Computing approximations and generalized solutions using moments and positive polynomials

Tillmann Weisser

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ FÉDÉRALE TOULOUSE MIDI-PYRÉNÉES

Délivré par :
l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

Présentée et soutenue le *03/10/2018* **par :**

Tillmann WEISSER

Computing Approximations and Generalized Solutions Using Moments and Positive Polynomials

---

JURY

Didier Henrion	DR LAAS-CNRS, Toulouse	Directeur de thèse
Salma Kuhlmann	PU Universität Konstanz	Rapporteur
Jean B Lasserre	DR LAAS-CNRS, Toulouse	Directeur de thèse
Jérôme Malick	DR LJK-CNRS, Grenoble	Examinateur
Mohab Safey el Din	PU Sorbonne Université, Paris	Examinateur
Emmanuel Trélat	PU Sorbonne Université, Paris	Rapporteur
Hasnaa Zidani	DR ENSTA ParisTech	Examinateur

École doctorale et spécialité :
MITT : Domaine Mathématiques : Mathématiques appliquées
Unité de Recherche :
Laboratoire d'analyse et d'architecture des systèmes
Directeur(s) de Thèse :
Jean Bernard LASSERRE et *Didier HENRION*
Rapporteurs :
Salma KUHLMANN et *Emmanuel TRELAT*

Résumé: Le problème généralisé des moments (PGM) est un problème d'optimisation linéaire sur des espaces de mesures. Il permet de modéliser simplement un grand nombre d'applications. En toute généralité il est impossible à résoudre mais si ses données sont des polynômes et des ensembles semi-algébriques alors on peut définir une hiérarchie de relaxations semidéfinies (SDP) – la hiérarchie *moments-sommes-de-carrés* (moments-SOS) – qui permet en principe d'approcher la valeur optimale avec une précision arbitraire. Le travail contenu dans cette thèse adresse deux facettes concernants le PGM et la hiérarchie moments-SOS:

Une première facette concerne l'évolution des relaxations SDP pour le PGM. Le degré des poids SOS dans la hiérarchie moments-SOS augmente avec l'ordre de relaxation. Lorsque le nombre de variables n'est pas modeste, on obtient rapidement des programmes SDP de taille trop grande pour les logiciels de programmation SDP actuels, sauf si l'on peut utiliser des symétries ou une parcimonie structurée souvent présente dans beaucoup d'applications de grande taille. On présente donc un nouveau certificat de positivité sur un compact semi-algébrique qui (i) exploite la parcimonie présente dans sa description, et (ii) dont les polynômes SOS ont un degré borné à l'avance. Grâce à ce nouveau certificat on peut définir une nouvelle hiérarchie de relaxations SDP pour le PGM qui exploite la parcimonie et évite l'explosion de la taille des matrices semidéfinies positives liée au degré des poids SOS dans la hiérarchie standard.

Une deuxième facette concerne (i) la modélisation de nouvelles applications comme une instance particulière du PGM, et (ii) l'application de la méthodologie *moments-SOS* pour leur résolution.

En particulier on propose des approximations déterministes de *contraintes probabilistes*, un problème difficile car le domaine des solutions admissibles associées est souvent non-convexe et même parfois non connecté. Dans notre approche moments-SOS le domaine admissible est remplacé par un ensemble plus petit qui est le sous-niveau d'un polynôme dont le vecteur des coefficients est une solution optimale d'un certain SDP. La qualité de l'approximation (interne) croît avec le degré du polynôme et la taille du SDP. On illustre cette approche dans le problème du calcul du flux de puissance optimal dans les réseaux d'énergie, une application stratégique où la prise en compte des contraintes probabilistes devient de plus en plus cruciale (e.g., pour modéliser l'incertitude liée á l'énergie éolienne et solaire). En outre on propose une extension des cette procedure qui est robuste à l'incertitude sur la distribution sous-jacente. Des garanties de convergence sont fournies.

Une deuxième contribution concerne l'application de la méthodologie moments-SOS pour l'approximation de *solutions généralisés en commande optimale*. Elle permet de capturer le comportement limite d'une suite minimisante de commandes et de la suite de trajectoires associée. On peut traiter ainsi le cas de phénomènes simultanés de concentrations de la commande et de discontinuités de la trajectoire.

Une troisième contribution concerne le calcul de *solutions mesures pour les lois de conservation hyperboliques scalaires* dont l'exemple typique est l'équation de Burgers. Cette classe d'EDP non linéaire peut avoir des solutions discontinues difficiles à approximer numériquement avec précision. Sous certaines hypothèses, la solution mesurepeut être identifiée avec la solution classique (faible) à la loi de conservation. Notre approche moment-SOS fournit alors une méthode alternative pour approcher des solutions qui contrairement aux méthodes existantes évite une discrétisation du domaine.

Mots-clés: moments – polynômes positifs – parcimonie – contraintes probabilistes – solutions mesures – relaxations semidéfinies

Abstract: The generalized moment problem (GMP) is a linear optimization problem over spaces of measures. It allows to model many challenging mathematical problems. While in general it is impossible to solve the GMP, in the case where all data are polynomial and semialgebraic sets, one can define a hierarchy of semidefinite relaxations – the *moment-sums-of-squares* (moment-SOS) *hierachy* – which in principle allows to approximate the optimal value of the GMP to arbitrary precision. The work presented in this thesis addresses two facets concerning the GMP and the moment-SOS hierarchy:

One facet is concerned with the scalability of relaxations for the GMP. The degree of the SOS weights in the moment-SOS hierarchy grows when augmenting the relaxation order. When the number of variables is not small, this leads quickly to semidefinite programs (SDPs) that are out of range for state of the art SDP solvers, unless one can use symmetries or some structured sparsity which is typically present in large scale applications. We provide a new certificate of positivity which (i) is able to exploit the structured sparsity and (ii) only involves SOS polynomials of fixed degree. From this, one can define a new hierarchy of SDP relaxations for the GMP which can take into account sparsity and at the same time prevents from explosion of the size of SDP variables related to the increasing degree of the SOS weights in the standard hierarchy.

The second facet focusses on (i) modelling challenging problems as a particular instance of the GMP and (ii) solving these problems by applying the moment-SOS hierarchy.

In particular we propose deterministic approximations of *chance constraints* a difficult problem as the associated set of feasible solutions is typically non-convex and sometimes not even connected. In our approach we replace this set by a (smaller) sub-level-set of a polynomial whose vector of coefficients is a by-product of the moment-SOS hierarchy when modeling the problem as an instance of the GMP. The quality of this inner approximation improves when increasing the degree of the SDP relaxation and asymptotic convergence is guaranteed. The procedure is illustrated by approximating the feasible set of an instance of the chance-constrained AC Optimal Power Flow problem (a nonlinear problem in the management of energy networks) which nowadays becomes more and more important as we rely increasingly on uncertain energy sources such as wind and solar power. Furthermore, we propose an extension of this framework to the case where the underlying distribution itself is uncertain and provide guarantees of convergence.

Another application of the moment-SOS methodology discussed in this thesis consider *measure valued solutions to optimal control problems*. We show how this procedure can capture the limit behavior of an optimizing sequence of control and its corresponding sequence of trajectories. In particular we address the case of concentrations of control and discontinuities of the trajectory may occur simultaneously.

In a final contribution, we compute *measure valued solutions to scalar hyperbolic conservation laws*, such as Burgers equation. It is known that this class of nonlinear partial differential equations has potentially discontinuous solutions which are difficult to approximate numerically with accuracy. Under some conditions the measure valued solution can be identified with the classical (weak) solution to the conservation law. In this case our moment-SOS approach provides an alternative numerical scheme to compute solutions which in contrast to existing methods, does *not* rely on discretization of the domain.

Key words: moments – positive polynomials – sparsity – chance constraints – measure valued solutions – semidefinite relaxations

Acknowledgments

It is a good tradition to use this place to pass the last years in review and thank the people who have accompanied this journey.

First of all I want to express my gratitude to my advisers *Jean Bernard Lasserre* and *Didier Henrion*. I think I could barley have found a better duo. Merci Jean de venir dans notre bureau et nous parler de tes dernières idées, pour ta patience avec moi pendant la redaction de mon manuscrit et bien sûr pour les soirées avec toi au Filochard. Merci Didier d'avoir toujours eu la vision d'ensemble et d'avoir pris le temps pour moi, même pendant des horaires et dans des lieux inhabituels.

I would not have dared to tackle a PhD without the encouragement and help of *Markus Schweighofer*, *Cordian Riener*, *Victor Magron* and *Benjamin Werner*. Markus und Cordian, ihr habt mich vor vier Jahren davon überzeugt, dass ich diesen Weg gehen kann und ohne euch wäre es wohl nie so weit gekommen. Merci Victor et Benjamin, avec mon stage chez vous à LIX vous m'avez préparé la voie pour mon doctorat au LAAS.

During the last three years I had the chance to collaborate with very experienced researchers from all over the world. Thank you *Kim-Chuan Toh*, *Martin Kruzik*, *Line Roald* and *Sidhant Misra*. I have learned a lot from each one of you. I am also proud to have worked together with two young researchers here at LAAS, *Swann Marx* and *Matteo Tacchi*. Je vous remercie de toutes nos discussions (pas seulement) scientifiques.

My daily life at LAAS would not have been the same without all the friendly face of the permanent and non-permanent members of my team MAC. Merci à vous pour tous, même si c'etait un petit sourire le matin ou des grands discours pendant les repas ou des soirées.

For me this PhD was not only a professional project but to a large extend a social and personal one. I found friends, personal challenges and most important strong support in the communities of the *Goethe-Institut* and my two choirs, the *Chœur Franco-Allemand de Toulouse* and the *Conférences Vocales*. Je te remercie en particulier Georgy pour m'avoir appris parler le francais et tes sages conseils. Isabelle et Eloy, vous m'avez accueilli si chaleureusement dans votre familie. Vous m'avez donné le sentiment d'être à la maison en France.

I would also like to thank my friends in Germany. Dank euch bin ich heute der Mensch, der ich bin. Ihr habt mich immer auf meinem Weg begleitet und ich weiß, dass ich mich immer auf euch verlassen kann.

Last but not least my thanks go to my family. Muma und Jogen, ihr habt mich immer unterstützt, egal welches Projekt ich gerade verfolgt habe. Danke!

Contents

Preface

This thesis is about applying and computing approximate solutions to the *generalized moment problem* (GMP). The underlying idea of basically all applications of the GMP is to embed a difficult (non-linear) problem in a bigger space where it can be described structurally simpler – i.e. as a linear problem – at the cost of now considering more feasible solutions in a more complicated space. The latter problem is then approximated by a hierarchy of programs that can be solved by a computer. This way the solution to the originally difficult problem can be approximated in a systematic manner.

As we will review in Chapter 1, the *moments* of a finite Borel measure are in duality with *positive polynomials*. This duality exposes a beautiful link between the mathematical fields of functional analysis and real algebraic geometry and indicates how a single thesis can be concerned with both the algebraic structure in representations of polynomials on the one hand (Chapter 2), and also solving (partial) differential equations, whose solutions are not at all polynomial functions (Chapter 4), on the other hand.

The third aspect of this thesis, or rather the initial one historically, is optimization. The GMP is an infinite dimensional linear optimization problem where the variables are (finite Borel) measures. The constraints and the criterion are linear combinations of the (unknown) moments of these measures. The modelling power of the GMP has already been described in [Lan87] before methods to solve the GMP have been known. For polynomial instances of the GMP, Lasserre in [Las10b] describes an efficient algorithm – the *moment-sums-of-squares-hierarchy* – to approximate solutions of the GMP as closely as desired via a hierarchy of *semidefinte programming (SDP) relaxations* of increasing size.

The increasing size of the SDPs in the hierarchy motivates our investigation in Chapter 2. Unfortunately little is known about the rate of convergence of the hierarchy to the optimal value of the GMP. Surprisingly often and in many applications, already the first relaxation provides a very good approximation or even the optimal value. If however relaxations of higher order are needed, the application of the moment-SOS approach is limited to problems of relatively small dimension as otherwise the size of the problem is too big for state of the art SDP solvers. Therefore one is interested in providing alternative hierarchies with a more favourable computational complexity. The approach we propose takes into account *structured sparsity*, which is typically present in large scale problems and has already been used in [Wak+06]. In addition we allow to prescribe an a priori bound on the size of the SDP variables, without loosing convergence properties of the moment-SOS hierarchy. We hence provide a new hierarchy for large scale problems.

While this first part is of algorithmic nature, the second part of this thesis has a different flavour. In the latter we focus on several applications of the GMP and its approximation via the moment-SOS hierarchy.

Motivated by the increasing interest in uncertainty quantification for optimization, we address the approximation of *chance constraints* in Chapter 3. Indeed many optimization problems are exposed to uncertainty. For example the power generation via wind and sun depends strongly on the weather which can be forecast only with a certain probability. To schedule generators, network operators need to take into account this uncertainty in the optimization process. In a case study of the *optimal power flow* (OPF) problem we

show how an approach based on the GMP can be used to replace the difficult chance constraints by easier polynomial constraints with strong asymptotic guarantees. In a first contribution we assume knowledge of the distribution of the uncertainty. However in many cases only some moments of this distribution are known, and sometimes even this knowledge is itself uncertain. Therefore, in a second contribution, we extend the approach and provide *distributionally robust inner approximations* of the feasible set associated with chance constraints. We only assume that the distribution can be any mixture of distributions in some parametrized family, e.g., Gaussian distributions with mean and variance known to be in some interval, respectively.

Another promising direction is to use the moment-SOS hierarchy to help solve ordinary and partial differential equations. The concept of occupation measures associated to a trajectory gives rise to a more general notion of solutions to a weak formulation of the original problem. This more general notion is called *measure-valued solutions*. We model the weak formulation as an instance of the GMP and approximate measure-valued solutions by the moment-SOS hierarchy. In particular, we investigate this approach in two different set ups: (i) optimal control with critical limit behaviour and (ii) the approximation of solutions to scalar hyperbolic conservation laws. In both applications we are interested in conditions that imply a well defined relation between solutions to the original problem posed on trajectories and their relaxed measure-valued version. Then, we use the moment-SOS hierarchy to approximate moments of the measure-valued solution. When the measure-valued solutions can be identified with the classical solution, this provides a numerical scheme complementary to existing methods. At this current stage however, we do not claim that this approach is mature and could replace sophisticated numerical schemes that have been developed for decades. The primary goal is to show that this method which avoids discretization has indeed some potential.

Organization

This manuscript is divided in four chapters. Chapter 1 introduces basic concepts relevant for all subsequent chapters. The remaining chapters can be read independently of each other.

Chapter 1: We briefly introduce the *generalized moment problem* (GMP) and how it can be approximated by hierarchies based on certificates for non-negativity and in particular by the *moment-SOS hierarchy*. The duality between moments and polynomial SOS is described.

Chapter 2: We describe how *structured sparsity*, typically present in large scale problems, can be exploited in order to provide *certificates of non-negativity* that are less computational demanding than the standard SOS certificates. While the first part of the chapter is based on [WLT17], the second part is unpublished work with a more practical focus on establishing sparsity patterns in problems that are only nearly sparse in their original formulation.

Chapter 3 is dedicated to the approximation of sets that are described by *chance constraints*. After an introduction on chance constraints we present results from [WRM18] where we approximate the feasible set of the chance constrained AC optimal power flow problem. The second contribution in this chapter is based on [LW18] and extends the framework to the approximation to *distributionally robust chance constraints*.

In Chapter 4 we consider *measure-valued solutions* to differential equations and their

approximation using the moment-SOS hierarchy. The first part of this chapter deals with critical limit behaviour of optimizing sequences in optimal control such as *oscillations, concentrations, and discontinuities* and is based on [HKW18]. In the second part we use the concept of measure valued solution to reformulate Cauchy problems to scalar hyperbolic non-linear conservation laws as instances of the GMP and provide approximate solutions by solving the moment-SOS hierarchy. In contrast to other methods in the literature, our approach does not rely on any discretization of time-space. This part of the chapter follows [Mar+18]

Contribution not included

In [Tac+18] we consider the problem of approximating the volume of a basic semialgebraic set of potentially large dimension, provided that its description has some structured sparsity similar to the one defined in Chapter 2. Contrary to a claim in [MHL15] the proposed numerical scheme adapted to sparsity *cannot* be used in the context of volume computation. Our contribution is to provide a different numerical scheme which exploits sparsity in a completely new manner. Remarkably one is able to approximate accurately volumes in large dimension. In particular an appropriate scaling allows to avoid numerical issues when approximating very small volumes which is typically the case for volumes in large dimensions.

Chapter 1

The Generalized Moment Problem and the Moment-SOS Hierarchy

The central object of this thesis is the *Generalized Moment Problem* (GMP) as it is discussed in [Las10b]: If not mentioned otherwise, throughout this work $f, g_1, \dots, g_m \in \mathbb{R}[\mathbf{x}]$ and $(h_\gamma)_{\gamma\in\Gamma} \subseteq \mathbb{R}[\mathbf{x}]$ denote (multivariate) polynomials, where Γ denotes a countable index set. Likewise we consider the *basic semialgebraic set* K defined by

$$K := \{\mathrm{x} \in \mathbb{R}^n : g_1(\mathrm{x}) \geq 0, \dots, g_m(\mathrm{x}) \geq 0\}. \tag{1.1}$$

By $\mathcal{M}(K)$ we denote the space of finite *signed* Borel measures supported on K and with $\mathcal{M}_+(K)$ the convex cone of finite (positive Borel) measures on K. Let $(b_\gamma)_{\gamma\in\Gamma} \subseteq \mathbb{R}$ be a family of real numbers. The GMP is the the following *linear problem* with potentially infinitely many constraints

$$\sup_{\phi} \int_K f \,\mathbf{d}\phi \quad \text{s.t.} \int_K h_\gamma \,\mathbf{d}\phi \leqq b_\gamma, \gamma \in \Gamma, \quad \phi \in \mathcal{M}_+(K). \tag{GMP}$$

In general optimization is with respect to finitely many measures $\phi_1, \dots, \phi_\ell$ supported on basic semialgebraic sets $K_1, \dots, K_\ell$, respectively. For ease of exposition however, we restrict the discussion to only one unknown measure. Everything discussed in this chapter generalizes directly to the case of more than one measure.

The spirit of this thesis is to consider mathematical problems that are typically challenging because of non-linearities in their original formulation, and reformulate them as instances of the GMP. In their GMP formulation these problems become linear and hence conceptually easier. However, we introduce another challenge, i.e., we now need to optimize over an infinite dimensional space of measures. So in general the GMP formulation is just a rephrasing of the original problem. However, because all data f, g_j, h_γ are polynomials and K is a basic semialgebraic set, one may exploit this algebraic feature and provide efficient schemes to approximate the solution numerically. Before we explain this in more detail in Section 1.4 we briefly show how global optimization problems can be rephrased as the simplest instance of the GMP.

In polynomial optimization one is interested in finding the *global optimum* (POP)* of the optimization problem

$$\inf_{\mathrm{x}} f(\mathrm{x}) \quad \text{s.t. } \mathrm{x} \in K, \tag{POP}$$

where K is assumed to be non-empty and defined by the polynomials $g_1, \dots, g_m$ as in (1.1). Note that (POP) is a non-linear problem and hence finding a global optimum – in contrast to local optima – is a difficult problem. In the following we show that (POP) can

be reformulated as the GMP:

$$\inf_{\phi} \int_K f \,\mathbf{d}\phi \quad \text{s.t.} \int_K 1 \,\mathbf{d}\phi = 1, \quad \phi \in \mathcal{M}_+(K). \tag{M-POP}$$

For illustration purpose we show equivalence of (POP) and (M-POP), i.e. that $(\text{POP})^* = (\text{M-POP})^*$. Throughout this thesis we will often need to show the equivalence of a problem and its moment formulation. The general strategy is aligned with the following argumentation. In a first step, we show that (M-POP) is a *relaxation* of (POP), i.e., we show that for every point x feasible for (POP), there is a feasible measure ϕ for (M-POP) which yields the same objective value. Let therefore x be feasible for (POP), i.e., $\mathrm{x} \in K$. Then the Dirac measure δ_{x} is a probability measure supported on K and moreover $\int f \,\mathbf{d}\delta_{\mathrm{x}} = f(\mathrm{x})$, showing that (M-POP) is a relaxation of (POP). The second step now is to show that $(\text{M-POP})^* \geq (\text{POP})^*$. In our case this is almost trivial, as by definition $(\text{POP})^* \leq f$ on K, and hence $\int_K f \,\mathbf{d}\phi \geq \int_K (\text{POP})^* \,\mathbf{d}\phi = f^*$. Consequently the problems (POP) and (M-POP) are equivalent.

Note that by passing from (POP) to (M-POP) we have relaxed the set of Dirac measures δ_{x} with $\mathrm{x} \in K$ to arbitrary probability measures ϕ supported on K. This idea is a very useful strategy to reformulate optimization problems as GMP and will occur several times throughout this manuscript.

In the rest of this chapter we describe a numerical scheme (the moment-SOS hierarchy) providing values that converge to the optimal value of the GMP.

1.1 Measures and Moments

The first important step to approach the GMP numerically is to rephrase the problem on measures as a problem on moments. The *moment sequence* $\mathrm{z} = (\mathrm{z}_\alpha)_{\alpha \in \mathbb{N}^n}$ of a measure $\phi \in \mathcal{M}_+(K)$ is defined by $\mathrm{z}_\alpha := \int_K \mathbf{x}^\alpha \,\mathbf{d}\phi$. Let $\mathbb{N}^n_d := \{\alpha \in \mathbb{N}^n : |\alpha| \leq d\}$, where $|\alpha|$ is the sum over all entries of α. Denote by $s(d) = \binom{n+d}{d}$ the number of elements of $\mathbb{N}^n_d$. A vector $\mathrm{p} \in \mathbb{R}^{s(\deg(p))}$ is called the *coefficient vector* (in the monomial basis) of a polynomial $p \in \mathbb{R}[\mathbf{x}]$ if

$$p = \sum_{\alpha \in \mathbb{N}^n_{\deg(p)}} \mathrm{p}_\alpha \mathbf{x}^\alpha.$$

With this notation integration of a polynomial p with respect to the measure ϕ can be expressed only using the first moments (up to $\deg(p)$) of ϕ:

$$\int_K p \,\mathbf{d}\phi = \int_K \sum_{\alpha \in \mathbb{N}^n_{\deg(p)}} \mathrm{p}_\alpha \mathbf{x}^\alpha \,\mathbf{d}\phi = \sum_{\alpha \in \mathbb{N}^n_{\deg(p)}} \mathrm{p}_\alpha \int_K \mathbf{x}^\alpha \,\mathbf{d}\phi = \sum_{\alpha \in \mathbb{N}^n_{\deg(p)}} \mathrm{p}_\alpha z_\alpha.$$

We use this observation to define a *pseudo-integration* with respect to any sequence $\mathrm{z} \in \mathbb{R}^{\mathbb{N}^n}$. Note briefly that a vector $\mathrm{p} \in \mathbb{R}^{s(d)}$ can be understood as an infinite sequence indexed by $\mathbb{N}^n$ by setting $\mathrm{p}_\alpha := 0$ if $|\alpha| > d$. The pseudo-integration with respect to a sequence $\mathrm{z} \in \mathbb{R}^{\mathbb{N}^n}$ defined by

$$L_{\mathrm{z}}(p) := \sum_{\alpha \in \mathbb{N}^n} \mathrm{p}_\alpha z_\alpha \tag{1.2}$$

is called the *Riesz functional*. With help of this functional it is possible to characterize moment sequences. Denote by $\mathrm{Pos}(K)$ the set of polynomials that are non-negative on K.

Theorem 1.1 (Riesz-Haviland)**.** *Let $K \subseteq \mathbb{R}^n$ be closed. A real sequence* z *is the moment sequence of some measure ϕ supported on K if and only if , $L_{\mathrm{z}}(p) \geq 0$ for all $p \in \mathrm{Pos}(K)$.*

Assuming K closed, we can hence reformulate (GMP) as a linear problem on sequences. Indeed, by Riesz-Haviland's theorem the following problem is an equivalent formulation of (GMP):

$$\sup_{\mathrm{z}} L_{\mathrm{z}}(f) \quad \text{s.t. } L_{\mathrm{z}}(h_\gamma) \leqq b_\gamma, \gamma \in \Gamma, \quad L_{\mathrm{z}} \geq 0 \text{ on } \mathrm{Pos}(K). \tag{1.3}$$

The key idea to approximate GMPs numerically now is to replace the convex cone $\mathrm{Pos}(K)$ of polynomials non-negative on K, by a subcone of *test-polynomials* $\mathrm{Cert}(K)$, which has the desirable property that testing non-negativity of L_{z} on $\mathrm{Cert}(K)$ can be done efficiently on a computer, and membership in $\mathrm{Cert}(K)$ provides a *certificate for non-negativity* on K.

1.2 Sums of Squares

Assume for a moment that $K = \mathbb{R}$ and that the maximal degree of the polynomials considered in (1.3) is $2d$. Notice, that aside from the non-negativity constraint on L_{z}, the numbers z_α with $\alpha > 2d$ are not constrained. We might therefore restrict the non-negativity constraint to hold only for polynomials in $\mathrm{Pos}(\mathbb{R})$ with degree at most $2d$. In the univariate case however, every such polynomial is a *sum of squares* of polynomials each of degree at most d.[1] We denote the set of sums of squares (SOS) of polynomials by

$$\Sigma[\mathbf{x}] := \{\sum_{k=1}^{\ell} p_k^2 : p_k \in \mathbb{R}[\mathbf{x}]\}.$$

Similarly we denote by $\Sigma[\mathbf{x}]_d \subseteq \mathbb{R}[\mathbf{x}]_{2d}$ the SOS of degree at most $2d$. Now, checking whether L_{z} is non-negative on $\Sigma[\mathbf{x}]_d$ (even if $n > 1$) is a semidefinite program (SDP) as we will see later. First note, that for polynomials $p_1, \ldots, p_\ell \in \mathbb{R}[\mathbf{x}]$, non-negativity of $L_{\mathrm{z}}(p_1^2 + \ldots + p_\ell^2)$ by linearity of L_{z} is equivalent to the non-negativity of each $L_{\mathrm{z}}(p_k^2)$ individually. Let $\mathrm{v}_d := (\mathbf{x}^\alpha)_{|\alpha| \leq d}$ be the vector of monomials of degree at most d, indexed by the exponents $\alpha \in \mathbb{N}_d^n$ of the respective monomials. This is the canonical basis of the linear space $\mathbb{R}[\mathbf{x}]_d$ of polynomials of degree at most d, i.e., every polynomial $p \in \mathbb{R}[\mathbf{x}]_d$ can be written as $\mathrm{p}^\top \mathrm{v}_d$ where $\mathrm{p} \in \mathbb{N}^{s(d)}$ is the vector of coefficients of p. In abuse of notation we will use the symbol L_{z} also for the entry wise application of L_{z} to a matrix polynomial

[1] This can be seen easily by applying the fundamental theorem of algebra and remarking the facts that real roots need to be of even multiplicity and that for complex conjugate numbers $x, \bar{x} \in C$, $x = a + ib$ the identity $(\mathbf{x} - x)(\mathbf{x} - \bar{x}) = (\mathbf{x} - a)^2 + b^2$ holds and the fact that sums of squares are closed under multiplication.

$P \in \mathbb{R}[\mathbf{x}]^{k\times\ell}$, i.e., $L_{\mathrm{z}}(P) := (L_{\mathrm{z}}(P_{ij}))_{ij}$. We have

$$\begin{aligned}\forall p \in \mathbb{R}[\mathbf{x}]_d : L_{\mathrm{z}}(p^2) \geq 0 &\Leftrightarrow \forall \mathrm{p} \in \mathbb{R}^{s(d)} : L_{\mathrm{z}}((\mathrm{p}^\top \mathrm{v}_d)(p^\top \mathrm{v}_d)) \geq 0\\ &\Leftrightarrow \forall \mathrm{p} \in \mathbb{R}^{s(d)} : L_{\mathrm{z}}((\mathrm{p}^\top \mathrm{v}_d)(\mathrm{v}_d^\top \mathrm{p})) \geq 0\\ &\Leftrightarrow \forall \mathrm{p} \in \mathbb{R}^{s(d)} : \mathrm{p}^\top L_{\mathrm{z}}(\mathrm{v}_d\mathrm{v}_d^\top)\mathrm{p} \geq 0\\ &\Leftrightarrow M_d(\mathrm{z}) := L_{\mathrm{z}}(\mathrm{v}_d\mathrm{v}_d^\top) \succeq 0.\end{aligned}$$

This shows that testing L_{z} for non-negativity on the set of SOS is equivalent to testing positive semidefiniteness of the so-called *moment matrix* $M_d(\mathrm{z})$. As the later is a linear expression of $(\mathrm{z}_\alpha)_{|\alpha|\leq 2d}$, requesting it to be positive semidefinite is a *linear matrix inequality* (LMI). Hence consider the following SDP (with potentially infinitely many constraints indexed by γ):

$$\sup_{\mathrm{z}} L_{\mathrm{z}}(f) \quad \text{s.t.}\, L_{\mathrm{z}}(h_\gamma) \leqq b_\gamma, \gamma \in \Gamma, \quad L_{\mathrm{z}}(\mathrm{v}_d\mathrm{v}_d^\top) \succeq 0. \tag{1.4}$$

As argued above this SDP is an equivalent reformulation of (1.3), in the case that $K = \mathbb{R}$ and the degree of the involved polynomials is less then $2d$. The same is actually true for $K = \mathbb{R}^n$ if $d = 1$ and for the particular case $K = \mathbb{R}^2$ and $d = 4$ [Hil88; Lau09]. However, in the general case the set of sums of squares is only a strict subset of the set of non-negative polynomials.

1.3 Putinar's Positivstellensatz

In this section we discuss a set of test-polynomials for the general case of a basic semialgebraic compact set $K \in \mathbb{R}^n$.

Definition 1.3.1 (Quadratic module)**.** Let $g_1, \ldots, g_m \in \mathbb{R}[\mathbf{x}]$ and $K := \{\mathrm{x} \in \mathbb{R}^n : g_1(\mathrm{x}) \geq 0, \ldots, g_m(\mathrm{x}) \geq 0\}$. Let $g_0 := 1$ The set

$$Q(K) := Q(g_1, \ldots, g_m) := \{\sum_{j=0}^{m} \sigma_j g_j : \sigma_j \in \Sigma[\mathbf{x}], j = 0, \ldots, m\}.$$

is called the *quadratic module* (associated to $g_1, \ldots, g_m$). The quadratic module is called *archimedean*, if there exists an $N \in \mathbb{N}$ such that $N - ||\mathbf{x}||^2 \in Q(K)$.

Membership in $Q(K)$ provides a certificate of non-negativity on K.

