



HAL
open science

Convergence of Lasserre's hierarchy: the general case

Matteo Tacchi

► **To cite this version:**

Matteo Tacchi. Convergence of Lasserre's hierarchy: the general case. *Optimization Letters*, 2022, 16 (3), pp.1015-1033. 10.1007/s11590-021-01757-6 . hal-03008545v2

HAL Id: hal-03008545

<https://laas.hal.science/hal-03008545v2>

Submitted on 3 Aug 2022 (v2), last revised 4 Aug 2022 (v3)

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

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

Convergence of Lasserre’s hierarchy: the general case

Matteo Tacchi^{1,2}

August 3, 2022

¹ École Polytechnique Fédérale, CH-1015 Lausanne, Switzerland.

² Réseau de Transport d’Electricité (RTE), 7c place du Dôme, Immeuble Window, 92073 Paris la Défense cedex, France.

✉ tacchi@laas.fr

Abstract

Lasserre’s moment-SOS hierarchy consists in approximating instances of the generalized moment problem (GMP) with moment relaxations and sums-of-squares (SOS) strengthening that boil down to convex semidefinite programming (SDP) problems. Due to the generality of the initial GMP, applications of this technology are countless, and one can cite among them the polynomial optimization problem (POP), the optimal control problem (OCP), the volume computation problem, stability sets approximation problems, and solving nonlinear partial differential equations (PDE). The solution to the original GMP is then approximated with finite truncatures of its moment sequence. For each application, proving convergence of these truncatures towards the optimal moment sequence gives valuable insight on the problem, including convergence of the relaxed values to the original GMP’s optimal value. This note proposes a general proof of such convergence, regardless the problem one is faced with, under simple standard assumptions. As a byproduct of this proof, one also obtains strong duality properties both in the infinite dimensional GMP and its finite dimensional relaxations.

Keywords

Infinite dimensional optimization; semidefinite programming; Lasserre hierarchy; numerical analysis; convex optimization; duality theory.

Declarations

Funding: This work was funded by the French Company Réseau de Transport d’Électricité.

Acknowledgements

The author would like to thank Didier Henrion and Jean Bernard Lasserre for fruitful discussions.

1 Introduction

- Let:
- $\mathbf{K} \subset \mathbb{R}^n$,
 - $c \in \mathbb{R}[\mathbf{x}]$,
 - \mathbb{A}, \mathbb{B} be countable index sets,
 - $\varphi_\alpha \in \mathbb{R}[\mathbf{x}], \alpha \in \mathbb{A}$,
 - $\psi_\beta \in \mathbb{R}[\mathbf{x}], \beta \in \mathbb{B}$,
 - $a_\alpha \in \mathbb{R}, \alpha \in \mathbb{A}$,
 - $b_\beta \in \mathbb{R}, \beta \in \mathbb{B}$,

and consider the generalized moment problem (GMP), as stated in [10, (1.1)]:

$$\begin{aligned}
 p_{\text{GM}}^* &:= \sup \int c \, d\mu & (1) \\
 \text{s.t.} \quad & \mu \in \mathcal{M}(\mathbf{K})_+ \\
 & \int \varphi_\alpha \, d\mu = a_\alpha \quad \alpha \in \mathbb{A} \\
 & \int \psi_\beta \, d\mu \leq b_\beta \quad \beta \in \mathbb{B},
 \end{aligned}$$

where $\mathcal{M}(\mathbf{K})_+$ denotes the cone of (nonnegative) Borel measures supported on \mathbf{K} , whose dual cone (w.r.t. the weak-* topology on measures) is the set $\mathcal{C}(\mathbf{K})_+$ of nonnegative continuous functions on \mathbf{K} .

Example (Generality of the problem).

- The \mathbf{K} -moment problem [10, (3.1),(3.2)] is a particular instance of Problem (1):

$$\begin{aligned}
 p_{\mathbf{z}}^* &:= \sup 0 \\
 \text{s.t.} \quad & \mu \in \mathcal{M}(\mathbf{K})_+ \\
 & \int \mathbf{x}^{\mathbf{k}} \, d\mu(\mathbf{x}) = z_{\mathbf{k}} \quad \mathbf{k} \in \Gamma,
 \end{aligned}$$

where $\Gamma \subset \mathbb{N}^n$. Feasibility of this problem means existence of a measure μ whose moments on \mathbf{K} coincide with the $z_{\mathbf{k}} \in \mathbb{R}, \mathbf{k} \in \Gamma$.

- The polynomial optimization problem (POP) [9, (4.1)] is another instance of Problem (1):

$$\begin{aligned}
 p_f^* &:= - \sup - \int f \, d\mu \\
 \text{s.t.} \quad & \mu \in \mathcal{M}(\mathbf{K})_+ \\
 & \int 1 \, d\mu = 1,
 \end{aligned}$$

where $f \in \mathbb{R}[\mathbf{x}]$. One can prove that $p_f^* = \inf\{f(\mathbf{x}) : \mathbf{x} \in \mathbf{K}\}$.

- The volume computation problem [5] is also an instance of Problem (1):

$$\begin{aligned}
 p_{\mathbf{K}}^* &:= \sup \int 1 \, d\mu \\
 \text{s.t.} \quad & \mu \in \mathcal{M}(\mathbf{K})_+ \\
 & \int p \, d\mu \leq \int_{\mathbf{B}} p(\mathbf{x}) \, d\mathbf{x} \quad p \in \mathbb{Q}[\mathbf{B}]_+,
 \end{aligned}$$

where $\mathbf{K} \subset \mathbf{B} \subset \mathbb{R}^n$ such that the $\int_{\mathbf{B}} \mathbf{x}^{\mathbf{k}} \, d\mathbf{x}$ are known and $\mathbb{Q}[\mathbf{B}]_+ := \{p \in \mathbb{Q}[\mathbf{x}] : \forall \mathbf{x} \in \mathbf{B}, p(\mathbf{x}) \geq 0\}$. One can prove that $p_{\mathbf{K}}^* = \text{vol } \mathbf{K}$ is the Lebesgue volume of \mathbf{K} .

In 2014, the authors of [6] proved that under specific conditions, the Lasserre hierarchy associated with the particular GMP instance known as the polynomial optimization problem (POP) has the strong duality property, which means that there is no duality gap between the moment relaxations and the sum-of-squares strengthenings that form the moment-SOS hierarchy.

Moreover, it is proved in [10, Theorem 5.6(b)] that if the POP has a unique minimizer \mathbf{x}^* , then the solutions to the corresponding moment hierarchy converge towards the moments of the Dirac measure $\delta_{\mathbf{x}^*}$ such that for all Borel set $\mathbf{A} \subset \mathbb{R}^n$, $\delta_{\mathbf{x}^*}(\mathbf{A}) = \begin{cases} 1 & \text{if } \mathbf{x}^* \in \mathbf{A} \\ 0 & \text{else.} \end{cases}$

This technical note is a follow-up that intends to close the subject of strong duality and convergence of the Lasserre hierarchy, by proving a general strong result at the abstraction level of the *generalized moment problem*.

First, it is possible to slightly generalize Problem (1) by allowing for multiple decision variables:

$$\begin{aligned} p_{\text{GM}}^* &:= \sup \int \mathbf{c} \cdot d\boldsymbol{\mu} & (2) \\ \text{s.t. } & \boldsymbol{\mu} \in \mathcal{M}(\mathbf{K}_1)_+ \times \dots \times \mathcal{M}(\mathbf{K}_N)_+ \\ & \int \boldsymbol{\varphi}_\alpha \cdot d\boldsymbol{\mu} = a_\alpha & \alpha \in \mathbb{A} \\ & \int \boldsymbol{\psi}_\beta \cdot d\boldsymbol{\mu} \leq b_\beta & \beta \in \mathbb{B}. \end{aligned}$$

where \mathbb{B} is finite, $\mathbf{c}, \boldsymbol{\varphi}_\alpha, \boldsymbol{\psi}_\beta \in \mathbb{R}[\mathbf{x}_1] \times \dots \times \mathbb{R}[\mathbf{x}_N]$ and

$$\int \mathbf{c} \cdot d\boldsymbol{\mu} := \sum_{i=1}^N \int c_i d\mu_i.$$

Remark 1 (Finitely many inequality constraints).

The introduction of multiple decision variables allows one to restrict to finitely many inequality constraints. Indeed, in the context of the moment-SOS hierarchy, infinite dimensional inequality constraints are always under the form $\nu - \Phi\boldsymbol{\mu} \in \mathcal{M}(\Omega)_+$ for some compact $\Omega \subset \mathbb{R}^n$ and some bounded operator $\Phi : \mathcal{M}(\mathbf{K}_1) \times \dots \times \mathcal{M}(\mathbf{K}_N) \rightarrow \mathcal{M}(\Omega)$ between vector spaces of signed measures. This means that these infinite dimensional inequality constraints can be recast as equality constraints with an additional decision variable $\bar{\mu} \in \mathcal{M}(\Omega)_+$: $\Phi\boldsymbol{\mu} + \bar{\mu} = \nu$. For example, the aforementioned volume computation problem can be recast as

$$\begin{aligned} p_{\mathbf{K}}^* &= \sup \int 1 d\mu \\ \text{s.t. } & \mu \in \mathcal{M}(\mathbf{K})_+, \quad \bar{\mu} \in \mathcal{M}(\mathbf{B})_+ \\ & \int \mathbf{x}^{\mathbf{k}} d\mu(\mathbf{x}) + \int \mathbf{x}^{\mathbf{k}} d\bar{\mu}(\mathbf{x}) = \int \mathbf{x}^{\mathbf{k}} d\mathbf{x} \quad \mathbf{k} \in \mathbb{N}^n. \end{aligned}$$

Example (A most general extension).

