A hierarchy of convex relaxations for the total variation distance

Jean-Bernard Lasserre

A HIERARCHY OF CONVEX RELAXATIONS FOR THE TOTAL VARIATION DISTANCE

JEAN B. LASSERRE

Abstract. Given two measures μ, ν on $\mathbb{R}^d$ that satisfy Carleman's condition, we provide a numerical scheme to approximate as closely as desired the total variation distance between μ and ν. It consists of solving a sequence (hierarchy) of convex relaxations whose associated sequence of optimal values converges to the total variation distance, an additional illustration of the versatility of the Moment-SOS hierarchy. Indeed each relaxation in the hierarchy is a semidefinite program whose size increases with the number of involved moments. It has an optimal solution which is a couple of degree-$2n$ pseudo-moments which converge, as n grows, to moments of the Hahn-Jordan decomposition of $\mu - \nu$.

Abstract. Given two measures μ, ν on $\mathbb{R}^d$ that satisfy Carleman's condition, we provide a numerical scheme to approximate as closely as desired the total variation distance between μ and ν. It consists of solving a sequence (hierarchy) of convex relaxations whose associated sequence of optimal values converges to the total variation distance, an additional illustration of the versatility of the Moment-SOS hierarchy. Each relaxation in the hierarchy is a semidefinite program whose size increases with the number of involved moments. It has an optimal solution which is a couple of degree-$2n$ pseudo-moments which converge, as n grows, to moments of the Hahn-Jordan decomposition of $\mu - \nu$. Illustrative examples are provided.
MSC: 46N30, 47N30, 60B10 60-08, 62-08, 90C22, 90C23
keywords: Total variation distance, Moment problem, Polynomial Optimization, Convex relaxations, Semidefinite programming

Contents

The author is supported by the AI Interdisciplinary Institute ANITI funding through the french program "Investing for the Future PI3A" under the grant agreement number ANR-19-PI3A-0004. This research is also part of the programme DesCartes and is supported by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme.

1. Introduction

Evaluating a "distance" between measures is an important topic with many applications, e.g. for homogeneity testing and independence testing as advocated in [16], and has also become increasingly important in Data Science and Machine Learning in particular. Among possible choices, the family of *integral probability metrics* (IPM) which includes the Kantorovich, Dudley, Kolmogorov and total variation (TV) metrics, is discussed in [16] where the authors provide several empirical estimators of the associated distances between two distributions, based on random i.i.d. samples. See also [13] for a discussion on relative merits of several distances.

In particular, the Kantorovich metric (dual to Wasserstein distance) has become popular and one reason is that its optimal transport formulation allows to define efficient specialized procedures (e.g. the Sinkhorn algorithm) for its computation [15]. On the other hand, as the TV distance is the same as the Wasserstein distance with (nasty) cost function $c(x,y) = 1_{x\neq y}(x,y)$, it is an indication that its effective computation is a computational challenge. For instance, in [16] where the authors provide several empirical estimators of integral probability metrics (IPMs), when specializing to TV distance the resulting estimator is not consistent, and for this reason the authors provide Lower bounds [16, Proposition 5.1]. The reason is that the set of bounded measurable functions of norm 1 is too large for efficient evaluation of evaluate $TV(\mathbb{P},\mathbb{Q}) = \sup_f\{\,\left|\int f\,d\mathbb{P} - \int f\,d\mathbb{Q}\right| : \|f\|_\infty \leq 1\,\}$ for two distributions $\mathbb{P}$ and $\mathbb{Q}$. In view of such difficulties, recent contributions have focused on providing analytical upper and/or lower bounds on $TV(\mathbb{P},\mathbb{Q})$ for $\mathbb{P},\mathbb{Q}$ in some classes of distributions, e.g. two high-dimensional gaussians with same mean in [6], or mixture of two Gaussians with same covariance matrix in [4], or two arbitrary measures with given means and variance in [14]; recently in [2] the authors provide a tight (up to a constant factor) lower bound on the TV distance for high-dimensional gaussians.

In another direction, in [3] the authors consider estimators of an unknown distribution μ and, in view of [5], advocate that some à priori information on μ is required if the estimators are required to be consistent in total variation. Then under the assumption that the non-atomic part of μ is absolutely continuous with respect to some à priori known σ-finite measure, they provide estimators which are consistent in total variation (a.s. and in expectation).

Contribution. In this paper we show that the *total variation* distance is amenable to practical computation under relatively weak assumptions and so could

provide an alternative to other distances when needed. In a rather general context, we provide a numerical scheme to approximate as closely as desired the total variation distance between two measures μ and ν. We do *not* assume that μ or ν has compact support, but we assume that all moments of μ and ν are finite, and that both μ and ν satisfy Carleman' condition. We formulate the problem as an infinite-dimensional linear program (LP) on a space of measures, with an important constraint of domination inherited from the Hahn-Jordan decomposition of $\mu - \nu$. This LP-formulation is then viewed as an instance of the Generalized Moment Problem (GMP) with polynomial data, so that the resulting GMP is amenable to practical computation via the Moment-SOS hierarchy [11, 7]. As a result, one may approximate as closely as desired $\|\mu - \nu\|_{TV}$ as more and more moments of μ and ν are taken into account. More precisely:

(i) Our numerical scheme consists of solving a sequence (hierarchy) of convex relaxations. Each convex relaxation of the hierarchy is a semidefinite program[1] whose size increases with the number of moments of μ and ν involved.

(ii) The associated sequence of optimal values is monotone non decreasing and converges from below to $\|\mu - \nu\|_{TV}$. Crucial for convergence is a domination constraint coming from a property of the Hahn-Jordan decomposition of $\mu - \nu$.

(iii) At last but not least, the associated sequence of optimal solutions of relaxations (a couple of pseudo-moment vectors whose size increases), converges to the unique couple of infinite moment vectors of the Hahn-Jordan decomposition (ϕ^*_+, ϕ^*_-) of the signed measure $\mu - \nu$.

(iv) Each semidefinite relaxation of the hierarchy has a dual semidefinite program, very much in the spirit of the classical TV-distance dual formulation

$$\|\mu - \nu\|_{TV} \;=\; \sup_f \{ \int f\, \hat{d}\mu - \int f\, \hat{d}\nu :\; \|f\|_\infty \leq 1 \} \tag{1.1}$$

where the "sup" is over bounded measurable functions. Our hierarchy of duals shows how the above classical formulation can be strengthened by (i) restricting to polynomials and (ii), including an additional penalized integral term (w.r.t. μ and ν) in the criterion. This term penalizes the unavoidable violation of the constraint $\|f\|_\infty \leq 1$ when f is a polynomial, and corresponds to the domination constraint in the primal formulation.

(v) It turns out that when μ and ν are measures on the real line, our first lower bound with $n = 1$ in the hierarchy (i.e. when one uses moments up to degree $2n = 2$ only) coincides with the analytical lower bound provided in [14] and based solely on the means and variances of μ and ν. As shown on some examples, the improvement is already significant with $n = 2$ (i.e. by now taking into account moments up to degree 4) and even better with $n = 3, 4$.

Moreover, and as a nice feature of our numerical scheme, we prove that for two atomic probability measures respectively supported on m_1 and m_2 atoms of the real line, the exact distance $\|\mu - \nu\|_{TV}$ is obtained as soon as the degree n of the semidefinite relaxation in the hierarchy, matches $\max[m_1, m_2]$, i.e., when the minimal information required is used. Hence, for instance, mutual singularity (if any) (i.e., $\|\mu - \nu\|_{TV} = 2$) is detected at $n = \max[m_1, m_2]$. In addition, in principle no geometric condition on a separation of the respective atoms of μ and

[1]A semidefinite program is a convex conic optimization problem that can be solved efficiently, up to arbitrary precision fixed in advance; see e.g. [1]

ν is required and this nice feature is illustrated on a toy example with μ the Dirac δ_0 at $x=0$ and ν the Dirac δ_ε at $x=\varepsilon$ (with arbitrary small $\varepsilon>0$). (However as in practice one uses a numerical semidefinite solver, this issue becomes relevant due to unavoidable potential numerical inaccuracies.)

(vi) We also provide a set of illustrative numerical experiments to illustrate (a) our result on discrete measures on the real line, and (b) the behavior of the algorithm when μ and ν are two univariate Gaussian $\mathcal{N}(m_1,\sigma_1)$ and $\mathcal{N}(m_2,\sigma_2)$.

(vii) Finally, it is worth emphasizing that the optimal value of each relaxation provides a *guaranteed* lower bound on the TV distance which increases with the degree of the relaxation. This information already provided at early steps of the hierarchy should be useful because in view of the current status of semidefinite solver software packages, one cannot expect to solve high degree relaxations, even for relatively modest dimensions.

At last but not least, the input data required at the n-th semidefinite relaxation of the hierarchy is the *finite* set of degree-$2n$ moments of μ and ν, assumed to be known[2] or estimated from random i.i.d. samples drawn from μ and ν. In the latter case, by the SLLN, such a finite set of degree-$2n$ moments can be estimated as closely as desired and almost surely, provided that the sample size is sufficiently large. Then the true moment matrices $\mathbf{M}_n(\mu)$ and $\mathbf{M}_n(\nu)$ of μ and ν needed in the n-th semidefinite relaxation of our numerical scheme, can be safely replaced with their analogues $\mathbf{M}_n(\mu^N)$ and $\mathbf{M}_n(\nu^N)$ obtained from the empirical measures μ^N and ν^N associated with a sample of size N. Of course, when n increases, the sample size N needs to be adjusted with the number of degree-$2n$ moments considered. This issues was also analyzed in [12] to analyze the respective behavior of the Christoffel functions respectively associated with a measure μ and its empirical version μ^N from a sample.