Indeed: Let $\sigma_0, \ldots, \sigma_m \in \Sigma[\mathbf{x}]$ and $\mathrm{x} \in K$. Then

$$\sum_{j=0}^{m} \underbrace{\sigma_j(\mathrm{x})}_{\in \mathbb{R}^2+\ldots+\mathbb{R}^2 \geq 0,\ \mathrm{x}\in K} \underbrace{g_j(\mathrm{x})}_{\geq 0,\ \mathrm{x}\in K} \geq 0.$$

The property of the quadratic module to be archimedean is slightly stronger than K being compact. Indeed, if $Q(K)$ is archimedean, K is a subset of $\{\mathrm{x} \in \mathbb{R}^n : N - ||\mathrm{x}||^2 \geq 0\}$ which is bounded. As by definition K is closed, it follows that K is compact. Conversely,

if K is known to be compact, it is often possible to compute a number $N \in \mathbb{N}$ sufficiently large such that $K \subseteq \{\mathrm{x} \in \mathbb{R}^n : N - ||\mathrm{x}||^2 \geq 0\}$. Then the polynomial $N - ||\mathbf{x}||^2$ can be added to the description of K, i.e., $g_{m+1} := N - ||\mathbf{x}||^2$ and $Q(K) = Q(g_1, \ldots, g_{m+1})$ is archimedean. Ensuring the archimedean property is in particular important because of the following theorem.

Theorem 1.2 (Putinar's Positivstellensatz [Put93])**.** *Let* $K := \{\mathrm{x} \in \mathbb{R}^n : g_1(\mathrm{x}) \geq 0, \ldots, g_m(\mathrm{x}) \geq 0\}$ *and the quadratic module* $Q(K)$ *be archimedean. Let* $p \in \mathbb{R}[\mathbf{x}]$ *be strictly positive on* K*. Then* $p \in Q(K)$*.*

Remark 1.3.2. Assume that in (1.3) one of the h_γ is the constant polynomial 1. Then in Theorem 1.1 it is actually sufficient to test non-negativity of L_z only for polynomials strictly positive on K. To see this assume $L_\mathrm{z}(p) \geq 0$ for all $p > 0$ on K. Let $p \geq 0$ on K. Then $p + \varepsilon > 0$ on K and therefore $L_\mathrm{z}(p + \varepsilon) \geq 0$, i.e., $L_\mathrm{z}(p) \geq -\varepsilon L_\mathrm{z}(1)$. As $L_\mathrm{z}(1)$ is bounded, $-\varepsilon L_\mathrm{z}(1) \to 0$ for $\varepsilon \to 0$, and hence $L_\mathrm{z}(p) \geq 0$.

Together with Theorem 1.2 this remark qualifies $Q(K)$ as a viable substitute for $\mathrm{Pos}(K)$. In Section 1.2 we have seen that non-negativity of L_z on $\Sigma[\mathbf{x}]_d$ can be shown by checking whether the moment matrix $M_d(\mathrm{z})$ is positive semidefinite. A similar result is true for the quadratic module.

Definition 1.3.3. In analogy with $\Sigma[\mathbf{x}]_d$ we define the *truncated quadratic module* $Q_d(K)$ for $2d \geq \max\{\deg(f), \deg(g_1), \ldots, \deg(g_m\}$. Let $d_0 = d$ and $d_j = \lfloor (2d - \deg(g_j))/2 \rfloor$, where $\lfloor r \rfloor := \max\{s \in \mathbb{N} : s \leq r\}$. Then the truncated quadratic module of degree at most $2d$ is defined as

$$Q_d(K) := Q_d(g_1, \ldots, g_m) := \{\sum_{j=0}^{m} \sigma_j g_j : \sigma_j \in \Sigma[\mathbf{x}]_{d_j}, j = 0, \ldots, m\} \subseteq \mathbb{R}[\mathbf{x}]_{2d}.$$

If it is clear from the context, we suppress the K in the notation and write Q_d instead of $Q_d(K)$. The truncated quadratic module is a convex cone in $\mathbb{R}[\mathbf{x}]_{2d}$. By the same argumentation as in Section 1.2, $L_\mathrm{z} \geq 0$ on $Q_d(K)$ if and only if

$$M_{d_j}(g_j\,\mathrm{z}) := L_\mathrm{z}(\mathrm{v}_{d_j}\mathrm{v}_{d_j}^\top g_j) \succeq 0, j = 0, \ldots, m.$$

The matrices $M_{d_j}(g_j\,\mathrm{z})$ for $j > 1$ are called *localization matrices*. Note that in contrast to Section 1.2 and despite Theorem 1.2 it is not true that every $p \in \mathbb{R}[\mathbf{x}]_{2d}$ which is strictly positive on K belongs to $Q_d(K)$, because of degree cancellations in the sum $\sum_{j=0}^m \sigma_j g_j$.

1.4 The Moment-SOS Hierarchy

Recall that for an optimization problem (OP) we denote its optimal value by $(OP)^*$. Consider

$$\sup_{\mathrm{z}} L_\mathrm{z}(f) \quad \text{s.t.}\, L_\mathrm{z}(h_\gamma) \leqq b_\gamma, \gamma \in \Gamma, \quad L_\mathrm{z} \geq 0 \text{ on } Q(K). \tag{PUT}$$

Then, under the assumptions that $Q(K)$ is archimedean and one of the $h_\gamma = 1$, $(\mathrm{GMP})^* = (\mathrm{PUT})^*$. Lasserre [Las01] proposed to approximate the solution to (PUT) by truncating the

quadratic module to a certain degree. A truncated version of (PUT) now can be formulated as the SDP

$$\begin{aligned} &\sup_{\mathrm{z}} L_{\mathrm{z}}(f) \\ &\text{s.t. } L_{\mathrm{z}}(h_\gamma) \leqq b_\gamma,\ \gamma \in \Gamma,\ \deg(h_\gamma) \leq 2d, \\ &\quad M_{d_j}(g_j\,\mathrm{z}) \succeq 0,\ j = 0, \ldots, m. \end{aligned} \tag{P_d}$$

Clearly (P_d) is a relaxation of (PUT) because fewer constraints $L_{\mathrm{z}}(h_\gamma) \leqq b_\gamma$ are considered and non-negativity of L_{z} is only asked for a subset of $Q(K)$. This implies $(P_d)^* \leq (\mathrm{GMP})^*$ for all d. In addition, since $Q_d(K) \subseteq Q_{d+1}$, $(P_d)^*_d$ is an increasing sequence. Consequently $(P_d)^*_d$ converges as soon as $(\mathrm{GMP})^*$ is finite. Actually we have the following result which is the basis for almost all work presented in this thesis.

Theorem 1.3 (Moment-SOS Hierarchy [Las10b])**.** *Let* $K := \{\mathrm{x} \in \mathbb{R}^n : g_1(\mathrm{x}) \geq 0, \ldots, g_m(\mathrm{x}) \geq 0\}$ *with* $g_1 := N - \|\mathbf{x}\|^2$ *for some* $N \in \mathbb{N}$ *and* $h_{\gamma_0} = 1$ *for some* $\gamma_0 \in \Gamma$*. Denote by* $d_{\min} := \max\{\deg f, \deg g_1, \ldots, \deg g_m\}$*. Assume that every semidefinite relaxation* $(P_d)_{d \geq d_{\min}}$ *has a feasible point. Then:*

(i) Every relaxation (P_d) *with* $d \geq d_{\min}$ *has an optimal solution. Moreover,* $(P_d)^*_d \to (\mathrm{GMP})^*$ *for* $d \to \infty$*.*

(ii) Let z^d *be an optimal solution of* (P_d)*. If* (GMP) *has a unique optimizer* ϕ^**, then*

$$\lim_{d\to\infty} \mathrm{z}^d_\alpha = \int \mathbf{x}^\alpha \,\mathbf{d}\phi^*, \quad \forall\, \alpha \in \mathbb{N}^n. \tag{1.5}$$

Proof. Without loss of generality, we may assume that $N = 1$ (possibly after scaling). Let $\mathrm{z} := (\mathrm{z}_\alpha)_{\alpha \in \mathbb{N}^n_{2d}}$ be a feasible solution of (P_d). We note that the expressions $L_{\mathrm{z}}(\mathbf{x}_i^{2k})$ are on the diagonal of the moment matrix and hence constrained to be non-negative. Further the constraint $M_{d_1}(g_1\,\mathrm{z}) \succeq 0$ implies that all its diagonal elements $L_{\mathrm{z}}(x_i^{2k}(1 - \sum_{j=1}^n \mathbf{x}_i^2))$, $k = 0, \ldots, d-1$, $i = 1, \ldots, n$, are non-negative. Fix i and d. By linearity of the Riesz functional and invoking the first statement, we conclude that $L_{\mathrm{z}}(\mathbf{x}_i^{2d}) \leq L_{\mathrm{z}}(\mathbf{x}_i^{2(d-1)})$. This implies that $L_{\mathrm{z}}(\mathbf{x}_i^{2k}) \leq b_{\gamma_0}$ for all i and all k, moreover [Las15, Proposition 2.38] now yields

$$|\mathrm{z}_\alpha| \leq b_{\gamma_0}, \quad \forall \alpha \in \mathbb{N}^n_{2d}.$$

This means that the feasible set of (P_d) is bounded, and hence compact as feasible sets of SDPs are closed. Hence for every $d \geq d_{\min}$ (P_d) has an optimal solution z^d which we understand as element of $\mathbb{R}^{\mathbb{N}^n}$ by the canonical embedding, i.e., by setting $\mathrm{z}^d_\alpha := 0$, for $|\alpha| > d$.

From now one we consider the sequence of sequences $(\mathrm{z}^d)_{d \geq d_{\min}}$. We have proved that this sequence is uniformly bounded by b_γ and hence contained in a ball $\mathbf{B}$ of the Banach space ℓ_∞ of bounded sequences. By Banach-Alaoglu theorem [Bre10, Theorem 3.16], $\mathbf{B}$ is weakly star compact. Hence there exists $\mathrm{z}^* \in \mathbf{B}$ and a subsequence $(\mathrm{z}^{d_k})_{k\in\mathbb{N}}$ such that $\mathrm{z}^{d_k} \overset{*}{\rightharpoonup} \mathrm{z}^*$. In particular, by considering sequences s^α defined by $s_\beta := 0$ for all $\beta \in \mathbb{N}^n \setminus \{\alpha\}$ and $s_\alpha := 1$,

$$\lim_{k\to\infty} \mathrm{z}^{d_k}_\alpha = \lim_{k\to\infty} \left\langle s^\alpha, \mathrm{z}^{d_k} \right\rangle = \lim_{k\to\infty} \langle s^\alpha, \mathrm{z}^* \rangle = \mathrm{z}^*_\alpha, \quad \forall \alpha \in \mathbb{N}^n. \tag{1.6}$$

Now, let $d \in \mathbb{N}$ be fixed, arbitrary. The convergence (1.6), implies $L_{z^*}(v_d v_d^\top) \succeq 0$ and $L_{z^*}(v_d v_d^\top g_j) \succeq 0$, for every $j = 1, \ldots, m$. Now this is equivalent to $L_{z^*} \geq 0$ on $Q(K)$, implying that there is a measure $\phi \in \mathcal{M}_+(K)$, such that $z^*_\alpha = \int \mathbf{x}^\alpha \, \mathbf{d}\phi$. Moreover, (1.6) implies that $\int h_\gamma \, \mathbf{d}\phi = L_{z^*}(h_\gamma) \leqq b_\gamma$ for all $\gamma \in \Gamma$ which proves that ϕ is a feasible solution of (GMP). In addition, as (P_d) is a relaxation of (GMP), we have

$$\int f \, \mathbf{d}\phi = \lim_{k\to\infty} L^*_{d_k}(f) \geq (\text{GMP})^* \geq \int f \, \mathbf{d}\phi, \tag{1.7}$$

which shows optimality of ϕ. Finally, if (GMP) has a unique minimizer ϕ^*, then $\phi = \phi^*$ and (1.6) holds for the whole sequence, which yields (1.5). □

Theorem 1.3 enables us in theory to approximate the value and the solution of (GMP) as closely as desired. This makes the moment-SOS hierarchy an important tool to solve problems that can be formulated as GMPs and builds the basis for the research presented in the subsequent chapters.

A Dual Point of View

The "SOS part" in the name moment-SOS hierarchy makes reference to the dual semidefinite programs associated to (P_d).

We review briefly the concept of duality in conic programming. Let $\mathbf{X}, \mathbf{V}, \mathbf{Y}$,and $\mathbf{W}$ be real finite-dimensional vector spaces such that there exist dual pairings $\langle \cdot, \cdot \rangle_{\mathbf{V},\mathbf{X}} : \mathbf{V} \times \mathbf{X} \to \mathbb{R}$ and $\langle \cdot, \cdot \rangle_{\mathbf{Y},\mathbf{W}} : \mathbf{Y} \times \mathbf{W} \to \mathbb{R}$. Let $D \subseteq \mathbf{W}$ and $E \subseteq \mathbf{X}$ be closed convex cones. Define the *dual cone* of D by $D^* := \{\mathrm{y} \in \mathbf{Y} : \langle \mathrm{y}, \mathrm{w} \rangle_{\mathbf{Y},\mathbf{W}} \geq 0, \forall \mathrm{w} \in D\}$ and $E^* \subseteq \mathbf{V}$ correspondingly. Let $A : \mathbf{X} \to \mathbf{W}$ be a linear mapping. Its *adjoint mapping* $A^* : \mathbf{Y} \to \mathbf{V}$ is defined by the relation $\langle A^*\mathrm{y}, \mathrm{x} \rangle_{\mathbf{V},\mathbf{X}} = \langle \mathrm{y}, A\mathrm{x} \rangle_{\mathbf{Y},\mathbf{W}}$ for all $\mathrm{x} \in \mathbf{X}$ and all $\mathrm{y} \in \mathbf{Y}$. Finally let $\mathrm{b} \in \mathbf{W}$ and $\mathrm{c} \in \mathbf{V}$. Then the following conic optimization problems are in duality:

$$\begin{aligned} \sup_{\mathrm{x}} \ & \langle \mathrm{c}, \mathrm{x} \rangle_{\mathbf{V},\mathbf{X}} \\ \text{s.t.} \ & \mathrm{b} - A\mathrm{x} \in D, \\ & \mathrm{x} \in E, \end{aligned} \tag{1.8}$$

$$\begin{aligned} \inf_{\mathrm{y}} \ & \langle \mathrm{y}, \mathrm{b} \rangle_{\mathbf{Y},\mathbf{W}} \\ \text{s.t.} \ & A^*\mathrm{y} - \mathrm{c} \in E^*, \\ & \mathrm{y} \in D^*. \end{aligned} \tag{1.9}$$

Problem (1.8) is called the *primal* and (1.9) its *dual* problem. A generic property of a pair of conic dual problems is *weak duality*, i.e., if both problems are feasible, then $(1.8)^* \leq (1.9)^*$. We say that *strong duality holds* when $(1.8)^* = (1.9)^*$. A sufficient condition for strong duality is that (1.8) is strictly feasible, i.e., there exists an $\mathrm{x} \in E$ such that $b - A\mathrm{x}$ is in the interior of D. If in addition $(1.8)^*$ is finite, then (1.9) has an optimal solution. Another sufficient condition for strong duality is that the feasible set of (1.8) is compact.

We can identify (P_d) with (1.8). To simplify notation let $\Gamma_d := \gamma \in \Gamma : \deg(h_\gamma) \leq 2d$ and assume that the constraints $L_z(h_\gamma) \leqq b_\gamma$ are equalities for all $\gamma \in \Gamma_d$. Let N be the number of elements of Γ_d. Recall that the semidefinite conditions in (P_d) are equivalent to the constraint $L_z(p) \geq 0$, for all $p \in Q_d(g_1, \ldots, g_m)$, i.e., $\mathrm{z} \in Q_d^*$. This means, that $z \in Q_d^*$. Now we identify (P_d) with (1.8) via

- $\mathbf{X} \leftarrow \mathbb{R}^{s(2d)}, \mathbf{V} \leftarrow \mathbb{R}[\mathbf{x}]_{2d}, \mathbf{Y} \leftarrow \mathbb{R}^N$, and $\mathbf{W} \leftarrow \mathbb{R}^N$,
- $\langle p, \mathrm{z} \rangle_{\mathbb{R}[\mathbf{x}]_{2d}, \mathbb{R}^{s(2d)}} \leftarrow L_z(p)$ and $\langle \mathrm{y}, \mathrm{w} \rangle_{\mathbb{R}^N, \mathbb{R}^N} \leftarrow \mathrm{y}^\top \mathrm{w}$,

- $D \leftarrow \{0\}^N$, $E \leftarrow Q_d^*$, $c \leftarrow f$, and $b \leftarrow (b_\gamma)_{\gamma\in\Gamma_d}$,
- $A : \mathbb{R}^{s(2d)} \to \mathbb{R}^N$, $Az \leftarrow (L_z(h_\gamma))_{\gamma\in\Gamma_d}$.

The dual cone of $\{0\}^N$ is $\mathbb{R}^N$. As Q_d is a closed convex cone in a finite dimensional linear space, $E^* = (Q_d^*)^* = Q_d$. By linearity of the Riesz functional we have $\langle y, Az\rangle = y^\top (L_z(h_\gamma))_{\gamma\in\Gamma_d} = L_z(y^\top(h_\gamma))_{\gamma\in\Gamma_d} = \langle A^* y, z\rangle$ defining A^*. Consequently the dual of (P_d) reads:

$$\begin{aligned} \inf_y \ & \sum_{\gamma\in\Gamma_d} y_\gamma b_\gamma \\ \text{s.t.} \ & \sum_{\gamma\in\Gamma_d} y_\gamma h_\gamma - f \in Q_d(g_1,\dots,g_m), \\ & y \in \mathbb{R}^N. \end{aligned} \tag{D_d}$$

If the $L_z(h_\gamma) \leqq b_\gamma$ is an inequality for some $\gamma \in \Gamma_d$ the corresponding y_γ will be constrained to be non-negative. When both problems are feasible we conclude $(P_d)^* \leq (D_d)^*$ by weak duality. Actually one can show, that strong duality holds under the conditions of Theorem 1.3 [Las01]. Finally we note that (D_d) indeed is an SDP. Therefore recall v_d denotes the vector of monomials of degree at most d, and any polynomial $p \in \mathbb{R}[\mathbf{x}]_d$ can be written as $p^\top v_d$ for a coefficient vector $p \in \mathbb{R}^{s(d)}$. Consequently for any $M \in \mathbb{N}$ and any matrix $P \in \mathbb{R}^{M\times s(d)}$ the polynomial $(Pv_d)^\top(Pv_d)$ is an SOS of degree at most $2d$. Now the cone of positive semidefinite matrices of size $s(d)$ is exactly the set of matrices $\{P^\top P : P \in \mathbb{R}^M \times s(d), \text{ for some } M \in \mathbb{N}\}$. As $\operatorname{trace}((Pv_d)^\top(Pv_d)) = \operatorname{trace}((P^\top P)v_d v_d^\top)$, the cone of sums of squares is in one to one correspondence to the cone of positive semidefinite matrices. This argumentation extends to the truncated quadratic module, showing that (D_d) is an SDP. It is also straightforward to see that a dual problem of (1.3) is

$$\begin{aligned} \inf_{y_\gamma} \ & \sum_{y\in\Gamma} y_\gamma b_\gamma \\ \text{s.t.} \ & \sum_{\gamma\in\Gamma} y_\gamma h_\gamma - f \in \operatorname{Pos}(K) \\ & y_\gamma \geq 0, \gamma \in \Gamma_+, \text{ finitely many } y_\gamma \neq 0. \end{aligned} \tag{1.10}$$

Indeed weak duality holds: let $y = (y_\gamma)_{\gamma\in\Gamma}$ feasible for (1.10) and $z = (z_\alpha)_{\alpha\in\mathbb{N}^n}$ feasible for (1.3). Then $L_z(f) \leq L_z(\sum_{\gamma\in\Gamma} y_\gamma h_\gamma) = \sum_{\gamma\in\Gamma} y_\gamma L_z(h_\gamma) \leq \sum_{y\in\Gamma} y_\gamma b_\gamma$, which shows that weak duality holds. We refer to [Las10b, Theorem 1.3] for conditions such that strong duality, i.e., $(1.3)^* = (1.10)^*$, holds. Finally note that if $\deg(h_\gamma) \leq 2d$ for all $\gamma \in \Gamma$, (D_d) is a strengthening of (1.10), because the non-negativity constraint is replaced by a certificate of non-negativity, i.e., by membership to the truncated quadratic module. As the latter is based on SOS weights, the semidefinite programs $(D_d)_d$ are called *SOS-strengthenings* of (1.10) – hence the name *moment-SOS* hierarchy for the sequence of semidefinite problems $(P_d)_d$ and the respective duals $(D_d)_d$.

We conclude the chapter by coming back to the example of polynomial optimization. In this context the dual to (M-POP) reads like (D-POP) with SOS strengthenings (S-POP$_r$).

$$\begin{array}{ll} \sup\limits_{\mathrm{t}} & t \\ \text{s.t.} & f - \mathrm{t} \geq 0 \text{ on } K, \end{array} \qquad \text{(D-POP)} \qquad \begin{array}{ll} \sup\limits_{t} & \mathrm{t} \\ \text{s.t.} & f - \mathrm{t} \in Q_r(K), \end{array} \qquad \text{(S-POP}_r\text{)}$$

where t is the real dual variable corresponding to the single moment constraint in (M-POP). We refer to the sequence of problems $(\text{S-POP}_r)_r$ as the *standard hierarchy for polynomial optimization*. In Chapter 2 we shall discuss alternative hierarchies for polynomial optimization, where the truncated quadratic module $Q_d(K)$ is replaced by other algebraic certificates for non-negativity on a basic semialgebraic compact set K.

Chapter 2

Sparse Certificates of Non-negativity

In this chapter we discuss in more detail the question of non-negativity of a polynomial on a basic semialgebraic compact set K, which was already mentioned in Chapter 1. Certificates of non-negativity are used in several fields of research. In the subsequent sections we discuss a natural use of certificates in global polynomial optimization. Among other field of applications for certificates of non-negativity are, e.g., control systems, where one is interested in proving sign conditions of so-called *Lyapunov functions*. In verification software certificates are used to automatically generate *proofs for inequalities* (see e.g.[Mag+15]). For a more exhaustive list of applications we refer to a nice introductory section in [AM14].

Polynomial Optimization Maybe the most intuitive use of certificates is *polynomial optimization*, where certificates for non-negativity are used to provide lower bounds on the global minimum of an optimization problem

$$\inf_{\mathrm{x}} f(\mathrm{x}) \quad \text{s.t.}\, \mathrm{x} \in K := \{\mathrm{x} \in \mathbb{R}^n : g_1(\mathrm{x}) \geq 0, \ldots, g_m(\mathrm{x}) \geq 0\}, \tag{POP}$$

where $f, g_1 \ldots, g_m \in \mathbb{R}[\mathbf{x}]$ are multivariate polynomials. In this chapter we will focus on its dual formulation

$$\sup_t t \quad \text{s.t.}\, f - t \geq 0 \text{ on } K, t \in \mathbb{R}. \tag{2.1}$$

We have seen in Chapter 1 that $(\text{POP})^* = (2.1)^*$. Replacing the non-negativity constraint on the parametrized polynomial $f - t$ by a certificate for non-negativity on K denoted by $\text{Cert}(K)$ or $\text{Cert}(g_1, \ldots, g_m)$ leads to the problem

$$\sup_t t \quad \text{s.t.}\, f - t \in \text{Cert}(g_1, \ldots, g_m), \tag{RES}$$

which is a strengthening of the original problem (POP). In Chapter 1 we have used the quadratic module associated to $g_1, \ldots, g_m$ as a set of certificates for non-negativity on K. In this chapter we are going to discuss some other certificates and their respective advantages. The individual sets of certificates are often parametrized by a parameter $r \in \mathbb{N}$, and

$$\ldots \subseteq \text{Cert}_r \subseteq \text{Cert}_{r+1} \subseteq \ldots.$$

As a consequence of this inclusion, certificates of higher order are less restrictive and hence passing from Cert_r to Cert_{r+1} in (RES) leads to a higher optimal value $(\text{RES})^*$, providing a better lower bound on $(\text{POP})^*$. Accordingly the sequence of certificates induces a sequence of problems $(\text{RES})_r$ with increasing optimal value. Such sequences are called *hierarchies*.

Positivstellensätze and Hierarchies Most certificates are derived from so-called Positivstellensätzen, i.e. theorems of positivity of the following form: *Let K satisfy some property. Then there exists a subset* $\mathrm{Cert} \subseteq \mathbb{R}[\mathbf{x}]$*, such that* $q \in \mathrm{Cert}$ *implies* $q \geq 0$ *on* K*, and for every* $p > 0$ *on* K *it holds that* $p \in \mathrm{Cert}$. Hierarchies then are obtained by restricting to subsets Cert of increasing size, for example at order r one might allow only for certificates of degree at most $2r$. The convergence of $(\mathrm{RES})_r^*$ to $(\mathrm{POP})^*$ when $r \to \infty$ holds whenever $\bigcup_{r\in\mathbb{N}} \mathrm{Cert}_r = \mathrm{Cert}$, because for any $\varepsilon > 0$ the polynomial $f - (\mathrm{POP})^* + \varepsilon$ is strictly positive on K and hence, by the Positivstellensatz, has a representation in Cert_r for r large enough.

We have already seen a first Positivstellensatz in Theorem 1.2 where the quadratic module associated with the polynomials $g_1, \ldots, g_m$ is the set of certificates guaranteeing non-negativity on K. Truncating the SOS weights to a certain degree that depends on the *truncation order* $r \in \mathbb{N}$, leads to a hierarchy of semidefinite programs $(\text{S-POP}_r)_r$. Note that there is an ambiguity in how the *truncation degree* $d(r)$ is to be obtained from the order r. In the literature one finds two divergent choices for $d(r)$, either $d(r) = r$ or $d(r) = 2r$. To avoid confusion we will often talk of the truncation degree $d(r)$, referring to the degree of σ_0 in (S-POP_r), rather than about the order r. In addition we will sometimes talk about the *first possible truncation*. With this we refer to the first feasible strengthening $(\mathrm{RES})_r$.

Discussion of Certificates In order to compare different positivity certificates we are going to focus on two aspects: (i) the quality of the lower bounds provided by different certificates, and (ii) the computational effort needed to compute them. Not surprisingly there is a trade-off between those two aspects, i.e., better quality bounds are usually computationally more demanding.

From this point of view, the certificate proposed in Putinar's Positivstellensatz Theorem 1.2 is very powerful, but expensive, i.e. usually already the first or second possible strengthenings provide values of $(\mathrm{RES})^*$ that are quite close to the actual global minimum $(\mathrm{POP})^*$. However, when the lower truncation orders $r = 1, 2, \ldots$ do not lead to a satisfying lower bound, increasing r results quickly in SDPs that become too large for state-of-the-art solvers and are not computable any more.

Although there is hope that more powerful SDP solvers will be developed in the future, overcoming this unsatisfying situation theoretically is an active field of research. From both practical and theoretical view-points two main research directions emerge. One branch is concerned with finding alternative certificates (e.g. [LTS17; AM14]), while the other tries to exploit some additional available information such as symmetry [GP04] or sparsity [Wak+06; Las06].

The main result of this chapter Theorem 2.7 links both research directions. It provides a sparse version to a variety of Positivstellensätze. The proof of the result has been published in [WLT17] in a less general context. The notion of sparsity that we use follows [Wak+06] and enables to consider problems of much higher dimension and compute more steps of the respective hierarchies.

Outline First we present the sparsity pattern introduced by Waki et al. and the sparse version of Putinar's Positivstellensatz. Then we discuss a list of alternative certificates of non-negativity and provide a sparse version for each of them. We conclude with two sections devoted to applications. In the first one we present some numerical experiments

from polynomial optimization and compare the sparse versions of the standard hierarchy and the so-called BSOS hierarchy to their dense versions. Finally we explain how sparsity can still be established in the case of only nearly sparse optimization problems.

2.1 Sparsity and the Question of Non-negativity

In this section we present what is called *correlative sparsity*. We show how this sparsity can be exploited in order to compute certificates of non-negativity more efficiently. The results in this section trace back to a group of researchers from Japan, namely H. Waki, S. Kim, M. Kojima, and M. Muramatsu [Wak+06] and J. B. Lasserre [Las06].

2.1.1 Sparsity Pattern

In the previous chapter we were interested in whether a polynomial $p \in \mathbb{R}[\mathbf{x}]$ is non-negative on the basic semialgebraic compact set $K = \{\mathrm{x} \in \mathbb{R}^n : g_1(\mathrm{x}) \geq 0, \ldots, g_m(\mathrm{x}) \geq 0\}$. In this chapter we slightly adapt the notation in order to be able to formulate results with sparsity. Let $I_1, \ldots, I_\ell$ be subsets of the set of variables $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$. We denote by $\mathbb{R}[I_k]$ the subring of polynomials only invoking monomials in the variables in I_k, i.e. $\mathbb{R}[I_k] := \mathbb{R}[(\xi)_{\xi \in I_k}]$. Note that $\mathbb{R}[I_k]$ can be seen as a linear subspace of $\mathbb{R}[\mathbf{x}]$, giving rise to the notation $\mathbb{R}[I_{k_1}] + \mathbb{R}[I_{k_2}] := \{p_1 + p_2 : p_1 \in \mathbb{R}[I_{k_1}],\ p_2 \in \mathbb{R}[I_{k_2}]\}$ and the evaluation $q(\mathrm{x})$ for a polynomial $q \in \mathbb{R}[I_k]$ and a point $\mathrm{x} \in \mathbb{R}^n$ in addition to the standard evaluation in a point $\mathrm{x} \in \mathbb{R}^{n_k}$, where n_k is the number of variables in I_k.

Definition 2.1.1 (Sparsity Pattern)**.** We say that (p, K) – or the polynomials $p, g_1, \ldots, g_m$ – respect a sparsity pattern $I_1, \ldots, I_\ell$, if

- $p \in \sum_{k=1}^{\ell} \mathbb{R}[I_k]$ and
- $g_j \in \mathbb{R}[I_k]$ for some $k \in \{1, \ldots, \ell\}$ for all $j = 1, \ldots, m$.