Problem (2) covers all existing applications of the Lasserre hierarchy so far, including for instance:

- The optimal control problem (OCP) [11],
- The controlled region of attraction problem [3] (as well as its inner approximation version [8] and its extension [7] to invariant sets),

- The reachable set problem [12] (as well as its continuous time counterpart, formulated as the maximal positively invariant set inner approximation problem in [15]),
- The nonlinear conservation PDE problem [13, (45),(46)].

Remark 2 (Duality).

In practice, instances of Problem (2) are formulated so that

$$\forall t_1, t_2 \in \mathbb{R}, \alpha_1, \alpha_2, \alpha_3 \in \mathbb{A}, \quad t_1 \varphi_{\alpha_1} + t_2 \varphi_{\alpha_2} = \varphi_{\alpha_3} \Rightarrow t_1 a_{\alpha_1} + t_2 a_{\alpha_2} = a_{\alpha_3}$$

(and same for the ψ_β and b_β). This compatibility with the linearity of optimization constraints allows to write a synthetic dual formulation to (2):

$$\begin{aligned} d_{\text{GM}}^* &:= \inf \mathbf{x} \cdot \mathbf{a} + \mathbf{y} \cdot \mathbf{b} & (3) \\ \text{s.t. } & \mathbf{v} + \mathbf{w} - \mathbf{c} \in \mathcal{C}(\mathbf{K}_1)_+ \times \dots \times \mathcal{C}(\mathbf{K}_N)_+ \\ & \mathbf{v} = \sum_{\alpha \in \mathbb{A}} x_\alpha \varphi_\alpha \quad \mathbf{x} = (x_\alpha)_{\alpha \in \mathbb{A}} \in \mathbb{R}^{\mathbb{A}} \text{ s.t. } \{\alpha \in \mathbb{A} : x_\alpha \neq 0\} \text{ is finite} \\ & \mathbf{w} = \sum_{\beta \in \mathbb{B}} y_\beta \psi_\beta \quad \mathbf{y} = (y_\beta)_{\beta \in \mathbb{B}} \in (\mathbb{R}_+)^{\mathbb{B}} \end{aligned}$$

where $\mathbf{a} := (a_\alpha)_{\alpha \in \mathbb{A}}$ and $\mathbf{b} := (b_\beta)_{\beta \in \mathbb{B}}$. In words, the decision variable \mathbf{v} is in the vector space spanned by the φ_α and \mathbf{w} is in the convex cone spanned by the ψ_β . \mathbf{x} (resp. \mathbf{y}) is the Lagrange multiplier corresponding to the equality (resp. inequality) constraints in (2).

The aim of this technical note is to state and prove a most general convergence theorem for the hierarchy corresponding to Problem (2) (and its dual (3)). The Lasserre hierarchy technology is most often deployed under the following standard working assumption (see [6, Assumptions 1 & 2]):

Assumption 1 (Ball constraints).

Suppose that there exists $\mathbf{g}_1 \in \mathbb{R}[\mathbf{x}_1]^{m_1}, \dots, \mathbf{g}_N \in \mathbb{R}[\mathbf{x}_N]^{m_N}$ such that for $i \in \mathbb{N}_N^* := \{1, \dots, N\}$,

$$\begin{aligned} g_{i,1}(\mathbf{x}_i) &= 1 \quad ; \quad g_{i,m_i}(\mathbf{x}_i) = 1 - \|\mathbf{x}_i\|^2 \\ \mathbf{K}_i &= \{\mathbf{x}_i \in \mathbb{R}^{n_i} : \mathbf{g}_i(\mathbf{x}_i) \geq \mathbf{0}\}, \end{aligned}$$

where if $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$, $\|\mathbf{x}\| = \sqrt{x_1^2 + \dots + x_n^2}$.

Remark 3 (Compactness).

Up to rescaling of the \mathbf{K}_i 's, it is always possible to enforce Assumption 1, as soon as they are compact basic semialgebraic sets.

Indeed let $\mathbf{g} \in \mathbb{R}[\mathbf{x}]^m$, $\mathbf{K} := \{\mathbf{x} \in \mathbb{R}^n : \mathbf{g}(\mathbf{x}) \geq \mathbf{0}\}$. For $R > 0$, define

$$\mathbf{K}_R := \{\mathbf{x} \in \mathbb{R}^n : \mathbf{g}(\mathbf{x}) \geq \mathbf{0}, \|\mathbf{x}\|^2 \leq R^2\} = \mathbf{K} \cap \mathbf{B}_R,$$

where $\mathbf{B}_R := \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| \leq R\}$ is the ball of radius R . In such setting, if \mathbf{K} is compact then it is bounded, so that there exists $R_0 > 0$ s.t. $\forall R \geq R_0$, $\mathbf{K} \subset \mathbf{B}_R$, and thus $\mathbf{K} = \mathbf{K}_R$. This shows that if \mathbf{K} is compact, it is always possible to add a redundant ball constraint to its description.

Remark 4 (Archimedean property).

More precisely, the standard condition for applying the moment-SOS hierarchy is actually that the \mathbf{K}_i 's have the Archimedean property, which we explain here. Considering the cone

$$\Sigma[\mathbf{x}] := \{p_1^2 + \dots + p_k^2 : k \in \mathbb{N}, p_1, \dots, p_k \in \mathbb{R}[\mathbf{x}]\}$$

of sums of squares of polynomials, we define the *quadratic module* of $\mathbf{g} \in \mathbb{R}[\mathbf{x}]^m$ as

$$\Sigma(\mathbf{g}) := \{\boldsymbol{\sigma} \cdot \mathbf{g} : \boldsymbol{\sigma} \in \Sigma[\mathbf{x}]^m\}.$$

Then, $\mathbf{K} := \{\mathbf{x} \in \mathbb{R}^n : \mathbf{g}(\mathbf{x}) \geq \mathbf{0}\}$ is said to have the *Archimedean property* if

$$\exists R > 0 \text{ s.t. } R^2 - \|\cdot\|^2 \in \Sigma(\mathbf{g}).$$

It is worth noticing, as SOS polynomials are by definition nonnegative, that by definition of \mathbf{K} one has $\Sigma(\mathbf{g}) \subset \mathbb{R}[\mathbf{K}]_+$ and thus the Archimedean property implies $\mathbf{K} \subset \mathbf{B}_R$ for some $R > 0$.

Remark 5 (Positivstellensatz).

The key instrument for the moment-SOS hierarchy is Putinar's Positivstellensatz [16, Theorem 1.3], which has a dual counterpart under the form of Putinar's Lemma [16, Lemma 3.2]. If the Archimedean property holds, then Putinar's Lemma gives a sufficient hierarchy of LMI conditions for a multi-indexed sequence $\mathbf{z} = (z_{\mathbf{k}})_{\mathbf{k} \in \mathbb{N}^n}$ to represent the moments of a measure $\mu \in \mathcal{M}(\mathbf{K}_i)_+$, which will be instrumental for the moment hierarchy definition and convergence.

Conversely, under the same assumption, Putinar's Positivstellensatz ensures that $\Sigma(\mathbf{g}_i)$ is dense in $\mathcal{C}(\mathbf{K}_i)_+$ w.r.t. the uniform topology, so that

$$\begin{aligned} d_{\text{GM}}^* &:= \inf \mathbf{x} \cdot \mathbf{a} + \mathbf{y} \cdot \mathbf{b} \\ &\text{s.t. } \mathbf{v} + \mathbf{w} - \mathbf{c} \in \Sigma(\mathbf{g}_1)_+ \times \dots \times \Sigma(\mathbf{g}_N)_+ \\ \mathbf{v} &= \sum_{\alpha \in \mathbb{A}} x_{\alpha} \varphi_{\alpha} & \mathbf{x} &= (x_{\alpha})_{\alpha \in \mathbb{A}} \in \mathbb{R}^{\mathbb{A}} \text{ s.t. } \{\alpha \in \mathbb{A} : x_{\alpha} \neq 0\} \text{ is finite} \\ \mathbf{w} &= \sum_{\beta \in \mathbb{B}} y_{\beta} \psi_{\beta} & \mathbf{y} &= (y_{\beta})_{\beta \in \mathbb{B}} \in (\mathbb{R}_+)^{\mathbb{B}} \end{aligned}$$

Then, the isomorphism between sums of squares and positive semidefinite matrices, provided by [10, Proposition 2.1], allows to parameterize inequality constraints as LMI constraints.

Note that this means that when restricting the constraints to bounded degree in the hierarchy, only a finite number of constraints need to be checked.

2 Notations and theorem statement

We briefly recall the usual notations and definitions that are used in the Lasserre hierarchy framework. Then, we proceed to introduce the last assumptions that will be needed for the proof of our theorem.

The basic notions that are used to formulate the moment hierarchy are the Riesz functional and the localizing matrix.

Definition (Riesz functional).