Hence in summary, our contribution is to provide an additional tool in the arsenal of algorithms available in applied probability, for approximating as closely as desired, the total variation distance $\|\mu-\nu\|_{TV}$ based on moment information. This tool can thus be applied

– not only in applications where moments of μ and ν are available inclosed form (e.g. for μ and ν Gaussian or exponentials (and their mixtures)), but also

– even in applications where only random i.i.d. samples from μ and ν are available. Indeed as already mentioned, with fixed n, the finite set of $2n$-degree empirical moments obtained from a sample, can approximate as closely as desired the same set of true degree-$2n$ moments, provided the sample size is sufficiently large (hence adapted to the degree n considered).

As a technical comment, we wish to also emphasize the relatively weak assumption on the measures μ,ν, namely that they satisfy Carleman's condition (no compact support is required). Crucial in our numerical scheme are the two domination constraints $\phi^+\le\mu$ and $\phi^-\le\nu$ where (ϕ^+,ϕ^-) is the Hahn-Jordan decomposition of the signed measure $\mu-\nu$. While redundant in the infinite-dimensional GMP

[2]For instance if μ and ν are two Gaussians $\mathcal{N}(\mathbf{m},\Sigma)$ and $\mathcal{N}(\mathbf{m}',\Sigma')$ respectively, their moments are known explicitly in terms of $\mathbf{m},\mathbf{m}'$ and the entries of Σ and Σ'. The same is true e.g. for pairs of exponential measures, or gaussian mixtures.

formulation, they become extremely useful (as a compactification tool) in the relaxation scheme. Interestingly, the effect of such domination constraints is also revealed in the dual problem at step n of the hierarchy when this dual is compared with the classical dual formulation (1.1) of the TV distance.

In a final remark, as an alternative to algorithms based on discretizations (like e.g. Sinkhorn algorithm), the Wasserstein distance $W_2(\mu,\nu)$ (with *polynomial* cost $c(x,y)$) can also be approximated as closely as desired in a mesh-free practical computation by (i) applying the Moment-SOS hierarchy [7, 10] for solving the associated optimal transport problem (OT), and (ii) extract the transport map from the moment vector solution of the OT, by a non-standard application of the Christoffel-Darboux kernel [8]. However, crucial in [7, 10] is the fact that the cost function is a polynomial (which of course excludes the nasty cost function $1_{x\neq y}(x,y)$ in the TV distance formulation).

2. Main result

2.1. **Notation and definitions.** Let $\mathbb{R}[\mathbf{x}]$ denote the ring of real polynomials in the variables $(x_1,\ldots,x_d)$ and $\mathbb{R}[\mathbf{x}]_n \subset \mathbb{R}[\mathbf{x}]$ be its subset of polynomials of total degree at most n. Let $\mathbb{N}^d_n := \{\boldsymbol{\alpha}\in\mathbb{N}^d : \sum_i \alpha_i \leq n\}$ with cardinal $s(n) = \binom{n+d}{n}$. Let $\mathbf{v}_n(\mathbf{x}) = (\mathbf{x}^{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^d_n}$ be the vector of monomials up to degree n, and let $\Sigma[\mathbf{x}]_n \subset \mathbb{R}[\mathbf{x}]_{2n}$ be the convex cone of polynomials of total degree at most $2n$ which are sum-of-squares (in short SOS). A polynomial $p\in\mathbb{R}[\mathbf{x}]_n$ can be identified with its vector of coefficients $\mathbf{p}=(p_{\boldsymbol{\alpha}})\in\mathbb{R}^{s(n)}$ in the monomial basis, and reads

$$\mathbf{x} \quad \mapsto p(\mathbf{x}) \,:=\, \langle \mathbf{p}, \mathbf{v}_n(\mathbf{x})\rangle\,, \quad \forall p\in\mathbb{R}[\mathbf{x}]\,.$$

Denote by $\mathscr{M}(\mathbb{R}^d)$ (resp. $\mathscr{M}(\mathbb{R}^d)_+$) the space of signed (resp. positive) Borel measures on $\mathbb{R}^d$. For two Borel measures $\mu,\nu\in\mathscr{M}(\mathbb{R}^d)_+$, the notation $\mu\leq\nu$ stands for $\mu(B)\leq\nu(B)$ for all Borel sets $B\in\mathcal{B}(\mathbb{R}^d)$. The support of a Borel measure μ on $\mathbb{R}^d$ is the smallest closed set A such that $\mu(\mathbb{R}^d\setminus A)=0$, and such a set A is unique. A Borel measure whose all moments are finite is said to be (moment) *determinate* if there is no other measure with same moments.

For a real symmetric matrix $\mathbf{A}=\mathbf{A}^T$, the notation $\mathbf{A}\succeq 0$ (resp. $\mathbf{A}\succ 0$) stands for $\mathbf{A}$ is positive semidefinite (p.s.d.) (resp. positive definite (p.d.)).

Hahn-Jordan decomposition. Given two finite Borel measures $\mu,\nu\in\mathscr{M}(\mathbb{R}^d)_+$, the signed measure $\mu-\nu$ has a unique Hahn-Jordan decomposition (ϕ^*_+,ϕ^*_-) such that $\phi^*_+-\phi^*_- = \mu-\nu$. That is, there exists a Borel set $A\in\mathcal{B}(\mathbb{R}^d)$ and two mutually singular positive measure ϕ^*_+,ϕ^*_- such that $\phi^*_+(\mathbb{R}^d)=\phi^*_+(A)$ while $\phi^*_-(A)=0$, and

$$(2.1)\;\; \phi^*_+(B) \,=\, (\mu-\nu)(B\cap A)\,; \quad \phi^*_-(B) \,=\, (\nu-\mu)(B\cap(\mathbb{R}^d\setminus A))\,, \quad \forall B\in\mathcal{B}(\mathbb{R}^d)\,.$$

In addition, and obviously, $\|\mu-\nu\|_{TV} \leq \mu(1)+\nu(1)$. Moreover, observe that $\phi^*_+\leq\mu$ and $\phi^*_-\leq\nu$. This property will turn out to be crucial for convergence of our numerical scheme.

Riesz linear functional and moment matrix. With a real sequence $\boldsymbol{\phi} = (\phi_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^d}$ (in bold) is associated the *Riesz* linear functional $\phi\in\mathbb{R}[\mathbf{x}]^*$ (not in bold)

defined by

$$p\,(=\sum_{\boldsymbol{\alpha}} p_{\boldsymbol{\alpha}}\mathbf{x}^{\boldsymbol{\alpha}}) \quad \mapsto \phi(p) \,=\, \langle \boldsymbol{\phi}, \mathbf{p}\rangle \,=\, \sum_{\boldsymbol{\alpha}} p_{\boldsymbol{\alpha}}\,\phi_{\boldsymbol{\alpha}}\,, \quad \forall p \in \mathbb{R}[\mathbf{x}]\,,$$

and the moment matrix $\mathbf{M}_n(\boldsymbol{\phi})$ with rows and columns indexed by $\mathbb{N}^d_n$ (hence of size $s(n)$), and with entries

$$\mathbf{M}_n(\boldsymbol{\phi})(\boldsymbol{\alpha},\boldsymbol{\beta}) \,:=\, \phi(\mathbf{x}^{\boldsymbol{\alpha}+\boldsymbol{\beta}}) \,=\, \phi_{\boldsymbol{\alpha}+\boldsymbol{\beta}}\,, \quad \boldsymbol{\alpha},\boldsymbol{\beta} \in \mathbb{N}^d_n\,.$$

Notice that one may write indifferently $\mathbf{M}_n(\boldsymbol{\phi})$ or $\mathbf{M}_n(\phi)$, i.e., referring to the sequence $\boldsymbol{\phi}$ truncated to degree-$2n$ moments or to the Riesz linear functional ϕ associated with $\boldsymbol{\phi}$.

A real sequence $\boldsymbol{\phi} = (\phi_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^d}$ has a representing mesure if its associated linear functional ϕ is a Borel measure on $\mathbb{R}^d$. In this case $\mathbf{M}_n(\boldsymbol{\phi}) \succeq 0$ for all n; the converse is not true in general.

Carleman's condition. A sequence $\boldsymbol{\mu} = (\mu_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^d}$ satisfies Carleman's condition if

$$\forall i = 1,\ldots,d: \quad \sum_{j=1}^{\infty} \mu(x_i^{2j})^{-1/2j} \,=\, +\infty\,. \tag{2.2}$$

The following theorem is due to Nussbaum:

Theorem 2.1. *([10, Theorem 3.5]) Let a sequence $\boldsymbol{\mu} = (\mu_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^d}$ be such that $\mathbf{M}_n(\boldsymbol{\mu}) \succeq 0$, for all $n \in \mathbb{N}$. If $\boldsymbol{\mu}$ satisfies Carleman's condition (2.2) then $\boldsymbol{\mu}$ has a representing measure μ on $\mathbb{R}^d$ and μ is determinate.*

A sufficient condition to ensure that a measure μ satisfies the multivariate Carleman's condition is that

$$\int \exp(c|x_i|)\,d\mu \,<\, \infty\,, \quad i = 1,\ldots,d\,, \tag{2.3}$$

if for some scalar $c > 0$.