We shall always assume that a sparsity pattern is a covering of $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, as otherwise there is a redundant variable among the $\mathbf{x}_i$. Note however that we do not assume that the intersections $I_i \cap I_j$ are empty, i.e., a sparsity pattern is not necessarily a partition of $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$. For a geometric interpretation of K respecting a sparsity pattern, denote by $\pi_k : \mathbb{R}^n \to \mathbb{R}^{n_k}$ the projection induced by $\mathbf{x} \mapsto (\xi)_{\xi \in I_k}$. Then we can write $K = \{\mathrm{x} \in \mathbb{R}^n : \pi_k(\mathrm{x}) \in K^{(k)}, k = 1, \ldots, \ell\}$, where

$$K^{(k)} := \{\mathrm{x} \in \mathbb{R}^{n_k} : g(\mathrm{x}) \geq 0, \text{ for all } g \in \{g_1, \ldots, g_m\} \cap \mathbb{R}[I_k]\}, \tag{2.2}$$

i.e., the set K is completely described by its "shadows" $\pi_k(K)$ and each of these projections is a basic semialgebraic set described by a subset of the constraint polynomials $g_1, \ldots, g_m$. The standard example for such a set K is the Steinmetz solid which is the intersection of two cylinders which are perpendicular to each other:

$$K = \{\mathrm{x} \in \mathbb{R}^3 : 1 - \mathrm{x}_1^2 - \mathrm{x}_2^2 \geq 0,\ 1 - \mathrm{x}_2^2 - \mathrm{x}_3^2 \geq 0\}$$

with sparsity pattern $I_1 = \{\mathbf{x}_1, \mathbf{x}_2\}$ and $I_2 = \{\mathbf{x}_2, \mathbf{x}_3\}$. The shadows $K^{(1)}$ and $K^{(2)}$ are the unit circles in the $\mathbf{x}_1, \mathbf{x}_2$- and $\mathbf{x}_2, \mathbf{x}_3$-plane, respectively. This easy example illustrates the

Figure 2.1: Left: RIP is not fulfilled because neither $3 \notin C_2$ nor $4 \notin C_1$. Right: RIP is fulfilled because $2, 4 \in C'_2$.

principle of sparsity. In practice however, the number n_k of variables in each I_k is assumed to be significantly smaller than the total number of variables n. Of course p and K always respect the trivial sparsity pattern $I_1 := \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$.

Definition 2.1.1 is coherent with the definition of correlative sparsity of [Wak+06], where a so-called csp[1] graph $G = (N, E)$ of $p, g_1, \ldots, g_m$ is considered. This undirected graph[2] with nodes $N = \{1, \ldots, n\}$ is constructed by putting an edge $(i_1, i_2) \in E$ if and only if

- p has a monomial involving both variables $\mathbf{x}_{i_1}$ and $\mathbf{x}_{i_2}$ or
- one of the g_j has a monomial involving $\mathbf{x}_{i_1}$ and a monomial involving $\mathbf{x}_{i_2}$.

Let $C_1, \ldots, C_\ell$ be the maximal cliques of G. Then, the sets of variables $\{\mathbf{x}_i : i \in C_k\}$, $k = 1, \ldots, \ell$ define a sparsity pattern respected by (p, K). Therefore we refer to the sets $I_1, \ldots, I_\ell$ as cliques, even if they are not constructed using the csp graph.

Waki et al. actually use the maximal cliques of a chordal extension of the csp graph because they can be obtained in time polynomial in the input size and (possibly after reordering) satisfy the so-called *running intersection property*:

Definition 2.1.2 (Running Intersection Property)**.** An ordering of sets $C_1, \ldots, C_\ell$ satisfies the running intersection property (RIP), if

$$\forall k > 1,\ \exists k_0 < k : C_k \cap \bigcup_{j<k} C_j \subseteq C_{k_0}.$$

In the sequel, a sparsity pattern respecting the running intersection property will be called a *RIP sparsity pattern*.

Note that the RIP is not necessarily a property of cliques, but of their ordering. The same set of cliques can both satisfy and violate the RIP, when ordered differently.

Example 2.1.3. Consider the following example visualized in Fig. 2.1. Define the cliques $C_1 := \{1, 2, 3\}$, $C_2 := \{2, 4, 5\}$, and $C_3 := \{2, 3, 4\}$. In this order, the RIP is not satisfied, as

$$C_3 \cap (C_2 \cup C_1) = C_3 \not\subseteq C_k,\ k = 1, 2.$$

However, when interchanging C_2 and C_3, i.e., defining $C'_1 := \{1, 2, 3\}$, $C'_2 := \{2, 3, 4\}$, and $C'_3 := \{2, 4, 5\}$, the RIP is satisfied. Indeed $C'_2 \cap C'_2 = \{2, 3\} \subseteq C'_1$ and $C'_3 \cap (C'_2 \cup C'_1) = \{2, 4\} \subseteq C'_2$.

[1] Correlative Sparsity Pattern

[2] We refer to [VA14, Chapter 1] for a brief introduction to graph theory.

From a computational point of view the RIP is crucial in order to compute efficiently Cholesky decompositions during the process of solving SDPs using interior point methods (see [VA14]). In addition and maybe even more importantly, it turns out that the same property is necessary in order to prove the following sparse version of Putinar's Positivstellensatz Theorem 1.2.

2.1.2 Sparse Putinar's Positivstellensatz

Theorem 2.1 (Sparse Putinar's Positivstellensatz)**.** *Let $p, g_1, \dots, g_m \in \mathbb{R}[\mathbf{x}]$ respect a RIP sparsity pattern $I_1, \dots, I_\ell$ and let the quadratic modules $Q^{(k)} \in \mathbb{R}[I_k]$ associated to the polynomials $\{g_1, \dots, g_m\} \cap \mathbb{R}[I_k]$, respectively, all be archimedean. If p is strictly positive on $K = \{\mathrm{x} \in \mathbb{R}^n : g_1(\mathrm{x}) \geq 0, \dots, g_m(\mathrm{x}) \geq 0\}$, then*

$$p \in \sum Q^{(k)}.$$

Note that similar to the non-sparse case Theorem 1.2 the Archimedean property of the $Q^{(k)}$ implies compactness of the $K^{(k)}$ and hence compactness of K. The proof of Theorem 2.1 makes extensive use of the RIP and can be found in [Las10b]. In [KP07; KP09] the result has been generalized to fibre products and projective limits.

In analogy to (S-POP$_r$), from Theorem 2.1 one constructs a hierarchy of problems which we call the sparse standard hierarchy in the following:

$$\sup_t t \quad \text{s.t. } f - t \in \sum Q_r^{(k)}. \tag{2.3}$$

For fixed r, problem (2.3) is an SDP (compare Section 1.4). We recall that an SOS of degree $2d$ in n variables can be represented via a positive semidefinite matrix of size $\binom{n+d}{d}$. As the SOS weights involved in the quadratic modules $Q^{(k)}$ in Theorem 2.1 are polynomials in the variables I_k only, the size of the involved positive semidefinite matrices is $\binom{n_k+d}{d}$. As soon as n_k is significantly smaller than n, replacing the quadratic module from Theorem 1.2 by the sum of quadratic modules in Theorem 2.1 results in drastic computational savings.

2.2 Exactness for SOS-convex Problems

In the case that $f, g_1, \dots, g_m$ are quadratic, the first possible relaxation of the standard hierarchy (S-POP$_r$) is the dual of the so-called *Shor relaxation* which is known to be exact in the case that the polynomials $f, -g_1, \dots, -g_m$ are convex [Sho98]. In this section we consider polynomials of maximal degree $2d$ that have the more restrictive property that $f, -g_1, \dots, -g_m$ are SOS-convex:

Definition 2.2.1. A polynomial p is SOS-convex if its Hessian[3] $\nabla^2 p$ is an SOS-matrix polynomial, i.e., if there exist $s \in \mathbb{N}$ and a matrix $L \in \mathbb{R}[\mathbf{x}]^{n \times s}$ such that $\nabla^2 p = LL^\top$.

[3]The Hessian of a polynomial p is the $n \times n$ matrix defined entry-wise by $(\nabla^2 p)_{ij} = \frac{\partial^2 p}{\partial \mathbf{x}_i \partial \mathbf{x}_j}$.

In the spirit of Shor we consider the following problem which can be seen as a further truncation of the first possible Putinar strengthening:

$$\sup_{t,\sigma_0,\lambda_j} \{t : f - t = \sigma_0 + \sum_{j=1}^{m} \lambda_j g_j\}, \tag{SH-D}$$

where σ_0 is an SOS of degree at most $2d$ and λ_j are SOS of degree 0, i.e., non-negative scalars. In the case of $d = 1$ (the quadratic case), (SH-D) coincides with the dual of the Shor relaxation [Las01].

As we will see in a minute, similar to the Shor relaxation, (SH-D) has the interesting property that, when f and all $-g_j$ are *SOS-convex* of degree at most $2d$, it is exact, i.e., the polynomial $f - (\text{POP})^*$ has a representation $\sigma_0 + \sum_{j=1}^{m} \lambda_j g_j$.

Theorem 2.2. *Let $K := \{\mathrm{x} \in \mathbb{R}^n : g_1(\mathrm{x}) \geq 0, \ldots, g_m(\mathrm{x}) \geq 0\}$ be compact with non-empty interior and $f, -g_1, \ldots, -g_m \in \mathbb{R}[\mathbf{x}]$ be SOS-convex of degree $2d$. Then (SH-D) is exact.*

The result is proved in [Sho98] for the case $d = 1$ where the SOS-convex condition simplifies to convexity. The presented proof builds on [Las08] where Theorem 2.2 is mentioned as a corollary.

Proof. Let x^* be a global minimizer of f on K, i.e., $f^* := f(\mathrm{x}^*) \leq f(\mathrm{x})$ for all $\mathrm{x} \in K$. By the optimality condition of Karush-Kuhn-Tucker there exist non-negative multipliers $\lambda_j \in \mathbb{R}$ such that

$$\nabla f(\mathrm{x}^*) - \sum_{j=1}^{m} \lambda_j \nabla g_j(\mathrm{x}^*) = 0 \tag{2.4}$$

$$\text{and} \quad \lambda_j g_j(\mathrm{x}^*) = 0, \ j = 1, \ldots, m \tag{2.5}$$

holds. Define the Lagrangian $L := f - f^* - \sum_{j=1}^{m} \lambda_j g_j$. Then by the SOS-convexity assumptions on the initial data, L is also SOS-convex. By (2.4) $\nabla L(\mathrm{x}^*) = 0$ and by (2.5) $L(\mathrm{x}^*) = 0$. A lemma from Helton and Nie [HN10] now yields that L is an SOS of degree $2d$, which finishes the proof. □

Having in mind the sparse version of Putinar's theorem it is immediate to define a sparse version of (SH-D):

$$\sup_{t,\sigma_0^k,\lambda_j} \{t : f - t = \sum_{k=1}^{\ell} \sigma_0^{(k)} + \sum_{j=1}^{m} \lambda_j g_j\}, \tag{SSH-D}$$

for (f, K) respecting an RIP sparsity pattern $I_1, \ldots, I_\ell$, with SOS weights $\sigma_0^{(k)} \in \mathbb{R}[I_k]$ of degree $2\lfloor \deg(f) \rfloor$ and non-negative scalars $\lambda_1, \ldots, \lambda_m$. It would be nice to prove exactness of (SSH-D) in the case of SOS-convex problems as we did for the non-sparse version. However, the lemma from Helton and Nie [HN10], which was used in the proof above, does not guarantee that sparsity in an SOS representation of the Hessian is transferred to an SOS representation of the initial polynomial. The following statement and its proof are based on a proof first presented in [WLT17].

Theorem 2.3. *Let $I_1, \dots, I_\ell$ be a (non-necessarily RIP) sparsity pattern for ($f := \sum_{k=1}^{\ell} f^k, K := \{\mathrm{x} \in \mathbb{R}^n : g_1(\mathrm{x}) \geq 0, \dots, g_m(\mathrm{x}) \geq 0\}$), where $f^1, \dots, f^k, -g_1, \dots, -g_m \in \mathbb{R}[\mathbf{x}]$ are SOS-convex polynomials of degree $2d$. Assume that for some positive integer $N \in \mathbb{N}$ the (SOS-concave) polynomials $N - \sum_{\xi \in I_k} \xi^{2d}$ are amongst the $g_1, \dots, g_m$ and that K is non-empty. Then (SSH-D) is exact.*

Note that the assumptions on K in the sparse assertion are quite different from the ones in the non-sparse case. The assumption on K being compact in the non-sparse version is replaced by the stronger assumption of the polynomials $N - \sum_{\xi \in I_k} \xi^{2d}$ being among the constraints. However, if K is already known to be compact, these polynomials can be computed and added to the description. The second assumption that K has non-empty interior in Theorem 2.2 is replaced by the weaker assumption that K itself is non-empty in Theorem 2.3. In particular, using the trivial sparsity pattern, this provides an alternative version of Theorem 2.2

Proof. Recall that (SSH-D) can be reformulated as an SDP

$$\sup_{t, X^k, \lambda_j} t \quad \text{s.t.} \, f - t - \sum_{k=1}^{\ell} \left\langle X^k, \mathrm{v}_d^k (\mathrm{v}_d^k)^\top \right\rangle - \sum_{j=1}^{m} \lambda_j g_j = 0 \tag{SSH-SDP}$$

where $t \in \mathbb{R}$, $X^k \in \mathbb{R}^{\mathrm{s}(n_k,d) \times \mathrm{s}(n_k,d)}$ positive semidefinite with $\mathrm{s}(n_k, d) := \binom{n_k+d}{d}$, and λ_j non-negative. The equality constraint is equivalent to equating all coefficients of the polynomial on the left hand side to zero. For each monomial $\mathbf{x}^\alpha$ with $|\alpha| \leq 2d$ let $\mathbf{z}_\alpha$ be the dual variable to this equality constraint on the coefficients of $\mathbf{x}^\alpha$. Then the dual to (SSH-SDP) reads

$$\inf_{\mathrm{z}} L_{\mathrm{z}}(f) \quad \text{s.t.} \quad L_{\mathrm{z}}(1) = 1, \quad L_{\mathrm{z}}(\mathrm{v}_d^k (\mathrm{v}_d^k)^\top) \succeq 0, \quad L_{\mathrm{z}}(g_j) \geq 0 \tag{SSH-P}$$

where z is a vector in $\mathbb{R}^{\mathrm{s}(n,2d)}$. The further outline of the proof is the following. In a first step we show that (SSH-P) is exact. Then we show that strong duality holds between (SSH-P) and (SSH-SDP), i.e., (SSH-P)* = (SSH-SDP)* and in conclusion (SSH-SDP)* = (POP)*

Note that as K is non-empty (SSH-P) has a feasible point defined by $\mathrm{z}_\alpha := \mathrm{x}^\alpha$ for some $\mathrm{x} \in K$. For this point it holds that $L_{\mathrm{z}}(f) = f(\mathrm{x})$. Consequently (SSH-SDP) is a relaxation of (POP), i.e.,(POP)$^* \geq$ (SSH-P)*. Next we show that the converse inequality holds, too.

Note that as the polynomials $N - \sum_{\xi \in I_k} \xi^{2d}$ are amongst the constraints, $L_{\mathrm{z}}(1) = 1$, and by linearity of L_{z} it holds that

$$N = L_{\mathrm{z}}(N) \geq N L_{\mathrm{z}}(\mathbf{x}_i^{2d}),$$

where we have used that $L_{\mathrm{z}}(\mathbf{x}_i^{2k}) = \mathrm{z}_{e_i 2k} \geq 0$ for every vector e_i of the standard basis of $\mathbb{N}^n$ and every $k = 1, \dots, d$. The latter is a consequence of the semidefiniteness of the moment matrix $L_{\mathrm{z}}(\mathrm{v}_d^k (\mathrm{v}_d^k)^\top)$. By [Las15, Proposition 2.38] this implies that $|\mathrm{z}_\alpha| \leq N$ for all $\alpha \in \mathbb{N}^n_{2d}$, i.e., the feasible set of (SSH-P) is compact. Consequently (SSH-P) attains its maximum at an optimal solution z^*. Define $\mathrm{x}^* := L_{\mathrm{z}^*}(\mathbf{x})$. By a Jensen type inequality valid for the Riesz functional and SOS-convex polynomials [Las15, Theorem 13.21] we have

$$f^k(\mathrm{x}^*) = f^k(L_{\mathrm{z}^*}(\mathbf{x})) \leq L_{\mathrm{z}^*}(f^k) \quad \text{and} \quad 0 \leq L_{\mathrm{z}^*}(g_j) \leq g_j(\mathrm{x}^*).$$

While the second statement above shows that x^* is feasible for (POP) from the first one we can conclude that

$$(\text{SSH-P})^* = L_{\mathrm{z}^*}(f) = L_{\mathrm{z}^*}(\sum_{k=1}^{\ell} f^k) = \sum_{k=1}^{\ell} L_{\mathrm{z}^*}(f^k) \geq \sum_{k=1}^{\ell} f^k(x^*) = f(\mathrm{x}^*),$$

showing that $(\text{SSH-P})^* = (\text{POP})^*$, i.e., (SSH-P) is exact. It now remains to show that $(\text{SSH-P})^* = (\text{SSH-SDP})^*$. For this note that we have just proved that (SSH-P) has an optimal solution. In addition, as the feasible set of (SSH-P) is bounded, the set of optimal solutions of (SSH-P) is bounded, too. Now [Trn05, Corollary 1] yields strong duality, i.e. $(\text{SSH-SDP})^* = (\text{SSH-P})^* = (\text{POP})^*$. □

As mentioned before we get the following alternative condition for exactness of (SH-D) by choosing the trivial sparsity pattern $I_1 := \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$.

Corollary 2.2.2. *Let $K := \{\mathrm{x} \in \mathbb{R}^n : g_1(\mathrm{x}) \geq 0, \ldots, g_m(\mathrm{x}) \geq 0\}$ be non-empty and $f, -g_1, \ldots, -g_m \in \mathbb{R}[\mathbf{x}]$ be SOS-convex of degree $2d$. For some positive integer $N \in \mathbb{N}$ let the polynomial $N - \sum_{i=1}^n \mathbf{x}_i^{2d}$ be amongst the g_j. Then (SH-D) is exact.*

With the proofs of exactness for (SH-D), in particular we have shown that the first steps of the (sparse and non-sparse) standard hierarchy are exact for SOS-convex problems, respectively.

Corollary 2.2.3. *In the situation of Theorem 2.2 or of Corollary 2.2.2 the first possible strengthening of the standard hierarchy (S-POP_r) is exact.*

Corollary 2.2.4. *In the situation of Theorem 2.3 the first possible strengthening of the sparse standard hierarchy (2.3) is exact.*

In the following we are going to discuss other certificates and corresponding sparse versions that permit to define other hierarchies for polynomial optimization and to prove their convergence to the optimal value. As it will turn out, the first steps of some of these hierarchies will be (SH-D) or (SSH-D) and will hence inherit the nice exactness results just proved in this section.

2.3 Positivstellensätze and their Sparse Versions

A diversity of certificates have been derived from alternative theorems of positivity, e.g. theorems by Schmüdgen Theorem 2.4 and Krivine Theorem 2.5. More recently the so-called BSOS certificate Theorem 2.6 has been introduced in order to reduce computational cost while maintaining a good convergence – meaning that already strengthenings of low order are able to represent the polynomial $f - (\text{POP})^*$.

2.3.1 Alternative Positivstellensätze

Putinar's Positivstellensatz Theorem 1.2 depends on the description of K. Even though we have already argued that it is always possible to add a redundant constraint in order to make the quadratic module archimedean, one might be interested in a Positivstellensatz

independent of the description of K. One such result is the following theorem due to Schmüdgen [Sch91].

Theorem 2.4 (Schmüdgen). *Let $K := \{\mathrm{x} \in \mathbb{R}^n : g_1(\mathrm{x}) \geq 0, \ldots, g_m(\mathrm{x}) \geq 0\}$ be compact and p strictly positive on K. Then*

$$p = \sum_{\alpha \in \{0,1\}^m} \sigma_\alpha \prod_{j=1}^{m} g_j^{\alpha_j},$$

with SOS weights $\sigma_\alpha \in \mathbb{R}[\mathbf{x}]$.

Note that this Schmüdgen-certificate is richer than the one from Putinar's theorem as we recover Putinar by only choosing exponents $\alpha \in \{0,1\}^m$ such that $|\alpha| \leq 1$. Once the degree of the SOS is fixed, a Schmüdgen-certificate can be computed by solving an SDP as we have already seen for Putinar in Section 1.4. The number of semidefinite variables however now is 2^m compared to $m+1$ for Putinar. This and the fact that Putinar usually works quite well are the reasons why, to the best of our knowledge, the theorem of Schmüdgen has never really been used for polynomial optimization purposes.

The following Positivstellensatz goes back to Krivine [Kri64], but can also be allocated to other authors [Ste74; Han88; Vas03]. Similar to (SH-D), it replaces SOS by non-negative variables and combines this with the idea of building products of the constraint polynomials.[4]

Theorem 2.5 (Krivine). *Let $g_1, \ldots, g_m$ be dominated by 1 on K, let $1, g_1, \ldots, g_m$ generate $\mathbb{R}[\mathbf{x}]$ as an algebra, and let K be compact. If p is strictly positive on K then*

$$p = \sum_{\alpha,\beta \in \mathbb{N}^m} \lambda_{\alpha\beta} \prod_{j=1}^{m} g_j^{\alpha_j} (1 - g_j)^{\beta_j},$$

for finitely many non-negative scalars $\lambda_{\alpha\beta}$.

To clarify the statement of the theorem we emphasize that the theorem establishes a finite number of positive scalars $\lambda_{\alpha\beta}$ but does not provide an a priori bound on $|\alpha|$ or $|\beta|$. Note further that $g_j \leq 1$ on K is not restrictive, as K is compact and the assumption can be satisfied by dividing each g_j by an upper bound of g_j on K. The assumption that $1, g_1, \ldots, g_m$ generate $\mathbb{R}[\mathbf{x}]$ is not restrictive either, as redundant constraints can always be added until the assumption is satisfied. Finally note, that when restricting to a finite number of α and β, computing a Krivine-certificate reduces to solving a linear program.

In view of this it is appropriate to ask why the certificate based on Putinar, which requires to solve an SDP, has become standard in polynomial optimization, whereas by the previous theorem there is a certificate which can be computed solving only an LP. The reason for this is that in general finite convergence cannot take place in a hierarchy based on Krivine. Assume for example that for some truncation $r \in \mathbb{N}$ the Krivine certificate is exact, i.e.,

$$f - f(\mathrm{x}^*) = \sum_{\substack{\alpha,\beta \in \mathbb{N}^m \\ |\alpha|+|\beta| \leq r}} \lambda_{\alpha\beta} \prod_{j=1}^{m} g_j^{\alpha_j} (1 - g_j)^{\beta_j},$$

[4] Actually Krivine's Theorem is older than the ones by Schmüdgen and Putinar and hence one should rather say that in the latter non-negative variables are replaced by SOS.

for a global minimizer x^* of f on K. Then, unless f is the zero polynomial, at least one $\lambda_{\alpha\beta}$ is strictly positive. Now, if x^* is in the interior of K, then evaluating the equation at x^* annihilates the left side, while the right side is strictly positive. This shows that in general a Krivine hierarchy cannot be exact at a finite degree of truncation. Another typical scenario is when a global minimizer x^* annihilates some of the g_j's (i.e., x^* is on the boundary of K) and there is a non-optimal point in K annihilating the same g_j's. Then finite convergence of a Krivine-hierarchy cannot take place neither. In addition to the lack of finite convergence, from a numeric perspective one observes, that computing lower bounds on f via Putinar certificates in general leads to closer bounds.

In the remaining section we discuss certificates that try to find a compromise between the powerful certificates of Schmüdgen and Putinar on the one side, and the *cheap*, i.e., computationally less demanding, Krivine certificate.

The following certificate has been proposed by Lasserre et al. in [LTS17] and is based on Krivine, but in order to establish at least finite convergence for the class of SOS-convex problems under some weak conditions.

Theorem 2.6 (BSOS)**.** *Let $g_1, \dots, g_m$ be dominated by 1 on K, let $1, g_1, \dots, g_m$ generate $\mathbb{R}[\mathbf{x}]$ as an algebra, and let K be compact. If p is strictly positive on K then*

$$p = \sigma_0 + \sum_{\alpha,\beta\in\mathbb{N}^m} \lambda_{\alpha\beta} \prod_{j=1}^{m} g_j^{\alpha_j}(1-g_j)^{\beta_j}, \tag{2.6}$$

for finitely many strictly positive scalars $\lambda_{\alpha\beta} \in \mathbb{R}_+$ and an SOS $\sigma_0 \in \mathbb{R}[\mathbf{x}]$ of fixed degree s.

There is nothing to prove in this statement, as it is a direct consequence of Theorem 2.5. The name BSOS comes from *bounded SOS* and makes reference to the way a hierarchy is constructed from this certificate. The truncation order for BSOS is the same as in the hierarchy from Krivine (called LP hierarchy in the sequel), i.e., the exponents α and β in (2.6) are restricted to be in the simplex $|\alpha| + |\beta| \leq r$. However, the degree of the SOS σ_0 is fixed by the user in advance and stays the same for all steps of the hierarchy. A typical choice for the degree of σ_0 is the smallest even integer greater or equal to $\deg(f)$, but also lower degrees for the SOS may improve the quality of lower bounds compared to the LP hierarchy.

An interesting feature of BSOS and an advantage over the pure Krivine hierarchy is the following Corollary.

Corollary 2.3.1. *Let $\deg(f) = 2d$ in (POP) and $f, -g_1, \dots, -g_m$ be SOS-convex. Choosing $s = 2d$ the first truncation of BSOS is exact.*

Proof. Note that the first BSOS truncation with the choice of $s = 2d$ coincides with the dual of the Shor relaxation (SH-D). Hence the proof is complete by Theorem 2.2. □

We have seen that both the standard hierarchy, based on Putinar, and the BSOS hierarchy are enhancements of the (SH-D). However, going higher in the truncation order with BSOS is computationally cheaper than increasing the degree of the SOS for the quadratic module in the standard hierarchy. While the latter may be prohibitively expensive in view of the performance of state-of-the-art SPD solvers, the former might provide better bounds or even the optimal value.

An Alternative Certificate Another approach to providing certificates that are less computationally demanding then Putinar has been proposed by Ahmadi and Majumdar [AM14]. Restricting the degree of some SOS to 0, i.e. replacing them by non-negative scalars, is a quite brutal method to enforce cheaper certificates and may be to restrictive to obtain good quality certified lower bounds. Ahmadi et al. use the fact that *(scaled) diagonal dominant* matrices are positive semidefinite. Conditioning a matrix to be (scaled) diagonal dominant can be done by linear or second order constraints, respectively. Hence restricting the matrices in Theorem 1.2 to be diagonal dominant or scaled diagonal dominant is a way to enforce positive semidefiniteness by only solving a linear program or second order cone program, respectively. Note that this procedure comes with no proof of convergence. Ahmadi et al. however establish extensions of these certificates where they can prove convergence. Replacing the non-negative weights in Theorem 2.5 or Theorem 2.6 by the weights proposed by Ahamdi and Majumar might be a good strategy to obtain a converging hierarchy that is less computationaly demanding than the one based on Putinar and at the same time better performing than the ones based on Krivine.

2.3.2 Sparse Positivstellensätze

The aim of this section is to provide sparse versions to the theorems of positivity just introduced. These sparse versions relate to the respective theorem as Sparse Putinar's Theorem 2.1 relates to its dense version Theorem 1.2. In fact the proof for the sparse versions, first presented in [WLT17], crucially relies on Theorem 2.1.

Theorem 2.7 (Sparse Positivstellensätze)**.** *Let $I_1, \dots, I_\ell$ be an RIP sparsity pattern for $(p, g_1, \dots, g_m)$ and $K := \{\mathrm{x} \in \mathbb{R}^n : g_1(\mathrm{x}) \geq 0, \dots, g_m(\mathrm{x}) \geq 0\}$ be compact. For $k = 1, \dots, \ell$ let $K^{(k)} := \{\mathrm{x} \in \mathbb{R}^{n_k} : g(\mathrm{x}) \geq 0,$ for all $g \in \{g_1, \dots, g_m\} \cap \mathbb{R}[I_k]\}$. Assume that if $p^{(k)} \in \mathbb{R}[I_k]$ is strictly positive on $K^{(k)}$ then $p^{(k)} \in \mathrm{Cert}^{(k)} \subseteq \mathbb{R}[I_k]$ for some certificate of non-negativity on $K^{(k)}$, respectively. Then, if p is strictly positive on K, p has a sparse representation*

$$p \in \sum \mathrm{Cert}^{(k)}.$$

Proof. As p is strictly positive on K there exists $\varepsilon > 0$ such that $p - \varepsilon > 0$ on K. As K is compact we can add redundant constraints to the description of K (compare Section 1.3) such that the quadratic modules $Q^{(k)}$ associated to each $K^{(k)}$ are archimedean. Consequently by Sparse Putinar Theorem 2.1 $p - \varepsilon = \sum_k p^{(k)}$ with $p^{(k)} \in Q^{(k)}$. In particular we have $\tilde{p}^{(k)} := p^{(k)} + \frac{\varepsilon}{\ell} > 0$ on $K^{(k)}$. By the Positivstellensatz we obtain that $\tilde{p}^{(k)} \in \mathrm{Cert}^{(k)}$ for each k. This finishes the proof as $p = \sum \tilde{p}^{(k)}$. □

From Theorem 2.7 we obtain immediately the sparse versions for the theorems of positivity from the previous section. Let $J_k := \{j \in \{1, \dots, m\} : g_j \in \mathbb{R}[I_k]\}$ and m_k the number of elements of J_k.

Corollary 2.3.2 (Sparse Schmüdgen)**.** *Let $I_1, \dots, I_\ell$ be an RIP sparsity pattern for $(p, g_1, \dots, g_m)$ and K be compact. If p is strictly positive on K, there exist SOS weights*

$\sigma_\alpha^{(k)} \in \mathbb{R}[I_k]$, *such that*

$$p = \sum_{k=1}^{\ell} \left(\sum_{\alpha \in \{0,1\}^{m_k}} \sigma_\alpha^{(k)} \prod_{j \in J_k} g_j^{\alpha_j} \right).$$

Corollary 2.3.3 (Sparse Krivine)**.** *Let $I_1, \dots, I_\ell$ be an RIP sparsity pattern for $(p, g_1, \dots, g_m)$, K be compact, all g_j be dominated by 1 on K, and let the polynomials $1, (g_j)_{j \in J_k}$ generate $\mathbb{R}[I_k]$ as algebras, respectively. If p is strictly positive on K, there exist finitely many non-negative weights $\lambda_{\alpha\beta}^{(k)}$, such that*

$$p = \sum_{k=1}^{\ell} \left(\sum_{\alpha,\beta \in \mathbb{N}^{m_k}} \lambda_{\alpha\beta} \prod_{j \in J_k} g_j^{\alpha_j} (1 - g_j)^{\beta_j} \right),$$

Corollary 2.3.4 (Sparse BSOS)**.** *In the same situation as in Corollary 2.3.3 there exist SOS $\sigma^{(k)} \in \mathbb{R}[I_k]$ of fixed degree $s \in \mathbb{N}$ and finitely many non-negative weights $\lambda_{\alpha\beta}^{(k)}$, such that*

$$p = \sum_{k=1}^{\ell} \left(\sigma^{(k)} + \sum_{\alpha,\beta \in \mathbb{N}^{m_k}} \lambda_{\alpha\beta} \prod_{j \in J_k} g_j^{\alpha_j} (1 - g_j)^{\beta_j} \right),$$

Remark 2.3.5. Though the proof of Theorem 2.7 is quite simple and the three corollaries are a direct application, their assertions are far from being obvious. While the constraint polynomials g_j only enter *linearly* in the Putinar certificate, and hence their variables are separated, in the other certificates one has to consider products of the g_j potentially mixing all variables.