Let $\mathbf{z} := (z_{\mathbf{k}})_{\mathbf{k} \in \mathbb{N}^n} \in \mathbb{R}^{\mathbb{N}^n}$ be a real sequence. We define, for $p(\mathbf{x}) := \sum_{|\mathbf{k}| \leq d} p_{\mathbf{k}} \mathbf{x}^{\mathbf{k}} \in \mathbb{R}[\mathbf{x}]$,

$$\mathbf{L}_{\mathbf{z}}(p) := \sum_{|\mathbf{k}| \leq d} p_{\mathbf{k}} z_{\mathbf{k}},$$

where for $\mathbf{k} = (k_1, \dots, k_n) \in \mathbb{N}^n$, $|\mathbf{k}| := k_1 + \dots + k_n$.

The linear map $\mathbf{L}_{\mathbf{z}} : p \mapsto \mathbf{L}_{\mathbf{z}}(p)$ is called the *Riesz functional* of \mathbf{z} .

Remark 6 (Link between $L_{\mathbf{z}}$ and integrals).

If $\mathbf{z} = (z_{\mathbf{k}})_{\mathbf{k} \in \mathbb{N}^n} \in \mathbb{R}^{\mathbb{N}^n}$ and $\mu \in \mathcal{M}(\mathbf{K})_+$ are such that

$$\forall \mathbf{k} \in \mathbb{N}^n \quad z_{\mathbf{k}} = \int \mathbf{x}^{\mathbf{k}} d\mu,$$

then by definition for all $p \in \mathbb{R}[\mathbf{x}]$

$$\int p d\mu = L_{\mathbf{z}}(p).$$

Notation (Bounded multi-indices).

For $n, d \in \mathbb{N}$, we use the following notations:

$$\mathbb{N}_d^n := \{\mathbf{k} \in \mathbb{N}^n : |\mathbf{k}| \leq d\} \quad \text{and} \quad N_{n,d} := \text{card}(\mathbb{N}_d^n) = \binom{n+d}{n},$$

so that the space $\mathbb{R}_d[\mathbf{x}]$ of polynomials in n variables with degree at most d is isomorphic to $\mathbb{R}^{\mathbb{N}_d^n}$ and $\mathbb{R}^{N_{n,d}}$.

Definition (Localizing matrix).

Let $d, d_g \in \mathbb{N}$, $g \in \mathbb{R}_{d_g}[\mathbf{x}]$. Let $\mathbf{e}_d(\mathbf{x}) := (e_i(\mathbf{x}))_{1 \leq i \leq N_{n,d}}$ be a base of $\mathbb{R}_d[\mathbf{x}]$. Let $\mathbf{z} = (z_{\mathbf{k}})_{|\mathbf{k}| \leq 2d+d_g} \in \mathbb{R}^{\mathbb{N}_{2d+d_g}^n}$.

The *degree d localizing matrix* $M_d(g, \mathbf{z})$ of \mathbf{z} in g is defined as the size $N_{n,d}$ matrix representation in base $\mathbf{e}_d(\mathbf{x})$ of the bilinear application

$$(p, q) \in \mathbb{R}_d[\mathbf{x}]^2 \mapsto L_{\mathbf{z}}(g p q).$$

The localizing matrix is defined so that if $p(\mathbf{x}) = \mathbf{p} \cdot \mathbf{e}_d(\mathbf{x})$ and $q(\mathbf{x}) = \mathbf{q} \cdot \mathbf{e}_d(\mathbf{x})$, $\mathbf{p}, \mathbf{q} \in \mathbb{R}^{N_{n,d}}$, then

$$L_{\mathbf{z}}(g p q) = \mathbf{p}^\top M_d(g, \mathbf{z}) \mathbf{q} \quad \text{and} \quad L_{\mathbf{z}}(g p^2) = \mathbf{p}^\top M_d(g, \mathbf{z}) \mathbf{p}.$$

Notation (Positive semidefinite matrices).

We denote by:

- $\mathbb{S}^n := \{M \in \mathbb{R}^{n \times n} : M^\top = M\}$ the space of symmetric matrices,
- $\mathbb{S}_+^n := \{M \in \mathbb{S}^n : \forall \mathbf{x} \in \mathbb{R}^n, \mathbf{x}^\top M \mathbf{x} \geq 0\}$ the closed convex cone of positive semidefinite matrices,
- if $M \in \mathbb{S}^n$, $M \succeq 0 \iff M \in \mathbb{S}_+^n$.

These definitions allow to formulate the moment hierarchy associated to Problem (2):

$$\begin{aligned} \text{pGM}^d := \sup & \sum_{i=1}^N L_{\mathbf{z}_i}(c_i) & (4) \\ \text{s.t. } & \mathbf{z}_i \in \mathbb{R}^{\mathbb{N}_{2d}^n} & i \in \mathbb{N}_N^* \\ & M_{d-d_{ij}}(g_{i,j}, \mathbf{z}_i) \succeq 0 & i \in \mathbb{N}_N^*, j \in \mathbb{N}_{m_i}^* \\ & \sum_{i=1}^N L_{\mathbf{z}_i}(\varphi_{\alpha,i}) = a_\alpha & \alpha \in \mathbb{A}_d \\ & \sum_{i=1}^N L_{\mathbf{z}_i}(\psi_{\beta,i}) \leq b_\beta & \beta \in \mathbb{B}_d, \end{aligned}$$

where $d_{ij} = \lceil \deg g_{i,j}/2 \rceil$, and

$$\mathbb{A}_d := \{\alpha \in \mathbb{A} : \forall i \in \mathbb{N}_N^*, \deg \varphi_{\alpha,i} \leq 2d\},$$

$$\mathbb{B}_d := \{\beta \in \mathbb{B} : \forall i \in \mathbb{N}_N^*, \deg \psi_{\beta,i} \leq 2d\}$$

are taken finite (which is possible using linearity of the Riesz functional, Remark 5 and finite dimensionality of $\mathbb{R}_d[\mathbf{x}] = \{p \in \mathbb{R}[\mathbf{x}] : \deg p \leq d\}$).

A vector $\mathbf{z}_i \in \mathbb{R}^{\mathbb{N}_{2d}^{n_i}}$ that is feasible for problem (4) is then called a *pseudo-moment sequence*.

Remark 7 (SOS hierarchy).

The dual to the moment hierarchy is called the SOS hierarchy, and is written as follows:

$$\begin{aligned} d_{\text{GM}}^d &:= \inf \mathbf{x} \cdot \mathbf{a} + \mathbf{y} \cdot \mathbf{b} & (5) \\ \text{s.t. } & \mathbf{v} + \mathbf{w} - \mathbf{c} \in \Sigma(\mathbf{g}_1) \times \dots \times \Sigma(\mathbf{g}_N) \\ & \mathbf{v} = \sum_{\alpha \in \mathbb{A}_d} x_\alpha \varphi_\alpha & \mathbf{x} = (x_\alpha)_{\alpha \in \mathbb{A}_d} \in \mathbb{R}^{\mathbb{A}_d} \\ & \mathbf{w} = \sum_{\beta \in \mathbb{B}_d} y_\beta \psi_\beta & \mathbf{y} = (y_\beta)_{\beta \in \mathbb{B}_d} \in (\mathbb{R}_+)^{\mathbb{B}_d}, \end{aligned}$$

where $\mathbf{a} := (a_\alpha)_{\alpha \in \mathbb{A}_d}$ and $\mathbf{b} := (b_\beta)_{\beta \in \mathbb{B}_d}$.

Remark 8 (Minimal degree for the hierarchy).

It appears in (4) that p_{GM}^d is only defined for $d \geq d_0$, where

$$d_0 := \max_{i \in \{1, \dots, N\}} \max_{j \in \{1, \dots, m_i\}} \max(\lceil \deg c_i/2 \rceil, d_{ij}).$$

We are finally able to formulate the last needed assumptions as well as our main theorem. First, convergence in the moment hierarchy is often obtained through an additional assumption on the mass of the involved measures:

Assumption 2 (Uniformly bounded mass).

For $i \in \{1, \dots, N\}$, denote by $\mathbf{z}_i \in \mathbb{R}^{\mathbb{N}_{2d}^{n_i}}$ the pseudo-moment sequence that represents μ_i in the Lasserre hierarchy, and suppose that there exists $C_i > 0$ s.t. if \mathbf{z}_i is feasible for the degree d relaxation of problem (2), then

$$z_{i,0} \leq C_i.$$

Eventually, our proof of convergence consists in proving that our hierarchy of pseudo-moment sequences has at least one accumulation point. So as to deduce actual convergence, we need a unique candidate for such accumulation point:

Assumption 3 (Existence of a unique optimal solution).

Suppose that there exists a unique $\mu^* \in \mathcal{M}(\mathbf{K}_1)_+ \times \dots \times \mathcal{M}(\mathbf{K}_N)_+$ feasible and optimal for problem (2):

$$\begin{cases} \forall \alpha \in \mathbb{A}, \int \varphi_\alpha \cdot d\mu^* = a_\alpha \\ \forall \beta \in \mathbb{B}, \int \psi_\beta \cdot d\mu^* \leq b_\beta \end{cases} \quad \text{and} \quad \int \mathbf{c} \cdot d\mu^* = p_{\text{GM}}^*.$$

Remark 9 (The question of uniqueness).

