 2014.
- [12] M. Korda, D. Henrion, J. B. Lasserre. The Moment-SOS hierarchy. World Scientific, 2020.
- [13] J. B. Lasserre. Moments, Positive Polynomials and Their Applications, Imperial College Press, 2010.
- [14] J. B. Lasserre. An introduction to Polynomial and Semi-Algebraic Optimization, Cambridge University Press, Cambridge, 2015.

- [15] S. Marx, E. Pauwels, T. Weisser, D. Henrion, J. B. Lasserre. Tractable semi-algebraic approximation using Christoffel-Darboux kernel, *Constructive Approximation*, to appear. HAL Report 02085835, April 2019.
- [16] S. Marx, T. Weisser, D. Henrion, J. B. Lasserre. A moment approach for entropy solutions to nonlinear hyperbolic PDEs. *Mathematical Control and Related Fields*, 10(1):113-140, 2020.
- [17] M. Putinar. Positive polynomial on compact semi-algebraic sets, *Indiana Univ. Math. J.* 42: 969–984, 1993
- [18] F. Santambrogio. *Optimal transport for applied mathematicians – Calculus of variations, PDEs and modeling*. Birkhäuser, 2015.
- [19] L. N. Trefethen. *Approximation theory and approximation practice*. SIAM, 2013.