As in the previous section we can define hierarchies of LPs and SDP form Corollary 2.3.2, Corollary 2.3.3, and Corollary 2.3.4 by restricting the degree of the SOS or the number of non-negative multipliers to some number depending on the truncation order r. When the number n_k of variables in each clique I_k is significantly larger than the number of total variables n, theses programs are much smaller and hence easier to solve. For the Sparse BSOS hierarchy at truncation order r we propose to restrict to $|\alpha| + |\beta| \leq r$ as in the dense case. With this convention we preserve the interesting feature of the BSOS hierarchy Corollary 2.3.1.

Corollary 2.3.6. *Let $\deg(f) = 2d$ in (POP), $f, g_1, \dots, g_m$ be SOS-convex and $I_1, \dots, I_\ell$ be a sparsity pattern for (f, K). Choosing $s = 2d$ the first step of the Sparse BSOS hierarchy is exact.*

The proof of this Corollary is just remarking that the first step of Sparse BSOS in this situation coincides with the (SSH-D) and applying Theorem 2.3.

2.4 Numerical Evaluation of the Sparse BSOS Hierarchy

As outlined earlier, theoretic results of convergence of the different hierarchies for polynomial optimization are important. However, from a practical point of view, the ability to retrieve

good bounds with a reasonable computational effort is a decisive factor, too. For this reason we show some numerical experiments how the Sparse BSOS hierarchy competes against its dense counterpart as well as against the sparse standard hierarchy. The results presented in this section are based on [WLT17].

2.4.1 Implementation

In order to compare the different hierarchies we developed a code for Sparse BSOS and Sparse Putinar [Wei17]. For the comparison to BSOS we used a code from [LTS17]. Both codes use the SDP solver SDPT3 [TTT12]. In this section we explain details of the implementations and discuss sufficient conditions to determine optimality from the solution.

BSOS

We explain how BSOS is implemented in [LTS17]. The polynomial equality (2.6) is implemented by sampling, i.e., a set of scalar equality constraints is generated by evaluating both sides of the polynomial equality in sufficiently many points. Let s be the fixed parameter for the degree of the sums of square σ_0 and fix a truncation order r, i.e., $|\alpha|+|\beta| \leq r$. Then the maximal degree of polynomials appearing in (2.6) is $d_{\max} := \max\{\deg(f), 2s, r\deg(g_j)\}$. Consequently we need $\mathrm{s}(d_{\max}) := \binom{n+d_{\max}}{n}$ point evaluations in order to generate enough scalar equality constraints. Let $(\mathrm{x}_i^\tau) \in [-1,1]^{n\times \mathrm{s}(d_{\max})}$ be a sample of generic points. For ease of notation define $h_{\alpha\beta} := \prod_{j=1}^m g_j^{\alpha_j}(1-g_j)^{\beta_j}$. Then the BSOS implementation solves the following SDP

$$\begin{aligned} \sup_{t,X,\lambda} t \quad \text{s.t.} \quad & f(\mathrm{x}^\tau) - \sum_{(\alpha,\beta)\in\mathbb{N}^{2m}_r} \lambda_{\alpha\beta}\, h_{\alpha\beta}(x^\tau) - \Big\langle X, \Big(\mathrm{v}_s(x^\tau)(\mathrm{v}_s(x^\tau))^\top\Big)\Big\rangle = 0, \\ & \forall \tau = 1,\dots,\mathrm{s}(d_{\max}), \quad X \in \mathcal{S}_+^{\mathrm{s}(n,s)},\ \lambda \in \mathbb{R}_+^{\mathrm{s}(2m,r)},\ t\in\mathbb{R}. \end{aligned} \tag{2.7}$$

where $\mathcal{S}_+^k$ denotes the convex cone of positive semidefinite matrices of size $k\times k$. As explained in [LTS17, Section 3], some of the constraints in the SDP stated above might be nearly redundant, i.e. they might be "almost" linearly dependent on others. Such constraints are removed before handing the problem over to the SDP solver. A close look to the code also reveals that the maximum number of point evaluations considered is limited to approximately 5000. This choice has been made to prevent the solver from running out of memory. As a consequence, when $\mathrm{s}(d_{\max}) > 5000$ the implemented SDP is a relaxation of the SDP stated above, as not enough point evaluation are considered to guarantee the polynomial equality constraint in (2.6).

Rank condition Optimality of BSOS at any step of the hierarchy r can be verified *a posteriori* by checking a rank condition on a matrix o from the dual solution to (2.7). When the rank of this matrix equals one, the value obtained from (2.7) coincides with $(\mathrm{POP})^*$ [LTS17, Lemma 1].

Sparse BSOS

In contrast to BSOS, for our code we decided to implement the polynomial equality constraint by comparing coefficients. Both strategies have drawbacks: To equate coefficients

one has to take powers of the polynomials g_j and $(1-g_j)$ which leads to an ill-conditioning of the coefficients of the polynomials $h_{\alpha\beta}$ (in the monomial basis) as some of them are multiplied by binomial coefficients which become large quickly when the truncation order r increases. On the other hand, when equating values the resulting linear system may become ill-conditioned because (depending on the points of evaluation and the g_j) the constraints may be nearly linearly dependent. The authors of [LTS17] chose point evaluation for the implementation of BSOS because SDPT3 is able to exploit the structure of the SDP generated in that way and hence problems with positive semidefinite variables of larger size can be solved. However, this feature cannot be used for Sparse BSOS. Indeed, equating coefficients is reasonable in the present context because we expect the number of variables n_k in each clique to be rather small. The drawback of this choice is that the resulting truncations with high order r can become time consuming (and even ill-conditioned as explained above). A crucial issue for the implementation of the Sparse BSOS hierarchy is how to equate the coefficients. We refer to [WLT17; Wei17] for more details on the implementation.

The actual implementation of the SDP for Sparse BSOS differs slightly from the description in Corollary 2.3.4. It turned out that SDPT3 performs better when adding some additional variables as follows. Remember the notation $J_k := \{j \in \{1,\dots,m\} : g_j \in \mathbb{R}[I_k]\}$ and define $N_r^k := \{(\alpha,\beta) \in \mathbb{N}^m \times \mathbb{N}^m : \operatorname{supp}(\alpha) \cup \operatorname{supp}(\beta) \subseteq J_k, |\alpha|+|\beta| \le r\}$ where $\operatorname{supp}(\alpha) := \{j \in \{1,\dots,m\} : \alpha_j \neq 0\}$.

$$\begin{aligned}\sup_{\substack{t,\lambda^{(k)},\\ X^{(k)},f^{(k)}}} \; t \quad \text{s.t.} \quad & t \in \mathbb{R}, \quad \Big(\sum_{k=1}^{\ell} f^{(k)}\Big)_{\mathbf{0}} = \mathrm{f}_{\mathbf{0}} - t, \quad \Big(\sum_{k=1}^{\ell} f^{(k)}\Big)_{\gamma} = \mathrm{f}_{\gamma}, \quad \forall \gamma \in \Gamma \\ & \Big(f^{(k)} - \sum_{(\alpha,\beta)\in N_r^k} \lambda_{\alpha\beta}^{(k)} h_{\alpha\beta} - \big\langle X^{(k)}, \big(\mathrm{v}_s^{(k)}(\mathrm{v}_s^{(k)})^\top\big)\big\rangle\Big)_{\gamma} = 0, \quad \forall \gamma \in \Gamma^{(k)} \\ & X^{(k)} \in \mathcal{S}_+^{\mathrm{s}(n_k,s)}, \quad \lambda^{(k)} \in \mathbb{R}_+^{|N_r^k|}, \quad f^{(k)} \in \mathbb{R}[I_k], \quad \forall k = 1,\dots,\ell,\end{aligned} \tag{2.8}$$

Here $\Gamma^{(k)}$ denotes the set of exponents of the monomials of $\mathbb{R}[I_k]_{d_{\max}}$ and $\Gamma = \bigcup \Gamma^{(k)}$. By $(p)_\gamma$ we refer to the coefficient p_γ of the polynomial p. With $\mathrm{v}_s^{(k)}$ we denote the vector consisting of the monomial basis of $\mathbb{R}[I_k]$. The additional variables, mentioned above, are the explicit polynomials $f^{(k)} \in \mathbb{R}[I_k]$ which appear as linear variables via their coefficients $\mathrm{f}_\gamma^{(k)}$.

A similar sufficient rank condition as for BSOS can be checked in order to determine that the truncation (2.8) has attained (POP)*. To see this consider the dual program to (2.8).

$$\begin{aligned}\inf_{\mathrm{z},\mathrm{z}^{(k)}} L_{\mathrm{z}}(f) \quad \text{s.t.} \quad & L_{\mathrm{z}^{(k)}}(\mathrm{v}_s^{(k)}(\mathrm{v}_s^{(k)})^\top) \succeq 0, \quad \forall k = 1,\dots,\ell, \\ & L_{\mathrm{z}^{(k)}}(h_{\alpha\beta}) \ge 0, \quad \forall (\alpha,\beta) \in N_r^k, \\ & \mathrm{z}_\gamma = \mathrm{z}_\gamma^{(k)}, \quad \forall \gamma \in \Gamma^{(k)}, \quad \forall k = 1,\dots,\ell, \\ & \mathrm{z}_{\mathbf{0}} = 1, \quad \mathrm{z} \in \mathbb{R}^{|\Gamma|}, \quad \mathrm{z}^{(k)} \in \mathbb{R}^{|\Gamma^{(k)}|}.\end{aligned} \tag{2.9}$$

Proposition 2.4.1. *Let $(\mathrm{z},\mathrm{z}^{(k)})$ be an optimal solution to (2.9) and $\omega \in \mathbb{N}$ such that*

$2\omega \geq \max\{\deg(f), \deg(g_j)\}$. *If* $\operatorname{rank}(L_{\mathrm{z}^{(k)}}(\mathrm{v}_\omega \mathrm{v}_\omega^\top)) = 1$ *for all* $k = 1, \ldots, \ell$, *then* $(2.9)^* = (2.8)^* = (\mathrm{POP})^*$ *and* $\mathrm{x}^* := (\mathrm{z}_\gamma)_{|\gamma|=1}$ *is an optimal solution to* (POP).

Note that in order to be able to compute the rank condition, the parameter s for the size of the SOS has to be chosen sufficiently large. The following proof was presented in [WLT17].

Proof. If $\operatorname{rank}(L_{\mathrm{z}^{(k)}}(\mathrm{v}_\omega \mathrm{v}_\omega^\top)) = 1$, $(\mathrm{z}_\gamma^{(k)})_{|\gamma|\leq 2\omega}$ is the vector of moments up to order 2ω of the Dirac measures $\delta_{\mathrm{x}^{(k)}}$ in the point $\mathrm{x}^{(k)} := (\mathrm{z}_\gamma^{(k)})_{|\gamma|=1}$ [Las10b, Theorem 4]. Let $p \in \mathbb{R}[I_k]_{2\omega}$. Then

$$L_{\mathrm{z}^{(k)}}(p) = \sum_\gamma \mathrm{p}_\gamma L_{\mathrm{z}^{(k)}}(\mathrm{x}^\gamma) = \sum_\gamma \mathrm{p}_\gamma (\mathrm{x}^{(k)})^\gamma = p(\mathrm{x}^{(k)}).$$

Let π_k be the projections induced by $\mathbf{x} \mapsto (\xi)_{\xi \in I_k}$. Then $\pi_k(\mathrm{x}^*) = \mathrm{x}^{(k)}$ for all k. Furthermore for all $p \in \mathbb{R}[I_k]_{2\omega}$ it holds that $p(\mathrm{x}^*) = p(\mathrm{x}^{(k)})$. In particular as $2\omega \geq \max\{\deg(f), \deg(g_j)\}$ and for particular choices of α, β and k, we obtain $g_j(\mathrm{x}^*) \geq 0$ for all j, i.e., x^* is feasible for (POP). Finally note that f can be written as sum of polynomials $f^{(k)} \in \mathbb{R}[I_k]_{2\omega}$. Consequently we have

$$(\mathrm{POP})^* \geq (2.8)^* \geq (2.9)^* = L_{\mathrm{z}}(f) = \sum L_{\mathrm{z}^{(k)}}(f^{(k)}) = \sum f^k(x^{(k)}) = f(\mathrm{x}^*) \geq (\mathrm{POP})^*,$$

where the first inequality is because (2.8) is a strengthening of (2.1), the second because of weak duality, and the last because $\mathrm{x}^* \in K$. □

Reducing problem size By looking at (2.8) more closely one may reduce the number of free variables and the number of constraints. It is likely that there are some indices $i \in \{1, \ldots, n\}$, that only appear in one of the I_k, say $i \in I_{k_i}$. Hence, for all $\gamma \in \Gamma$ such that $\gamma_i \neq 0$ the second equality constraint in (2.8) reduces to $\mathrm{f}_\gamma^{k_i} = f_\gamma$. Consequently, there is a number of variables that can be fixed from the beginning. We do this in our implementation. However, in order to be able to certify optimality by Proposition 2.4.1 one needs to trace back these substitutions to recover the moment sequences $\mathrm{z}^{(k)}$ from the solution of the dual problem. Removing these fixed variables occasionally leads to equality constraints $0 = 0$ in the SDP. We remove those constraints for better conditioning.

Sparse Putinar

Our implementation of the sparse standard certificates (Sparse PUT in the sequel) is quite similar to the one of Sparse BSOS. For ease of notation let $g_0 := 1 \in \mathbb{R}[I_k]$ for all k, and define $d_j := \lfloor \frac{2r - \deg(g_j)}{2} \rfloor$. We implemented the following SDP

$$\begin{aligned} \sup_{t, X_j^{(k)}, f^{(k)}} \quad t \quad \text{s.t.} \quad & t \in \mathbb{R}, \quad \left(\sum_{k=1}^{\ell} f^{(k)}\right)_{\mathbf{0}} = \mathrm{f}_{\mathbf{0}} - t, \quad \left(\sum_{k=1}^{\ell} f^{(k)}\right)_{\gamma} = \mathrm{f}_\gamma, \quad \forall \gamma \in \Gamma \\ & \left(f^{(k)} - \sum_{j \in J_k} \left\langle X_j^{(k)}, \left(\mathrm{v}_{d_j} \mathrm{v}_{d_j}^\top g_j\right)\right\rangle\right)_\gamma = 0, \quad \forall \gamma \in \Gamma^{(k)} \\ & X_j^{(k)} \in \mathcal{S}_+^{\mathrm{s}(n_k, d_j)}, \quad \forall j \in J_k, \quad f^{(k)} \in \mathbb{R}[I_k], \quad \forall k = 1, \ldots, \ell, \end{aligned} \tag{2.10}$$

where we used the same notation as in (2.8). Similar to Sparse BSOS we can check a rank condition to state optimality. Before solving the SDP with SDPT3 we apply the same techniques as discussed for Sparse BSOS in order to get a better conditioned program.

2.4.2 Introductory Remarks on the Numeric Comparison

The aim of the experiments presented in the following is twofold. On the one hand we compare Sparse BSOS with BSOS on small and medium size problems (as BSOS cannot handle large scale problems) with different sparsity patterns. In particular we are interested in the following issues:

- When the size of overlaps between blocks of variables is fixed. How does the clique sizes n_k (depending on the sparsity pattern) influence the performance?
- How do various clique and overlap sizes for a fixed number of variables ($n = 90$) influence the performance?
- Does the finite convergence of the dense version occur systematically earlier than for the sparse version? (As it cannot occur later.)

On the other hand we compare Sparse BSOS with Sparse PUT on high degree small size and lower degree medium and large scale problems. This comparison requires some care because the feasible set for Sparse PUT is $K = \{\mathrm{x} : g_j(\mathrm{x}) \geq 0\}$ while for Sparse BSOS (and BSOS) it is $K = \{\mathrm{x} : 0 \leq g_j(\mathrm{x}) \leq 1\}$. Hence we code the information about the feasible set in the constraints $0 \leq g_j(\mathrm{x})$ and scale the constraint polynomials g_j to be less than 1 on the feasible set. In general one expects that if Sparse PUT gives a good result, Sparse BSOS will not do better. However we have identified at least three scenarii where Sparse BSOS can beat Sparse PUT (and does it at least in some examples). These scenarii are:

- The first possible Sparse PUT strengthening yields the optimal value of the polynomial optimization problem, and some degrees d_j of the SOS weights are potentially greater than 0. This happens, when the degree of the objective function is larger than $\deg(g_j) + 2$ for some j. Then setting the parameter $s = \deg(f)/2$, the first Sparse BSOS strengthenings ($r = 1, 2, \ldots$) are faster than Sparse PUT and may also reach the optimal value; this is illustrated in Table 2.5 and Table 2.6.
- The first possible Sparse PUT strengthening does not reach the optimal value of the POP and the second strengthening cannot be solved (because its size is too large and/or is too costly to implement). If the SOS weights in the first Sparse PUT strengthening are all of degree 0, then again setting the parameter $s = \deg(f)/2$, the first Sparse BSOS strengthening gives the same result and it is possible to obtain better bounds by going higher in the truncation order. In particular, this is the case for the important class of quadratic/quadratically constrained programs (that is when $\max\{\deg(f), \deg(g_j)\} \leq 2$); this is illustrated in Table 2.10.
- The first possible Sparse PUT strengthening cannot be solved. Then setting the parameter $k < \deg(f)/2$, the first steps of the Sparse BSOS hierarchy ($r = 1, 2, \ldots$) may be solvable and so provide lower bounds on the optimal value of the polynomial optimization problem whereas Sparse PUT cannot; this is illustrated in Table 2.11.

To summarize the above cases, in Sparse BSOS we take full advantage of the facts that (a) the constraints enter the certificate only with non-negative weights in contrast to SOS weights in Sparse PUT, (b) the size of the positive semidefinite variables is fixed and does not increase with the truncation order. In particular, while the minimal size of the largest positive semidefinite variable in Sparse PUT is determined by the polynomial data and strictly increases when augmenting in the hierarchy, in Sparse BSOS one can always set the parameter s to 1 which implies that the maximum size of the semidefinite matrices in the SDP (2.8) is always at most $O(n^*)$ where $n^* = \max_k n_k$, for all r. This is because by assumption, the g_j generate the algebra $\mathbb{R}[\mathrm{x}]$ and so a polynomial of arbitrary degree and positive on K, can be obtained as a positive linear combination of the $g_j^{\alpha_j}(1-g_j)^{\beta_j}$ (with no SOS involved). So even if $\max\{\deg(f), \deg(g_j)\} > 2$, the optimal value of (2.8) (with $s = 1$) is finite as soon as r is large enough, and so provides a non-trivial lower bound.

The results on numerical experiments described in the next sections are biased by the (limited) sample of examples that we have considered. Therefore they should be understood as partial indications rather than definite conclusions. The latter would require much more computational experiments.

All experiments were performed on an Intel Core i7-5600U CPU @ 2.60GHz × 4 with 16GB RAM. Scripts are executed in Matlab 8.5 (R2015b) 64bit on Ubuntu 14.04 LTS operating system. The SDP solver used is SDPT3-4.0 [TTT12].

The results are presented in tables below. They provide the following information:

- A pattern or problem code to identify the example.
- The truncation order r and the chosen parameter s for the positive semidefinite constraints.
- The maximal degree $d_{\max}$, appearing in the certificate.
- The numbers of non-negative variables (corresponding to $\lambda_{\alpha\beta}$), unrestricted (free) variables (corresponding to t and the coefficients of the $f^{(k)}$), and the number (and size) of the positive semidefinite variables (corresponding to the SOS).
- The number of (equality) constraints in the SDP.
- The (primal) solution of the SDP.
- The time in seconds, including the times to generate and solve the SDP as well as computing the optimality condition.
- The abbreviation rk stands for the rank of the moment matrices according to Proposition 2.4.1 and its equivalents for BSOS and Sparse PUT. In the case of Sparse BSOS and Sparse PUT, rk is the average rank of all moment matrices and can hence be a decimal number. When reporting an integer the rank is actually integer, i.e. if we write rk= 1, the rank is actually 1, if however we write 1.0 the rank is strictly bigger than 1.
- If the primal solution is written in bold, it was certified by the rank condition and coincides[5] with the global optimum of the POP.

[5]In this context we consider two numbers to be equal if there difference is less than 10^{-8}.

			BSOS			Sparse BSOS		
problem	(r,s)	$d_{\max}$	solution	rk	time	solution	rk	time
P4_2	(1,1)	2	**-5.7491e-01**	1	0.8s	**-5.7491e-01**	1	1.6s
P4_4	(1,2)	4	-6.5919e-01	7	0.3s	-6.5919e-01	3	0.4s
	(2,2)	8	**-4.3603e-01**	1	0.7s	**-4.3603e-01**	1	0.5s
P4_6	(1,3)	6	-6.2500e-02*	27	1.0s	-6.2500e-02	15	0.5s
	(2,3)	12	-6.0937e-02	7	0.7s	-6.0937e-02*	6	0.6s
	(3,3)	18	-6.0693e-02	4	2.6s	-6.0693e-02*	4	4.7s
P4_8	(1,4)	8	-9.3381e-02*	39	9.2s	-9.3355e-02	15	1.7s
	(2,4)	16	-8.5813e-02*	9	3.0s	-8.5813e-02	4	1.3s
	(3,4)	24	-8.5813e-02	4	4.3s	-8.5814e-02*	4	4.1s
P6_2	(1,1)	2	**-5.7491e-01**	1	0.2s	**-5.7491e-01**	1	0.3s
P6_4	(1,2)	4	-5.7716e-01	13	0.7s	-5.7716e-01	4	0.4s
	(2,2)	8	-5.7696e-01	4	4.4s	-5.7696e-01	3	0.7s
	(3,2)	12	-5.7696e-01	3	25.0s	-5.7765e-01*	3	16.6s
P6_6	(1,3)	6	-6.5972e-01*	35	6.6s	-6.5972e-01	7	2.7s
	(2,3)	12	-6.5972e-01*	32	21.5s	-6.5972e-01	4	4.4s
	(3,3)	18	**-4.1288e-01***	1	44.5s	**-4.1288e-01***	1	82.1s
P8_2	(1,1)	2	**-5.7491e-01**	1	0.2s	**-5.7491e-01**	1	0.3s
P8_4	(1,2)	4	-6.5946e-01	21	1.5s	-6.5946e-01	5	0.8s
	(2,2)	8	**-4.3603e-01***	1	17.7s	**-4.3603e-01***	1	2.7s
P10_2	(1,1)	2	**-5.7491e-01**	1	0.3s	**-5.7491e-01**	1	0.3s
P10_4	(1,2)	4	-6.5951e-01	31	5.5s	-6.5951e-01	6	2.2s
	(2,2)	8	**-4.3603e-01***	1	23.4s	**-4.3603e-01***	1	8.4s
P20_2	(1,1)	4	**-5.7492e-01***	1	0.8s	**-5.7491e-01**	1	0.4s

Table 2.1: Comparison BSOS vs. Sparse BSOS on non-sparse examples

- Primal solutions were marked with * when the solver stopped because *steps were too short*, the *maximum number of iterations* was achieved, or *lack of progress*. In these cases one has to consider the result carefully.

2.4.3 BSOS vs. Sparse BSOS

Dense small size examples

In Table 2.1 we compare the sparse and the dense version of BSOS on a set of examples introduced in [LTS17]. In the problem description the first number of the name indicates the number of variables, the second the degree of the problem. The examples are relatively small size, i.e. $n \leq 20$. The degree of the objective function and the constraints is between 2 and 8. As the test sample is from the dense version, no sparsity pattern is present and we pass the information $I_1 = \{1, \ldots, n\}$ and $J_1 = \{1, \ldots, m\}$ to Sparse BSOS. Consequently both hierarchies compute the same certificate. The only difference comes from the implementation of the equality constraints and the different handling of the positive semidefinite variable in SDPT3.

We see that Sparse BSOS is able to solve the same problems as BSOS. As the problems are dense, Sparse BSOS uses a trivial sparsity pattern and both certificates are the same.

Consequently the optimal value coincides unless one of the SDPs stopped because of numerical issues. In most cases Sparse BSOS is faster than BSOS proving that our implementation is efficient.

Sparse quadratic examples (medium and large scale)

For the remaining examples we consider sparsity patterns having some banded structure. The patterns are described by a vector $\mathbf{n} \in \mathbb{N}^\ell$ and a natural number $\mathbf{o}$. The vector $\mathbf{n}$ determines the size of the blocks I_k whereas $\mathbf{o}$ defines the number of overlapping variables between two consecutive blocks. More formally defining $c_1 := n_1$ and $c_k := c_{k-1} + n_k - \mathbf{o}$ we construct

$$I_k := \{c_k - n_k + 1, \ldots, c_k\}.$$

Note that the total number of variables in pattern I is c_ℓ. We call those sparsity pattern banded, because the RIP (Definition 2.1.2) is satisfied by

$$\left(I_{k+1} \cap \bigcup_{j=1}^{k} I_j \right) \subseteq I_k.$$

Informally for $\mathbf{n}$ we use notation like (7×5) instead of $(5,5,5,5,5,5,5)$ or $(2 \times 17, 13)$ instead of $(17, 17, 13)$ without any misunderstanding.

We analyse the impact of different sparsity pattern on instances of the following sparse quadratic optimization problem. Given a sparsity pattern $I = \{I_1, \ldots, I_\ell\}$ we consider

$$\min_{\mathrm{x}} \left\{ x^\top A x + b^\top x \ : \ 1 - \sum_{i \in I_\ell} x_i^d \geq 0, \quad \ell = 1, \ldots, p, \quad x_i \geq 0 \quad i = 1, \ldots, n, \right\}, \qquad \text{(QP)}$$

where b is a random vector and the symmetric matrix A is randomly generated according to I[6]. We verify that A has positive and negative eigenvalues to make sure, that our problem is non-convex. Depending on the choice of $d \in \{1, 2\}$ we call the constraints *linear* or *quadratic*, although in the latter case we still have the linear constraints $x_i \geq 0$. Note that the constraints $x_i \geq 0$ imply that $1 - \sum_{i \in I_\ell} x_i^d \leq 1$ and vice versa.

To compare the respective SDPs arising from BSOS and Sparse BSOS in Table 2.2 we fix $\mathbf{n} = (2 \times 50)$ and create different sparsity patterns by varying $\mathbf{o}$. From these patterns we generate instances of (QP) with $d = 2$. Choosing parameter $s = 1$ for the size of the SOS we compute and solve the first step $r = 1$ of BSOS and Sparse BSOS. As $\mathbf{n}$ is fixed for all examples the number of variables n grows when the overlap $\mathbf{o}$ decreases.

Both BSOS and Sparse BSOS are able to solve all instances of this problem and provide the same lower bounds. In contrast to the previous example the certificates and the corresponding SDP handed over to the solver are different: Consider the example with $\mathbf{o} = 40$ and $n = 60$ variables. As $s = 1$ the positive semidefinite variable in BSOS is of size $\binom{n+s}{s} = 61$ and grows with the number of variables. As the sparsity pattern in all examples consists of two blocks of 50 variables, Sparse BSOS always has 2 positive

[6] By this we means that a random value between -1 and 1 is assigned to an entry a_{ij} of A if and only if both i and j are contained in the same I_k for some k. Otherwise $a_{ij} = 0$. The values of b are randomly generated between -1 and 1, too.

$\mathbf{o}/n$		# n-neg. var.	# free var.	# positive semidefinite var. (size)	# cons.	sol.	time
40/60	BSOS	125	1	1(61)	1891	-1.1123e+01	13.8s
	Sp. BSOS	206	1723	2(51)	3513	-1.1123e+01	14.0s
30/70	BSOS	145	1	1(71)	2556	-1.2753e+01	24.3s
	Sp. BSOS	206	993	2(51)	3148	-1.2753e+01	11.2s
20/80	BSOS	165	1	1(81)	3321	-1.3376e+01	48.1s
	Sp. BSOS	206	463	2(51)	2883	-1.3376e+01	10.5s
10/90	BSOS	185	1	1(91)	4186	-1.5406e+01	73.6s
	Sp. BSOS	206	133	2(51)	2718	-1.5406e+01	9.3s
5/95	BSOS	195	1	1(96)	4656	-1.5665e+01	89.5s
	Sp. BSOS	206	43	2(51)	2673	-1.5665e+01	9.2s
1/99	BSOS	203	1	1(100)	5050	-1.5658e+01 *	152.2s
	Sp. BSOS	206	7	2(51)	2655	-1.5658e+01	10.8s

Table 2.2: QPI $\mathbf{n} = (50, 50)$, quadratic constraints: $d = 2$, maximal degree of the certificate $d_{\max} = 2$, time to compute the first step $r = 1$ with $s = 1$

semidefinite variables of size $\binom{n_k+s}{s} = 51$, independently of the total number of variables. The unrestricted variable in BSOS corresponds to the optimizing variable t. Sparse BSOS also has this optimizing variable. The other unrestricted variables correspond to the non-fixed coefficients of the polynomials $f^{(k)}$. This is easy to see in the case of $\mathbf{o} = 1$: The maximal degree of the certificate is $d_{\max} = 2$. Hence, the non-fixed coefficients are the coefficients of the monomials $1, x_{50}$ and x_{50}^2. Consequently after removing the unrestricted variables for the fix coefficients, we have 3 unrestricted variables for $f^{(1)}$ and 3 for $f^{(2)}$. Together with the optimizing variable, we end up with 7 unrestricted variables as presented in the table. The non-negative variables in Table 2.2 correspond to the $\lambda_{\alpha\beta}$ in the description of BSOS and Sparse BSOS. Note that the number of constraints for each block is $m_1 = m_2 = 51$ for Sparse BSOS and $m = n + 2$ for BSOS (the linear constraints plus the two quadratic constraints). Consequently there are $206 = \binom{2m_1+r}{r} + \binom{2m_2+r}{r}$ non-negative variables for Sparse BSOS and $\binom{2(n+2)+r}{r}$ non-negative variables for BSOS depending on the number of variables n. The number of constraints in BSOS corresponds to the number of point evaluations needed to guarantee the equality constraint in the BSOS formulation, i.e $\mathrm{s}(n, d_{\max}) = \binom{n+d_{\max}}{d_{\max}}$, and hence increases with n. For Sparse BSOS three equalities have to be considered. As they are implemented by comparing coefficients we expect $|\Gamma^{(1)}| + |\Gamma^{(2)}| + |\Gamma| = 2 \times \binom{50+2}{2} + (2 \times \binom{50+2}{2} - \binom{\mathbf{o}+2}{2})$ many constraints. The difference to the reported numbers in Table 2.2 comes from the fact that with every removed unrestricted variable, we also remove a constraint.