Assumption 3 is crucial for the proof of convergence. However, if one removes it, *existence* of the optimal μ^* could still be obtained from Assumption 2, through an infinite dimensional strong duality proof similar to what we present in the next section (which would bring no additional theoretical insight, so that we do not display it). The actual stake of this assumption is indeed the *uniqueness* assumption. Note that in most of the moment-SOS hierarchy applications, the GMP is designed in a way that enforces Assumption 3.

Our contribution consists in stating and proving a general theorem for convergence of the Lasserre hierarchy (4) corresponding to problem (2):

Theorem 4 (Convergence of the general Lasserre hierarchy).

Under Assumptions 1, 2 and 3, there exist sequences $(\mathbf{z}_i^d)_{d \geq d_0}$ of feasible pseudo-moment sequences for the moment hierarchy (4) s.t. $\sum_{i=1}^N L_{\mathbf{z}_i^d}(c_i) = p_{\text{GM}}^d$. Then, for all $\mathbf{k} \in \mathbb{N}^n$,

$$z_{i,\mathbf{k}}^d \xrightarrow{d \rightarrow \infty} \int \mathbf{x}^{\mathbf{k}} d\mu_i^*(\mathbf{x}).$$

In particular, one has $p_{\text{GM}}^d \xrightarrow{d \rightarrow \infty} p_{\text{GM}}^*$.

Moreover, this automatically yields strong duality $p_{\text{GM}}^d = d_{\text{GM}}^d$ and $p_{\text{GM}}^* = d_{\text{GM}}^*$, where d_{GM}^* and d_{GM}^d are the values of the duals (3) and (5), respectively, so that $d_{\text{GM}}^d \xrightarrow{d \rightarrow \infty} d_{\text{GM}}^*$.

Remark 10 (Contribution w.r.t. [6] and [10]).

Whereas [6] and [10, Theorem 5.6(b)] respectively prove strong duality in the moment-SOS hierarchy and convergence of the resulting pseudo-moments towards the unique solution to the original POP, this contribution extends these results from the particular POP problem to *all* possible instances of the GMP, including the various examples mentioned in introduction, which should facilitate future contributions in the field of moment-SOS hierarchies.

Such generality is obtained by enforcing, in addition to the standard Assumptions 1 and 3 that were already introduced in [6] and [10, Theorem 5.6(b)], of the boundedness Assumption 2 that guarantees strong duality in the GMP, a property that is trivially established in the context of POP but can become difficult to check in other situations, such as [12] and [15].

3 Proof of Theorem 4

For the sake of simplicity and without loss of generality, we work on the simple case (1) where $N = 1$, with $\mathbb{B} = \emptyset^1$, whose corresponding hierarchy writes:

$$\begin{aligned} p_{\text{GM}}^d &:= \sup L_{\mathbf{z}}(c) & (6) \\ \text{s.t. } \mathbf{z} &\in \mathbb{R}^{\mathbb{N}_{2d}^n} \\ M_{d-d_j}(g_j \mathbf{z}) &\succeq 0 & j \in \mathbb{N}_m^* \\ L_{\mathbf{z}}(\varphi_\alpha) &= a_\alpha & \alpha \in \mathbb{A}_d. \end{aligned}$$

All the proofs are the same for $N > 1$, at the price of additional notations that do not bring any theoretical insight, so that we stick to the case $N = 1$. We first prove an easy lemma that gives all its importance to Assumption 1:

Lemma 5 (Pseudo moment sequences boundedness).

Let $d \in \mathbb{N} \setminus \{0\}$, $R > 0$, $\mathbf{z} \in \mathbb{R}^{\mathbb{N}_{2d}^n}$ s.t. $M_d(\mathbf{z}) \succeq 0$ and $M_{d-1}((R^2 - \|\cdot\|^2) \mathbf{z}) \succeq 0$. Then,

$$\min_{|\mathbf{k}| \leq d} z_{2\mathbf{k}} \geq 0 \quad \text{and} \quad \max_{|\mathbf{k}| \leq 2d} |z_{\mathbf{k}}| \leq z_{\mathbf{0}} \max(1, R^{2d}).$$

¹Indeed as \mathbb{B} is finite, there exists a finite value of d that captures all inequality constraints; arbitrarily setting this value to 0 does not change our proof for convergence.

Proof. $M_d(\mathbf{z}) \succeq 0$ is equivalent to

$$\forall p \in \mathbb{R}_d[\mathbf{x}], L_{\mathbf{z}}(p^2) \geq 0, \quad (\text{a})$$

while $M_{d-1}((R^2 - \|\cdot\|^2) \mathbf{z}) \succeq 0$ means that

$$\forall p \in \mathbb{R}_{d-1}[\mathbf{x}], L_{\mathbf{z}}((R^2 - \|\cdot\|^2) p^2) \geq 0. \quad (\text{b})$$

(a) with $p(\mathbf{x}) = \mathbf{x}^{\mathbf{k}}$, $|\mathbf{k}| \leq d$ yields $z_{2\mathbf{k}} \geq 0$.

(b) with $p(\mathbf{x}) = 1$ yields $R^2 z_{\mathbf{0}} \geq \sum_{|\mathbf{k}|=1} z_{2\mathbf{k}}$, since $\|\mathbf{x}\|^2 = \sum_{j=1}^n x_j^2 = \sum_{|\mathbf{k}|=1} \mathbf{x}^{2\mathbf{k}}$. Hence, since the $z_{2\mathbf{k}}$ are nonnegative, one has $|\mathbf{k}| = 1 \Rightarrow z_{2\mathbf{k}} \leq R^2 z_{\mathbf{0}}$.

Going forward, if $|\mathbf{k}| = 1$, (b) with $p(\mathbf{x}) = \mathbf{x}^{\mathbf{k}}$ yields $R^2 z_{2\mathbf{k}} \geq \sum_{|\mathbf{k}'|=1} z_{2(\mathbf{k}+\mathbf{k}')} with $z_{2(\mathbf{k}+\mathbf{k}')} \geq 0$ by (a), so that $R^4 z_{\mathbf{0}} \geq R^2 z_{2\mathbf{k}} \geq z_{2(\mathbf{k}+\mathbf{k}'})$ as long as $|\mathbf{k}| = |\mathbf{k}'| = 1$, and thus, if $|\mathbf{k}| = 2$, $R^4 z_{\mathbf{0}} \geq z_{2\mathbf{k}}$. By induction, one has for $\mathbf{k} \in \mathbb{N}_d^n$ that$

$$0 \leq z_{2\mathbf{k}} \leq R^{2|\mathbf{k}|} z_{\mathbf{0}} \leq z_{\mathbf{0}} \max(1, R^{2d}). \quad (\text{c})$$

Let $\mathbf{k}, \mathbf{k}' \in \mathbb{N}_d^n$. Then, (a) with $p(\mathbf{x}) = \mathbf{x}^{\mathbf{k}} \pm \mathbf{x}^{\mathbf{k}'}$ yields $0 \leq L_{\mathbf{z}}(p^2) = z_{2\mathbf{k}} \pm 2z_{\mathbf{k}+\mathbf{k}'} + z_{2\mathbf{k}'}$ so that

$$|z_{\mathbf{k}+\mathbf{k}'}| \leq \frac{z_{2\mathbf{k}} + z_{2\mathbf{k}'}}{2} \leq \max(z_{2\mathbf{k}}, z_{2\mathbf{k}'}) \stackrel{(\text{c})}{\leq} z_{\mathbf{0}} \max(1, R^{2d}). \quad (\text{d})$$

Eventually, any $\mathbf{k} \in \mathbb{N}_{2d}^n$ can be written $\mathbf{k} = \mathbf{k}' + \mathbf{k}''$ with $\mathbf{k}', \mathbf{k}'' \in \mathbb{N}_d^n$, so that it satisfies (d), and $|z_{\mathbf{k}}| \leq z_{\mathbf{0}} \max(1, R^{2d})$. \square

This lemma proves several important facts, among which any nonzero feasible pseudo-moment vector \mathbf{z}_i for (4) satisfies $z_{i,\mathbf{0}} > 0$ if Assumption 1 holds, and under Assumptions 1 and 2, for all $d \in \mathbb{N}$, \mathbf{z}_i feasible for (4) satisfies $|z_{i,\mathbf{k}}| \leq C_i \forall \mathbf{k}$.

Remark 11 (Relaxing Assumption 1).

Actually, this lemma still holds if one replaces Assumption 1 with the Archimedean property. Indeed, suppose that $R^2 - \|\cdot\|^2 \in \Sigma(\mathbf{g})$ for some $R > 0$. Then, there exists $\boldsymbol{\sigma} \in \Sigma[\mathbf{x}]^m$ such that

$$R^2 - \|\cdot\|^2 = \boldsymbol{\sigma} \cdot \mathbf{g} = \sum_{i=1}^m \sigma_i g_i = \sum_{i=1}^m \sum_{j=1}^{k_i} p_{i,j}^2 g_i.$$

Then, let $D := \max_{i,j} \lceil \deg(p_{i,j}^2)/2 \rceil$ and for $i \in \mathbb{N}_m^*$ let $d_i := d + D - 1$ and suppose $M_{d_i}(g_i \mathbf{z}) \succeq 0$. In that case, for $q \in \mathbb{R}_{d-1}[\mathbf{x}]$, one recovers (b):

$$L_{\mathbf{z}}((R^2 - \|\cdot\|^2) q^2) = \sum_{i=1}^m \sum_{j=1}^{k_i} L_{\mathbf{z}}(g_i p_{i,j}^2 q^2) \geq 0.$$

However, we chose to keep Assumption 1 instead of the Archimedean property, as the former is much easier to verify in practice and can always be enforced when the latter holds, at the price of adding a redundant ball constraint to the \mathbf{K}_i 's descriptions (see remarks 3 and 4).