2.2. **A preliminary result.**

Lemma 2.2. *Let $\mu, \varphi \in \mathscr{M}(\mathbb{R}^d)_+$ have finite moments and assume that μ satisfies Carleman's condition (2.2). Then*

$$\varphi \le \mu \quad \Leftrightarrow \quad \mathbf{M}_n(\varphi) \preceq \mathbf{M}_n(\mu)\,, \quad \forall n \in \mathbb{N}\,. \tag{2.4}$$

Proof. $\Rightarrow$ is straightforward. Indeed:

$$\mu \ge \varphi \Rightarrow \left[\int p^2\,d\mu \,\ge\, \int p^2\,d\varphi\,,\, \forall p \in \mathbb{R}[\mathbf{x}]\right] \quad \Rightarrow \quad \mathbf{M}_n(\mu) \,\succeq\, \mathbf{M}_n(\varphi)\,,\, \forall n \in \mathbb{N}\,.$$

$\Leftarrow$ Assume that $\mathbf{M}_n(\varphi) \preceq \mathbf{M}_n(\mu)$ for all $n \in \mathbb{N}$, and consider the sequence $\boldsymbol{\gamma} = (\gamma_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^d}$, with $\gamma_{\boldsymbol{\alpha}} = \mu_{\boldsymbol{\alpha}} - \varphi_{\boldsymbol{\alpha}}$, for all $\boldsymbol{\alpha} \in \mathbb{N}^d$. Then $\int x_i^{2n} d\varphi \le \int x_i^{2n} d\mu$ for all n, and as Carleman's condition (2.2) holds for μ, we infer $\gamma(x_i^{2n}) \le \mu(x_i^{2n})$ for all n, and all $i = 1\ldots,d$. This implies that $\boldsymbol{\gamma}$ satisfies Carleman's condition (2.2) and therefore, as $\mathbf{M}_n(\boldsymbol{\gamma}) = \mathbf{M}_n(\mu) - \mathbf{M}_n(\varphi) \succeq 0$ for all n, we deduce that $\boldsymbol{\gamma}$ has a determinate representing measure γ on $\mathbb{R}^d$. In particular:

$$\int \mathbf{x}^{\boldsymbol{\alpha}}\,d(\gamma+\varphi) \,=\, \gamma_{\boldsymbol{\alpha}} + \varphi_{\boldsymbol{\alpha}} \,=\, \mu_{\boldsymbol{\alpha}} \,=\, \int \mathbf{x}^{\boldsymbol{\alpha}}\,d\mu\,, \quad \forall \boldsymbol{\alpha} \in \mathbb{N}^d \quad \Rightarrow \gamma + \varphi \,=\, \mu\,,$$

where the last statement follows from determinateness of μ. Hence $\varphi \le \mu$. □

2.3. **Main result.** Given two finite Borel measures μ and ν on $\mathbb{R}^d$, introduce the infinite-dimensional LP:

$$\tau = \inf_{\phi^+,\phi^- \in \mathscr{M}(\mathbb{R}^d)_+} \{\, \phi^+(1) + \phi^-(1) : \quad \phi_+ - \phi_- \,=\, \mu - \nu \,\}\,. \tag{2.5}$$

Proposition 2.3. *The LP* (2.5) *has a unique optimal solution* (ϕ_+^*, ϕ_-^*) *which is the Hahn-Jordan decomposition of the signed measure* $\mu - \nu$*, and therefore* $\tau = \phi_+^*(1) + \phi_-^*(1) = \|\mu - \nu\|_{TV}$*.*

Proof. Let (ϕ^+, ϕ^-) be an arbitrary feasible solution of (2.5). Then as $\phi^+ - \phi^- = \mu - \nu$ one obtains $\phi^+(1) + \phi^-(1) \geq \|\phi^+ - \phi^-\|_{TV} = \|\mu - \nu\|_{TV}$. On the other hand, the Hahn-Jordan decomposition (ϕ_+^*, ϕ_-^*) of $\mu - \nu$ is feasible for (2.5), with value $\|\mu - \nu\|_{TV}$, whence the result. □

Unfortunately the LP (2.5) is not very useful as its stands. It is just a particular rephrasing of the total variation distance between μ and ν. However we next see the a slight reinforcement of (2.5) will turn out to be very useful when passing to some hierarchy of convex relaxations. Indeed:

Proposition 2.4. *The infinite-dimensional linear program*

$$\inf_{\phi^+,\phi^- \in \mathscr{M}(\mathbb{R}^d)_+} \{\, \phi^+(1) + \phi^-(1) : \quad \phi_+ - \phi_- \,=\, \mu - \nu\,; \quad \phi^+ \leq \mu\,;\, \phi^- \leq \nu \,\} \tag{2.6}$$

has same optimal value $\tau = \|\mu - \nu\|_{TV}$*, and optimal solution* (ϕ_+^*, ϕ_-^*) *as* (2.5)*.*

Proof. By construction, the optimal value ρ of (2.6) satisfies $\rho \geq \tau = \|\mu - \nu\|_{TV}$. On the other hand, with (ϕ_+^*, ϕ_-^*) being the Hahn-Jordan decomposition of $\mu - \nu$, observe that $\phi_+^* \leq \mu$, and $\phi_-^* \leq \nu$. Therefore (ϕ_+^*, ϕ_-^*) is an optimal solution of (2.6). Equivalently, the constraints $\phi^+ \leq \mu$ and $\phi^- \leq \nu$ are automatically satisfied at the optimal solution (ϕ_+^*, ϕ_-^*) of (2.5) and therefore (2.5) and (2.6) have same optimal value and same optimal solution. □

Next, from now on we make the following assumption:

Assumption 2.5. *(i) All moments of* μ *and* ν *are finite, and*
(ii) μ *and* ν *satisfy* (2.3) *(hence satisfy Carleman's condition* (2.2)*) for some scalar* $c > 0$*.*

Consider the optimization problem

$$\begin{aligned} \hat{\tau} = \min_{\phi^+,\phi^- \in \mathscr{M}(\mathbb{R}^d)_+} \quad & \{\, \phi^+(1) + \phi^-(1) : \quad \phi^+ - \phi^- \,=\, \mu - \nu\,; \\ & \mathbf{M}_n(\phi^+) \preceq \mathbf{M}_n(\mu)\,; \quad \mathbf{M}_n(\phi^-) \preceq \mathbf{M}_n(\nu)\,, \quad \forall n \in \mathbb{N}\,\}. \end{aligned} \tag{2.7}$$

Corollary 2.6. *Let Assumption 2.5 hold. Then the Hahn-Jordan decomposition* (ϕ_+^*, ϕ_-^*) *of the signed measure* $\mu - \nu$*, is the unique optimal solution of* (2.7)*, and* $\hat{\tau} = \tau = \|\mu - \nu\|_{TV}$*.*

Proof. By Lemma 2.2, (2.6) and (2.7) are equivalent. □

The nice feature of the LP (2.7) when compared to its equivalent formulation (2.6), is that the cost as well as the constraints of (2.7) can next be formulated in

terms of moments of (μ,ν,ϕ^+,ϕ^-), so as to yield the optimization problem:

$$\begin{aligned}\rho = \min_{\phi^+,\phi^-\in\mathscr{M}(\mathbb{R}^d)_+} \quad & \{\,\phi^+(1)+\phi^-(1):\\ & \int \mathbf{x}^{\boldsymbol{\alpha}} d(\phi^+-\phi^-) \;=\; \int \mathbf{x}^{\boldsymbol{\alpha}}\, d(\mu-\nu)\,, \quad \forall \boldsymbol{\alpha}\in\mathbb{N}^d\,;\\ & \mathbf{M}_n(\phi^+)\succeq \mathbf{M}_n(\mu)\,;\; \mathbf{M}_n(\phi^-)\succeq \mathbf{M}_n(\nu)\,,\;\forall n\in\mathbb{N}\,\}\,,\end{aligned} \tag{2.8}$$

which is an instance of the Generalized Moment Problem (GMP); see e.g. [10].