Summarizing, when the overlap $\mathbf{o}$ decreases Sparse BSOS benefits from having less unrestricted variables and constraints while the size of the positive semidefinite variables remains the same; the SDP becomes easier. In contrast to this in BSOS the size of the positive semidefinite variable and the number of non-negative variables and constraints increases; the SDP becomes harder. The solving time reported in the last column of the table reflects this nearly perfectly.

We employ Problem (QP) a second time. The aim of this example is to observe the influence of using different sparsity patterns that all are valid for the same optimization

	$\mathbf{n}$	# n-neg. var.	# free var.	# positive semidefinite var. (size)	# cons.	time
BSOS	none	20706	1	1(91)	4186	882.4s
Sp. BSOS	(90)	20706	2	1(91)	4187	49.0s
	$(50, 42)$	11001	13	1(51)/1(43)	2278	10.5s
	$(50, 26, 18)$	9072	24	1(51)/1(27)/1(19)	1905	7.8s
	$(50, 2 \times 18, 10)$	8439	35	1(51)/2(19)/1(11)	1788	6.7s
	$(50, 5 \times 10)$	7821	57	1(51)/5(11)	1682	6.6s
	$(2 \times 34, 26)$	7776	24	2(35)/1(27)	1649	5.3s
	$(3 \times 26, 18)$	6171	35	3(27)/1(19)	1340	3.2s
	$(34, 3 \times 18, 10)$	5862	46	1(35)/3(19)/1(11)	1287	3.4s
	$(2 \times 26, 2 \times 18, 10)$	5538	46	2(27)/2(19)/1(11)	1223	2.8s
	$(5 \times 18, 10)$	4581	57	5(19)/1(11)	1042	1.7s
	(11×10)	3036	112	11(11)	777	0.8s

Table 2.3: QPII $n = 90$, overlap 2, linear constraints: $d = 1$, $(r, s) = (2, 1)$, maximal degree of the certificate $d_{\max} = 2$, same optimal solution verified by rank one condition in all cases

problem. For Table 2.3 we create the sparsity pattern I with $\mathbf{n} = (11 \times 10)$ and $\mathbf{o} = 2$ and consider *one* instance of (QP) with $d = 1$. Building on this sparsity pattern we can construct coarser patterns by building unions of smaller cliques to create bigger ones. By a similar argument as presented in the proof of Theorem 2.3 we can guarantee the existence of a dual solution of Sparse BSOS by choosing $r \geq 2$. Again, we choose parameter $s = 1$ and compute the second BSOS and Sparse BSOS truncation $r = 2$.

The dense and the sparse hierarchy are able to solve this sparse problem and certify optimality by the rank one condition. We do not repeat the discussion on the number of variables and constraints in this second example. We only remark that the additional unrestricted variable and constraints in in Sparse BSOS for $\mathbf{n} = (90)$ compared to the BSOS case without sparsity pattern comes from the equality $\mathrm{f}_{\mathbf{0}}^1 = \mathrm{f}_{\mathbf{0}} - t$. Because t is variable $\mathrm{f}_{\mathbf{0}}^1$ is not fixed and hence cannot be removed like all the other coefficients in this case (cf. p 29).

The reason for the big difference of computing times between BSOS and Sparse BSOS is double. On the one hand side, searching for linearly dependent constraints in BSOS takes a lot of time. Generating the SDP with BSOS took over 70 seconds whereas the SDP for Sparse BSOS was generated in less than 5 seconds. The main reason however is hidden in the constraints. Indeed the constraints in Sparse BSOS are sparse whereas the constraints in BSOS are dense and therefore the SDP solver is much slower in the dense case.[7] With regard to the computing times for Sparse BSOS one can see that the size of the biggest positive semidefinite variable is a more important factor than the number of non-negative and free variables or the number of constraints.

We next use the quadratic problem (QP) a third time to investigate the range of Sparse BSOS on large scale examples in Table 2.4. In this sample we generate sparsity patterns with 400 to 1000 cliques of size 3 to 9 and small overlap between 1 and 3. As in the previous example we chose $d = 1$ and $r = 2$. The size of those examples is by far too large to run

[7] Regarding this example the implementation of equalities using sampling might look questionable. Its positive effect comes into play when considering larger positive semidefinite variables. In [LTS17] the authors were able to handle problems with positive semidefinite variables of size up to 861 ($n = 40$, $s = 2$) due to the special handling of the constraints associated to the positive semidefinite variable in the case of sampling. This is by far out of the range of what can be done by comparing coefficients.

n	o	n	# n-neg.var.	# unrest.var.	# positive semidefinite var.(size)	# cons.	rnk	time
100x4	1	301	6 600	497	100(5)	1 699	1	1.8s
400x4	1	1201	26 400	1 997	400(5)	6 799	1.03	11.8s
700x4	1	2102	46 200	3 497	700(5)	11 899	1.02	28.2s
1000x4	1	3001	66 000	4 997	1000(5)	16 999	1.03	49.1s
100x5	2	302	9 100	1 091	100(6)	2 596	1.06	2.4s
400x5	2	1202	36 400	4 391	400(6)	10 396	1.04	16.2s
700x5	2	2102	63 700	7 691	700(6)	18 196	1.10	35.1s
1000x5	2	3002	91 000	10 991	1000(6)	25 996	1.08	74.8s
50x8	2	302	9 500	541	50(9)	2 496	1	4.3s
200x8	2	1202	38 000	2 191	200(9)	9 996	1.08	14.9s
350x8	2	2102	66 500	3 841	350(9)	17 496	1.01	35.3s
500x8	2	3002	95 000	5 491	500(9)	24 996	1.02	68.7s
50x9	3	303	11 550	933	50(10)	3 192	1.08	3.2s
200x9	3	1203	46 200	3 783	200(10)	12 792	1.07	18.2s
350x9	3	2103	80 850	6 633	350(10)	22 392	1.06	44.6s
500x9	3	3003	115 500	9 483	500(10)	31 992	1.04	133.4s

Table 2.4: QPLS, linear constraints: $d = 1$, $(r, s) = (2, 1)$, maximal degree of the certificate $d_{\max} = 2$

BSOS and so we only display the results obtained by Sparse BSOS. They show that Sparse BSOS is able to compute lower bounds for sparse large scale problems in reasonable time.

Summarizing the computational results of this section, we saw that Sparse BSOS is competitive with the dense version on dense examples. On sparse examples the Sparse BSOS outperforms BSOS as it can use the additional information. The advantage becomes bigger, when the block size n_k is small with respect to the total number of variables n. In addition, the number of variables in the intersection of at least two blocks I_k influences the performance of Sparse BSOS. Although the certificates depend on the information, known about the sparsity pattern, we did not encounter that the value computed by BSOS or by Sparse BSOS with a coarse sparsity pattern, was better than the one computed with the actual pattern.

2.4.4 Sparse BSOS vs. Sparse PUT

As already mentioned, the sparse version Sparse PUT [Wak+06] of the standard hierarchy has been proved to be efficient in solving several large scale problems; see for instance its successful application to some Optimal Power Flow problems [MH15]. However there are a number of cases where only the first truncation of Sparse PUT can be implemented because the second one is too costly to implement. Also if $t := \deg(f) > 2$ then the first SDP truncation is already very expensive (or cannot be implemented) because some moment matrices of size $\binom{n^*+t}{t} \times \binom{n^*+t}{t}$ are constrained to be positive semidefinite (where $n^* := \max_\ell n_\ell$).

Test Problems from the Literature

In this section we present experiments on some test problems considered to be challenging in non-linear optimization. All test functions are SOS and share the global minimum 0. Hence, it would be possible to compute the minimum in the unconstrained case. However, if not using constraints both Sparse BSOS and Sparse PUT reduce to searching for SOS.

Hence, we restrict the problems to the set

$$K = \{\mathrm{x} \in \mathbb{R}^n \ : \ 1 - \sum_{\xi \in I_k} \xi \geq 0, \quad k = 1, \ldots, \ell; \quad \mathrm{x}_i \geq 0, \quad i = 1, \ldots, n\},$$

which depends on the sparsity pattern of the specific function. Consequently, the optimal values of the considered functions are strictly greater than zero, when the minimizer of the unconstrained problem is not in K. We consider the following test functions of degree 4:

- The *Chained Wood Function*:

$$f := \sum_{j \in H} \left(100(\mathrm{x}_{j+1} - \mathrm{x}_j^2)^2 + (1 - \mathrm{x}_j)^2 + 90(\mathrm{x}_{j+3} - \mathrm{x}_{j+2}^2)^2\right.$$

  where $H := \{2i - 1 \ : \ i = 1, \ldots, n/2 - 1\}$ and $n \equiv 0 \mod 4$. The sparsity pattern is given by $\mathbf{n} = (\ell \times 4)$ and $\mathbf{o} = 2$.

- The *Chained Singular Function*:

$$f := \sum_{j \in H} \left((\mathrm{x}_j + 10\mathrm{x}_{j+1})^2 + 5(\mathrm{x}_{j+2} - \mathrm{x}_{j+3})^2 + (\mathrm{x}_{j+1} - 2\mathrm{x}_{j+2})^4 + 10(\mathrm{x}_j - \mathrm{x}_{j+3})^4\right)$$

  where $H := \{2i - 1 \ : \ i = 1, \ldots, n/2 - 1\}$ and $n \equiv 0 \mod 4$. The sparsity pattern is given by $\mathbf{n} = (\ell \times 4)$ and $\mathbf{o} = 2$.

- The *Generalized Rosenbrock Function*:

$$f := \sum_{i=2}^{n} \left(100(\mathrm{x}_i - \mathrm{x}_{i-1}^2)^2 + (1 - \mathrm{x}_i)^2\right).$$

  The sparsity pattern is given by $\mathbf{n} = (\ell \times 2)$ and $\mathbf{o} = 1$.

Table 2.5 and Table 2.6: We solve the Chained Wood and the Chained Singular Function for $n = 500, \ldots, 1000$ with Sparse BSOS and Sparse PUT. For Sparse BSOS we fix $s = 2$ and compute the first and the second truncation. For Sparse PUT we only compute the first feasible truncation.

Both Sparse BSOS and Sparse PUT are able to find and certify the optimal value (up to numerical errors) for both functions. For BSOS we go up to $r = 2$ because the solver reported irregularities when solving the SDPs for $r = 1$. In these examples Sparse PUT is slower than Sparse BSOS because for every constraint Sparse PUT ($r = 1$) introduces a positive semidefinite variable of size 5×5 corresponding to an SOS of degree 2. Sparse BSOS ($r = 2$) only introduces a non-negative variable for all products and squares of constraints. As the number of non-negative variables is not too big, Sparse BSOS beats Sparse PUT.

At this point we noticed a rather strange phenomenon. For the first Sparse BSOS truncation $r = 1$, the SDP solver runs into numerical problems. Yet, from the definition of the Chained Functions we know that they are SOS of degree 4, which in principle Sparse BSOS is able to represent with $s = 2$ for any d. One possible explanation is that the solver may be influenced by the additional non-negative variables. However, we could reproduce the same behaviour when omitting the constraints and explicitly only searching for an SOS. This phenomena is not specific to our implementation or to SDPT3. It also

		Sparse BSOS			Sparse PUT		
ChainedWood	rel.	solution	rk	time	solution	rk	time
$n = 500$	$r = 1$	**3.8394e+03***	1	16.7s	**3.8394e+03**	1	16.7s
	$r = 2$	**3.8394e+03**	1	10.4s	-	-	-
$n = 600$	$r = 1$	**4.6104e+03***	1	20.6s	**4.6104e+03**	1	21.0s
	$r = 2$	**4.6104e+03**	1	13.2s	-	-	-
$n = 700$	$r = 1$	**5.3813e+03***	1	24.1s	**5.3813e+03**	1	26.0s
	$r = 2$	**5.3813e+03**	1	15.8s	-	-	-
$n = 800$	$r = 1$	**6.1523e+03***	1	27.3s	**6.1523e+03**	1	31.1s
	$r = 2$	**6.1523e+03**	1	19.4s	-	-	-
$n = 900$	$r = 1$	**6.9232e+03***	1	30.8s	**6.9232e+03**	1	36.5s
	$r = 2$	**6.9232e+03**	1	22.3s	-	-	-
$n = 1000$	$r = 1$	7.6942e+03*	3	28.6s	**7.6942e+03**	1	42.3s
	$r = 2$	**7.6942e+03**	1	26.1s	-	-	-

Table 2.5: Comparison Sparse BSOS ($s = 2$) and Sparse PUT on the Chained Wood Function

		Sparse BSOS			Sparse PUT		
ChainedSingular	rel.	solution	rk	time	solution	rk	time
$n = 500$	$r = 1$	-1.4485e-02*	1.0	19.6s	**-2.0271e-10**	1	22.6s
	$r = 2$	**-9.7833e-10**	1	17.8s	-	-	-
$n = 600$	$r = 1$	-2.7372e-03*	1.0	40.1s	**-1.9613e-10**	1	27.8s
	$r = 2$	**-1.2640e-09**	1	21.4s	-	-	-
$n = 700$	$r = 1$	-1.7548e-03*	1.0	41.6s	**-2.4628e-10**	1	34.1s
	$r = 2$	**-1.7613e-09**	1	25.3s	-	-	-
$n = 800$	$r = 1$	-1.9438e-03*	1.0	58.9s	**-2.3398e-10**	1	41.0s
	$r = 2$	**2.1935e-09**	1	29.0s	-	-	-
$n = 900$	$r = 1$	-1.8924e-02*	1.0	43.5s	**-3.5871e-10**	1	47.3s
	$r = 2$	**-2.6072e-09**	1	33.5s	-	-	-
$n = 1000$	$r = 1$	-4.4914e-02*	1.0	35.5s	**-1.7329e-10**	1	54.9s
	$r = 2$	**-9.3508e-10**	1	39.5s	-	-	-

Table 2.6: Comparison Sparse BSOS ($s = 2$) and Sparse PUT on the Chained Singular Function

Generalized Rosenbrock	rel.	Sparse BSOS solution	rk	time	Sparse PUT solution	rk	time
$n = 100$	$r = 1$	4.8496e+01	2.0	2.8s	**9.6197e+01**	1	2.4s
	$r = 2$	9.6145e+01	2.2	1.7s	-	-	-
	$r = 3$	9.6184e+01	2.1	4.5s	-	-	-
	$r = 4$	9.6195e+01*	1.3	18.2s	-	-	-
$n = 200$	$r = 1$	9.7496e+01	2.0	2.7s	**1.9519e+02**	1	4.6s
	$r = 2$	1.9512e+02	2.1	3.2s	-	-	-
	$r = 3$	1.9516e+02	2.1	5.9s	-	-	-
	$r = 4$	1.9395e+02*	3	565.6s	-	-	-
$n = 300$	$r = 1$	1.4650e+02	2.0	3.9s	**2.9418e+02**	1	6.9s
	$r = 2$	2.9410e+02	2.1	4.8s	-	-	-
	$r = 3$	2.9414e+02	2.0	9.3s	-	-	-
	$r = 4$	2.9176e+02*	3	695.6s	-	-	-
$n = 400$	$r = 1$	1.9550e+02	2.0	5.2s	**3.9317e+02**	1	9.4s
	$r = 2$	3.9308e+02	2.1	6.5s	-	-	-
	$r = 3$	3.9312e+02*	2.0	27.6s	-	-	-
	$r = 4$	-8.1403e+05*	3	801.9s	-	-	-
$n = 500$	$r = 1$	2.4450e+02	2.0	6.8s	**4.9216e+02**	1	12.4s
	$r = 2$	4.9206e+02	2.0	8.3s	-	-	-
	$r = 3$	4.9210e+02*	2.0	31.8s	-	-	-
	$r = 4$	4.9215e+02*	3	1144.6s	-	-	-
$n = 600$	$r = 1$	2.9350e+02	2.0	8.1s	**5.9115e+02**	1	15.5s
	$r = 2$	5.9104e+02	2.0	10.4s	-	-	-
	$r = 3$	5.9108e+02*	2.0	22.3s	-	-	-
	$r = 4$	5.9114e+02*	1.0	111.6s	-	-	-

Table 2.7: Comparison Sparse BSOS ($s = 2$) and Sparse PUT on the Generalized Rosenbrock Function

occurs when searching for an SOS representation of the Chained Wood Function ($n = 4$) with Gloptipoly3 [HLL09b] using SeDuMi1.3 [Stu99] and with Yalmip [Löf04] using Mosek [MOS17].

In Table 2.7 we show results from solving the Generalized Rosenbrock Function for $n = 100, \ldots, 600$ with Sparse BSOS and Sparse PUT. As in the previous examples for Sparse BSOS we fix $s = 2$ and compute the first and the second truncation. For Sparse PUT again we only compute the first feasible truncation.

As in the previous examples we find a unique minimizer with Sparse PUT, certified by the rank condition. This time Sparse BSOS is not able to find the optimum even when going up to the truncation $r = 4$. However, even though Sparse BSOS does not obtain the optimal value at an early truncation, its optimal value at step $r = 2$ is already in the right order of magnitude and can be computed faster than the optimal value provided by Sparse PUT.

Note that for $n = 100$ the truncation $r = 1$ is slower than the truncation $r = 2$. The same happens in Table 2.9 for some values of n. This is an issue related to our configuration and could not be reproduced when using another SDP solver or another operating system,

Discrete		Sparse BSOS			Sparse PUT		
Boundary	rel.	solution	rk	time	solution	rk	time
$n = 15$	$r = 1$	**9.8705e-04**	1	1.4s	**9.8705e-04**	1	2.0s
$n = 20$	$r = 1$	**4.4893e-04**	1	1.8s	**4.4893e-04**	1	2.6s
$n = 25$	$r = 1$	**2.4060e-04**	1	2.3s	**2.4060e-04**	1	3.3s
$n = 30$	$r = 1$	1.4358e-04	2.1	2.7s	**1.4359e-04**	1	3.9s
	$r = 2$	1.4359e-04	1.1	3.2s	-	-	-
	$r = 3$	**1.4358e-04**	1	4.0s	-	-	-
$n = 35$	$r = 1$	9.2438e-05	4	3.9s	**9.2441e-05**	1	4.5s
	$r = 2$	9.2438e-05*	3.8	4.3s	-	-	-
	$r = 3$	**9.2439e-05**	1	4.8s	-	-	-

Table 2.8: Comparison Sparse BSOS ($s = 3$) and Sparse PUT on the Discrete Boundary Value Function

respectively.

To close this section we consider the following test functions of degree 6:

- The *Discrete Boundary Value Function*:

$$f := \sum_{i=1}^{n} (2\mathrm{x}_i - \mathrm{x}_{i-1} - \mathrm{x}_{i+1} + \frac{1}{2}h^2(\mathrm{x}_i + ih + 1)^3)^2,$$

  where $h := \frac{1}{n+1}$, $\mathrm{x}_0 := 9 =: \mathrm{x}_{n+1}$. The sparsity pattern is $\mathbf{n} = (\ell \times 3)$ and $\mathbf{o} = 2$.

- The *Broyden Banded Function*:

$$f := \sum_{i=1}^{n} \left(\mathrm{x}_i(2 + 10\mathrm{x}_i^2) + 1 - \sum_{j \in H_i} (1 + \mathrm{x}_j)\mathrm{x}_j \right)^2,$$

  where $H_i := \{j : j \neq i, \max(1, i-5) \leq j \leq \min(n, i+1)\}$. The sparsity pattern is $\mathbf{n} = (\ell \times 7)$ and $\mathbf{o} = 6$.

Table 2.8: Solving the Discrete Boundary Value Function for $n = 15, \ldots, 35$. For Sparse BSOS we choose $k = 3$ and compute the first truncations until we get the certified optimal value. The first possible truncation for Sparse PUT involves SOS of degree up to 6.

As for the Chained Functions in Table 2.5 and Table 2.6 both Sparse BSOS and Sparse PUT are able to certify the minimum in all cases. When Sparse BSOS succeeds to do so at an early step of the truncation it is faster and if not then it takes approximately the same time. Note that Sparse BSOS and Sparse PUT certify different optimal values in the case $n = 30$ and $n = 35$ and that in contrast to the theory, the series of lower bounds computed by Sparse BSOS for $n = 35$ is not monotonously increasing. The difference however is less than 10^{-8} and can be considered to be zero "numerically".

Table 2.9: Solving the Broyden Banded Function for $n = 7, \ldots, 15$. Again for Sparse BSOS we let $s = 3$ and compute the first truncations.

As for the Generalized Rosenbrock Function (Table 2.7) Sparse PUT is able to find and certify the optimal solution at the first possible truncation step whereas Sparse BSOS

		Sparse BSOS			Sparse PUT		
BroydenBanded	rel.	solution	rk	time	solution	rk	time
$n = 7$	$r = 1$	2.1371	2	11.5s	**3.4233**	1	15.2s
	$r = 2$	2.7522	2	9.2s	-	-	-
	$r = 3$	3.1161	2	11.0s	-	-	-
$n = 9$	$r = 1$	2.2171	3	77.0s	**3.3941**	1	105.6s
	$r = 2$	2.8313	3	72.7s	-	-	-
	$r = 3$	3.1354	3	87.1s	-	-	-
$n = 11$	$r = 1$	2.2968	3	160.3s	**3.3924**	1	215.1s
	$r = 2$	3.0108	2	159.7s	-	-	-
	$r = 3$	3.2638	4	190.7s	-	-	-
$n = 13$	$r = 1$	2.3353	3	282.9s	**3.4120**	1	357.7s
	$r = 2$	3.0963	2.9	301.6s	-	-	-
	$r = 3$	3.3268	4.4	367.5s	-	-	-
$n = 15$	$r = 1$	2.3555	3	445.2s	**3.4243**	1	545.2s
	$r = 2$	3.1514	3.8	466.3s	-	-	-
	$r = 3$	3.3617	3.8	509.1s	-	-	-

Table 2.9: Comparison Sparse BSOS ($s = 3$) and Sparse PUT on the Broyden Banded Function

does not succeed to do so in the first three steps. However up to truncation order $r = 3$ Sparse BSOS is faster than Sparse PUT and hence provides lower bounds for the objective function in less time and reasonably close to the optimal value.

Random Medium Scale Quadratic and Quartic Test Problems

So far we have compared Sparse BSOS and Sparse PUT on examples where the first possible truncation of Sparse PUT is exact. We now present examples where this is not the case or the first truncation cannot even be computed because it is already too large. To this end we choose sparsity patterns with 40 to 80 variables in blocks of 10 to 40 and fixed overlap $\mathbf{o} = 5$. Note that the crucial parameter for the sparse hierarchies is not so much the total number of variables but rather the maximum clique size of the sparsity pattern.

We change Problem (QP) slightly to generate the following sample of problems. Given a sparsity pattern $I = \{I_1, \ldots, I_\ell\}$ we now consider:

$$\min_{\mathrm{x}} \left\{ \sum_{i=1}^{n} a_i \mathrm{x}_i^4 + \mathrm{x}^\top A \mathrm{x} + b^\top \mathrm{x} \ : \ 1 - \mathrm{x}_i^2 \geq 0, \quad \mathrm{x}_i \geq 0 \quad i = 1, \ldots, n, \right\}, \tag{QP'}$$

where b is a random vector and the symmetric matrix A is randomly generate according to I. We verify that A has positive and negative eigenvalues to make sure that our problem is non-convex again. When $a = 0 \in \mathbb{R}^n$ we refer to (QP') as a *quadratic* problem. When mentioning the *quartic* problem (QP'), it means that we chose a randomly in $[-1, 1]^n$.

Table 2.10: We consider the quadratic instance of Problem (QP') and compute, whenever possible, the first two truncations of BSOS, Sparse BSOS and Sparse PUT.

We were able to solve the first truncation for all algorithms. However, as BSOS cannot

(QP') Quadratic	n	rel.	BSOS solution	BSOS rk	BSOS time	Sparse BSOS solution	Sparse BSOS rk	Sparse BSOS time	Sparse PUT solution	Sparse PUT rk	Sparse PUT time
(7×10)	40	$r = 1$	-1.2496e+02	4	3.5s	-1.2496e+02	4.0	1.8s	-1.2496e+02	4.0	0.8 s
		$r = 2$	-4.4436e+01*	15	574.4s	-4.4457e+01	3.9	8.8s	**4.4326e+01**	1	45.8s
$(2 \times 20, 10)$	40	$r = 1$	-1.6197e+02	5	3.4s	-1.6197e+02	5.0	0.7s	-1.6197e+02	5.0	0.7 s
		$r = 2$	-5.9412e+01*	16	592.9s	-5.9447e+01	9.3	6.5s	-	-	-
(15×10)	80	$r = 1$	-2.7552e+02*	8	71.7s	-2.7557e+02	4.0	0.8s	-2.7557e+02	4.0	0.8 s
		$r = 2$	-	-	-	-1.0837e+02	2.4	2.7s	**1.0825e+02**	1	143.7s
(5×20)	80	$r = 1$	-3.4782e+02*	5	86.6s	-3.4782e+02	5.0	1.6s	-3.4782e+02	5.0	1.6s
		$r = 2$	-	-	-	-1.2536e+02	9.0	26.7s	-	-	-
$(2 \times 40, 10)$	80	$r = 1$	-4.8983e+02*	5	69.3s	-4.8988e+02	5.0	3.9s	-4.8988e+02	5.0	4.2 s
		$r = 2$	-	-	-	**-1.8564e+02**	1	66.8s	-	-	-
(23×10)	120	$r = 1$	-3.8765e+02*	8	849.2s	-3.8765e+02	4.0	0.9s	-3.8765e+02	4.0	1.0 s
		$r = 2$	-	-	-	-1.5884e+02	2.2	4.6s	-	-	-
$(7 \times 20, 15)$	120	$r = 1$	-5.4921e+02*	5	772.0s	-5.4920e+02	4.6	3.6s	-5.4920e+02	4.6	3.8 s
		$r = 2$	-	-	-	-2.3846e+02*	6.5	85.2s	-	-	-
$(3 \times 40, 15)$	120	$r = 1$	-7.1721e+02 *	6	581.0s	-7.1720e+02	6.0	9.8s	-7.1720e+02	6.0	11.1s
		$r = 2$	-	-	-	-2.3079e+02	12.2	143.8s	-	-	-

Table 2.10: Comparison BSOS($k = 1$), Sparse BSOS($k = 1$), and Sparse PUT on Quadratic Problem (QP'), $\mathbf{o} = 5$

benefit from the sparsity structure, it runs into numerical problems, in particular for the examples with $n = 120$ variables. Note that in [LTS17] problems comparable to this one have been solved by BSOS. There the authors were able to solve the second truncation for a quadratic problem with 100 variables. Indeed we were able to compute the second truncation for some examples with 80 variables. However, this depends strongly on which constraints BSOS deletes before handing over the system to the solver. We decided not to search for examples where we can compute the second truncation.

When the solver does not run into numerical problems all solutions of the first truncations coincide. This is because of the degree of the constraints and the objective function, the first truncations are more or less the same. In fact the SOS weights of the constraints in Sparse BSOS are all of degree 0, i.e. they are non-negative scalar variables (and are implemented as such). BSOS and Sparse BSOS use twice as many constraints because they not only consider the constraints $g_j(x) \geq 0$ but also the constraints $g_j(x) \leq 1$. This explains why Sparse PUT is faster for the first truncation.

In a number of cases the second truncation of BSOS and Sparse PUT could not be implemented because the positive semidefinite variables become to big for the solver. Sparse PUT is able to solve the second truncation in two cases where the block size is 10 and we could actually certify optimality by the rank condition. The examples with same block size but $n = 120$ variables could not be solved because the solver runs out of memory. The same happened for examples with larger block size.

Only Sparse BSOS was able to compute the second truncation in all cases. Of course this second truncation is weaker than the second truncation of Sparse PUT and Sparse BSOS could only certify optimality in one case. However, when comparing the results with the certified values from Sparse PUT, we see that they are actually quite close and much less time was spent to compute them. In all cases where Sparse PUT could not solve the second truncation, Sparse BSOS could provide a lower bound that is much better that than the one provided by the the first truncation of Sparse PUT.

Table 2.11: The quartic version ($a \neq 0$) of Problem (QP'). As the degree of the objective function is now 4, the first feasible truncation of Sparse PUT involves SOS of degree 4. Choosing $s = 1$ as fixed parameter, the respective truncations with $r = 1$ of BSOS and Sparse BSOS are not feasible, therefore one computes the second and the third truncation, whenever possible.

Sparse PUT could solve the first feasible truncation only for the patterns (7×10) and (15×10) in which case the optimal solution was attained and certified. With $s = 1$, i.e., with SOS weights of degree at most 2 which is less than the degree $(= 4)$ of the objective function one still may solve higher truncations with the Sparse BSOS hierarchy. Note that the values of the second truncation of Sparse BSOS for the patterns (7×10) and (15×10) are not so far from being optimal, which suggests that for the other test problems they are not so bad either. In any case Sparse BSOS is able to provide lower bounds for larger block size, whereas the other hierarchies already overpassed their limit.

2.4.5 Conclusions from the Numerical Experiments

The experiments have shown that the Sparse BSOS hierarchy has its place between BSOS ans Sparse Putinar. In particular for problems with large clique size Sparse BSOS is able to provide bounds for small choices of the parameter s, influencing the degree of the SOS,

(QP')			BSOS			Sparse BSOS			Sparse PUT		
Quartic	n	rel.	solution	rk	time	solution	rk	time	solution	rk	time
(7×10)	40	$r = 1$	inf	-	-	inf	-	-	**-5.2576e+01**	1	49.6s
		$r = 2$	-5.4736e+01*	16	466.3s	-5.4802e+01	6.0	8.9s	-	-	-
		$r = 3$	-	-	-	-5.3047e+01*	3.1	319.0s	-	-	-
$(2 \times 20, 10)$	40	$r = 1$	inf	-	-	inf	-	-	out of memory	-	-
		$r = 2$	-6.4481e+01*	17	606.3s	-6.4528e+01	8.7	5.9s	-	-	-
(15×10)	80	$r = 1$	inf	-	-	inf	-	-	**-1.1897e+02**	1	150.1s
		$r = 2$	-	-	-	-1.2037e+02	3.8	3.2s	-	-	-
		$r = 3$	-	-	-	-1.1950e+02	1.7	37.9s	-	-	-
(5×20)	80	$r = 1$	inf	-	-	inf	-	-	out of memory	-	-
		$r = 2$	-	-	-	-1.2725e+02	8.4	23.5s	-	-	-
$(2 \times 40, 10)$	80	$r = 1$	inf	-	-	inf	-	-	out of memory	-	-
		$r = 2$	-	-	-	-1.9476e+02	12.3	78.7s	-	-	-
(23×10)	120	$r = 1$	inf	-	-	inf	-	-	out of memory	-	-
		$r = 2$	-	-	-	-1.6791e+02	4.0	8.7s	-	-	-
		$r = 3$	-	-	-	-1.6375e+02	1.2	103.0s	-	-	-
$(7 \times 20, 15)$	120	$r = 1$	inf	-	-	inf	-	-	out of memory	-	-
		$r = 2$	-	-	-	2.1229e+02	9.5	54.7s	-	-	-
$(3 \times 40, 15)$	120	$r = 1$	inf	-	-	inf	-	-	out of memory	-	-
		$r = 2$	-	-	-	-2.3994e+02	11.8	137.1s	-	-	-

Table 2.11: Comparison BSOS($k = 1$), Sparse BSOS($k = 1$), and Sparse PUT on Quartic Problem (QP'), $\mathbf{o} = 5$

while Sparse Putinar runs out of memory even in the first possible truncation. A second advantage of Sparse BSOS over Sparse Putinar is that the steps between two truncations are somehow *closer* in terms of computational effort. This can be exploited in multi-ordered hierarchies such as proposed in [JM18].