A first consequence of Lemma 5 is the strong duality property in the hierarchy.

Proposition 6 (Strong duality in the hierarchy).

Suppose that each relaxation of Problem (1) has a feasible solution. In that case, under Assumptions 1 and 2,

$$\forall d \geq d_0, \quad \text{p}_{\text{GM}}^d = \text{d}_{\text{GM}}^d.$$

Moreover, the degree d relaxation (6) of (1) has an optimal solution $\mathbf{z}_d \in \mathbb{R}^{\mathbb{N}_{2d}^n}$.

Proof. We rely on [1, Chapter IV: Theorem (7.2), Lemma (7.3)] to prove our result. Consider the cone

$$\mathcal{K}_d := \{(A_d \mathbf{z}, L_{\mathbf{z}}(c)) : \mathbf{z} \in \mathcal{X}_d\},$$

where $A_d \mathbf{z} := (L_{\mathbf{z}}(\psi_\beta))_{\beta \in \mathbb{B}_d}$ and

$$\mathcal{X}_d := \left\{ \mathbf{z} \in \mathbb{R}^{\mathbb{N}_{2d}^n} : \forall j \in \{1, \dots, m\}, M_{d-d_j}(g_j \mathbf{z}) \succeq 0 \right\}.$$

According to [1, Theorem (7.2)] a sufficient condition for our result to hold requires that $p_{\text{GM}}^d < \infty$ and \mathcal{K}_d is closed. Clearly,

$$p_{\text{GM}}^d \leq N_{n,d_0} \left(\max_{\mathbf{k} \in \mathbb{N}^n} c_{\mathbf{k}} \right) \left(\max_{\mathbf{z}} \max_{\mathbf{k} \in \mathbb{N}_{2d}^n} z_{\mathbf{k}} \right) \leq N_{n,d_0} \left(\max_{\mathbf{k} \in \mathbb{N}^n} c_{\mathbf{k}} \right) C < \infty.$$

Besides, [1, Lemma (7.3)] states that for \mathcal{K}_d to be closed, it is sufficient to prove that \mathcal{X}_d has a compact, convex base, and that

$$\forall \mathbf{z} \in \mathcal{X}_d, \quad (A_d \mathbf{z}, L_{\mathbf{z}}(c)) = (\mathbf{0}, 0) \implies \mathbf{z} = \mathbf{0}. \quad (*)$$

We first exhibit a compact convex base for \mathcal{X}_d . A base for a cone is defined as follows:

Definition (Convex base of a cone). Let \mathcal{X} be a cone, *i.e.* a subset of a real vector space that is invariant under multiplication by nonnegative real numbers. $\mathcal{P} \subset \mathcal{X}$ is said to be a *base* of \mathcal{X} if the map

$$\chi : \begin{cases} (0, +\infty) \times \mathcal{P} & \longrightarrow \mathcal{X} \setminus \{\mathbf{0}\} \\ (t, \mathbf{p}) & \longmapsto t \mathbf{p} \end{cases}$$

is a bijection.

We are looking for such a base, with the additional properties that it should be convex (*i.e.* stable by barycenter operation) and compact (*i.e.*, in this finite dimensional context, closed and bounded). Let

$$\mathcal{P}_d := \{\mathbf{z} = (z_{\mathbf{k}})_{\mathbf{k}} \in \mathcal{X}_d : z_{\mathbf{0}} = 1\}.$$

\mathcal{P}_d is a base of \mathcal{X}_d in the sense that $\mathcal{X}_d \setminus \{\mathbf{0}\}$ is isomorphic to $(0, +\infty) \times \mathcal{P}_d$ through the bijective application $\chi_d : (t, \mathbf{p}) \mapsto t \mathbf{p}$, with $\chi_d^{-1}(\mathbf{z}) = (z_{\mathbf{0}}, \mathbf{z}/z_{\mathbf{0}})$ (using Lemma 5 and Assumption 1, for any $M_d(\mathbf{z})$ and $M_{d-1}((1 - \|\cdot\|^2) \mathbf{z})$ to be simultaneously positive semi-definite with $\mathbf{z} \neq \mathbf{0}$, it is necessary that $z_{\mathbf{0}} > 0$).

\mathcal{P}_d is convex. Indeed, let $\mathbf{p}_1, \mathbf{p}_2 \in \mathcal{P}_d$, $t \in [0, 1]$, $\tilde{\mathbf{p}} := t \mathbf{p}_1 + (1-t) \mathbf{p}_2$. Then, by linearity of the localizing matrix operator,

$$\forall g \in \mathbb{R}_d[\mathbf{x}] \quad M_{d-d_g}(g \tilde{\mathbf{p}}) = t M_{d-d_g}(g \mathbf{p}_1) + (1-t) M_{d-d_g}(g \mathbf{p}_2),$$

where $d_g = \lceil \deg g/2 \rceil$, so that its semidefinite positivity is preserved by convex combination, by convexity of \mathbb{S}_+^n for any $n \in \mathbb{N}$. Thus, $\tilde{\mathbf{p}} \in \mathcal{X}_d$. Besides,

$$\tilde{p}_{\mathbf{0}} = t p_{1\mathbf{0}} + (1-t) p_{2\mathbf{0}} = t + 1 - t = 1$$

so that $\tilde{\mathbf{p}} \in \mathcal{P}_d$, which proves convexity.

Eventually, we move on to showing compactness of \mathcal{P}_d . According to Lemma 5, the ball constraint in the description of \mathbf{K} (Assumption 1) and the upper bound $z_{\mathbf{0}} \leq C$ (Assumption 2) yield boundedness of \mathcal{P}_d for the distance $\text{dist}(\mathbf{z}, \mathbf{z}') := \max_{|\mathbf{k}| \leq 2d} |z_{\mathbf{k}} - z'_{\mathbf{k}}|$.

Finally, \mathcal{P}_d is closed as the intersection between the level-1 set of the continuous function $\mathbf{z} \mapsto z_{\mathbf{0}}$ and the closed cone \mathcal{X}_d . Indeed \mathcal{X}_d is closed as the pre-image of the closed cone $\left(\mathbb{S}_+^{N_n, d}\right)^m$ by the (continuous) linear map

$$\mathbf{z} \mapsto \left(\mathbb{M}_{d-d_j}(g_j \mathbf{z})\right)_{j \in \{1, \dots, m\}}.$$

Since finite dimensional closed bounded sets are compact, this proves that \mathcal{P}_d is compact.

It remains to prove (*). Let $\mathbf{z} \in \mathcal{X}_d$ s.t. $\mathbb{A}_d \mathbf{z} = \mathbf{0}$ and $\mathbb{L}_{\mathbf{z}}(c) = 0$. We want to prove that $\mathbf{z} = \mathbf{0}$ so that (*) holds.

Let $\mathbf{z}_0 \in \mathcal{X}_d$ s.t. $\forall \beta \in \mathbb{B}_d, \mathbb{L}_{\mathbf{z}_0}(\psi_\beta) \leq b_\beta$. Define for $t \geq 0$ $\mathbf{z}_t := \mathbf{z}_0 + t \mathbf{z}$. Let $t \geq 0$.

Since \mathcal{X}_d is a convex cone, $\mathbf{z}_t \in \mathcal{X}_d$. In addition, by construction of \mathbf{z} and linearity of the operator $\mathbf{z} \mapsto \mathbb{L}_{\mathbf{z}}(\cdot), \forall \beta \in \mathbb{B}_d, \mathbb{L}_{\mathbf{z}_t}(\psi_\beta) = \mathbb{L}_{\mathbf{z}_0}(\psi_\beta) \leq b_\beta$, so that Assumption 2 ensures that $z_{t, \mathbf{0}} \leq C$. However, $z_{t, \mathbf{0}} = z_{0, \mathbf{0}} + t z_{\mathbf{0}}$.

Combined with the fact that $z_{0, \mathbf{0}} \geq 0$ (with Lemma 5), this yields that for all $t \geq 0$,

$$t z_{\mathbf{0}} \leq C,$$

which is only possible if $z_{\mathbf{0}} = 0$, *i.e.* if $\mathbf{z} = \mathbf{0}$ using again Lemma 5. \square

Eventually, we can end this section by proving Theorem 4. As announced at the beginning of this section, w.l.o.g we actually only do the proof for the case $N = 1$:

Theorem 7 (Convergence of the pseudo-moment sequences).