Corollary 2.7. *Let Assumption 2.5 hold. Then the Hahn-Jordan decomposition (ϕ^*_+,ϕ^*_-) of the signed measure $\mu-\nu$, is the unique optimal solution of* (2.8)*, and $\rho=\|\mu-\nu\|_{TV}$.*

Proof. Let (ϕ^+,ϕ^-) be an arbitrary feasible solution of (2.8). By Lemma 2.2, $\phi^+\le\mu$ and $\phi^-\le\nu$. Hence $\phi^++\nu\le\mu+\nu$, and $\phi^-+\mu\le\mu+\nu$. As Assumption 2.5(ii) holds,

$$\begin{aligned}\int \exp(c\,|x_i|)\,d(\phi^++\nu) \;&<\; \int \exp(c|x_i|)\,d(\mu+\nu) \;<\; \infty\\ \int \exp(c\,|x_i|)\,d(\phi^-+\mu) \;&<\; \int \exp(c|x_i|)\,d(\mu+\nu) \;<\; \infty\,,\end{aligned}$$

and therefore the measure $\phi^++\nu$ (resp. $\phi^-+\mu$) is determinate. But then the constraint $\int x^{\boldsymbol{\alpha}}d(\phi^+-\phi^-)=\int x^{\boldsymbol{\alpha}}d(\mu-\nu)$ for all $\boldsymbol{\alpha}\in\mathbb{N}^d$ reads:

$$\int \mathbf{x}^{\boldsymbol{\alpha}}\,d(\phi^++\nu) \;=\; \int \mathbf{x}^{\boldsymbol{\alpha}}\,d(\phi^-+\mu)\,,\quad \forall\boldsymbol{\alpha}\in\mathbb{N}^d\,,$$

which implies $\phi^++\nu=\phi^-+\mu$ by determinacy of the measures. Therefore (ϕ^+,ϕ^-) is a feasible solution of (2.7) with same value. In other words, (2.8) is equivalent to (2.7), whence the result. □

3. A convergent hierarchy of semidefinite relaxations

As (2.8) is an instance of the GMP, it is natural to apply the Moment-SOS hierarchy [7, 11]. With each fixed $n\in\mathbb{N}$, consider the optimization problem

$$\begin{aligned}\rho_n = \min_{\boldsymbol{\phi},\boldsymbol{\psi}} \quad & \{\,\phi(1)+\psi(1):\quad \phi_{\boldsymbol{\alpha}}-\psi_{\boldsymbol{\alpha}} \;=\; \mu_{\boldsymbol{\alpha}}-\nu_{\boldsymbol{\alpha}}\,,\quad \forall\boldsymbol{\alpha}\in\mathbb{N}^d_{2n}\,;\\ & 0\preceq \mathbf{M}_n(\boldsymbol{\phi})\preceq \mathbf{M}_n(\boldsymbol{\mu})\,;\quad 0\preceq \mathbf{M}_n(\boldsymbol{\psi})\preceq \mathbf{M}_n(\boldsymbol{\nu})\,\}\,,\end{aligned} \tag{3.1}$$

where now the optimization is over degree-$2n$ pseudo-moment vectors $\boldsymbol{\phi}=(\phi_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^d_{2n}}$ and $\boldsymbol{\psi}=(\psi_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^d_{2n}}$ (hence not necessarily coming from measures ϕ and ψ on $\mathbb{R}^d$). Of course (3.1) is an obvious relaxation of (2.8) and therefore $\rho_n\le\rho=\|\mu-\nu\|_{TV}$ for all $n\in\mathbb{N}$.

Observe that for each fixed $n\in\mathbb{N}$, (3.1) is a semidefinite program that can be solved by off-the-shelf solvers like GloptiPoly [9] or Jump [17] (package of the Julia programming language).

Theorem 3.1. *Let Assumption 2.5 hold.*

(i) For every fixed $n\in\mathbb{N}$, the optimization problem (3.1) *has an optimal solution denoted $(\boldsymbol{\phi}^{(n)},\boldsymbol{\psi}^{(n)})$.*

(ii) In addition, $\rho_n\uparrow\|\mu-\nu\|_{TV}$ as $n\to\infty$, and moreover,

$$\lim_{n\to\infty}\phi^{(n)}_{\boldsymbol{\alpha}} \;=\; \int \mathbf{x}^{\boldsymbol{\alpha}}\,d\phi^*_+\,;\quad \lim_{n\to\infty}\psi^{(n)}_{\boldsymbol{\alpha}} \;=\; \int \mathbf{x}^{\boldsymbol{\alpha}}\,d\phi^*_-\,,\quad \forall\boldsymbol{\alpha}\in\mathbb{N}^d\,, \tag{3.2}$$

where (ϕ^*_+,ϕ^*_-) is the Hahn-Jordan decomposition of the signed measure $\mu-\nu$.

Proof. (i) Let $(\boldsymbol{\phi}, \boldsymbol{\psi})$ be an arbitrary feasible solution of (3.1). As $\mathbf{M}_n(\boldsymbol{\phi}) \preceq \mathbf{M}_n(\mu)$ one obtains

$$\phi(1) \,\leq\, \mu(1)\,; \quad \phi(x_i^{2n}) \,\leq\, \mu(x_i^{2n})\,, \quad \forall i = 1, \ldots, d\,,$$

and therefore, as $\mathbf{M}_n(\phi^+) \succeq 0$, by [10, Proposition 3.6],

$$|\phi_{\boldsymbol{\alpha}}| \,\leq\, \max[\mu(1)\,, \max_i \mu(x_i^{2d})]\,, \quad \forall \boldsymbol{\alpha} \in \mathbb{N}^d_{2n}\,. \tag{3.3}$$

Similarly, as $0 \preceq \mathbf{M}_n(\boldsymbol{\psi}) \preceq \mathbf{M}_n(\nu)$,

$$|\psi_{\boldsymbol{\alpha}}| \,\leq\, \max[\nu(1)\,, \max_i \nu(x_i^{2d})]\,, \quad \forall \boldsymbol{\alpha} \in \mathbb{N}^d_{2n}\,. \tag{3.4}$$

Therefore the feasible set of (3.1) is compact. Hence (3.1) has an optimal solution.

(ii) For each fixed $n \in \mathbb{N}$, and since $\mathbf{M}_k(\boldsymbol{\phi}^{(n)})$ is a submatrix of $\mathbf{M}_n(\boldsymbol{\phi}^{(n)})$ for all $k = 1, \ldots, n$, again by [10,],

$$\forall \boldsymbol{\alpha} : 2k{-}1 \leq |\boldsymbol{\alpha}| \leq 2k : \quad |\phi^{(n)}_{\boldsymbol{\alpha}}| \,\leq\, \max[\mu(1)\,, \max_i \mu(x_i^{2k})] \,=:\, a_k, ; \quad k = 1, \ldots, n\,,$$

and similarly

$$|\psi^{(n)}_{\boldsymbol{\alpha}}| \,\leq\, \max[\nu(1)\,, \max_i \nu(x_i^{2k})] \,=:\, b_k, \quad \forall \boldsymbol{\alpha} \,:\, 2k-1 \leq |\boldsymbol{\alpha}| \leq 2k\,;\, k = 1, \ldots, n\,.$$

Next, introduce the new infinite peudo-moment sequences:

$$\hat{\phi}^{(n)}_{\boldsymbol{\alpha}} \,:=\, \phi^{(n)}_{\boldsymbol{\alpha}}/a_k\,, \quad \forall \boldsymbol{\alpha} :\, 2k-1 \leq |\boldsymbol{\alpha}| \leq 2k\,; \quad k = 1, \ldots, n\,, \tag{3.5}$$

and $\hat{\phi}^{(n)}_{\boldsymbol{\alpha}} = 0$ for all $\boldsymbol{\alpha} \in \mathbb{N}^d$ with $|\boldsymbol{\alpha}| > 2n$. Similarly,

$$\hat{\psi}^{(n)}_{\boldsymbol{\alpha}} \,:=\, \psi^{(n)}_{\boldsymbol{\alpha}}/b_k\,, \quad \forall \boldsymbol{\alpha} :\, 2k-1 \leq |\boldsymbol{\alpha}| \leq 2k\,; \quad k = 1, \ldots, n\,, \tag{3.6}$$

and $\hat{\psi}^{(n)}_{\boldsymbol{\alpha}} = 0$ for all $\boldsymbol{\alpha} \in \mathbb{N}^d$ with $|\boldsymbol{\alpha}| > 2n$.

Both sequences $\hat{\boldsymbol{\phi}}^{(n)}$ and $\hat{\boldsymbol{\psi}}^{(n)}$ are considered as elements of the unity ball $\mathbf{B}(0,1)$ of the Banach space ℓ_∞ of uniformly bounded sequences, which is sequentially compact in the $\sigma(\ell_\infty, \ell_1)$ weak topology. Therefore there exist $\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\psi}} \in \mathbf{B}(0,1)$ and a subsequence $(n_k)_{k\in\mathbb{N}}$ such that

$$\lim_{k\to\infty} \hat{\phi}^{(n_k)}_{\boldsymbol{\alpha}} \,=\, \hat{\phi}_{\boldsymbol{\alpha}}\,; \quad \lim_{k\to\infty} \hat{\psi}^{(n_k)}_{\boldsymbol{\alpha}} \,=\, \hat{\psi}_{\boldsymbol{\alpha}}\,, \quad \forall \boldsymbol{\alpha} \in \mathbb{N}^d\,. \tag{3.7}$$

By doing the reverse scaling of (3.5)-(3.6), one obtains:

$$\forall \boldsymbol{\alpha} \in \mathbb{N}^d : \quad \lim_{k\to\infty} \phi^{(n_k)}_{\boldsymbol{\alpha}} \,=\, \phi_{\boldsymbol{\alpha}}\,; \quad \lim_{k\to\infty} \psi^{(n_k)}_{\boldsymbol{\alpha}} \,=\, \psi_{\boldsymbol{\alpha}}\,, \tag{3.8}$$

where

$$\phi_{\boldsymbol{\alpha}} \,:=\, a_k \cdot \hat{\phi}_{\boldsymbol{\alpha}}\,; \quad \psi_{\boldsymbol{\alpha}} \,:=\, b_k \cdot \hat{\psi}_{\boldsymbol{\alpha}}\,; \quad \forall \boldsymbol{\alpha} :\, 2k-1 \leq |\boldsymbol{\alpha}| \leq 2k\,; \quad k \,\in\, \mathbb{N}\,.$$

Fix $t \in \mathbb{N}$ arbitrary. As $\mathbf{M}_t(\boldsymbol{\phi}^{(n)}) \succeq 0$ for all $n \geq t$, then by (3.8), $0 \preceq \mathbf{M}_t(\boldsymbol{\phi}) \preceq \mathbf{M}_t(\mu)$, and as t was arbitrary, $0 \preceq \mathbf{M}_n(\boldsymbol{\phi}) \preceq \mathbf{M}_n(\mu)$ for all n, and similarly $0 \preceq \mathbf{M}_n(\boldsymbol{\psi}) \preceq \mathbf{M}_n(\nu)$ for all n.