2.5 Nearly Sparse Polynomial Optimization

In the previous sections we have shown how sparsity can be exploited in order to attack problems that are far out of range of the non-sparse versions. In this section we demonstrate that even in the case when (POP) is not completely sparse one may recover sparsity by introducing a reasonable number of additional variables. Consider the polynomial optimization problem

$$\inf_{\mathrm{x}} f(\mathrm{x}) \quad \text{s.t.}\, g_1(\mathrm{x}) \geq 0, \ldots, g_m(\mathrm{x}) \geq 0, h(\mathrm{x}) \geq 0, \tag{POP-ns}$$

where $f, g_1, \ldots, g_m, h \in \mathbb{R}[\mathbf{x}]$. We say that (POP-ns) is *nearly sparse* if $(f, g_1, \ldots, g_m)$ respect a sparsity pattern $I_1, \ldots, I_\ell$ and $h \in \sum \mathbb{R}[I_k]$ (compare condition on f in Definition 2.1.1). Indeed, in some applications a natural structured sparsity is apparent but violated by a few constraints. For instance, one constraint might state that $\sum_{i=1}^n \mathrm{x}_i \leq M$ for some $M > 0$. Although the problem is nearly sparse, the sparsity-adapted hierarchies [Wak+06] and [WLT17] cannot be applied because the constraint $h(\mathrm{x}) = M - \sum_{i=1}^n \mathrm{x}_i \geq 0$ links all variables. Hence, solving and even relaxing a nearly sparse polynomial optimization problem is limited to problems with a rather small number of variables as if the problem was dense.

We propose a systematic procedure to replace the nearly sparse optimization problem (POP-ns) by an *equivalent sparse problem*. The price to pay is (i) the introduction of $\ell - 1$ additional variables, and (ii) the introduction of ℓ new constraints in lieu of the sparsity-violating constraint $h(\mathrm{x}) \geq 0$. In contrast to the nearly sparse problem, its equivalent can be attacked by the previously discussed sparsity-adapted hierarchies [Wak+06] and [WLT17] and hence may be solved much more efficiently. In addition, if there are several constraints violating the sparsity pattern, this scheme can be repeated and eventually results in an equivalent sparse problem.

Illustrative Example

As the construction of the equivalent sparsity pattern is quite technical we send ahead an illustrative example in a more restricted set up. Assume $I_1, \ldots, I_\ell$ is a sparsity pattern for f and $K = \{\mathrm{x} \in \mathbb{R}^n : 1 - x_i^2 \geq 0, i = 1, \ldots, n\}$ (compare Definition 2.1.1) such that $\bigcup_{t<k} I_t \cap I_k \subseteq I_{k-1}$ (compare Definition 2.1.2). When we add the constraint $h(\mathrm{x}) := \sum_{i=1}^n \mathrm{x}_i = 0$ the problem $\inf\{f(\mathrm{x}) : \mathrm{x} \in K, h(\mathrm{x}) = 0\}$ is nearly sparse but not sparse any more. The basic idea described in the sequel is to introduce slack variables in the sum h and separate $h(\mathrm{x}) = 0$ into several constraints. More precisely let $\mathbf{y}_1, \ldots, \mathbf{y}_{\ell-1}$ be additional variables and define a sparsity pattern $\tilde{I}_1 = I_1 \cup \{\mathbf{y}_1\}$, $\tilde{I}_k = I_k \cup \{\mathbf{y}_{k-1}, \mathbf{y}_k\}$, and $\tilde{I}_\ell = I_\ell \cup \{\mathbf{y}_{\ell-1}\}$. Then $\tilde{I}_1, \ldots, \tilde{I}_\ell$ is a sparsity pattern for f and K, where we understand

$\mathbb{R}[\mathbf{x}] \subseteq \mathbb{R}[\mathbf{x}, \mathbf{y}]$. In particular $\tilde{I}_1, \dots, \tilde{I}_\ell$ is a sparsity pattern for

$$\begin{aligned} \inf_{\mathrm{x}\in K, \mathrm{y}\in\mathbb{R}^{\ell-1}} f(\mathrm{x}) \quad \text{s.t.} \sum_{\xi\in I_1} \xi - \mathrm{y}_1 = 0 \\ \mathrm{y}_{k-1} + \sum_{\xi\in I_k} \xi - \mathrm{y}_k = 0, \quad k = 2, \dots, \ell-1 \\ \mathrm{y}_{\ell-1} + \sum_{\xi\in I_\ell} \xi = 0 \end{aligned}$$

In this case it is easy that the system of equations above is equivalent to the original constraint $h(\mathrm{x}) = 0$.

2.5.1 Construction of the Equivalent Sparse Problem

In the following we describe a procedure to construct a sparse polynomial optimization problem from the nearly sparse problem (POP-ns). In particular, if $I_1, \dots, I_\ell$ has the Running Intersection Property, the resulting sparsity pattern for the sparse problem will respect the RIP, too.

New Sparsity Pattern Consider the sparsity pattern $I_1, \dots, I_\ell$ for $(f, g_1, \dots, g_m)$. For $k > 2$ define the index sets

$$\mathcal{I}_k := \{t_0 : \bigcup_{t<k} I_t \cap I_k \subseteq I_{t_0}\}. \tag{2.11}$$

Since the RIP holds, the sets $\mathcal{I}_k$ are non-empty and for all $t \in \mathcal{I}_k$ it holds that $t < k$. For a fixed choice function $\phi : \{2, \dots, p\} \to \bigcup_{k=2}^{\ell} \mathcal{I}_k$ define the set $N := N(\phi) := \{\alpha \in \mathbb{N}^2 : \alpha_1 = \phi(\alpha_2)\}$.

We introduce $\ell - 1$ new variables $\mathbf{y} = (\mathbf{y}_\alpha)_{\alpha\in N}$ and construct a new pattern $\tilde{I}$ by

$$\tilde{I}_k := I_k \cup \{\mathbf{y}_\alpha \ : \ \alpha_1 = k \text{ or } \alpha_2 = k\}, \quad k = 1, \dots, \ell.$$

By construction $\tilde{I}$ respects the RIP. To see this, fix $k \in \{2, \dots, \ell\}$. Then

$$\begin{aligned} \bigcup_{t<k} \tilde{I}_t \cap \tilde{I}_k &= \left(\bigcup_{t<k} I_t \cup \{\mathbf{y}_\alpha : \alpha_1 < k \text{ or } \alpha_2 < k\}\right) \cap (I_k \cup \{\mathbf{y}_\alpha : \alpha_1 = k \text{ or } \alpha_2 = k\}) \\ &= \left(\bigcup_{t<k} I_t \cap I_k\right) \cup \{\mathbf{y}_\alpha : (\alpha_1 < k \text{ or } \alpha_2 < k) \text{ and } (\alpha_1 = k \text{ or } \alpha_2 = k)\}. \end{aligned}$$

Now note, that $\alpha_1 < \alpha_2$ for all $\alpha \in N$. Hence, the case $(\alpha_1 < k \text{ or } \alpha_2 < k)$ and $\alpha_1 = k$ does not contribute to the above equation. As further $\alpha_1 = \phi(\alpha_2)$ for all $\alpha \in N$, we can reduce $\{\mathbf{y}_\alpha : (\alpha_1 < k \text{ or } \alpha_2 < k) \text{ and } (\alpha_1 = kl \text{ or } \alpha_2 = k)\} = \{\mathbf{y}_{(\phi(k),k)}\}$ and hence,

$$\bigcup_{t<k} \tilde{I}_t \cap \tilde{I}_k = \left(\bigcup_{t<k} I_t \cap I_k\right) \cup \left\{\mathbf{y}_{(\phi(k),k)}\right\} \subseteq I_{\phi(k)} \cup \{\mathbf{y}_{(\phi(k),k)}\} \subseteq \tilde{I}_{\phi(k)}.$$

As $\phi(k) < k$, $\tilde{I}$ respects the RIP (see Definition 2.1.2). Since I is a sparsity pattern for $(f, g_1, \dots, g_m)$ and $I_k \subseteq \tilde{I}_k$ for all k, $\tilde{I}$ is also a sparsity pattern for $(f, g_1, \dots, g_m)$.

New inequalities We are now constructing a system of inequalities, equivalent to $h(\mathrm{x}) \geq 0$ and respecting the new pattern $\tilde{I}$.

Remember $h = \sum_{k=1}^{\ell} h_k$ with $h_k \in \mathbb{R}[I_k]$. Define the following polynomials over the extended set of variables $\mathbb{R}[\mathbf{x}, \mathbf{y}] \supseteq \mathbb{R}[\mathbf{x}]$:

$$\tilde{h}_k := \sum_{\substack{\alpha \in N \\ \alpha_2 = k}} \mathbf{y}_\alpha + h_k - \sum_{\substack{\alpha \in N \\ \alpha_1 = k}} \mathbf{y}_\alpha, \quad k = 1, \dots, \ell. \tag{2.12}$$

Note that the first sum is empty for $k = 1$, and equals $\mathrm{y}_{(\phi(k),k)}$ for $k > 1$. Furthermore, since $\tilde{h}_k \in \mathbb{R}[\tilde{I}_k]$, the polynomial optimization problem (POP-s) respects the new sparsity pattern $\tilde{I}$

$$\inf_{\mathrm{x},\mathrm{y}} f(\mathrm{x}) \quad \text{s.t.} \quad g_1(\mathrm{x}) \geq 0, \dots, g_m(\mathrm{x}) \geq 0, \tilde{h}_1(\mathrm{x}, \mathrm{y}) \geq 0, \dots, \tilde{h}_\ell(\mathrm{x}, \mathrm{y}) \geq 0, \tag{POP-s}$$

Equivalence It remains to show that (POP-s) is equivalent to (POP-ns). To that end, let $(\mathrm{x}^*, \mathrm{y}^*)$ be feasible for (POP-s). Then, since by construction $\sum \tilde{h}_\ell = h$, x^* is feasible for (POP-ns). Now, let x^* be feasible for (POP-ns). Define recursively

$$\mathrm{y}^*_{(\phi(k),k)} := -h_k(\mathrm{x}^*) + \sum_{\substack{\alpha \in N \\ \alpha_1 = k}} \mathrm{y}^*_\alpha.$$

To see that this definition terminates, notice that the sum on the right hand side is empty for $k = \ell$ and that for $k < \ell$ the sum is either empty, or only involves y^*_α with $\alpha_2 > k$, which hence can already be defined. Now it is easy to see that this choice of $(\mathrm{x}^*, \mathrm{y}^*)$ is feasible for (POP-s), since $\tilde{h}_1(\mathrm{x}^*, \mathrm{y}^*) =$

$$h_1(\mathrm{x}^*) - \sum_{\substack{\alpha \in N \\ \alpha_1 = 1}} \mathrm{y}^*_\alpha = h_1(\mathrm{x}^*) - \sum_{\substack{\ell \in \mathbb{N} \\ (1,\ell) \in N}} \left(-h_\ell(\mathrm{x}^*) + \sum_{\substack{\alpha \in N \\ \alpha_1 = \ell}} \mathrm{y}^*_\alpha \right) = \sum_{k=1}^{\ell} h_k(\mathrm{x}^*) = h(\mathrm{x}^*) \geq 0$$

and $\tilde{h}_k(\mathrm{x}^*, \mathrm{y}^*) = \mathrm{y}^*_{(\phi(k),k)} + h_k(\mathrm{x}^*) - \sum_{\substack{\alpha \in N \\ \alpha_1 = k}} \mathrm{y}^*_\alpha = 0$ for $k = 2, \dots, \ell$, by definition of y^*.

When constructing the sparsity pattern $\tilde{I}$ from the original pattern I, the procedure allows for some freedom when defining the choice function ϕ. There are two obvious choices, $\phi(k) := \max \mathcal{I}_k$ or $\phi(k) := \min \mathcal{I}_k$. However, depending on the sparsity pattern, there might be better ways to choose. It is not clear that one of the two choices is always the better one. Consider the following example

Example 2.5.1 (Choice function)**.** The maximal block size of a sparsity pattern plays a significant role in sparse polynomial optimization. As the choice function determinates the number of variables added to each block of the sparsity pattern, it is advisable to choose in a way that minimizes the maximal block size of the new sparsity pattern. Consider for

example the sparsity pattern

$$\begin{aligned} I_1 &:= \{\mathbf{x}_1, \mathbf{x}_2\}, \\ I_2 &:= \{\mathbf{x}_1, \mathbf{x}_3\},\ I_3 := \{\mathbf{x}_1, \mathbf{x}_4\}, \\ I_4 &:= \{\mathbf{x}_2, \mathbf{x}_5\},\ I_5 := \{\mathbf{x}_2, \mathbf{x}_6\}. \end{aligned}$$

The sets $\mathcal{I}_k$ defined in (2.11) are $\mathcal{I}_2 = \{1\}, \mathcal{I}_3 = \{1, 2\}, \mathcal{I}_4 = \{1\}, \mathcal{I}_5 = \{1, 4\}$. If we use min as a choice function in the procedure proposed in the proof of Section 2.5.1, we end up with the pattern

$$\begin{aligned} I_1^{\min} &:= \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}_{12}, \mathbf{y}_{13}, \mathbf{y}_{14}, \mathbf{y}_{15}\}, \\ I_2^{\min} &:= \{\mathbf{x}_1, \mathbf{x}_3, \mathbf{y}_{12}\},\ I_3^{\min} := \{\mathbf{x}_1, \mathbf{x}_4, \mathbf{y}_{13}\}, \\ I_4^{\min} &:= \{\mathbf{x}_2, \mathbf{x}_5, \mathbf{y}_{14}\},\ I_5^{\min} := \{\mathbf{x}_2, \mathbf{x}_6, \mathbf{y}_{15}\}. \end{aligned}$$

Note that all additional variables have been added to I_1. In contrast to that, the max function leads to the pattern

$$\begin{aligned} I_1^{\max} &:= \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}_{12}, \mathbf{y}_{14}\}, \\ I_2^{\max} &:= \{\mathbf{x}_1, \mathbf{x}_3, \mathbf{y}_{12}, \mathbf{y}_{23}\},\ I_3^{\max} := \{\mathbf{x}_1, \mathbf{x}_4, \mathbf{y}_{23}\}, \\ I_4^{\max} &:= \{\mathbf{x}_2, \mathbf{x}_5, \mathbf{y}_{14}, \mathbf{y}_{45}\},\ I_5^{\max} := \{\mathbf{x}_2, \mathbf{x}_6, \mathbf{y}_{45}\}. \end{aligned}$$

Here, the additional variables have been added to different groups. As the maximal clique size determines the performance of the solver, in the given example the second pattern is preferable to the first one. However, it is not true that max always leads to better patterns than min. To see this, one can replace $I_2, \ldots, I_4$ by

$$\begin{aligned} I_2 &:= \{\mathbf{x}_1, \mathbf{x}_3, \ldots, \mathbf{x}_6\}, I_3 := \{\mathbf{x}_1, \mathbf{x}_7, \ldots, \mathbf{x}_{10}\}, \\ I_4 &:= \{\mathbf{x}_2, \mathbf{x}_{11}, \ldots, \mathbf{x}_{14}\}, I_5 := \{\mathbf{x}_2, \mathbf{x}_{15}, \ldots, \mathbf{x}_{18}\}. \end{aligned}$$

Now min leads to a sparsity pattern with a maximal clique size of 6 ($I_1^{\min}$, $I_2^{\min}$, and $I_4^{\min}$) while max leads to cliques of size 7 ($I_2^{\max}$ and $I_4^{\max}$).

We conclude this section about nearly sparse problems by treating two examples in more detail. The first example stems from an *optimal control* problem which has been discritized in order to obtain a polynomial optimization problem. A constraint on the total energy consumption however prevents a purely sparse formulation. The second example is from graph theory. Computing the so-called *stability number* can be achieved by computing the solution to a nearly sparse polynomial optimization problem. We reformulate this problem as a sparse problem and solve truncations to it using the hierarchy described in [WLT17].

2.5.2 Optimal Control with Limited Total Energy Consumption

Let $T > 0$, $X \subseteq \mathbb{R}^n$ and $U \subseteq \mathbb{R}^m$ be compact semialgebraic sets, $\xi_0, \xi_T \in X$ and $L \in \mathbb{R}[t, \mathbf{x}, \mathbf{u}]$, $F \in \mathbb{R}[t, \mathbf{x}, \mathbf{u}]^n$. For simplicity we assume $n = m = 1$. The example generalizes directly to arbitrary n and m. Let $\mathrm{u}_{\max} > 0$. We consider the polynomial optimal control

with limited total energy consumption

$$\begin{aligned} &\inf_u \int_0^\top L(t, x(t), u(t))\,\mathbf{dt} \\ &\text{s.t. } \dot{x}(t) = F(t, x(t), u(t)) \quad \text{for almost all } t \in [0, T] \\ &\quad (x(t), u(t)) \in X \times U, \quad \text{for all } t \in [0, T], \\ &\quad x(0) = \xi_0,\ x(T) = \xi_T, \\ &\quad \int_0^\top u(t)^2\,\mathbf{dt} \le u_{\max}. \end{aligned} \tag{2.13}$$

We are searching for a continuous function $x : [0, T] \to X$ and a function $u : [0, T] \to U$ that respect the dynamic $\dot{x}(t) = F(t, x(t), u(t))$ and minimize the integral over the cost function L, where the total energy of the control is limited by $u_{\max} > 0$. One way to approach (2.13) is by discretization. Here we approximate x and u by piecewise linear and piecewise constant functions respectively, supported on a grid on $[0, T]$. Choosing $N + 1 \in \mathbb{N}$ discretization points $t_0 := 0$ and $t_{i+1} := t_i + \Delta t$ with $\Delta t := \frac{T}{N}$ we replace the unknown functions x and u by vectors $\mathrm{x} \in X^{N+1}$, and $\mathrm{u} \in U^N$ and the dynamic by $x_{i+1} = x_i + f(t_i, x_i, u_i)\Delta t$. For large N, problem (2.13) then is approximated by the problem

$$\begin{aligned} &\inf_{\mathrm{u}} \sum_{i=0}^{N-1} l(t_i, \mathrm{x}_i, \mathrm{u}_i)\Delta t \\ &\text{s.t. } \mathrm{x}_{i+1} = \mathrm{x}_i + f(t_i, \mathrm{x}_i, \mathrm{u}_i)\Delta t \\ &\quad \mathrm{x} \in X^{N+1}, \mathrm{u} \in U^N \\ &\quad \mathrm{x}_0 = \xi_0,\ \mathrm{x}_N = \xi_T, \\ &\quad \sum_{i=1}^{N} \mathrm{u}_i^2 \Delta t \le \mathrm{u}_{\max}. \end{aligned} \tag{2.14}$$

Of course this discretization by the Euler method is quite simple and has its own drawbacks. However this example extends easily to more sophisticated discretization schemes.

If one ignores the last constraint on the total energy consumption in (2.14), one immediately identifies the sparsity pattern $I_0, \dots, I_{N-1}$ where

$$I_k := \{\mathbf{x}_k, \mathbf{x}_{k+1}, u_k\}, \quad k = 0, \dots, N-1.$$

The sparsity pattern satisfies the RIP Definition 2.1.2. As the violating constraint is of the form $\sum_k h_k$ with $h_k \in \mathbb{R}[I_k]$, we can apply the procedure from Section 2.5.1. A close look on the pattern reveals $\mathcal{I}_k = \{k-1\}$ for all $k = 1, \dots, N-1$. Hence, there is only one choice function $\phi(k) := k - 1$. We introduce new variables $\mathbf{y} = (\mathbf{y}_k)_{1 \le k < N}$ and define

$$\begin{aligned} \tilde{I}_0 &:= \{\mathbf{x}_0, \mathbf{x}_1, \mathbf{u}_0, \mathbf{y}_1\}, \\ \tilde{I}_k &:= \{\mathbf{x}_k, \mathbf{x}_{k+1}, \mathbf{u}_k, \mathbf{y}_k, \mathbf{y}_{k+1}\} \text{ for } k = 1, \dots, N-2 \text{ and} \\ \tilde{I}_{N-1} &:= \{\mathbf{x}_{N-1}, \mathbf{x}_N, \mathbf{u}_{N-1}, \mathbf{y}_{N-1}\} \end{aligned}$$

With the help of the newly introduced variables, the last constraint in (2.14) can be replaced

by

$$\begin{aligned} \mathrm{u}_{\max} - \mathrm{u}_1^2 \Delta t - \mathrm{y}_1 &\geq 0, \\ \mathrm{y}_k - \mathrm{u}_k^2 \Delta t - \mathrm{y}_{k+1} &\geq 0 \text{ for } k = 1, \ldots, N-2 \text{ and} \\ \mathrm{y}_{N-1} - \mathrm{u}_N^2 \Delta t &\geq 0. \end{aligned}$$

Hence, we can handle the sparsification of Problem (2.14) with the sparsity pattern $\tilde{I}$. Note that we included $N-1$ additional variables $\{\mathbf{y}_1, \ldots, \mathbf{y}_{N-1}\}$ to the problem. However, the size n_ℓ of each block of the sparsity pattern remains small (≤ 5), which is essential for solving non-convex sparse polynomial optimization problems.

2.5.3 Stability Number of a Sparse Graph

We consider a graph $G := (E, V)$, where $E = \{1, \ldots, n\}$ for some natural number n, and V is a (finite) set of (distinct) pairs $(i, j) \in E^2$. For simplicity, we limit ourselves to undirected graphs (i.e. $(i, j) \in V \Rightarrow (j, i) \in V$) that don't have any self loops (i.e. $(i, i) \notin V$). A subset S of E is called an independent set (or maximal clique) of G, if $V \cap S^2 = \emptyset$. A maximal clique of largest cardinality is called a maximum clique. The graph's independence number (also stability number) $\alpha(G)$ is defined as the cardinality of a maximum clique. It can be computed by the following quadratic optimization problem [KP02]

$$\begin{aligned} \frac{1}{\alpha(G)} = \min_{\mathrm{x}} \; & \sum_{(i,j) \in V} \mathrm{x}_i \mathrm{x}_j + \sum_{i=1}^{n} \mathrm{x}_i^2 \\ \text{s.t. } & 0 \leq \mathrm{x}_i \leq 1, \quad \sum_{i=1}^{n} \mathrm{x}_i = 1. \end{aligned} \tag{2.15}$$

If $l^* > 0$ is a lower bound to (2.15), $(l^*)^{-1}$ is an upper bound for $\alpha(G)$. As the stability number is an integer, the result can be rounded down to the next integer while maintaining an upper bound. As this procedure is sensitive to numerical errors (especially when the upper bound is tight), we define the more conservative upper bound

$$l' := \lfloor (l^* - 10^{-6})^{-1} \rfloor$$

for $l^* > 10^{-6}$ and $l' = n$ for $l^* \leq 10^{-6}$. From now on we assume that the graph respects some RIP sparsity pattern $(I_1, \ldots, I_\ell)$. Then problem (2.15) is of the form (POP-ns) and nearly sparse.

Numeric Computation

In the following we apply our implementation [Wei17] of [WLT17] to problem (2.15) and its sparsification. When computing the truncations to the original (nearly sparse and hence dense) problem, we transmit the trivial sparsity pattern $I := \{1, \ldots, n\}$ to [Wei17]. In contrast to Section 2.4, we are using the SDP solver SeDuMi1.32 [Stu99] instead of SDPT3, because we encountered that SDPT3 is violating the constraints too aggressively. Note that the lower bounds computed by truncation are only reliable if the SDP is solved accurately. In this section we consider the SDP to be solved if the solver doesn't report any numerical

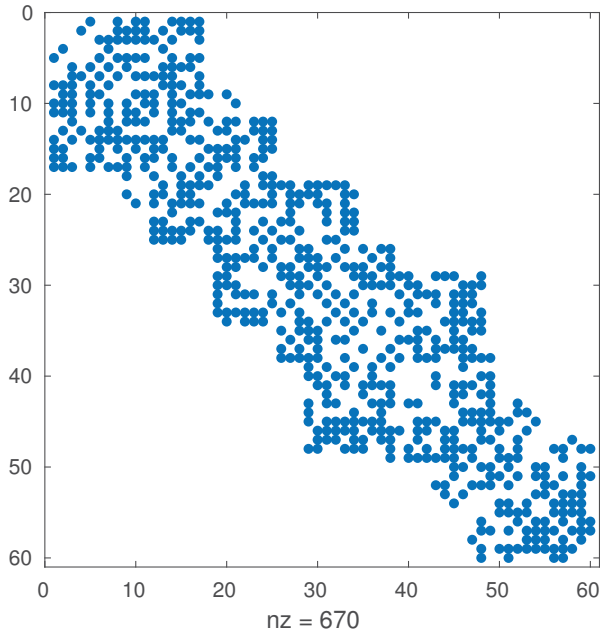

Figure 2.2: The set V of a randomly generated sparse graph with $n = 60$, $c = [10, 20]$, $o = [5, 10]$, and $p = 50\%$..

problem or stops with precision at least 10^{-6} and duality gap less than 10^{-6}. We then perform the rounding technique discussed above in order to compute the stability number.

Graph Generation To study the problem of computing stability numbers via (2.15), we randomly generate sparsity patterns respecting the RIP by $\left(\bigcup_{j<k} I_j\right) \cap I_k \subseteq I_{k-1}$ for every $k = 2, \ldots, \ell$. Once such a sparsity pattern has been generated, we randomly generate a graph with such a fixed pattern. The sparsity pattern is generated from :

- the total number n of nodes,
- an interval c of integers defining the possible number of nodes in each block[8], and
- an interval o defining the possible overlaps of two consecutive blocks of the sparsity pattern.

The actual block size and the overlap for each block are chosen randomly in the intervals c and o, respectively. A random graph is generated from such a randomly generated sparsity pattern by connecting two nodes in a block with a a certain probability p. In Fig. 2.2, we plot a matrix representing the vertices V of such a graph given parameters $n = 60$, $c = [10, 20]$, $o = [5, 10]$, and $p = 50\%$.

Fixed number of variables To compare the nearly sparse version (POP-ns) and its sparsification (POP-s), we consider an example where the total number of nodes n is small

[8]In the following we use the term "blocks" when referring to the cliques of the sparsity pattern in order to distinguish from the cliques of the constructed graph.

block size	10	15	20	25	30
overlap [min, max]	[5, 7]	[7, 9]	[8, 10]	[10, 12]	[13, 15]
Time (nearly sparse) sec.	38.7	41.7	41.6	40.8	40.1
Time(sparsification) sec.	2.6	11.6	6.0	6.2	10.6

Table 2.12: 50 instances of Problem (2.15) with $n = 60$ and increasing clique size and overlaps.

enough such that second truncation $d = 2$ with SOS parameter $k = 1$ of [Wei17] can be computed. All examples were run 50 times and the average computing time spent by the SDP solver is reported. The optimal values of both truncations are not reported as (i) they coincided for both truncations and (ii) we have no mean to verify that they are optimal for (2.15).[9] However the interest of this experiment is to show the speed up when using the sparsification of the problem instead of the initial nearly sparse formulation, rather than to demonstrate accuracy (which will is the subject of another series of examples).

In Table 2.12 we report results for graphs of fixed total size $n = 60$ where the sizes of cliques and overlaps increases. The nodes in each block are connected with a probability of 0.5. SeDuMi could solve all examples to the desired precision. The SDP solver took about 40 seconds to solve the second truncation $d = 2$ with SOS parameter $k = 1$ of the original problem (2.15) for all examples considered. This is coherent as for the nearly sparse problems only the total number of variables is decisive for the complexity of the SDP. In contrast to that, the number of variables in each clique is the important parameter for the performance of the sparsified problem. This number is approximately the number reported in the first row of Table 2.12 *plus* two $\mathbf{y}$- variables. Solving the equivalent sparse problem with same truncation order, but exploiting sparsity leads to a significant speed up, especially when the clique size is small. Note that solving the sparse version for $c = 15$ is the slowest in this sample. This behaviour could not be reproduced using the SDP solver SDPT3 on the same problems (perhaps it is due to the choice of default parameters for SeDuMi).

Fixed block size In Table 2.13 we consider graphs with fixed block size c and fixed overlap o while increasing the total number of variables n. The SDPs become more and more time consuming to solve as the number of nodes increases. Nonetheless almost all instances of the sparsification could be solved by SeDuMi. For the dense (nearly sparse) version the ratio of failures becomes significant at $n = 80$.[10] As in the previous set of examples the respective bounds of the stability number computed by the sparse and the nearly sparse version coincide whenever both problems could be solved. The computing time needed to solve the nearly sparse (dense) problems increases fast with the *number of nodes* of the graph. This is in contrast with the time needed to solve the sparsified problem which increases rather linearly with the *number of blocks*.

Known stability number In the following series of examples we construct random graphs in such a way that we know a lower bound on the stability number in order to have

[9] We could check for optimality using Proposition 2.4.1 if the maximal clique was unique in this examples. However generically this is not the case here.

[10] SeDuMi is capable to solve instances of this problem for $n = 100$ nodes to optimality. However this exceeds a timeout of 1000 seconds and hence was not included when testing 50 problems.

total number of nodes	20	40	60	80	100	120	140
average number of blocks	5.1	10.0	16.0	21.7	27.15	32.0	38.5
Solved (nearly sparse)	100%	100%	98%	90%	–	–	–
Solved (sparsification)	100%	100%	100%	100%	98%	100%	100%
Time(nearly sparse) sec.	0.4	3.7	36.7	231.0	–	–	–
Time(sparsification) sec.	0.4	1.6	2.0	2.3	2.7	2.8	3.7

Table 2.13: 50 instances of Problem (2.15) with $c = [6, 10]$ and $o = [2, 5]$ for an increasing number of variables

a criterion for the tightness of the upper bound computed by the truncations.