Under Assumptions 1, 2 and 3, there exists a sequence $(\mathbf{z}_d)_{d \geq d_0}$ of feasible pseudo-moment sequences for the hierarchy (6) corresponding to Problem (1), s.t. $\mathbb{L}_{\mathbf{z}_d}(c) = \mathbb{p}_{\text{GM}}^d$. Then, for all $\mathbf{k} \in \mathbb{N}^n$,

$$z_{d, \mathbf{k}} \xrightarrow{d \rightarrow \infty} \int \mathbf{x}^{\mathbf{k}} d\mu^*(\mathbf{x}).$$

In particular, one has $\mathbb{p}_{\text{GM}}^d \xrightarrow{d \rightarrow \infty} \mathbb{p}_{\text{GM}}^$.*

Moreover, this automatically yields strong duality $\mathbb{p}_{\text{GM}}^d = \mathbb{d}_{\text{GM}}^d$ and $\mathbb{p}_{\text{GM}}^ = \mathbb{d}_{\text{GM}}^*$ and then dual convergence $\mathbb{d}_{\text{GM}}^d \xrightarrow{d \rightarrow \infty} \mathbb{d}_{\text{GM}}^*$.*

Proof. Existence of $(\mathbf{z}_d)_{d \geq d_0}$ follows from Proposition 6 (using Assumptions 1 and 2), so we focus on the proof of convergence. Let $d \in \mathbb{N}$. For $\mathbf{k} \in \mathbb{N}^n$, define

$$\hat{z}_{d, \mathbf{k}} := \begin{cases} z_{d, \mathbf{k}} & \text{if } |\mathbf{k}| \leq 2d, \\ 0 & \text{else,} \end{cases}$$

so that $\hat{\mathbf{z}}_d \in \mathbb{R}^{\mathbb{N}^n}$ with

$$\|\hat{\mathbf{z}}_d\|_{\ell^\infty(\mathbb{N}^n)} := \max_{\mathbf{k} \in \mathbb{N}^n} |\hat{z}_{d, \mathbf{k}}| = \max_{|\mathbf{k}| \leq 2d} |z_{d, \mathbf{k}}| \stackrel{\text{Lemma 5}}{\leq} z_{d, \mathbf{0}} \leq C,$$

using again Assumptions 1 and 2 (recall that in particular Assumption 1 sets $R = 1$ in the application of Lemma 5). Then, $(\hat{\mathbf{z}}_d)_{d \in \mathbb{N}}$ is a uniformly bounded sequence of

$$\ell^\infty(\mathbb{N}^n) := \left\{ \mathbf{u} \in \mathbb{R}^{\mathbb{N}^n} : \max_{\mathbf{k} \in \mathbb{N}^n} |u_{\mathbf{k}}| < \infty \right\},$$

which is the topological dual of $\ell^1(\mathbb{N}^n) := \{\mathbf{u} \in \mathbb{R}^{\mathbb{N}^n} : \sum_{\mathbf{k} \in \mathbb{N}^n} |u_{\mathbf{k}}| < \infty\}$. Thus, the Banach-Alaoglu theorem [2, Theorem 3.16] yields a weak-* converging subsequence $(\hat{z}_{d_r})_{r \in \mathbb{N}} : \exists \mathbf{z}_\infty \in \ell^\infty(\mathbb{N}^n)$ s.t. $\forall \mathbf{u} \in \ell^1(\mathbb{N}^n)$,

$$\sum_{\mathbf{k} \in \mathbb{N}^n} u_{\mathbf{k}} z_{d_r, \mathbf{k}} \xrightarrow{r \rightarrow \infty} \sum_{\mathbf{k} \in \mathbb{N}^n} u_{\mathbf{k}} z_{\infty, \mathbf{k}}.$$

In particular, if $\mathbf{k} \in \mathbb{N}^n$, $u_{\mathbf{l}} := 0$ if $\mathbf{l} \neq \mathbf{k}$, $u_{\mathbf{k}} = 1$ defines a $\mathbf{u} \in \ell^1(\mathbb{N}^n)$ for which the above statement becomes $z_{d_r, \mathbf{k}} \xrightarrow{r \rightarrow \infty} z_{\infty, \mathbf{k}}$.

Thus, what we want to show is that for $\mathbf{k} \in \mathbb{N}^n$, $z_{\infty, \mathbf{k}} = \int \mathbf{x}^{\mathbf{k}} d\mu^*(\mathbf{x}) =: z_{\mathbf{k}}^*$, where we recall that μ^* is the unique solution to (1) given by Assumption 3.

Let $\alpha \in \mathbb{A}$, $j \in \mathbb{N}_m^*$. Then, for all $d \geq d_0$ and $r \in \mathbb{N}$ big enough, by feasibility of \mathbf{z}_{d_r} for the relaxation of degree d_r , one has

$$\bullet \quad 0 \preceq M_{d-d_j}(g_j \mathbf{z}_{d_r}) = M_{d-d_j}(g_j \hat{\mathbf{z}}_{d_r}) \xrightarrow{r \rightarrow \infty} M_{d-d_j}(g_j \mathbf{z}_\infty)$$

$$\bullet \quad a_\alpha = L_{\mathbf{z}_{d_r}}(\varphi_\alpha) = L_{\hat{\mathbf{z}}_{d_r}}(\varphi_\alpha) \xrightarrow{r \rightarrow \infty} L_{\mathbf{z}_\infty}(\varphi_\alpha)$$

so that according to Putinar's Lemma [16, Lemma 3.2], whose conditions hold by Assumption 1, \mathbf{z}_∞ is the actual moment sequence of a measure μ_∞ that is feasible for problem (1). Then, one directly has

$$p_{\text{GM}}^* \geq \int c d\mu_\infty = L_{z_\infty}(c) = \lim_{r \rightarrow \infty} L_{\hat{z}_{d_r}}(c) = \lim_{r \rightarrow \infty} p_{\text{GM}}^{d_r} \geq p_{\text{GM}}^*$$

since for any $d \geq d_0$ $p_{\text{GM}}^d \geq p_{\text{GM}}^*$ (by construction of the moment *relaxation*).

Hence, $\int c d\mu_\infty = p_{\text{GM}}^*$, *i.e.* μ_∞ is *optimal* for problem (1). By Assumption 3, this yields $\mu_\infty = \mu^*$, *i.e.* $\mathbf{z}_\infty = \mathbf{z}^*$. Thus, $(\hat{\mathbf{z}}_d)_d$ is bounded and has a unique weak-* accumulation point \mathbf{z}^* , which means that for any $\mathbf{k} \in \mathbb{N}^n$,

$$z_{d, \mathbf{k}} \xrightarrow{d \rightarrow \infty} z_{\mathbf{k}}^* = \int \mathbf{x}^{\mathbf{k}} d\mu^*(\mathbf{x}).$$

Eventually, since Assumptions 1 and 2 hold, Proposition 6 ensures strong duality $p_{\text{GM}}^d = d_{\text{GM}}^d$, so that putting together weak GMP duality and the strenghtening property, one has

$$p_{\text{GM}}^* \stackrel{\text{weak duality}}{\leq} d_{\text{GM}}^* \stackrel{\text{strengthening}}{\leq} d_{\text{GM}}^d \stackrel{\text{strong duality}}{=} p_{\text{GM}}^d \stackrel{\text{convergence}}{\xrightarrow{d \rightarrow \infty}} p_{\text{GM}}^*$$

and the sandwich rule yields strong GMP duality $p_{\text{GM}}^* = d_{\text{GM}}^*$. □

4 Discussing the uniqueness Assumption 3

In practice, it is quite easy to enforce Assumptions 1 and 2, up to rescaling and addition of mass constraints. However, Assumption 3 may not always hold (especially in the case of polynomial optimization or PDE solutions), so that it is worth discussing what can be done without this Assumption. If Assumption 3 does not hold, then our contribution reduces to the following result:

Corollary 8 (to Theorem 7).

Under Assumptions 1 and 2, there exists a subsequence $(\mathbf{z}_{d_r})_{r \in \mathbb{N}}$, where $(d_r)_{r \in \mathbb{N}} \in ([d_0, \infty) \cap \mathbb{N})^{\mathbb{N}}$ is strictly increasing, of feasible pseudo-moment sequences for the hierarchy (6), as well as

a measure $\mu_\infty \in \mathcal{M}(\mathbf{K})_+$ feasible for Problem (1), s.t. $L_{\mathbf{z}_{d_r}}(c) = p_{\text{GM}}^{d_r}$, $\int c d\mu_\infty = p_{\text{GM}}^*$ and for all $\mathbf{k} \in \mathbb{N}^n$,

$$z_{d_r, \mathbf{k}} \xrightarrow{r \rightarrow \infty} \int \mathbf{x}^{\mathbf{k}} d\mu_\infty(\mathbf{x}).$$

In particular, one still has $p_{\text{GM}}^d \xrightarrow{d \rightarrow \infty} p_{\text{GM}}^*$.

Moreover, this automatically yields strong duality $p_{\text{GM}}^d = d_{\text{GM}}^d$ and $p_{\text{GM}}^* = d_{\text{GM}}^*$ and then dual convergence $d_{\text{GM}}^d \xrightarrow{d \rightarrow \infty} d_{\text{GM}}^*$.