Next, as $\mathbf{M}_n(\boldsymbol{\phi}) \preceq \mathbf{M}_n(\mu)$, and $\boldsymbol{\mu}$ satisfies Carleman's condition, then so does $\boldsymbol{\phi}$, and as $\mathbf{M}_n(\boldsymbol{\phi}) \succeq 0$ for all n, it follows that $\boldsymbol{\phi} = (\phi_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^d}$ has a representing measure ϕ on $\mathbb{R}^d$. Similarly, for same reasons, $\boldsymbol{\psi}$ has a representing measure ψ on $\mathbb{R}^d$.

In addition , by (3.8),

$$\|\mu - \nu\|_{TV} \,\geq\, \lim_{k\to\infty} \rho_{n_k} \,=\, \lim_{k\to\infty} \phi^{(n_k)}(1) + \psi^{(n_k)}(1) \,=\, \phi(1) + \psi(1)\,,$$

and

$$\forall \boldsymbol{\alpha} \in \mathbb{N}^d : \quad \mu_{\boldsymbol{\alpha}} - \nu_{\boldsymbol{\alpha}} = \lim_{k\to\infty} \phi_{\boldsymbol{\alpha}}^{(n_k)} - \psi_{\boldsymbol{\alpha}}^{(n_k)} = \phi_{\boldsymbol{\alpha}} - \psi_{\boldsymbol{\alpha}} .$$

Hence (ϕ, ψ) is an optimal solution of (2.7) (hence of (2.5) as well), and by Corollary 2.6, $(\phi, \psi) = (\phi_+^*, \phi_-^*)$, the Hahn-Jordan decomposition of $\mu - \nu$.

Finally, as the $(n_k)_{k\in\mathbb{N}}$ was an arbitrary converging subsequence and the limit is independent of the subsequence, the whole sequence converges. □

3.1. **A dual of** (3.1)**.** In this section we describe a dual of (3.1) and compare this dual to the standard dual formulation

$$\|\mu - \nu\|_{TV} = \sup_{f\in\mathscr{B}(\mathbb{R}^d)} \Big\{ \int f\, d(\mu - \nu) : \|f\|_\infty \le 1 \Big\}, \tag{3.9}$$

of the TV distance. Problem (3.9) is very difficult to solve, especially if at least $\mathrm{supp}(\mu)$ and/or $\mathrm{supp}(\nu)$ is unbounded. In fact we are not aware of any convergent sequence of semidefinite relaxations to approach the optimal value of (3.9).

On the other hand, with $n \in \mathbb{N}$ fixed, consider the optimization problem

$$\begin{aligned} \rho_n^* := \sup_{p,\sigma_i,\psi_j} \quad & \Big\{ \int p\, d(\mu - \nu) - \int \sigma_1\, d\mu - \int \psi_1\, d\nu : \\ & 1 - p = \sigma_0 - \sigma_1 ;\ 1 + p = \psi_0 - \psi_1 ; \\ & p \in \mathbb{R}[\mathbf{x}]_{2n} ;\ \sigma_i, \psi_i \in \Sigma[\mathbf{x}]_n ,\ i = 1, 2 \Big\} . \end{aligned} \tag{3.10}$$

As σ_i, ψ_i are all SOS polynomials, the constraints of (3.10)

$$p \le 1 + \sigma_1 \quad \text{and} \quad -p \le 1 + \psi_1 , \quad \forall \mathbf{x} \in \mathbb{R}^d , \tag{3.11}$$

imply

$$|\, p(\mathbf{x})\,| \le 1 + \max[\sigma_1(\mathbf{x}), \psi_1(\mathbf{x})] , \quad \forall \mathbf{x} \in \mathbb{R}^d . \tag{3.12}$$

and in (3.10), $\int \sigma_1\, d\mu + \int \psi_1\, d\nu$ is penalized in the criterion which maximizes $\int p d(\mu - \nu)$. So as the constraint $\|p\|_\infty < 1$ cannot be satisfied by a polynomial $p \in \mathbb{R}[\mathbf{x}]_n$, one may see (3.11) as a polynomial relaxation of the restrictive constraint $\|f\|_\infty \le 1$, $f \in \mathscr{B}(\mathbb{R}^d)$. However

Proposition 3.2. (3.10) *is a dual of* (3.1)*, i.e.,* $\rho_n \ge \rho_n^*$ *for every* n*.*

Proof. Let $(\boldsymbol{\phi}, \boldsymbol{\psi})$ and $p \in \mathbb{R}[\mathbf{x}]_{2n}$) be arbitrary feasible solutions of (3.1) and (3.10) respectively. As $\int \sigma_1\, d\mu \ge \phi^+(\sigma_1)$ and $\int \psi_1\, d\nu \ge \phi^-(\psi_1)$,

$$\begin{aligned} \int p\, d(\mu - \nu) - \int \sigma_1\, d\mu - \int \psi_1\, d\nu &\le \phi^+(p) - \phi^+(\sigma_1) - \phi^-(p) - \phi^-(\psi_1) \\ &\le \phi^+(1 - \sigma_0) + \phi^-(1 - \psi_0) \le \phi^+(1) + \phi^-(1) , \end{aligned}$$

where we have used that $\phi^+(\sigma_0) \ge 0$ (as $\mathbf{M}_n(\boldsymbol{\phi}^+) \succeq 0$ and $\sigma_0 \in \Sigma[\mathbf{x}]_n$). This proves weak duality, i.e., $\rho_n \ge \rho_n^*$. □

We next prove that strong duality holds, i.e., there is no duality gain between (3.1) and its dual (3.10). Recall that if (ϕ^+, ϕ^-) is the Hahn-Jordan decomposition of $\mu - \nu$, then $\phi^+ \le \mu$ and $\phi^- \le \nu$. Therefore

$$\phi^+ = f^+\, d\mu \quad \text{and} \quad \phi^- = f^-\, d\nu , \tag{3.13}$$

for some nonnegative measurable functions f^+, f^- with $f^+ \le 1$, μ-a.e., and $f^- \le 1$, ν-a.e.

Lemma 3.3. *Let (ϕ^+, ϕ^-) be the Hahn-Jordan decomposition of $\mu-\nu$ and suppose that with f^+, f^- as in* (3.13)*, $f^+ < 1$ (resp. $f^- < 1$) on some open set O_+ (resp. O^-). Then there is no duality gap between* (3.1) *and its dual* (3.10)*, i.e., $\rho_n = \rho_n^*$ for all n and in addition,* (3.10) *has an optimal solution $(p^*, \sigma_i^*, \psi_i^*)$.*

Proof. Let $\boldsymbol{\phi}^+ = (\phi^+_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^d_{2n}}$ and $\boldsymbol{\phi}^- = (\phi^-_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^d_{2n}}$ be the respective moment vectors of ϕ^+ and ϕ^- up to degree $2n$. Then $(\boldsymbol{\phi}^+, \boldsymbol{\phi}^-)$ is an obvious feasible solution of (3.1), and we next prove it is a strictly feasible solution. Then by our assumption $\mathbf{M}_n(\phi^+) \prec \mathbf{M}_n(\mu)$; indeed otherwise suppose that $\mathrm{Ker}(\mathbf{M}_n(\mu) - \mathbf{M}_n(\phi^+)) \neq \emptyset$, i.e., there exists $p \in \mathbb{R}[\mathbf{x}]_n$ such that

$$0 \,=\, \int p^2\, d(\mu - \phi^+) \,=\, \int p^2\,(1 - f^+)\, d\mu\,,$$

But then one obtains the contradiction

$$0 \,=\, \int p^2\, d(\mu - \phi^+) \,\geq\, \int_{O^+} p^2\,(1 - f^+)\, d\mu \,>\, 0\,,$$

as $p \neq 0$ cannot vanish on an open set. For the same reasons, $\mathbf{M}_n(\phi^+) \succ 0$, and similarly $0 \prec \mathbf{M}_n(\phi^-) \prec \mathbf{M}_n(\nu)$. But this strict feasibility of $(\boldsymbol{\phi}^+, \boldsymbol{\phi}^-)$ in (3.1) implies that Slater's condition holds for (3.1). Hence by a standard results of duality for conic convex programs, $\rho_n^* = \rho_n$ and since $2 \geq \rho_n \geq 0$, (3.10) is solvable, i.e., it has an optimal solution $(p^*, \sigma_i^*, \psi_i^*)$.

□

3.2. **Computational remarks.**

Moment information. To implement the semidefinite relaxation (3.1) with fixed degree n, knowledge of the two moment sequences $(\mu_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^d_{2n}}$ and $(\nu_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathbb{N}^d_{2n}}$, that is, all moments of μ and ν up to degree $2n$. In some cases, all moments of μ and ν can be obtained exactly in explicit form. This is the case if μ and ν are Gaussian, or a mixture of Gaussians, or an exponential (or a mixture of exponentials). On the other hand, if the only information available is some sample of i.i.d. random vectors $(X_i)_{i\leq N}$ and $(Y_i)_{i\leq N}$ drawn according to μ and ν respectively, then for any fixed degree n, we may invoke the strong law of large numbers, and consider the moment matrices $\mathbf{M}_n(\mu^N)$ and $\mathbf{M}_n(\nu^N)$ associated with the corresponding empirical measures μ^N and ν^N. By continuity of the eigenvalues, $\|\mathbf{M}_n(\mu) - \mathbf{M}_n(\mu^N)\|$ can be made as small a desired provided that N is sufficiently large. Of course when n increases the sample size N needs to be adjusted accordingly.