As seen in Proposition 2.4.1 there is a criterion to detect global optimality: If a certain rank condition on the dual SDP (2.9) is satisfied then the optimal value of the truncation is the global optimum of the original problem. If the original polynomial optimization problem has a unique optimizer (in particular if there is a unique maximal clique in the graph) then the rank condition is likely to be satisfied in practice.

As in the previous examples we create a random sparsity pattern. Now two nodes of the same block are connected with probability $p = 99\%$. Then a subset of nodes is chosen randomly (uniformly distributed over all blocks of the sparsity pattern) and all edges within this clique are deleted. The size of this clique is a lower bound for the stability number of the graph. Indeed it is very likely that the clique constructed this way is a maximum clique. However, the probability that the maximum clique is unique decreases with the number of nodes of the graph, as the number of disconnections in the rest of the graph then increases and hence alternative maximum cliques might pop up.

For fixed block sizes $c = [7, 13]$ and overlaps $o = [3, 6]$, we solved the problem for $n = 100, \ldots, 500$ nodes. As in the previous examples we solved the second step of the sparse BSOS hierarchy with SOS parameter $k = 1$. For each n we ran 100 examples. In Table 2.14 we report the average computing time needed. In addition, we report the percentage of examples where the known lower bound was attained by the truncation (and hence the stability number could be found). We also show the percentage of examples where optimality could be certified by the rank condition. All SDPs could be solved to the desired accuracy. In all cases the upper bound computed by the SDP coincides with lower bound known in advance. The number of problems where optimality could be certified by the rank condition is very high for $n = 100, 200$. For larger values of n the percentage of certified solutions decreases in accordance with a higher probability of having multiple maximum cliques. The computing time needed to solve the SDP increases linearly in the number of blocks of the sparsity pattern as already seen in previous examples.

2.6 Conclusion

In this chapter we have discussed how sparse certificates for non-negativity can be used in polynomial optimization. In particular we generalized the notion of the dual of the Shor relaxation from quadratic problems to generic polynomial optimization problems (SH-D) and provided a generalized sparse version (SSH-D). Whence the former was already known to be exact in the case of SOS-convex problems, we provided the corresponding results for

number of nodes	100	200	300	400	500
average number of blocks	20.6	42.9	64.4	86.8	107.5
Solved	100%	100%	100%	100%	100%
Stability number reached	100%	100%	100%	100%	100%
Optimality certified	100%	99%	87%	56%	12%
Time(sparse) sec.	1.8	3.3	5.1	6.7	7.7

Table 2.14: 100 instances of Problem (2.15) with known lower bound for the stability number. $c = [7, 13]$, $o = [3, 6]$, increasing number of variables

the sparse version (Theorem 2.3, Corollary 2.2.2).

Many certificates for non-negativity are derived from theorems of positivity. We hence discussed some of these results from real algebraic geometry and how to obtain computable tractable certificates from them. In Theorem 2.7 we proved that every Positivstellensatz valid for a compact set K has a sparse version. We used this result to prove convergence of what we called the Sparse BSOS hierarchy (Corollary 2.3.4). By pointing out its relation to the Shor relaxation we could show finite convergence of this hierarchy for SOS-convex problems.

A major contribution of the work presented in this chapter is the implementation of Sparse BSOS and Sparse Putinar [Wei17]. We therefore presented a numerical evaluation of this code in Section 2.4. In particular we illustrated the advantages of Sparse BSOS over its dense version and compared the new hierarchy to the sparse standard hierarchy based on Putinar. It turned out that Sparse BSOS is favourable when the clique size of the sparsity pattern is rather big ($n_k \approx 20 - 40$).

In the final section of this chapter we argued that sparsity can be exploited even if the original problem slightly violates the sparsity conditions (Definition 2.1.1). We proposed a systematic procedure to reformulate nearly sparse problems as sparse ones and demonstrated the method on an example from optimal control. Finally, as a numeric application, we used the procedure in conjunction with our implementation of Sparse BSOS in order to compute stability numbers of sparse graphs.

Perspectives

The research presented in this section motivates to investigate further numerically computational certificates for non-negativity. The BSOS certificate Theorem 2.6 is a crossover of Putinar Theorem 1.2 and Krivine Theorem 2.5. Similarly one could construct converging hierarchies by replacing the non-negative coefficients in Theorem 2.5 by SOS of fixed degree. Theorem 2.7 provides convergence of sparse hierarchies constructed in this manner. Additionally, one could define sparse hierarchies using different certificates for each clique.

While from a theoretical point of view these questions do not seem very interesting, they might be of importance in practise. Where limited computational time and power is prohibitive for going far in the hierarchies, *crossover certificates* might provided the missing accuracy. First steps in this direction have been presented in [MH15].

In this thesis we only talk about theorems of positivity of real polynomials. It would be interesting to see how our results can be used with certificates involving *complex numbers* such as [JM18].

These questions are in particular interesting for practical applications. It would be interesting to provide a *software package* being able to compute at least all the certificates discussed above and possibly mix them.

Chapter 3

Representation and Approximation of Chance Constraints

The ability to describe engineering problems by mathematical models that can be solved numerically is key to answer questions of decision and control. Recently there has been an increasing demand for models taking into account that some of the parameters determining the system are uncertain. A prominent example is the Optimal Power Flow problem (OPF), which will be presented in Section 3.2. With an increasing amount of power generated by wind and solar, the power grid is strongly influenced by deviations from the weather forecast. These fluctuations cannot be omitted in the model. As an answer to this demand, *chance constraints* describe mathematically constraints influenced by uncertainty. The idea is that a constraint on the decision variable $\mathbf{x}$ might change depending on the realization of some uncertain parameter or noise $\boldsymbol{\omega}$. More precisely, let $g \in \mathbb{R}[\mathbf{x}, \boldsymbol{\omega}]$. Then for each x, a chance constraint constrains the probability of the set of realizations ω satisfying $g(\mathrm{x}, \omega) \geq 0$ from below, i.e.,

$$\mathbb{P}(\{\omega \in \Omega : g(\mathrm{x}, \omega) \geq 0\}) \geq 1 - \varepsilon.$$

In general such constraints are numerically intractable (see Section 3.1.4). In this chapter we propose methods to approximate a chance constrained feasible set by an explicit basic semialgebraic set. In particular, we use results from [Las17b] to take into account *uncertainty in physical network flow problems*. Then, we extend the results from [Las17b; Las17a] further to approximate *distributionally robust chance constraints*, i.e., chance constraints where the distribution of the noise itself is uncertain but belongs to a certain family of distributions.

All work presented in this chapter can be seen as a follow up to the seminal paper [HLS09]. The results have been extended to *computing probabilities* with respect to some probability measures in [Las17a]. In this paper Lasserre also proposes a technique called *Stokes constraints* which significantly accelerates the convergence when approximating volumes via the generalized moment approach. Finally in [Las17b] the framework was used to provide *approximations to chance constraints*.

The outline of this chapter is the following. In Section 3.1 we summarize the state of the art for volume approximation and chance constraints. Then we apply these results to physical network flow problems and in particular to uncertainty in the OPF problem in Section 3.2. Motivated by this application we extend the method of [Las17b] to distributionally robust probability approximations in Section 3.3.

3.1 Preliminaries

We start by reviewing existing methods which build the basis for our contributions in the subsequent sections.

3.1.1 Volume of Semialgebraic Compact Sets

In this section we present how the volume or the probability of a basic semialgebraic compact set can be approximated as closely as desired via the moment-SOS approach, i.e., by reformulating the problem as a GMP and then solving semidefinite programs in order to approximate this value as explained in Chapter 1. In general this is a hard problem as mentioned in [BF87; DF88]. To the best of our knowledge, all other existing methods rely on convexity of the set in question (e.g.,[AS86; DFK91]) and/or are based on randomized strategies [BF08]. In contrast, the approach proposed in [HLS09] is based on a natural formulation of the volume as solution to a linear problem on measures which can be written as a GMP. In addition its dual problem has a very intuitive interpretation in the space of continuous functions.

Linear Problems on Measures

As usual denote by $K = \{\mathrm{x} \in \mathbb{R}^n : g_1(\mathrm{x}) \geq 0, \ldots, g_m(\mathrm{x}) \geq 0\}$ our generic basic semialgebraic compact set. Assume further that K has non-empty interior and is contained in some set $B \subseteq \mathbb{R}^n$ such that we know all moments of a reference (Borel-) measure $\mu \in \mathcal{M}_+(B)$. For technical reasons we will also assume that $B \setminus K$ has non-empty interior. We are interested in the volume (if $\mu = \lambda_B$ is the Lebesgue measure on B normalized to have mass 1) or the probability (if μ is any probability measure) of K with respect to μ denoted by $\mathrm{vol}(K)$ or $\mu(K)$. The following relation between (signed) measures is essential for the reformulation of the volume as optimal value to a linear problem on measures.

Definition 3.1.1. Let $\phi, \mu \in \mathcal{M}(B)$. Then we say that ϕ is *dominated by* μ, if

$$\mu - \phi \in \mathcal{M}_+(B).$$

Equivalently we can ask that for all Borel sets $A \in \mathcal{B}(B)$ it holds that $\phi(A) \leq \mu(A)$. We therefore use the intuitive notation $\phi \leq \mu$ to express that ϕ is dominated by μ.

Note that domination is a partial ordering on $\mathcal{M}(B)$. With help of this definition it is quite direct to reformulate $\mathrm{vol}(K)$ as the optimal value of two linear problems that are in duality to each other[1]:

$$\begin{aligned} \sup_{\phi} \ & \int_K \mathbf{d}\phi \\ \text{s.t.} \ & \phi \leq \mu, \\ & \phi \in \mathcal{M}_+(K) \end{aligned} \tag{3.1}$$

$$\begin{aligned} \inf_{f} \ & \int_B f \, \mathbf{d}\mu \\ \text{s.t.} \ & f \geq 1 \text{ on } K, \\ & f \in C_+(B), \end{aligned} \tag{3.2}$$

where $C_+(B)$ denotes the cone of non-negative continuous functions on B. It is easy to see that the optimal measure in (3.1) is the restriction of μ to K, denoted by $\mu|_K$ yielding the

[1] For a compact set $\mathcal{X}$ we understand the duality of $\mathcal{M}(\mathcal{X})$ and $C(\mathcal{X})$ with respect to the weak topologies $\sigma(\mathcal{M}(\mathcal{X}), C(\mathcal{X}))$ and $\sigma(C(\mathcal{X}), \mathcal{M}(\mathcal{X}))$ induced by the duality pairing $\langle \mu, f \rangle := \int_{\mathcal{X}} f \, \mathbf{d}\mu$, respectively.

objective value $\int_B f \,\mathbf{d}\mu|_K = \mu(K)$: Indeed this value is optimal as, due to the domination constraint, the optimal value cannot be larger than $\int_K \mathbf{d}\mu = \mu(K)$. In the dual, we see that any feasible function $f \in C_+(B)$ has to be a point wise over-approximation of the indicator function $\mathbf{1}_K$ on B. An optimizing sequence $(f^\ell)_{\ell\in\mathbb{N}} \subseteq C_+(B)$ of (3.2) converges in L^1 norm to $\mathbf{1}_K$. As the latter is discontinuous the infimum in (3.2) is not attained. However, as shown in [HLS09], if K and $B \setminus K$ have non-empty interior there is no duality gap, i.e., the optimal values of (3.1) and (3.2) coincide.

Moment Formulation and Polynomial Approximations

In order to apply the moment-SOS hierarchy to approximate $\mu(K)$ we need to formulate the constraint in (3.1) as constraints on moments. Let therefore $\nu \in \mathcal{M}_+(B)$ be such that $\mu - \phi = \nu$ and assume B to be compact. Then, as consequence of the Theorem of Stone-Weierstrass, the equality of measures can be expressed by a system of equations on moments, more precisely by

$$\int_K \mathbf{x}^\alpha \,\mathbf{d}\phi + \int_B \mathbf{x}^\alpha \,\mathbf{d}\nu = \int_B \mathbf{x}^\alpha \,\mathbf{d}\mu, \quad \forall \alpha \in \mathbb{N}^n. \tag{3.3}$$

Replacing the domination constraint in (3.1) by (3.3) we end up with GMP (3.4) whose dual (3.5) now reads like (3.2), where $f \in C_+(B)$ is replaced by $p \in \mathbb{R}[\mathbf{x}]$, $p \geq 0$ on B:

$$\begin{aligned} &\sup_{\phi,\nu} \langle \phi, \mathbf{1}\rangle \\ &\text{s.t. } \langle \phi, \mathbf{x}^\alpha\rangle + \langle \nu, \mathbf{x}^\alpha\rangle = \langle \mu, \mathbf{x}^\alpha\rangle, \forall \alpha \in \mathbb{N}^n, \\ &\quad \phi \in \mathcal{M}_+(K), \\ &\quad \nu \in \mathcal{M}_+(B) \end{aligned} \tag{3.4}$$

$$\begin{aligned} &\inf_{p} \langle \mu, p\rangle \\ &\text{s.t. } p \geq 1 \text{ on } K, \\ &\quad p \geq 0 \text{ on } B, \\ &\quad p \in \mathbb{R}[\mathbf{x}]. \end{aligned} \tag{3.5}$$

Now we can apply the hierarchy (P_d) of SDP relaxations to (3.4) and SDP strengthenings to (3.5) in order to compute approximations of $\mu(K)$. Note that due to the constraints in (3.4) all moments of ϕ and ν are bounded by the moments of μ. This implies that for each d the moment matrices corresponding to ϕ and ν are bounded by the moment matrix of μ and hence all moment relaxations have optimal solutions $(\mathrm{z}_\phi^{(d)}, \mathrm{z}_\nu^{(d)})$. In presence of the constraint $N - \|\mathrm{x}\|^2 \geq 0$ in the description of K and B, the sequence of solutions $(\mathrm{z}_\phi^{(d)}, \mathrm{z}_\nu^{(d)})_{d\in\mathbb{N}}$ converges towards the moments of the optimal measures $(\mu|_K, \mu|_{B\setminus K})$ when $d \to \infty$ as described in Theorem 1.3. If K and $B \setminus K$ have non-empty interior the d-th SDP stregthening of (3.5) has an optimal solution $p^{(d)} \in \mathbb{R}[\mathbf{x}]_{2d}$. Its coefficients are the dual variables associated with the constraints in (3.4).

3.1.2 Stokes Constraints for Faster Convergence

When looking at the problems (3.2) or (3.5) we face a classical issue in approximation theory, the so-called *Gibbs phenomenon*: When approximating a discontinuous function, like the indicator function, by a continuous (or even polynomial) one, the L^1 approximation will always *overshoot* the values of the approximated function at the points of discontinuity [Car25]. This effect is displayed in Fig. 3.1. As a consequence of this effect, the volume approximations via the relaxations of (3.4) are rather bad on early levels of the hierarchy

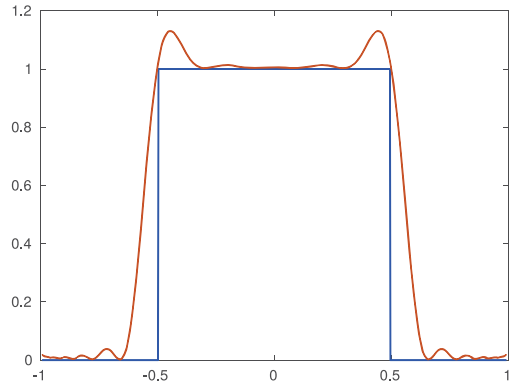

Figure 3.1: Polynomial L^1 approximation (red) of the indicator function (blue) with Gibbs' phenomenon at the discontinuities.

although they are guaranteed to converge when the relaxation degree tends to infinity. For this reason in [Las17a] Lasserre introduced so-called *Stokes constraints* that can be added to (3.4) without changing the optimal value and at the same time significantly accelerating the speed of convergence of the hierarchy. In addition to the before mentioned paper the following exposition is based on discussions in [Las17b; WRM18; LW18; Tac+18].

Let $\vartheta_1, \dots, \vartheta_n \in \mathbb{R}[\mathbf{x}]$ be polynomials such that $\vartheta_i \mathbf{n}_i$ *vanishes on the boundary* of K, where $\mathbf{n}_i$ is the i-th entry of the outer normal vector $\mathbf{n}$ for K. For instance if $K = \{\mathrm{x} \in \mathbb{R}^n : g(\mathrm{x}) \geq 0\}$ we can choose $\vartheta_i = g$ for all $i = 1, \dots, n$. Further assume that $\mathbf{d}\mu = \exp(\rho(\mathbf{x}))\,\mathbf{dx}$ for some polynomial $\rho \in \mathbb{R}[\mathbf{x}]$. Generic examples for such measures are the *Lebesgue measure* ($\rho = 0$), the *Exponential measure* ($\rho = -\sum \mathbf{x}_i$), and the *Gaussian measure* ($\rho = -\frac{1}{2}\mathbf{x}^\top \mathbf{x}$). Now Stokes theorem [HB07, Section 6.10] yields for all $\alpha \in \mathbb{N}^n$ and all $i \in \{1, \dots, n\}$ that

$$\int_K \tfrac{\partial}{\partial \mathbf{x}_i} \left(\mathbf{x}^\alpha \vartheta_i \exp(\rho)\right) \mathbf{dx} = \int_{\partial K} \mathbf{x}^\alpha \vartheta_i \exp(\rho) \mathbf{n}_i \,\mathbf{d}S = 0, \tag{3.6}$$

where $\mathbf{d}S$ denotes integration with respect to the Hausdorff measure [EG92, Chapter 2] on the boundary ∂K of K. Remember that the optimal measure in (3.1) and (3.4) is $\phi^* = \mu|_K$. A short calculation shows that (3.6) can be rewritten as

$$\forall \alpha \in \mathbb{N}^n : \left\langle \phi^*, \frac{\partial}{\partial \mathbf{x}_i} \left(\mathbf{x}^\alpha \vartheta_i\right) + \mathbf{x}^\alpha \vartheta_i \frac{\partial}{\partial \mathbf{x}_i} \rho \right\rangle = 0. \tag{3.7}$$

Since these equations are true for the optimal measure we can add (3.7) to (3.4) as redundant constraints on the optimization variable ϕ. As they are a consequence of Stokes' theorem, we call (3.7) *Stokes constraints*. After adding the Stokes constraints (3.7) to (3.4), its dual now reads

$$\begin{aligned} \inf_{p, p_1 \dots, p_n} & \int_B p \,\mathbf{d}\mu \\ \text{s.t. } & p + \sum_{i=1}^n \frac{\partial}{\partial \mathbf{x}_i} \left(p_i \vartheta_i\right) + p_i \vartheta_i \frac{\partial}{\partial \mathbf{x}_i} \rho \geq 1 \text{ on } K, \\ & p \in \mathbb{R}[\mathbf{x}], \quad p \geq 0 \text{ on } B, \quad p_1, \dots, p_n \in \mathbb{R}[\mathbf{x}]. \end{aligned} \tag{3.8}$$

It is evident from the constraint in (3.8) that in contrast to (3.5) now p is not an approximation of the (discontinuous) indicator function any more. However, as the Stokes constraints are redundant, the optimal value of (3.8) is still equal to the optimal value of (3.5).

Adding redundant constraints to an optimization problem before solving relaxations to the problem is a common strategy to improve the quality of the relaxation. As we will see in several examples in the subsequent sections, adding Stokes constraints significantly accelerates the convergence towards the volume of K.

3.1.3 Gaussian and Exponential Measures on Unbounded Sets

Up to now we have described methods that apply to basic semialgebraic sets K that are compact. In [Las17a] the method was extended to compute Gaussian or exponential measures of *unbounded sets*. More generally, the extension can take into account any reference measure whose moments can be computed efficiently and do not grow too fast for $|\alpha| \to \infty$, more precisely, whose moments satisfy Carleman's condition:

Definition 3.1.2 (Carleman's Condition)**.** A sequence $\mathrm{z} \subseteq \mathbb{R}$ satisfies the (multivariate) Carleman condition, if for all $i \in \{1, \dots, n\}$ the sum $\sum_{k=1}^{\infty} \left(L_{\mathrm{z}}(\mathbf{x}_i^{2k})\right)^{-\frac{1}{2k}}$ diverges.

The crucial difficulty here is that because K is not compact the quadratic module Definition 1.3.1 cannot be archimedean. In consequence we cannot rely on Putinar's Theorem in order to prove that the hierarchy converges towards a moment sequence of a measure supported on K as we did in Section 1.4. Here this defect is compensated by Carleman's condition [Las13, Thm. 2.2].

3.1.4 Chance Constraints Approximation

The probability approximation via the moment approach described in the precedent sections has a neat application in the approximation of *chance constraints*, which are (in-)equalities parametrised by some parameters $\omega \in \Omega \subseteq \mathbb{R}^\ell$: Let $\mu^{\boldsymbol{\omega}}$ be a probability measure on Ω, and $B_{\mathbf{x}} \subseteq \mathbb{R}^n$ be a compact set such that the moments of the normalized Lebesgue measure $\lambda_{B_{\mathbf{x}}}$ on $B_{\mathbf{x}}$ can be computed efficiently. Further for this section let $K \subseteq B_{\mathbf{x}} \times \Omega$ be a basic semialgebraic (compact) set[2] described by polynomials $g_1, \dots, g_m \in \mathbb{R}[\mathbf{x}, \boldsymbol{\omega}]$, i.e.,

$$K := \{(\mathrm{x}, \omega) \in \mathbb{R}^n \times \Omega : g_1(\mathrm{x}, \omega) \geq 0, \dots, g_m(\mathrm{x}, \omega) \geq 0\}. \tag{3.9}$$

A chance constraint asks that the probability with respect to $\mu^{\boldsymbol{\omega}}$ of the event $K_{\mathrm{x}} := \{\omega \in \Omega : (\mathrm{x}, \omega) \in K\}$ is sufficiently large. More precisely, let $\varepsilon > 0$ and denote by $\mathbb{P}(A) := \int \mathbf{1}_A \, \mathbf{d}\mu$ the probability of a measurable event $A \subseteq \Omega$ with respect to $\mu^{\boldsymbol{\omega}}$. Then a chance constraint is a constraint of the form:

$$\mathbb{P}(K_{\mathrm{x}}) \geq 1 - \varepsilon. \tag{3.10}$$

Note that the feasible set for (3.10) $X^\varepsilon := \{\mathrm{x} \in \mathbb{R}^n : \mathbb{P}(K_{\mathrm{x}}) \geq 1 - \varepsilon\}$ is usually non-convex or not even connected. Moreover and in contrast to K and K_{x}, X^ε is not a basic semialgebraic set. In [Las17b] Lasserre proposed to approximate X^ε by a nested sequence of basic semialgebraic sets X_d^ε which are *outer approximations* of X^ε such that $X_d^\varepsilon \supseteq X_{d+1}^\varepsilon$.

[2] In the case that Ω is non-compact $\mu^{\boldsymbol{\omega}}$ needs to satisfy Carleman's condition.

Lasserre also establishes the strong asymptotic guarantee that the Lebesgue volume of $X_d^\varepsilon \setminus X^\varepsilon$ converges to zero when $\mathbf{d} \to \infty$. In addition, as the complement of K in $B_\mathbf{x} \times \Omega$ can be described as uion of basic semialgebraic set whose intersections have measure zero, this procedure also provides a sequence of *inner approximations*, a typically desired feature in applications that need to deal with randomness in the data.

The approach nicely exploits the dual (3.5) of (3.4), where now K is defined as in (3.9), the reference measures μ is replaced by $\mu := \lambda_{B_\mathbf{x}} \otimes \mu^{\boldsymbol{\omega}} \in \mathcal{M}_+(B)$, and $B := B_\mathbf{x} \times \Omega$. Again it can be shown that the unique optimal measure for (3.4) is the restriction of μ to K, that there is no duality gap between (3.4) and (3.5) as long as K and $B \setminus K$ have non-empty interior, and that the moment relaxations (P_d) have optimal solutions $(P_d)^*$ converging to the moment sequences of $\mu|_K$ and $\mu|_{B\setminus K}$.

Let $p^{(d)} \in \mathbb{R}[\mathbf{x}, \boldsymbol{\omega}]$ be part of an optimal solution to the d-th SOS-strengthening (D_d) of (3.5) (see Chapter 1). Then $p^{(d)}$ is a polynomial over-approximation of the indicator function of K on B as explained above. Now observe that integrating $p^{(d)}$ partially, only over Ω, yields a polynomial $h_d \in \mathbb{R}[\mathbf{x}]$ with the interesting property

$$h_d(\mathrm{x}) := \int_\Omega p^{(d)}(\mathrm{x}, \boldsymbol{\omega})\, \mathbf{d}\mu^{\boldsymbol{\omega}} \geq \int_\Omega \mathbf{1}_K\, \mathbf{d}\mu^{\boldsymbol{\omega}} = \mathbb{P}(K_\mathrm{x}).$$

Consequently $X_d^\varepsilon := \{x \in B_\mathbf{x} : h_d \geq 1 - \varepsilon\}$ is an outer approximation of (3.10) as desired. The convergence of $\mathrm{vol}(X_d^\varepsilon \setminus X^\varepsilon)$ to zero is a consequence of the convergence of the optimal values $(D_d)^* = \int h_d\, \mathbf{d}\lambda_{B_x}$ towards $\mu(K)$.

Stokes for Chance Constraints

As it has been pointed out in Section 3.1.2, the convergence $(D_d)^* \to \mu(K)$ and hence $\mathrm{vol}(X_d^\varepsilon \setminus X^\varepsilon) \to 0$ is expected to be slow due to the Gibbs phenomenon and one might want to use Stokes constraints in order to accelerate the convergence. However, as already seen, Stokes constraints destroy the meaning of the dual variable as an over-approximation of the indicator of K – the key fact used to define the polynomial h_d and the approximations X_d^ε. Though, when only enforcing Stokes constraints coming from derivatives in the directions of $\boldsymbol{\omega}$ in (3.6) or (3.7), the polynomial $p^{(d)}$ can still be used. The reasoning behind this was tacitly assumed to be evident in [Las17b] and first mentioned explicitly in [WRM18]: Let $\mu = \exp(\rho(\boldsymbol{\omega}))\lambda_\Omega$, $\vartheta_{\boldsymbol{\omega}_1}, \dots \vartheta_{\boldsymbol{\omega}_\ell} \in \mathbb{R}[\mathbf{x}, \boldsymbol{\omega}]$ be polynomials such that $\vartheta_{\boldsymbol{\omega}_k} \mathbf{n}_{n+k}$ vanishes where now $\mathbf{n}$ is the outer normal vector of K written in the standard basis of $\mathbb{R}^n \times \mathbb{R}^\ell \supset B_\mathbf{x} \times \Omega$. Then (3.8) with the new notation reads

$$\begin{aligned} \inf_{p, p_1 \dots, p_n} & \int_B p(\mathbf{x}, \boldsymbol{\omega})\, \mathbf{d}(\lambda_{B_\mathbf{x}} \otimes \mu^{\boldsymbol{\omega}}) \\ \text{s.t. } & p + \sum_{k=1}^{\ell} \frac{\partial}{\partial \boldsymbol{\omega}_k} (p_k \vartheta_{\boldsymbol{\omega}_k}) + p_k \vartheta_{\boldsymbol{\omega}_k} \frac{\partial}{\partial \boldsymbol{\omega}_k} \rho \geq 1 \text{ on } K, \\ & p \in \mathbb{R}[\mathbf{x}, \boldsymbol{\omega}], \mathrm{p} \geq 0 \text{ on } B, \quad p_1, \dots, \mathrm{p}_n \in \mathbb{R}[\mathbf{x}, \boldsymbol{\omega}]. \end{aligned}$$

Consequently the polynomials $p^{(d)}, p_1^{(d)}, \ldots, \mathrm{p}_\ell^{(d)}$ obtained from the d-th strengthening satisfy

$$h_d := \int_\Omega p^{(d)} \,\mathbf{d}\mu^{\boldsymbol{\omega}} \geq \int_\Omega \mathbf{1}_K \,\mathbf{d}\mu^{\boldsymbol{\omega}} - \sum_{k=1}^{\ell} \underbrace{\int_\Omega \frac{\partial}{\partial \boldsymbol{\omega}_k} \left(p_k^{(d)} \vartheta_{\boldsymbol{\omega}_k} \right) + p_k^{(d)} \vartheta_{\boldsymbol{\omega}_k} \frac{\partial}{\partial \boldsymbol{\omega}_k} \rho \,\mathbf{d}\mu^{\boldsymbol{\omega}}}_{=0} = \mathbb{P}(K_{\mathbf{x}}),$$

where $\mathbb{P}(K_{\mathbf{x}})$ denotes the function $\mathrm{x} \mapsto \mathbb{P}(K_{\mathrm{x}})$ and the integrals in the sum are zero exactly by construction of the Stokes constraints (see (3.6) and use linearity of the integral). This shows that Stokes constraints can be used in the context of chance constraint approximations, if they are added in the directions of $\boldsymbol{\omega}$, only.

3.2 Chance-Constrained Optimization for Non-Linear Network Flow Problems

In this section we apply the concept of chance constraint approximation via the moment approach discussed in Section 3.1.4 to Chance-Constrained Optimal Physical Network Flow problems.

Many engineered systems, such as energy and transportation infrastructures, are networks governed by non-linear physical laws. A primary challenge for operators of these networks is to achieve optimal utilization while maintaining safety and feasibility, especially in the face of uncertainty regarding the system model. To address this problem, we formulate a Chance Constrained Optimal Physical Network Flow (CC-OPNF) problem that attempts to optimize the system while satisfying safety limits with a high probability. However, the non-linear equality constraints representing the network physics introduce modelling and optimization challenges which make the chance constraints numerically intractable in their original form. The main difficulty tackled in this contribution is to reformulate the initial problem in a way such that we can apply the methods discussed in Section 3.1.4 in order to approximate chance constraints by polynomial constraints. In addition, we develop a two-step procedure to improve computational speed and to enable the use of Stokes constraints. While the method is applicable to general physical network flow problems with polynomial constraints, we use the AC optimal power flow problem for electric grids as an example to demonstrate the method numerically. This section is based on [WRM18].

Motivation Networked systems are ubiquitous and include critical infrastructure networks such as the power grid, gas transmission pipelines, water networks and district heating systems. In such systems, optimization is often leveraged to maximize technical performance or economic efficiency, giving rise to what we will call Optimal Physical Network Flow (OPNF) problems. Optimization of system operation requires a mathematical model of the system. However, in practical systems, imperfect information and forecast errors introduce uncertainty in system operation and planning. If the uncertainty is not accounted for properly during the design and optimization process, the optimized system solution might be vulnerable to uncertainty, with potentially detrimental impacts on system risk.