Proof. Going back to the proof of Theorem 7, most arguments are still valid, which lets us with a subsequence $(\mathbf{z}_{d_r})_{r \in \mathbb{N}}$ that converges towards the moments of an *optimal* μ_∞ . The only part that is not valid anymore without Assumption 3 is the identification between μ_∞ and μ^* , and thus the convergence of the whole sequence $(\mathbf{z}_d)_{d \geq d_0}$. Indeed, in such context, the pseudo-moment sequence might have several accumulation points μ_1^*, \dots, μ_M^* (or even an infinity of accumulation points), and either converge to one of them, or oscillate between them. The strong duality and value convergence results being independent from this, the rest of Theorem 7 is unchanged. \square

To illustrate this result, consider the POP problem $f^* := \inf_{x \in [0,1]} (1-x)x$, whose value is obviously $f^* = 0$ with minimizers $x^* = 0$ and $x^* = 1$. We then write the corresponding GMP instance as well as its order d moment relaxation:

$$\begin{aligned} p_f^* &:= \inf \int (1-x)x d\mu(x) & (7a) & & p_f^d &:= \inf z_1 - z_2 & (7b) \\ \text{s.t. } & \mu \in \mathcal{M}([0,1])_+ & & & \text{s.t. } & \mathbf{z} = (z_k)_{0 \leq k \leq 2d} \in \mathbb{R}^{\mathbb{N}_{2d}} \\ & \int 1 d\mu(x) = 1, & & & & M_d(\mathbf{z}) \succeq 0 \\ & & & & & M_{d-1}(x(1-x)\mathbf{z}) \succeq 0 \\ & & & & & M_{d-1}((1-x^2)\mathbf{z}) \succeq 0 \\ & & & & & z_0 = 1, \end{aligned}$$

where $[0,1]$ is described as $[0,1] = \{x \in \mathbb{R} : 1 \geq 0, (1-x)x \geq 0 \text{ and } 1-x^2 \geq 0\}$ so as for Assumption 1 to hold. Note that constraint $\int 1 d\mu = 1$ automatically enforces Assumption 2 through its moment writing $z_0 = 1$. Now, the obvious solution to problem (7a) is $p_f^* = 0$ with minimizing set

$$\mathbf{M}^* := \{\mu^* = t\delta_1 + (1-t)\delta_0 : t \in [0,1]\}.$$

Indeed, \mathbf{M}^* parameterizes all possible probability measures supported on the set $\{0,1\}$ of minimizers of $(1-x)x$ (see figure 1). Thus, in this case, Assumption 3 does not hold.

We then proceed to study the moment relaxation (7b).

Proposition 9 (Finite convergence).

For $d \geq 1$, the minimizers of problem (7b) are in correspondance with the minimizers of problem (7a):

$$\mathbf{M}^d = \left\{ \left(\int x^k d\mu^*(x) \right)_{k \in \mathbb{N}_{2d}} : \mu^* \in \mathbf{M}^* \right\}.$$

This is called a finite convergence phenomenon, and it is very common with POP moment hierarchies [14].

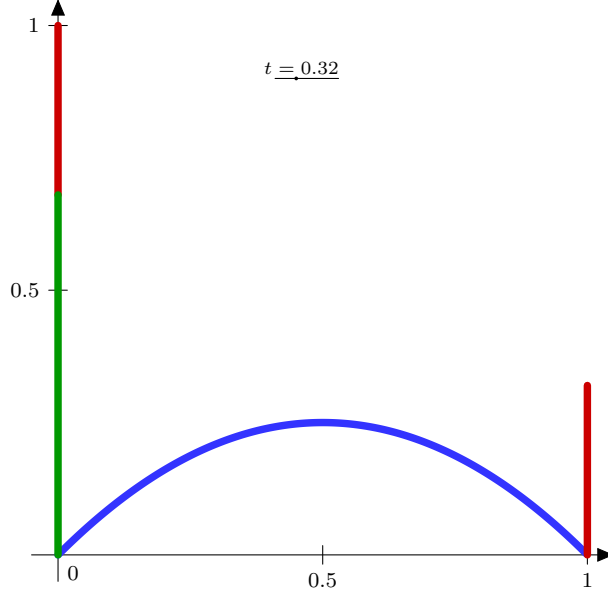


Figure 1: Graph of the function $x \mapsto (1-x)x$ that we minimize.

We represent a possible optimal measure $\mu^* = t \delta_1 + (1-t) \delta_0$ where $t = 0.32$ is the length of the red segment and $(1-t) = 0.68$ is the length of the green one.

Proof. Let $d \geq 1$. Let $\mathbf{z} = (z_k)_{0 \leq k \leq 2d} \in \mathbb{R}^{\mathbb{N}_{2d}}$ be feasible for (7b). For $k \in \mathbb{N}_{2d-2}$, we define

$$u_k := z_{k+1} - z_{k+2},$$

which yields a new vector $\mathbf{u} = (u_k)_{0 \leq k \leq 2d-2} \in \mathbb{R}^{2d-1}$.

Then, $M_{d-1}(\mathbf{u}) = M_{d-1}(x(1-x)\mathbf{z}) \succeq 0$. We next prove that $M_{d-2}((1-x^2)\mathbf{u}) \succeq 0$. Let $p \in \mathbb{R}_{d-2}[\mathbf{x}]$.

$$\begin{aligned} L_{\mathbf{u}}((1-x^2)p^2) &= L_{\mathbf{z}}((1-x^2)p^2(1-x)x) \\ &= L_{\mathbf{z}}((1+x)(1-x)^2p^2x) \\ &= L_{\mathbf{z}}((1-x+2x)(1-x)^2p^2x) \\ &= \underbrace{L_{\mathbf{z}}(x(1-x)(1-x)^2p^2)}_{\geq 0 \text{ since } M_{d-1}(x(1-x)\mathbf{z}) \succeq 0} + 2 \underbrace{L_{\mathbf{z}}(x^2(1-x)^2p^2)}_{\geq 0 \text{ since } M_d(\mathbf{z}) \succeq 0} \geq 0, \end{aligned}$$

which is the definition of $M_{d-2}((1-x^2)\mathbf{u}) \succeq 0$. Then, applying Lemma 5 to \mathbf{u} , we deduce that

$$0 \leq \max_{1 \leq k \leq 2d-1} |z_k - z_{k+1}| = \max_{0 \leq k \leq 2d-2} |u_k| \leq u_0 = z_1 - z_2. \quad (8)$$

This proves that $p_f^d \geq 0$. On the other hand, $\mathbf{z} = (1, t, \dots, t) \in \mathbb{R}^{\mathbb{N}_{2d}}$, with $t \in [0, 1]$, defines a feasible \mathbf{z} (corresponding to the moments of optimal measure $\mu^* = t \delta_1 + (1-t) \delta_0$) such that $z_1 - z_2 = 0$, which proves that $p_f^d \leq 0$ and thus $p_f^d = 0$.

Eventually, let $\mathbf{z} = (z_k)_{0 \leq k \leq 2d}$ be feasible and optimal for (7b): $z_1 - z_2 = 0$. Then, equation (8) instantly yields that for all $k \geq 2$, $z_k = z_1$. Moreover, according to Lemma 5 applied to \mathbf{z} ,

for all $k \geq 1$, $z_k = z_2 \in [0, 1]$. In other words, for $d \geq 1$ the set of minimizers for problem (7b) is exactly the set of truncated moment sequences of minimizers for problem (7a):

$$\mathbf{M}^d = \{(z_k)_{0 \leq k \leq 2d} : z_0 = 1, z_1 \in [0, 1], \forall k \geq 1, z_k = z_1\} = \left\{ \left(\int x^k d\mu^*(x) \right)_{k \in \mathbb{N}_{2d}} : \mu^* \in \mathbf{M}^* \right\},$$

no matter the size of $d \geq 1$. \square

Remark 12 (Notions of convergence).

Here it is important to distinguish between two different notions of convergence:

- *Finite convergence* of the hierarchy denotes the aforementioned phenomenon, when the minimizers of a finite degree relaxation all correspond to minimizers of the original GMP;
- One talks about *pseudo-moment sequence convergence* when any sequence $(\mathbf{z}_d)_d$ of minimizers for (6) converges to the sequence of moments of a minimizer for (1).

Theorem 4 ensures pseudo-moment sequence convergence under Assumptions 1–3. Corollary 8 states that removing Assumption 3 leads to having pseudo-moment sequence convergence *only up to a subsequence* $(\mathbf{z}_{d_r})_r$.

In our illustrative example one indeed has finite convergence, but there is no *a priori* reason that each step of the hierarchy will return an optimal pseudo-moment sequence corresponding to the same optimal measure, and one could have for example that for $d \geq 1$,

$$\mathbf{z}_d = (1, d \bmod 2, \dots, d \bmod 2) \in \mathbb{R}^{\mathbb{N}_{2d}},$$

where $d \bmod 2 = \begin{cases} 1 & \text{if } d \text{ is odd} \\ 0 & \text{if } d \text{ is even} \end{cases}$ so that \mathbf{z}_d would be the truncated moment sequence corresponding to optimal solution $\mu_d^* = (d \bmod 2) \delta_1 + (1 - d \bmod 2) \delta_0 = \delta_{d \bmod 2}$.

Such a $(\mathbf{z}_d)_d$ would obviously not converge when the degree d tends to infinity.

A natural question that arises then is: how does one modify the original GMP or the moment relaxations in order to retrieve pseudo-moment sequence convergence?