If μ and ν are two probability measures, mutually singular, then $\|\mu-\nu\|_{TV}$=2. A perfect case to check whether (3.1) is efficient, is to test (3.1) with the toy univariate example where $\mu = \delta_0$ and $\nu = \delta_\varepsilon$ for small value of $\varepsilon > 0$. Indeed, one might expect that the convergence $\rho_n \uparrow \|\mu - \nu\|_{TV}$ as n grows, could depend on ε (the smaller ε, the slower the convergence), or suffer from some numerical difficulties for small $\varepsilon > 0$.

3.3. **Discrete (univariate) measures.**

If the optimal value of (3.1) satisfies $\rho_n = 2$ then obviously μ and ν are mutually singular. Indeed since (3.1) has an optimal solution $(\boldsymbol{\phi}^*, \boldsymbol{\psi}^*)$ with $\rho_n = \phi^*(1) + \psi^*(1) = 2$, and since $\mathbf{M}_n(\boldsymbol{\phi}^*) \leq \mathbf{M}_n(\boldsymbol{\mu})$ (resp. $\mathbf{M}_n(\boldsymbol{\psi}^*) \preceq \mathbf{M}_n(\boldsymbol{\nu})$), then necessarily $1 = \phi^*(1) = \psi^*(1)$. This implies that the vector $\boldsymbol{\phi}^*$ (resp. $\boldsymbol{\psi}^*$) of pseudo-moments up to degree $2n$, is identical to $\boldsymbol{\mu}$ (i.e. that of μ) (resp. $\boldsymbol{\nu}$, i.e., that of ν). However one may ask whether such a

situation happens for a finite degree n. We show that this is indeed the case for atomic probability measures on the real line with finite supports, in which case $n = \max[m_1, m_2]$ where $m_1 = \#\mathrm{supp}(\mu)$, and $m_2 = \#\mathrm{supp}(\nu)$.

Theorem 3.4. *Let μ and ν be two probability measures on the real line, supported on $X := (x(i))_{i=1,\ldots,m_1}$ and $Y := (y(j))_{j=1,\ldots,m_2}$ respectively. Then with ρ_n as in* (3.1)*, $\rho_n = \|\mu - \nu\|_{TV}$ for all $n \geq \max[m_1, m_2]$. In particular if $X \cap Y = \emptyset$ (i.e., if μ and ν are mutually singular) then $\rho_n = 2$ for all $n \geq \max[m_1, m_2]$.*

For clarity of exposition, the proof is postponed to Section 5.

Notice that in Theorem 3.4, there is *no* assumption on the "distance" between points of the respective supports X and Y of the discrete measures μ and ν. However in practice, the behavior of (numerical) semidefinite software packages needed to solve (3.1) is sensitive to this parameter for numerical reasons.

Example 1. *To illustrate Theorem 3.4 for two mutually singular measures, consider the toy example with $d = 1$, $\mu = \delta_0$, $\nu = \delta_\varepsilon$, $\varepsilon \neq 0$, so that $\|\mu - \nu\|_{TV} = 2$ and $(\phi_+^*, \phi_-^*) = (\mu, \nu)$. The semidefinite relaxation* (3.1) *with $n = 1$ reads:*

$$\rho_1 = \min_{\boldsymbol{\phi},\boldsymbol{\psi}} \quad \{\, \phi_0 + \psi_0 : \phi_0 = \psi_0\,;\ \phi_1 - \psi_1 = -\varepsilon\,;\ \phi_2 - \psi_2 = -\varepsilon^2$$
$$0 \preceq \begin{bmatrix} \phi_0 & \phi_1 \\ \phi_1 & \phi_2 \end{bmatrix} \preceq \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix};\ 0 \preceq \begin{bmatrix} \psi_0 & \psi_1 \\ \psi_1 & \psi_2 \end{bmatrix} \preceq \begin{bmatrix} 1 & \varepsilon \\ \varepsilon & \varepsilon^2 \end{bmatrix} \}\,.$$

The constraint $0 \preceq \mathbf{M}_1(\boldsymbol{\phi}) \preceq \mathbf{M}_n(\delta_0)$ combined with $(0,1) \in \mathrm{Ker}(\mathbf{M}_1(\mu))$ implies $(0,1) \in \mathrm{Ker}(\mathbf{M}_1(\boldsymbol{\phi}))$, which in turn implies $\phi_1 = \phi_2 = 0$. Hence $\psi_1 = \varepsilon$ and $\psi_2 = \varepsilon^2$. But then $\mathbf{M}_1(\boldsymbol{\psi}) \succeq 0$ implies $\varepsilon^2\psi_0 \geq \varepsilon^2$, which with $\psi_0 \leq 1$, implies $\psi_0 = 1$ and so $\phi_0 = \psi_0 = 1$, and $\rho_1 = 2$.

This toy example illustrates that in principle the first semidefinite relaxation (3.1) *provides the optimal solution (ϕ_+^*, ϕ_-^*), no matter how close is ε to 0 (see Theorem 3.4). One can see here (and also in the proof of Theorem 3.4) how crucial for the relaxations* (3.1) *are the domination constraints $\mathbf{M}_n(\boldsymbol{\phi}) \preceq \mathbf{M}_n(\boldsymbol{\mu})$ and $\mathbf{M}_n(\boldsymbol{\psi}) \preceq \mathbf{M}_n(\boldsymbol{\nu})$, whereas they are not needed in the infinite-dimensional LP* (2.5).

Theorem 3.4 shows that (at least in the univariate case) the semidefinite relaxations (3.1) obtain the exact value $\|\mu - \nu\|_{TV}$ as soon as $n \geq \max[m_1, m_2]$, that is, as soon as the minimal required moment information is used. Moreover, nowhere in the proof was a condition on some minimum distance between atoms of μ and ν. In fact Theorem 3.4 and the toy illustrative example of Example 1 above, show that the atoms can be as close as desired without affecting the result. Of course this assertion is only theoretical in nature and must be mitigated by the numerical behavior of the semidefinite solver in charge of solving the semidefinite program (3.1). Indeed if some atoms are too close one should reasonably expect to encounter some numerical issues.

3.4. **Numerical examples.** In this section we provide some illustrative examples that give a first idea of the behavior of the moment-relaxations (3.1).

Discrete measures. To illustrate Theorem 3.4, we first consider the simple case of two discrete measures

$$\mu = \frac{1}{m_1}\sum_{i=1}^{m_1} \delta_{x(i)}\,, \quad \nu = \frac{1}{m_2}\sum_{i=1}^{m_2} \delta_{y(i)}\,.$$

Example 2. *With no point in common, i.e.,* $X := \{x(i)\} \cap \{y(j)\} =: Y = \emptyset$ *so that* $\|\mu - \nu\|_{TV} = 2$ *as* μ *and* ν *are mutually singular. Let* $X = \{-1.0, 0.0, 1.0, 2.0\}$*;* $Y = \{-0.7, 0.3, 1.3, 2.3\}$*. Then in solving* (3.1) *with* $n = 4$ *(i.e. with* 8 *moments of* μ *and* ν*), we obtain* $\rho_4 = 1.9999$ *which up to machine precision is considered to be* 2*, as predicted by Theorem 3.4.*

Example 3. *With one point in common. If we now consider* $X = \{-1.0, 0.0, 1.0, 2.0\}$ *and* $Y = \{-2.0, -1.0, 0.1, 1.5\}$ *so that* $X \cap Y = \{-1.0\}$ *and as the weights are all equal, one obtains* $\|\mu - \nu\|_{TV} = 1.5$*. Then with* $n = 4$ *we obtain* $\rho_4 = 1.499$*, which again up to machine precision can be considered as* 1.5*.*

Example 4. *In this example,* $X = \{-1.0, 0.0, 1.0, 2.0\}$ *and* $Y = \{-0.7, 0.3, 1.3, 2.3\}$ *(so that the points of* Y *are "closer" to those of* X*. From results displayed in*

TABLE 1. $\|\mu - \nu\|_{TV}$ for two discrete measures; $X \cap Y = \emptyset$
$X = \{-1.0, 0.0, 1.0, 2.0\}$; $Y = \{-0.7, 0.3, 1.3, 2.3\}$

n	4	5
ρ_n	1.9999	1.9999

Table 1, one can see that even if some points are relatively close to each other, the semidefinite relaxation (3.1) *still provide a value* ρ_n *very close to* 2*, as soon as* $n \geq 4$ *(i.e., with* $2n = 8$ *moments), as predicted by Theorem 3.4. But now due to numerical inaccuracies of the semidefinite solver, the resulting value is less precise (but one can still extract a solution* (ϕ^+, ϕ^-) *very close to* (μ, ν)*.*

Example 5. *With one point in common, i.e.,* $\#(X \cap Y) = 1$*. Let* $X = \{0.0, 0.3, 0.4, 0.9\}$ *and* $Y = \{0.3, 0.6, 0.7, 1.2\}$ *and let the weights be equal so that one must find* $\|\mu - \nu\|_{TV} = 1.5$*. From results displayed in Table 2, again one can see that even*

TABLE 2. $\|\mu - \nu\|_{TV}$ for two discrete measures; $X \cap Y = \{0.3\}$.

n	4	5	6
ρ_n	1.4879	1.4993	1.4997

if some points are relatively close to each other (and with 1 *point in common), the semidefinite relaxation* (3.1) *still provide a value* ρ_n *close to* 1.5*, as soon as* $n \geq 5$ *(i.e., with* $2n = 10$ *moments).*

Two Gaussian measures. We next consider the case where $\mu = \mathcal{N}(m_1, \sigma_1)$ and $\nu = \mathcal{N}(m_2, \sigma_2)$, and we fix the number of moments that we consider to be $2n = 4, 6, 8$. From results in Table 3 we can see the influence of a small variance, which tends to provide ρ_4 with a value close to 2, as expected since μ and ν behave almost like the two Dirac measures δ_{m_1} and δ_{m_2}, which are mutually singular whenever $m_1 \neq m_2$. It also turns out that ρ_1 coincide with the analytical lower bound provided in [14] on two arbitrary measures with given means and variances (m_1, σ_1) and (m_2, σ_2), namely

$$\|\mu - \nu\|_{TV} \;\geq\; 2\,\frac{(m_1 - m_2)^2}{(\sigma_1 + \sigma_2)^2 + (m_1 - m_2)^2}\,.$$

(See [14].) Notice that already with $n = 2$, i.e., with moments up to degree 4, ρ_n provides with a significant improvement in all cases.

TABLE 3. $\|\mu - \nu\|_{TV}$ for Gaussian measures $\mathcal{N}(m_1, \sigma_1)$ and $\mathcal{N}(m_2, \sigma_2)$

$(m_1\,,\,s_1)$	$(m_2\,,\,s_2)$	ρ_1	ρ_2	ρ_3	ρ_4
(0 , 0.1)	(1 , 0.1)	1.9231	1.9936	1.9991	1.9997
(0 , 0.2)	(1 , 0.2)	1.7241	1.9049	1.9376	1.939
(0 , 0.1)	(1 , 0.5)	1.4706	1.6267	1.6283	1.7032
(0 , 0.5)	(1 , 0.5)	1.0000	1.0000	1.1653	1.1897
(0.5 , 0.1)	(1 , 0.1)	1.7241	1.9049	1.9375	1.9378
(0.5 , 0.1)	(1 , 0.5)	0.8197	0.8497	1.1249	1.1294
(0.8 , 0.1)	(1 , 0.1)	1.000	1.0000	1.1645	1.1709
(0.8 , 0.05)	(1 , 0.1)	1.2800	1.3507	1.4123	1.4290
(0.8 , 0.05)	(1 , 0.01)	1.8349	1.9616	1.9785	1.9852

4. Conclusion

We have provided a numerical scheme to approximate as closely as desired the total variation distance between two measures μ and ν on $\mathbb{R}^d$. We have addressed this problem under fairly general assumptions on μ and ν (Carleman's condition or the easier to check sufficient condition (2.3)). Moreover, in case where μ and ν are only accessible via i.i.d. samples, and for a fixed value of the degree n, the SLLN ensures that empirical moments obtained from a sufficiently large sample, are enough for the step-n semidefinite relaxation to provide accurate results. Finally, even before convergence takes place, the optimal value of each semidefinite relaxation provides a useful guaranteed lower bound on the TV-distance, the larger n, the better. Of course this numerical scheme is sensitive to the dimension and so far is restricted to small dimension problems if good quality lower bounds are expected. (On the other hand, even crude lower bounds might be interesting in higher dimensional problems.) Therefore a topic of further investigation is to provide alternative and computationally cheaper lower bounds, possibly at the price of loosing convergence.

5. Appendix

Proof of Theorem 3.4.

Proof. Define the (monic) polynomials

$$x \mapsto p(x) := \prod_{i=1}^{m_1}(x - x(i))\,; \quad x \mapsto q(x) := \prod_{j=1}^{m_2}(x - y(j))\,, \tag{5.1}$$

with respective vector of coefficients $\mathbf{p} \in \mathbb{R}^{m_1+1}$ and $\mathbf{q} \in \mathbb{R}^{m_2+1}$ in the usual monomial basis. Observe that $q(x(i)) \neq 0$ for all $i = 1, \ldots, m_1$, and $p(y(j)) \neq 0$ for all $j = 1, \ldots, m_2$.

Let $(\boldsymbol{\phi}^*, \boldsymbol{\psi}^*)$ be an optimal solution of (3.1) with $n = n_0 := \max[m_1, m_2]$, and w.l.o.g. suppose that $n_0 = m_1$. Then from $\int p^2 \hat{d}\mu$ one deduce that $\mathbf{M}_{m_1}(\boldsymbol{\mu})\mathbf{p} = 0$ and combining with $0 \preceq \mathbf{M}_{m_1}(\boldsymbol{\phi}^*) \preceq \mathbf{M}_{m_1}(\boldsymbol{\mu})$, one also obtains $\mathbf{M}_{m_1}(\boldsymbol{\phi}^*)\mathbf{p} = 0$. Hence $\text{rank}(\mathbf{M}_{m_1}(\boldsymbol{\phi}^*)) = \text{rank}(\mathbf{M}_{m_1-1}(\boldsymbol{\phi}^*))$ because

– to every zero-eigenvector $\mathbf{h} \in \mathbb{R}^{m_1}$ of $\mathbf{M}_{m_1-1}(\boldsymbol{\phi}^*)$ (if any exists) corresponds a zero-eigenvector $(\mathbf{h}, 0) \in \mathbb{R}^{m_1+1}$ of $\mathbf{M}_{m_1}(\boldsymbol{\phi}^*)$. Indeed

$$0 \,=\, \mathbf{h}^T \mathbf{M}_{m_1-1}(\boldsymbol{\phi}^*)\mathbf{h} \,=\, \begin{pmatrix} \mathbf{h} \\ 0 \end{pmatrix}^T \mathbf{M}_{m_1}(\boldsymbol{\phi}^*) \begin{pmatrix} \mathbf{h} \\ 0 \end{pmatrix} \Rightarrow \mathbf{M}_{m_1}(\boldsymbol{\phi}^*) \begin{pmatrix} \mathbf{h} \\ 0 \end{pmatrix} \,=\, 0\,,$$

– the vector $\mathbf{p} \in \mathbb{R}^{m_1+1}$ of the polynomial $p \in \mathbb{R}[x]_{m_1}$ (and $p \notin \mathbb{R}[x]_{m_1-1}$) is in the kernel of $\mathbf{M}_{m_1}(\boldsymbol{\phi}^*)$ and not in the kernel of $\mathbf{M}_{m_1-1}(\boldsymbol{\phi}^*)$.

Then by Curto and Fialkow's flat extension theorem [10, Theorem 3.7, p. 62], $\boldsymbol{\phi}^*$ has a an atomic representing measure ϕ^* supported on at most $\text{rank}(\mathbf{M}_{m_1}(\boldsymbol{\phi}^*))$ points. In addition $\text{supp}(\phi^*) \subset X$ as $\int p^2 d\phi^* = 0$.

Next, with $m_2 \leq n = m_1$, and considering the sub-matrices $\mathbf{M}_{m_2-1}(\boldsymbol{\psi}^*)$ and $\mathbf{M}_{m_2}(\boldsymbol{\psi}^*)$ as principal submatrices of $\mathbf{M}_n(\boldsymbol{\psi}^*)$, a similar argument as above (but with q instead of p) yields $\text{rank}(\mathbf{M}_{m_2}(\boldsymbol{\psi}^*)) = \text{rank}(\mathbf{M}_{m_2-1}(\boldsymbol{\psi}^*))$. In addition, if $m_2 < n$ then consider the polynomials $x^k q \in \mathbb{R}[x]_{m_2+k}$, with respective vectors $\mathbf{q}_k \in \mathbb{R}^{m_2+k+1}$, $1 \leq k \leq n - m_2$. Observe that for every k, $\mathbf{M}_{m_2+k}(\boldsymbol{\psi}^*)\mathbf{q}_k = 0$ because $\mathbf{M}_{m_2+k}(\boldsymbol{\psi}^*) \preceq \mathbf{M}_{m_2+k}(\boldsymbol{\nu})$, and $\int q_k^2 d\nu = 0$).