A trivial way to fix this issue is to fix the value of z_1 in the constraints of the GMP:

$$\begin{aligned} \mathbf{p}_t^* &:= \inf \int (1-x)x d\mu(x) & (9a) & \quad \mathbf{p}_t^d := \inf z_1 - z_2 & (9b) \\ \text{s.t. } & \mu \in \mathcal{M}([0, 1])_+ & & \text{s.t. } & \mathbf{z} = (z_k)_{0 \leq k \leq 2d} \in \mathbb{R}^{\mathbb{N}_{2d}} \\ & \int 1 d\mu(x) = 1 & & & \mathbf{M}_d(\mathbf{z}) \succeq 0 \\ & \int x d\mu(x) = t, & & & \mathbf{M}_{d-1}(x(1-x)\mathbf{z}) \succeq 0 \\ & & & & \mathbf{M}_{d-1}((1-x^2)\mathbf{z}) \succeq 0 \\ & & & & z_0 = 1 \\ & & & & z_1 = t, \end{aligned}$$

for a given $t \in [0, 1]$. Then, the only solution to (9a) is

$$\mu^* = t \delta_1 + (1-t) \delta_0,$$

so that Assumption 3 holds and one retrieves pseudo-moment sequence convergence (and finite convergence still holds), as the unique minimizer for (9b) is

$$\mathbf{z}_d = (1, t, \dots, t) = \left(\int x^k d\mu^*(x) \right)_{0 \leq k \leq 2d} \in \mathbb{R}^{\mathbb{N}_{2d}}.$$

Remark 13 (Enforcing uniqueness).

It is not always possible to enforce Assumption 3 as we did here without knowing the GMP solution in advance. Another common heuristics to obtain a unique solution in the relaxations consists in adding a penalty in the cost function: instead of minimizing $z_1 - z_2$ in (7b), one could minimize $z_1 - z_2 + \epsilon \operatorname{Tr} M_d(\mathbf{z})$, for some $\epsilon \in \mathbb{R} \setminus \{0\}$. Then, depending on the sign of ϵ , one would obtain a unique solution $\mathbf{z}_d = (1, \dots, 1)$ ($\epsilon < 0$) corresponding to $\mu^* = \delta_1$, or $\mathbf{z}_d = (1, 0, \dots, 0)$ ($\epsilon > 0$) corresponding to $\mu^* = \delta_0$.

We finally get to numerically implement the moment hierarchy corresponding to our illustrating example. In the “fixed” cases, the behavior of the SDP solver is fixed, up to numerical errors. The interesting case is then problem (7). We implemented the corresponding hierarchy of moment problems (7b) with increasing relaxation degree $d \in \mathbb{N}_{20}^*$. First, not displayed is the fact that we indeed invariably obtained $z_0 = 1.0000$ and $z_k = z_1$ (up to solver precision) for all $k \geq 2$. We implemented the hierarchy in two different frameworks:

- GloptiPoly [4], a Matlab toolbox that models moment problems, interfaced with the SDP solvers SeDuMi (see the numerically computed values of z_1 in Table 1a) and Mosek (see Table 1b);
- MomentOpt [17], an open source Julia module, interfaced with the SDP solver CSDP (see the numerically computed values of z_1 in Table 1c).

d	1	2	3	4	5	6	7	8	9	10
z_1	0.3974	0.3094	0.2228	0.2020	0.2103	0.1960	0.1440	0.1400	0.1369	0.0953
d	11	12	13	14	15	16	17	18	19	20
z_1	0.1088	0.1116	0.0982	0.0937	0.0851	0.0762	0.0716	0.0674	0.0649	0.0620

(a) with GloptiPoly and SeDuMi.

d	1	2	3	4	5	6	7	8	9	10
z_1	0.3259	0.3034	0.1782	0.1312	0.1092	0.1236	0.0976	0.0898	0.0861	0.0830
d	11	12	13	14	15	16	17	18	19	20
z_1	0.0763	0.0744	0.0666	0.0611	0.0564	0.0528	0.0499	0.0473	0.0450	0.0428

(b) with GloptiPoly and Mosek.

d	1	2	3	4	5	6	7	8	9	10
z_1	0.3741	0.0553	0.0772	0.0340	0.0418	0.0280	0.0169	0.0148	0.0121	0.0137
d	11	12	13	14	15	16	17	18	19	20
z_1	0.0079	0.0107	0.0082	0.0078	0.0071	0.0074	0.0075	0.0071	0.0059	0.0057

(c) with MomentOpt and CSDP.

Table 1: Numerical solutions of problem (7b)

At first glance it would seem that the sequence $(z_{d,1})_d$ decreases when d increases, but looking at specific values shows that it is not the case. What we can deduce from Table 1 is that the sequence $(z_{d,1})_d$ seems to converge to 0, so that $(\mathbf{z}_d)_d$ would converge towards $\mu_\infty = \delta_0$, but we have no proof that it is actually the case. Moreover, the sequence $\mathbf{z}_d = (1, 0, \dots, 0)$ is already feasible for the degree $d = 1$ relaxation, so that there is no reason, if the tendency were indeed to go to $z_1 = 0$, that such constraint is not satisfied already in the first steps of the hierarchy. Also,

it is important to notice that the results we obtained are stable, meaning that running the same code always returns the same values for z_1 . In other words, the obtained values *only depend on the solver*. These tendencies then depend on how the interior point algorithm of the solver is implemented, which is out of the scope of the present work.

5 Conclusion

While usually it is only proved that the optimal values for the moment relaxations of the GMP monotonically converge towards the optimal value of the GMP, this note proves that the solutions of these moment relaxations converge to the actual solution of the GMP in ℓ^∞ 's weak-* topology, provided that some elementary assumptions hold. So far, such proof has only been provided for very specific instances of the GMP (see e.g. [10, Theorem 5.6(b)], [5]), while here the generic case is completely dealt with. Up to rescaling the problem (which also precludes ill behaviours in the numerical implementations), one can usually enforce quite easily that these conditions are met. As a byproduct of this proof, one also obtains strong duality both in the GMP and its corresponding hierarchy, generalizing the results of [6]. Among the open questions this note answers, one can cite the weak-* convergence of the moment sequences associated to the OCP [11] and set approximation problems in [3, 15, 12], as well as the strong duality between the GMP formulation in [13] and its dual, regardless of the formulation of such dual. Finally, this note should facilitate most proofs of convergence and strong duality related to future applications of Lasserre's moment-SOS hierarchy.

References

- [1] Alexander Barvinok. A Course in Convexity. Number 54. American Mathematical Society, Providence, Rhode Island, 2002.
- [2] Haim Brezis. Functional Analysis, Sobolev Spaces and Partial Differential Equations. 2011.
- [3] Didier Henrion and Milan Korda. Convex computation of the region of attraction of polynomial control systems. IEEE Transactions on Automatic Control, 59(2):297–312, 2014.
- [4] Didier Henrion, Jean B. Lasserre, and Johan Loefberg. GloptiPoly 3: moments, optimization and semidefinite programming. Optimization Methods and Software, 24(4-5):761–779, 2009.
- [5] Didier Henrion, Jean Bernard Lasserre, and Carlo Savorgnan. Approximate volume and integration for basic semialgebraic sets. Siam Review, 51:722–743, 2009.
- [6] Cédric Jozs and Didier Henrion. Strong duality in lasserre's hierarchy for polynomial optimization. Optimization Letters, 10:3–10, 2016.
- [7] Milan Korda, Didier Henrion, and Colin Jones. Convex computation of the maximum controlled invariant set for polynomial control systems. SIAM Journal on Control and Optimization, 2014.
- [8] Milan Korda, Didier Henrion, and Colin N. Jones. Inner approximations of the region of attraction for polynomial dynamical systems. In IFAC Symposium on Nonlinear Control Systems (NOLCOS), Toulouse, France, 2013.
- [9] Jean Bernard Lasserre. Global optimization with polynomials and the problem of moments. SIAM Journal on Optimization, 11:796–817, 2001.

- [10] Jean Bernard Lasserre. Moments, Positive Polynomials and Their Applications. 2010.
- [11] Jean-Bernard Lasserre, Didier Henrion, Christophe Prieur, and Emmanuel Trélat. Nonlinear optimal control via occupation measures and lmi-relaxations. SIAM Journal on Control and Optimization, 47(4):1643–1666, 2008.
- [12] Victor Magron, Pierre-Loïc Garoche, Didier Henrion, and Xavier Thirioux. Semidefinite approximations of reachable sets for discrete-time polynomial systems. SIAM Journal on Control and Optimization, 2019.
- [13] Swann Marx, Tillmann Weisser, Didier Henrion, and Jean Bernard Lasserre. A moment approach for entropy solutions to nonlinear hyperbolic partial differential equations. Mathematical Control and Related Fields, 10:113–140, 2020.
- [14] Jiawang Nie. Optimality conditions and finite convergence of Lasserre’s hierarchy. Mathematical Programming, 146:97–121, 2014.
- [15] Antoine Oustry, Matteo Tacchi, and Didier Henrion. Inner approximations of the maximal positively invariant set for polynomial dynamical systems. IEEE Control System Letters, 2019.
- [16] Mihai Putinar. Positive polynomials on compact semi-algebraic sets. Indiana University Mathematics Journal, 42, 1993.
- [17] Tillmann Weisser, Benoît Legat, Chris Coey, Lea Kapelevich, and Juan Pablo Vielma. Polynomial and moment optimization in Julia and JuMP. In JuliaCon, 2019.