Hence $\mathbf{q}_k \in \text{Ker}(\mathbf{M}_{m_2+k}(\boldsymbol{\psi}^*))$, for every $1 \leq k \leq n - m_2$, and repeating the arguments that we have used for ϕ^* and μ, one obtains $\text{rank}(\mathbf{M}_{m_2+k}(\boldsymbol{\psi}^*)) = \text{rank}(\mathbf{M}_{m_2}(\boldsymbol{\psi}^*))$ for every $k \leq n - m_2$. Therefore invoking again Curto and Fialkow's flat extension theorem, $\boldsymbol{\psi}^*$ has an atomic representing measure ψ^* supported on at most $\text{rank}(\mathbf{M}_{m_2}(\boldsymbol{\psi}^*))$ points with $\text{supp}(\psi^*) \subset Y$. Next, write

$$\mu = \sum_{i=1}^{m_1} \alpha_i \, \delta_{x(i)}\,; \quad \nu = \sum_{j=1}^{m_2} \beta_j \, \delta_{y(j)}\,, \quad \text{with } \alpha_i, \beta_j > 0,\, \forall i, j\,; \, \sum_i \alpha_i = \sum_j \beta_j = 1\,,$$

and from $\text{supp}(\phi^*) \subset X$ and $\text{supp}(\psi^*) \subset Y$, we can also write

$$\phi^* \,=\, \sum_{i=1}^{m_1} \alpha'_i \, \delta_{x(i)}\,; \quad \psi^* \,=\, \sum_{j=1}^{m_2} \beta'_j \, \delta_{y(j)}\,, \quad \text{with } \alpha'_i, \beta'_j \geq 0,\, \forall i, j.$$

Next, consider the interpolation polynomials

$$p_i(x) \,:=\, \frac{\prod_{\ell \neq i}(x - x(\ell))}{\prod_{\ell \neq i}(x(i) - x(\ell))}\,, \quad q_j(x) \,:=\, \frac{\prod_{\ell \neq j}(x - y(\ell))}{\prod_{\ell \neq j}(y(j) - y(\ell))}\,,$$

so that $p_i \in \mathbb{R}[x]_{m_1-1}$ and $q_j \in \mathbb{R}[x]_{m_2-1}$ for all $i = 1, \ldots, m_1$, $j = 1, \ldots, m_2$. With $n \geq \max[m_1, m_2]$, and using $0 \preceq \mathbf{M}_n(\boldsymbol{\phi}^*) \preceq \mathbf{M}_n(\boldsymbol{\mu})$, observe that

$$\alpha_i \,=\, \int p_i^2 \, d\mu \,\geq\, \int p_i^2 \, d\phi^* \quad (= \langle \mathbf{p}_i, \mathbf{M}_n(\boldsymbol{\phi}^*)\mathbf{p}_i \rangle) \,=\, \alpha'_i\,, \quad \forall i = 1, \ldots, m_1\,.$$

Similarly, using $\mathbf{M}_n(\boldsymbol{\psi}^*) \preceq \mathbf{M}_n(\boldsymbol{\nu})$,

$$\beta_j \,=\, \int q_j^2 \, d\nu \,\geq\, \int q_j^2 \, d\psi^* \quad (= \langle \mathbf{q}_j, \mathbf{M}_n(\boldsymbol{\psi}^*)\mathbf{q}_j \rangle) \,=\, \beta'_j\,, \quad \forall j = 1, \ldots, m_2\,.$$

Hence we may deduce that $\phi^* \leq \mu$ and $\psi^* \leq \nu$. In addition, since $2m_1 \leq 2n$, and as $\phi^*_j - \psi^*_j = \mu_j - \nu_j$ for all $j \leq 2n$,

$$0 \,=\, \int p^2 \, d(\mu - \phi^*) \,=\, \int p^2 \, d(\nu - \psi^*) \quad \Rightarrow \text{supp}(\nu - \psi^*) \subset X\,.$$

In particular this implies

$$\int x^k \, p \, d(\nu - \psi^*) \,=\, 0\,, \quad \forall k \in \mathbb{N}\,. \tag{5.2}$$

We want to prove that $\mu - \phi^* = \nu - \psi^*$. Indeed if true then (ϕ^*, ψ^*) is a feasible solution of (2.6) with value $\rho_n \leq \|\mu - \nu\|_{TV}$, which implies that (ϕ^*, ψ^*) is an optimal solution of (2.6) hence with $\rho_n = \|\mu - \nu\|_{TV}$, the desired result.

To prove that $\mu - \phi^* = \nu - \psi^*$ we first prove that given $j \in \mathbb{N}$,

$$\begin{aligned} \mu_k - \phi^*_k &= \nu_k - \psi^*_k\,, \quad \forall k \leq m_1 + j \\ \Rightarrow \mu_k - \phi^*_k &= \nu_k - \psi^*_k\,, \quad \forall k \leq m_1 + j + 1\,. \end{aligned} \tag{5.3}$$

This is already true for all $k \leq 2n$, i.e., $k \leq m_1 + j$ with $j = 2n - m_1$. With p as in (5.1), write $p(x) = x^{m_1} - \sum_{k=0}^{m_1-1} p_k\, x^k$, so that

$$x^{m_1+j+1} \,=\, x^{j+1}\, p(x) + \sum_{k=0}^{m_1-1} p_k\, x^{k+j+1}\,, \tag{5.4}$$

and therefore, integrating with respect to $\mu - \phi^*$, yields

$$\begin{aligned} \mu_{m_1+j+1} - \phi^*_{m_1+j+1} &= \underbrace{\int x^{j+1}\, p(x)\, d(\mu - \phi^*)}_{[=0 \text{ as } \mathrm{supp}(\mu), \mathrm{supp}(\phi^*) \subset X]} + \sum_{k=0}^{m_1-1} p_k\, (\mu_{k+j+1} - \phi^*_{k+j+1}) \\ &= \sum_{k=0}^{m_1-1} p_k\, (\mu_{k+j+1} - \phi^*_{k+j+1}) \\ &= \sum_{k=0}^{m_1-1} p_k\, (\nu_{k+j+1} - \psi^*_{k+j+1}) \text{ [by induction hypothesis]} \\ &= \int x^{m_1+j+1}\, d(\nu - \psi^*) - \underbrace{\int x^{j+1} p(x)\, d(\nu - \psi^*)}_{=0 \text{ by (5.2)}} \quad [\text{ using (5.4) }] \\ &= \int x^{m_1+j+1}\, d(\nu - \psi^*) \,=\, \nu_{m_1+j+1} - \psi^*_{m_1+j+1}\,, \end{aligned}$$

which proves that (5.3) is true. As $j \in \mathbb{N}$ was arbitrary, we have proved that

$$\mu_k - \phi^*_k \,=\, \nu_k - \psi^*_k\,, \quad \forall k \,\in\, N\,,$$

and as $\mu - \phi^* \geq 0$, $\nu - \psi^* \geq 0$, and their respective support are compact, one must have $\mu - \phi^* = \nu - \psi^*$, which implies the desired result. □

References

1. M. Anjos and J. B. Lasserre (eds.), *Handbook on Semidefinite, Conic and Polynomial Optimization*, Internat. Ser. Oper. Res. Management Sci., vol. 166, Springer, New York, 2012.
2. J. Arbas, H. Ashtiani, and C. Liaw, *Polynomial time and private learning of unbounded gaussian mixture models*, Tech. report, 2023, `arXiv.2303.04288`.
3. A. Barron, L. Györfi, and E.C. van der Meulen, *Distribution estimation consistent in total variation and in two types of information divergence*, IEEE Trans. Info. Theory **38** (1992), no. 5, 1437–1454.
4. S. Davies, A. Mazumdar, S. Pal, and C. Raschtchian, *Lower bounds on the total variation distance between mixtures of two gaussians*, Proc. Machine Learning Research, vol. 132, 2022, pp. 1–23.
5. L. Devroye and L. Györdi, *No empirical measure can converge in total variation sense for all distributions*, Ann. Stat. **18** (1990), 1496–1499.
6. L. Devroye, A. Mehrabian, and T. Redid, *The total variation distance between high-dimensional gaussians with the same mean*, Tech. report, 2018, `arXiv:1810.08693`.

7. D. Henrion, M. Korda, and J.B. Lasserre, *The moment-sos hierarchy: Lectures in probability, statistics, computational geometry, control and nonlinear pdes*, World Scientific, Singapore, 2020.
8. D. Henrion and J. B. Lasserre, *Graph recovery from incomplete moment information*, Constr. Approx. **56** (2022), 165–187.
9. D. Henrion, J. B. Lasserre, and J. Lofberg, *Gloptipoly 3: moments, optimization and semidefinite programming*, Optim. Methods and Softwares **24** (2009), 761–779.
10. J. B. Lasserre, *Moments, positive polynomials and their applications*, Imperial College Press, London, UK, 2009.
11. ———, *The moment-sos hierarchy*, Proceedings of the International Congress of Mathematicians (ICM 2018) (B. Sirakov, P. Ney de Souza, and M. Viana, eds.), vol. 4, World Scientific, 2019, pp. 3773–3794.
12. J. B. Lasserre, E. Pauwels, and M. Putinar, *The christoffel-Darboux Kernel for Data Aanalysis*, Cambridge University Press, Cambridge, UK, 2022.
13. M. Markatou and Yang Chen, *Non-quadratic distances in model assessment*, Entropy **20** (2018), no. 6, 464.
14. T. Nishiyama, *Lower bounds for the total variation distance given means and variances of distributions*, Tech. report, 2022, `arXiv:2212.05820`.
15. G. Peyré and M. Cuturi, *Computational Optimal Transport: With applications to Data Sscience*, Found. Trends in Machine Learning **11** (2019), no. 5-6, 355–607.
16. B. K. Sriperumbudur, K. Fukumizu, B. Schölkopf, and G.R.G. Landkriet, *On the empirical estimation of integral probability metrics*, Electr. J. Stat. **6** (2012), 1550–1599.
17. T. Weisser, B. Legat, C. Coey, L. Kapelevich, and J.P. Vielma, *Polynomial and moment optimization in Julia and Jump*, Juliacon, 2019.

LAAS-CNRS and Toulouse School of Economics (TSE), University of Toulouse, LAAS, 7 avenue du Colonel Roche, BP 54200, 31031 Toulouse Cédex 4, France
Email address: `lasserre@laas.fr`