A prominent example is the Optimal Power Flow (OPF) problem in electric power grids, which minimizes operational cost subject to technical constraints, and is used to clear

electricity markets, perform security assessment and guide system expansion planning. The most significant source of uncertainty in the OPF problem is due to imperfect forecasts in renewable generation and loading conditions. System security must be maintained by ensuring that all variables are kept within acceptable values for a range of uncertainty realizations. Problems with similar structure also arise in other infrastructure networks such as natural gas and water networks.

A typical approach to account for uncertainty is to formulate the OPNF as a robust or stochastic program. However, for many of the above mentioned systems, the physics governing the network flows are given by a set of non-linear equations, such as branch flow equations and nodal conservation laws. This gives rise to non-linear equality constraints, which are inherently non-convex and thus challenging for both deterministic and stochastic optimization algorithms. In addition, the non-linearity significantly complicates the characterization of the uncertainty propagation throughout the system.

Most existing robust and stochastic programming methods rely on assumptions of convexity. For practical problems such as the OPF problem, solution methods for robust or stochastic problem formulations typically use linear approximations [RA17; DBS17] or convex relaxations [Vra+13; LS17; Nas+16] to circumvent the problem of non-convexity. This enables the application of well-known methods for robust [BGN09] or chance-constrained [CC06; CE06] programming, at the expense of a reduction in model fidelity and less comprehensive feasibility guarantees for the underlying problem.

Here, we take a different approach. Instead of approximating or relaxing the non-linear network flow equations, we aim at treating the non-convex problem directly using the technique discussed in Section 3.1.4. The method is applicable for problems where the equality and inequality constraints can be represented as polynomials in both the decision variables and uncertain parameters.

We first formulate the uncertainty-aware problem as a Chance-Constrained OPNF (CC-OPNF) to guarantee that the constraints are satisfied with a high probability. In the formulation, we explicitly distinguish between violations of the network flow (equality) constraints, which indicate global instability of the system, and violations of engineering (inequality) constraints, which have a more local impact.

Due to the chance-constraints, the CC-OPNF is intractable in its original form. Our main contribution is to develop conservative, tractable approximations of the chance constraints in the form of polynomial constraints. Apart from the methods already discussed, we also build on [MHL15] to deal with projections of semialgebraic sets. We provide two crucial extensions:

1. While [Las17b] allow for outer approximations of the chance constraints using a hierarchy of SDPs, in practice it is not straight-forward to obtain an *inner approximation*, which is typically of interest in our setting. Therefore, we use a series of set manipulations to extend the existing methods towards practical *inner* approximations.

2. To improve computational performance, we develop a *two-step approximation procedure*, which allows for better approximations at lower computational overhead.

Replacing the chance constraints by their respective polynomial approximations yields an Approximate CC-OPNF (ACC-OPNF) problem that is still non-convex, but readily solvable to local optimality by state-of-the-art non-linear programming solvers. Note that

the polynomial chance-constraint approximations, which can be computationally heavy to compute, are determined a-priori in a *pre-processing* step to the ACC-OPNF.

Based on a small case study for the AC OPF problem, we demonstrate the practical performance of the method. In particular, we demonstrate the value of the extensions to inner approximations and the benefit of the two-step procedure.

3.2.1 Problem Formulation

We now present the problem formulation in abstract form for a generic physical network flow problem, as the method can be applied to any problem that has the structure described below. For a concrete example, we refer the reader to Section 3.2.5, where the method is applied to the AC OPF problem.

Deterministic Optimal Physical Network Flow

In this section we consider problem variables $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ and $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_m)$. For multivariate polynomials $f_1^0, \ldots, f_m^0,\ g_1^0, \ldots, g_k^0 \in \mathbb{R}[\mathbf{x}, \mathbf{y}]$ we consider the following Deterministic Optimal Physical Network Flow (D-OPNF) problem:

$$\begin{aligned} \min_{\mathrm{x},\mathrm{y}}\ & c(\mathrm{x},\mathrm{y}) \quad \text{s.t.} \\ & f_i^0(\mathrm{x},\mathrm{y}) = 0,\ i = 1, \ldots, m, && \text{(3.11a)} \\ & g_j^0(\mathrm{x},\mathrm{y}) \geq 0,\ j = 1, \ldots, k. && \text{(3.11b)} \end{aligned}$$

Here, the cost function is given by a polynomial $c \in \mathbb{R}[\mathbf{x}, \mathbf{y}]$. The polynomial equality constraints $f_i^0(\mathrm{x},\mathrm{y}) = 0$ represent the network flow physics. The polynomial inequality constraints $g_j^0(\mathrm{x},\mathrm{y}) \geq 0$ represent engineering limits.

To explicitly describe the degree of freedom in the system, we have separated the variables into $\mathbf{x}$ and $\mathbf{y}$. Since the equality constraints $f_i^0(\mathrm{x},\mathrm{y}) = 0$ eliminate m degrees of freedom, the variables $\mathbf{y}$ are intuitively implicit functions of the independent variables $\mathbf{x}$. Note that due to the non-linearity of the f_i^0, in general y might not be determined uniquely by a choice of x. In this paper we make a practical assumption stated below.

Assumption 1. *The engineering constraints $g_j^0(\mathrm{x},\mathrm{y}) \geq 0$ are such that the solution* y *to the system of equalities $f_i^0(\mathrm{x},\mathrm{y}) = 0$, whenever it exists is unique. As a result, we therefore can write* $\mathbf{y}$ *as a function of* $\mathbf{x}$*, i.e.* $\mathbf{y} = \mathrm{y}_{\mathbf{x}} := \mathrm{y}(\mathbf{x})$.

The above assumption reflects a feature often encountered in engineered networks. Even though, mathematically the network physics described by the non-linear system $f_i^0 = 0$ can have multiple solutions, there is only one solution that is physically meaningful within the region in which the system is operated. As soon as the variables $\mathbf{x}$ are set, the state of the system is fully determined. Assumption 1 allows us to formalize this notion.

Chance-Constrained Optimal Physical Network Flow

As we are aiming to account for uncertainty in the D-OPNF (3.11), we formulate the problem as a chance-constrained optimization problem. Chance constraints limit the probability

of constraint violations, and can be enforced either as joint chance constraints (several equations hold jointly with a given probability) or separate chance constraints (each constraint is assigned its own probability). Due to the underlying physics of the problem, the network flow constraints f_i^0 must be satisfied jointly: If one of them is violated, the solution is not physically valid and the remaining constraints are meaningless. The probability of not jointly satisfying the network flow constraints can be understood as the probability that the uncertainty realization will lead to a situation where the flow problem is unstable and there exist no steady-state operating point (e.g., voltage instability in electric power grids). The engineering limits g_j^0 can be satisfied either jointly or separately, depending on the preferred method for risk management. In the following, we provide a method for enforcing the engineering limits as separate chance constraints.

As in the previous section let $(\Omega, \mu^{\boldsymbol{\omega}})$ be a probability space. The random variables $\boldsymbol{\omega} = (\boldsymbol{\omega}_1, \dots, \boldsymbol{\omega}_\ell)$ have zero mean $\mathbf{0} \in \mathbb{R}^\ell$. Finally, let $f_1, \dots, f_m, g_1, \dots, g_k \in \mathbb{R}[\mathbf{x}, \mathbf{y}, \boldsymbol{\omega}]$ be multivariate polynomials. The notation is motivated by the idea that $f_i^0(\mathrm{x}, \mathrm{y}) = f_i(\mathrm{x}, \mathrm{y}, \mathbf{0})$ and $g_j^0(\mathrm{x}, \mathrm{y}) = g_j(\mathrm{x}, \mathrm{y}, \mathbf{0})$. Define $f = \sum_{i=1}^m f_i^2$. Then enforcing the system of equations $f_i(\mathrm{x}, \mathrm{y}, \omega) = 0, i = 1, \dots, m$, is equivalent to imposing $f(\mathrm{x}, \mathrm{y}, \omega) = 0$. We will use the later for better readability, although our implementation is based on the system of equalities rather than the single constraint $f(\mathrm{x}, \mathrm{y}, \omega) = 0$. We state the CC-OPNF problem:

$$
\begin{aligned}
\min_{\mathrm{x}, \mathrm{y}_\mathrm{x}, y(\boldsymbol{\omega})} \; & c(\mathrm{x}, \mathrm{y}_\mathrm{x}) \quad \text{s.t.} \\
& f_i^0(\mathrm{x}, \mathrm{y}_\mathrm{x}) = 0, \; i = 1, \dots, m, && \text{(3.12a)} \\
& g_j^0(\mathrm{x}, \mathrm{y}_\mathrm{x}) \geq 0, \; j = 1, \dots, k, && \text{(3.12b)} \\
& \mathbb{P}\left(f(\mathrm{x}, y(\boldsymbol{\omega}), \boldsymbol{\omega}) = 0\right) \geq 1 - \varepsilon_1, && \text{(3.12c)} \\
& \mathbb{P}\left(g_j(\mathrm{x}, y(\boldsymbol{\omega}), \boldsymbol{\omega}) \geq 0\right) \geq 1 - \varepsilon_2, \; j = 1, \dots, k. && \text{(3.12d)}
\end{aligned}
$$

Here $0 < \varepsilon_1, \varepsilon_2 < 1$ are the accepted violation probabilities of the respective constraints. In addition to the chance-constraints to account for uncertainty, we also keep the constraints (3.12a), (3.12b) from problem (3.11). These constraints give a precise meaning to the cost function $c(\mathbf{x}, \mathrm{y}_\mathbf{x})$, which is expressed as the operation cost for the expected realization $\omega = \mathbf{0}$.

We note that the problem as presented in (3.12) is a variational optimization problem. The controllable variables $\mathbf{x}$ however do not depend on the realization ω of $\boldsymbol{\omega}$, which means that once they are chosen, they cannot be modified in response to uncertainty. Note that in (3.11) we are optimizing over functions $y : \Omega \to \mathbb{R}^m$. However, similar to the DNF (3.11), the equality constraints eliminate the degrees of freedom for $y(\boldsymbol{\omega})$, and by a direct generalization of Assumption 1, one can think of y in (3.13) as a function of $(\mathbf{x}, \boldsymbol{\omega})$, within the region defined by the g_j. As a result, the constraints in (3.12d) are simply constraints on x, a property that we exploit in our approach to convert the variational problem in (3.13) into a standard optimization problem in x.

Unlike as in (3.11), where the inequalities (3.11b) along with Assumption 1 guarantee uniqueness and physical interpretability, eliminating $\mathrm{y}(\boldsymbol{\omega})$ in (3.12) is not trivial, as (3.12c) and (3.12d) are coupled through y. Eliminating y would allow for different choices of $\mathbf{y}$ for the same realization of $\boldsymbol{\omega}$ in the constraints (3.12c) and (3.12d), respectively. To circumvent this issue, we introduce a set Y and make the following assumption:

Assumption 2. *Restricting the range of y to a set $Y \subseteq \mathbb{R}^m$ the solution $y(\boldsymbol{\omega})$ to the system of equalities $f(\mathrm{x}, y(\boldsymbol{\omega}), \boldsymbol{\omega}) = 0$ in (3.12c) is unique whenever it exists.*

The set Y can be interpreted as domain specific knowledge about the system introduced in order to reduce the feasible space to a region where our physical model is valid and exclude physically meaningless solutions to $f(\mathrm{x}, y(\omega), \omega) = 0$. We propose the abstract formulation of the CC-OPNF problem below:

$$\min_{\mathrm{x},\mathrm{y_x}} c(\mathrm{x}, \mathrm{y_x}) \quad \text{s.t.}$$

$$f_i^0(\mathrm{x}, \mathrm{y_x}) = 0,\ i = 1, \ldots, m, \tag{3.13a}$$

$$g_j^0(\mathrm{x}, \mathrm{y_x}) \geq 0,\ j = 1, \ldots, k, \tag{3.13b}$$

$$\mathbb{P}\left(\exists \mathrm{y} \in Y,\ f(\mathrm{x}, \mathrm{y}, \boldsymbol{\omega}) = 0\right) \geq 1 - \varepsilon_1, \tag{3.13c}$$

$$\mathbb{P}\left(\exists \mathrm{y} \in Y,\ f(\mathrm{x}, \mathrm{y}, \boldsymbol{\omega}) = 0 \wedge g_j(\mathrm{x}, \mathrm{y}, \boldsymbol{\omega}) \geq 0\right) \geq 1 - \varepsilon_2,\ j = 1, \ldots, k. \tag{3.13d}$$

By Assumption 2 the constraints $f(\mathrm{x}, \mathrm{y}(\boldsymbol{\omega}), \boldsymbol{\omega}) = 0$ in (3.13d) implicitly specify y as a function of $\boldsymbol{\omega}$ given x.

Our contribution is to provide *tractable approximations* to the chance constraints (3.13c), (3.13d). more precisely these constraints will be replaced by polynomial constraints. The details of how we obtain the polynomial approximations will be explained over the next sections.

3.2.2 Polynomial Approximations of Chance Constraints

In this section, we first review another result from the literature which further extends the methods discussed in Section 3.1. Building on this we then develop inner and outer approximations of the chance constraint formulation in (3.13).

Volume of Projections of Semialgebraic Sets

Comparing the generic representation of chance constraints in (3.10) to the one presented in (3.13), we see that in many applications such as the OPNF, the presence of equality constraints introduce additional dependent variables $\mathbf{y}$ that are needed to describe the system. It is straightforward to extend the framework described in Section 3.1 by appending the additional variables $\mathbf{y}$ to form the set K in $(\mathbf{x}, \mathbf{y}, \boldsymbol{\omega})$-space and apply the same procedure outlined above. However, since y is fully specified by a choice (x, ω) the volume of the set K is zero. Consequently we cannot establish strong duality between the problems (3.1) and (3.2). This problem can be addressed by approximating the *projection* of K onto the $(\mathbf{x}, \boldsymbol{\omega})$-space where the volume is non-zero, instead of the original set K, using the method in Magron et al. [MHL15]. To approximate the indicator function of the projection

$$\pi^{\mathbf{x}\boldsymbol{\omega}}(K) := \{(\mathrm{x}, \omega) : \exists \mathrm{y} \in \mathbb{R}^m, (\mathrm{x}, \mathrm{y}, \omega) \in K\}$$

of K onto $(\mathbf{x}, \boldsymbol{\omega})$-space, consider the variant of (3.5):

$$\begin{aligned} \min_{p \in \mathbb{R}[\mathbf{x},\boldsymbol{\omega}]} & \int_B p(\mathbf{x}, \boldsymbol{\omega})\, \mathbf{d}\mu \\ \text{s.t.} \quad & p - 1 \geq 0 \text{ on } K, \\ & p \geq 0 \text{ on } B, \end{aligned} \tag{3.14}$$

where now $B := B_{\mathbf{x}} \times B_{\mathbf{y}} \times \Omega$ for some sets $B_{\mathbf{x}}$ and $B_{\mathbf{y}}$ for which it is easy to compute the moments of the Lebesgue measure, and $\mu := \lambda_{B_{\mathrm{x}}} \otimes \lambda_{B_{\mathrm{y}}} \otimes \mu^{\boldsymbol{\omega}}$. Note that the optimizing (polynomial) variable p is restricted to be invariant in $\mathbf{y}$-direction. The constraints guarantee that p is an overestimator of the indicator function of $\pi^{\mathbf{x}\boldsymbol{\omega}}(K)$ on $\pi^{\mathbf{x}\boldsymbol{\omega}}(B) = B_{\mathbf{x}} \times \Omega$. Similar to the results reviewed in Section 3.1, Magron et al. prove convergence results for the corresponding SDP hierarchy, i.e., $p^{(d)}$ converges to the indicator function of $\pi^{\mathbf{x}\boldsymbol{\omega}}(K)$ and the optimal value of the stregthenings of (3.14) converges to the volume of $\pi^{\mathbf{x}\boldsymbol{\omega}}(K)$ with respect to the marginal $\mu^{\mathbf{x}\boldsymbol{\omega}}$ of μ on $B_{\mathbf{x}} \times \Omega$.

Approximations of the CC-OPNF

We are now able to describe how to use the methods outlined in Section 3.1 and Section 3.2.2 to provide outer and inner approximations of the CC-OPNF (3.13). To that end we specify the feasible set of the chance constraints that we want to approximate by

$$\begin{aligned} \mathcal{L}^{\mathbf{x}} := \{ & \mathrm{x} \in B_{\mathrm{x}} : \\ & \mathbb{P}(\exists \mathrm{y} \in Y, f(\mathrm{x}, \mathrm{y}, \boldsymbol{\omega}) = 0) \geq 1 - \varepsilon_1, && \text{(3.15a)} \\ & \mathbb{P}(\exists \mathrm{y} \in Y, f(\mathrm{x}, \mathrm{y}, \boldsymbol{\omega}) = 0 \wedge g_j(\mathrm{x}, \mathrm{y}, \boldsymbol{\omega}) \geq 0) \geq 1 - \varepsilon_2, \quad j = 1, \ldots, k\}, && \text{(3.15b)} \end{aligned}$$

where we assume that $\mathcal{L}^{\mathbf{x}} \subseteq B_{\mathrm{x}}$. As mentioned in Section 3.2.1, our goal is to approximate the set $\mathcal{L}^{\mathbf{x}}$ by replacing the intractable chance constraints by polynomial constraints.

Outer Approximation of the Feasible Set We define the sets for which the constraints remain satisfied as

$$\begin{aligned} K_0 &:= \{(\mathrm{x}, \mathrm{y}, \omega) \in B : f(\mathrm{x}, \mathrm{y}, \omega) = 0\}, \\ K_j &:= \{(\mathrm{x}, \mathrm{y}, \omega) \in B : f(\mathrm{x}, \mathrm{y}, \omega) = 0 \wedge g_j(\mathrm{x}, \mathrm{y}, \omega) \geq 0\}, \quad j = 1, \ldots, k. \end{aligned} \tag{3.16}$$

An outer approximation of the set $\mathcal{L}^{\mathbf{x}}$ can be obtained by applying the method outlined in Section 3.1.4 to each of the sets K_j for $j = 0, \ldots, k$. For each K_j, we get a polynomial $h_j^* \in \mathbb{R}[\mathbf{x}]$ which approximates the function $\mathrm{x} \mapsto \mathbb{P}(\pi^{\mathbf{x}\boldsymbol{\omega}}(K_j))$ from above, leading to an overestimation of the satisfaction probability and an outer approximation of the chance constraints. Consequently the set

$$\{\mathrm{x} \in B_{\mathrm{x}} : h_0^*(\mathrm{x}) \geq 1 - \varepsilon_1, h_j^*(\mathrm{x}) \geq 1 - \varepsilon_2, j = 1, \ldots, k\}$$

is an outer approximation of $\mathcal{L}^{\mathbf{x}}$, and the corresponding ACC-OPNF provides a lower bound to the optimal cost of the OPNF.

Inner Approximation of the Feasible Set In applications where system security is of primary concern, obtaining feasible solutions to (3.13) is more important than obtaining lower bounds to the cost, hence motivating an investigation of inner approximations to the chance constraints. However, as opposed to the outer approximation, obtaining an inner approximation of $\mathcal{L}^{\mathbf{x}}$ is more involved. In principle one would like to obtain inner approximations by considering the complements of K_j defined in (3.16). Note however that in (3.15) we are considering projections of theses sets. Hence we would actually need to use the compliments of the projections. However we do not have a basic semialgebraic description of the latter.

In the following, we propose a modification of $\mathcal{L}^{\mathbf{x}}$, which we can use to approximate this set (almost) from the interior. For $\varepsilon_1 < \varepsilon_2$ define

$$\begin{aligned}\mathcal{K}^{\mathbf{x}} := \{&\mathrm{x} \in B_{\mathrm{x}} : \\ &\mathbb{P}(\exists \mathrm{y} \in Y, f(\mathrm{x},\mathrm{y},\boldsymbol{\omega}) = 0) \geq 1-\varepsilon_1, && \text{(3.17a)}\\ &\mathbb{P}(\exists \mathrm{y} \in Y, (f(\mathrm{x},\mathrm{y},\boldsymbol{\omega}) = 0 \wedge g_j(\mathrm{x},\mathrm{y},\boldsymbol{\omega}) \leq 0)) \leq \varepsilon_2 - \varepsilon_1, \quad j = 1,\dots,k\}. && \text{(3.17b)}\end{aligned}$$

The essential difference between $\mathcal{L}^{\mathbf{x}}$ and $\mathcal{K}^{\mathbf{x}}$ is that the probabilities (3.17b) in $\mathcal{K}^{\mathbf{x}}$ are bounded from above whereas the probabilities (3.15b) in $\mathcal{L}^{\mathbf{x}}$ are bounded from below. Since the methods discussed in the beginning of Section 3.1.4 lead to overestimators of the probability, the reversal of the inequality in the formulation in $\mathcal{K}^{\mathbf{x}}$ now enables us to approximate the sets described by the chance constraints in (3.17b) from the interior. The following proposition relates the approximation $\mathcal{K}^{\mathbf{x}}$ to $\mathcal{L}^{\mathbf{x}}$.

Proposition 3.2.1. *The set $\mathcal{K}^{\mathbf{x}}$ is an inner approximation of $\mathcal{L}^{\mathbf{x}}$.*

Proof. For the proof it will be handy to introduce some formulas. Define

$$\begin{aligned}A :=& \exists \mathrm{y} \in Y, f(\mathrm{x},\mathrm{y},\omega) = 0,\\ B :=& \exists \mathrm{y} \in Y, (f(\mathrm{x},\mathrm{y},\omega) = 0 \wedge g_j(\mathrm{x},\mathrm{y},\omega) \leq 0,\\ B' :=& \forall \mathrm{y} \in Y, (f(\mathrm{x},\mathrm{y},\omega) \neq 0 \vee g_j(\mathrm{x},\mathrm{y},\omega) > 0),\\ B'' :=& \forall \mathrm{y} \in Y, (f(\mathrm{x},\mathrm{y},\omega) \neq 0 \vee g_j(\mathrm{x},\mathrm{y},\omega) \geq 0),\\ C :=& \exists \mathrm{y} \in Y, (f(\mathrm{x},\mathrm{y},\omega) = 0 \wedge g_j(\mathrm{x},\mathrm{y},\omega) \geq 0).\end{aligned}$$

Note that $\neg B = B'$ and $B' \Rightarrow B''$. Therefore $\mathrm{x} \in \mathcal{K}^{\mathbf{x}}$ is a stronger condition than

$$\mathrm{x} \in K^{\mathbf{x}} := \{\mathrm{x} \in B_{\mathrm{x}} : \mathbb{P}(A) \geq 1-\varepsilon_1, \mathbb{P}(B'') \geq 1-\varepsilon_2+\varepsilon_1,\ j = 1,\dots,k\},$$

i.e., $\mathcal{K}^{\mathbf{x}} \subseteq K^{\mathbf{x}}$. To see that $K^{\mathbf{x}} \subseteq \mathcal{L}^{\mathbf{x}}$, note that $B'' \Leftrightarrow (A \wedge B'') \vee (\neg A \wedge B'')$, $(A \wedge B'') \Rightarrow C$, and $(\neg A \wedge B'') \Leftrightarrow \neg A$. Hence, if $\mathrm{x} \in \mathcal{K}^{\mathbf{x}}$, $1-\varepsilon_2+\varepsilon_1 \leq \mathbb{P}(B) \leq \mathbb{P}(\neg A) + \mathbb{P}(C) \leq \varepsilon_1 + \mathbb{P}(C)$. Consequently, $\mathbb{P}(C) \geq 1-\varepsilon_2$, i.e., $\mathrm{x} \in \mathcal{L}^{\mathbf{x}}$. □

Instead of directly dealing with $\mathcal{L}^{\mathbf{x}}$, we attempt to approximate the set $\mathcal{K}^{\mathbf{x}}$ from the interior. Using the same procedure as before we compute polynomials $h_0^*, \dots, h_m^*$ approximating

the functions $\mathrm{x} \mapsto \mathbb{P}(\pi^{\mathbf{x}\boldsymbol{\omega}}(K_j))$ where K_j now is defined by

$$\begin{aligned} K_0 :=& \{(\mathrm{x},\mathrm{y},\omega) \in B : f(\mathrm{x},\mathrm{y},\omega) = 0\}, \\ K_j :=& \{(\mathrm{x},\mathrm{y},\omega) \in B : f(\mathrm{x},\mathrm{y},\omega) = 0 \wedge g_j(\mathrm{x},\mathrm{y},\omega) \leq 0\}. \end{aligned} \tag{3.18}$$

Note that though we are aiming for an inner approximation of $\mathcal{K}^{\mathbf{x}}$, the polynomials h_j^* are over-approximations of the probability. The set $\mathcal{K}^{\mathbf{x}}$ is then approximated by the set

$$\tilde{\mathcal{K}}^{\mathbf{x}} := \{x \in B_{\mathrm{x}} : h_0^*(\mathrm{x}) \geq 1 - \varepsilon_1 \tag{3.19a}$$

$$h_j^*(\mathrm{x}) \leq \varepsilon_2 - \varepsilon_1, j = 1, \ldots, k\}. \tag{3.19b}$$

Since the polynomials $h_j^*(\mathrm{x})$ over approximate the probabilities in (3.17b), the set defined by the inequalities in (3.19b) are inner approximations of the corresponding sets defined by (3.17b). Unfortunately the same relation is not true for the sets defined by (3.19a) and (3.17a) that correspond to the probability of joint violation of the equality constraints $f_i(\mathrm{x},\mathrm{y},\omega) = 0$. Therefore, $\tilde{\mathcal{K}}^{\mathbf{x}}$ is an *approximate* inner approximation to $\mathcal{L}^{\mathbf{x}}$.

3.2.3 Improved Approximations through Stokes Constraints

Partial Stokes Constraints for Projection of Sets

Stokes constraints cannot directly be applied to the setting where the feasible set is described by the projection of a semialgebraic set. This is because in order to be able to add Stokes constraints to the problem in (3.14), we must first find a polynomial $\vartheta \in \mathbb{R}[\mathbf{x}, \boldsymbol{\omega}]$ that vanishes on the boundary of the projection $\pi^{\mathbf{x}\boldsymbol{\omega}}(K)$ of K. Note that in Section 3.1.4, where there is no projection involved, the polynomial $\vartheta \in \mathbb{R}[\mathbf{x}, \boldsymbol{\omega}]$ can be readily obtained as the product of the polynomials that define the semialgebraic set K. For the projection $\pi^{\mathbf{x}\boldsymbol{\omega}}(K)$ of a semialgebraic set K in $(\mathbf{x}, \mathbf{y}, \boldsymbol{\omega})$-space, this trick is not applicable.

Our solution to this issue is a two-step-procedure: In a first step we approximate the projection $\pi^{\mathbf{x}\boldsymbol{\omega}}(K)$ by the super-level-set S of some polynomial $p^{(1)} \in \mathbb{R}[\mathbf{x}, \boldsymbol{\omega}]$. In a second step, we use this set S to compute a second polynomial $p^{(2)} \in \mathbb{R}[\mathbf{x}, \boldsymbol{\omega}]$ approximating the indicator function of S. From this we obtain the polynomial h approximating the chance constraint (see Section 3.1.4). We explain the procedure in more detail:

Step 1: Approximating the Projection of K We first apply the method in Section 3.1.4 and solve the problem in (3.14) to obtain a polynomial $p^{(1)}$ that is an overestimator of the indicator function of $\pi^{\mathbf{x}\boldsymbol{\omega}}(K)$, i.e. $p^{(1)} \geq 1$ on $\pi^{\mathbf{x}\boldsymbol{\omega}}(K)$. In particular the super-level-set given by

$$S := \{(\mathrm{x}, \omega) \in B_{\mathbf{x}} \times \Omega : p^{(1)}(\mathrm{x}, \omega) - 1 \geq 0\} \tag{3.20}$$

is an outer approximation of $\pi^{\mathbf{x}\boldsymbol{\omega}}(K)$. Figure 3.2 illustrates this step. Numerical experiments have shown that the 1-super-level-set of the optimizing polynomial is quite accurate already for low relaxation degrees.

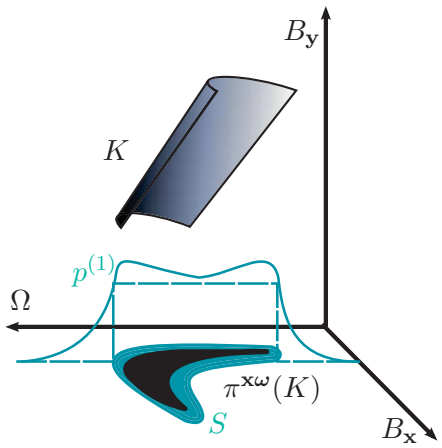

Figure 3.2: Step 1: Projection step. The projection of K is approximated as S, which is defined by the 1-super-level set of $p^{(1)}$.

Step 2: Probability Approximation After the first step we replace the actual projection $\pi^{\mathbf{x}\boldsymbol{\omega}}(K)$ by its approximation S defined in (3.20). In doing so we lose information about $\pi^{\mathbf{x}\boldsymbol{\omega}}(K)$ but we gain two important advantages. First, moving from K to S we get a significant reduction in the number of variables, as we eliminate the whole $\mathbf{y}$-space. This allows us to afford computational capacity for higher levels in the SDP relaxation hierarchy and get better volume approximations. Second, we now have a polynomial, specifically $p^{(1)} - 1$, that vanishes on the boundary of S. This crucial difference enables us to use Stokes constraints to improve the volume approximation. Applying the method in Section 3.1.4, we obtain a polynomial $p^{(2)} \in \mathbb{R}[\mathbf{x}, \boldsymbol{\omega}]$ that still preserves the desired over-approximation property:

$$h^*(\mathrm{x}) := \int_\Omega p^{(2)}(\mathrm{x}, \omega)\, \mathbf{d}\mu \overset{(a)}{\geq} \mathbb{P}(S) \overset{(b)}{\geq} \mathbb{P}(\pi^{\mathbf{x}\boldsymbol{\omega}}(K)),$$

where (a) follows from Section 3.1.4 and (b) follows because $\pi^{\mathbf{x}\boldsymbol{\omega}}(K) \subseteq S$. This step is summarized in Fig. 3.3.

3.2.4 The Overall Approach

To summarize the overall approach, we first recall the problem formulation (3.13). Our aim is to eliminate the chance constraints (3.13c) and (3.13d) and replace them by tractable polynomial constraints. The challenge is to (i) ensure existence of solution to the equality constraints, (ii) compute inner approximations to the chance constraints, and (iii) enable the use of Stokes constraints to speed up convergence. We address the challenges in the following steps:

1. We reformulate the feasible set $\mathcal{L}^{\mathbf{x}}$ of the chance constraints by the set $\mathcal{K}^{\mathbf{x}}$ that allows us to obtain inner approximations.

2. We eliminate the dependent $\mathbf{y}$ variables by approximating the projection of each K_j defining $\mathcal{K}^{\mathbf{x}}$ as the super-level set S of a polynomial $p_j^{(1)}$.

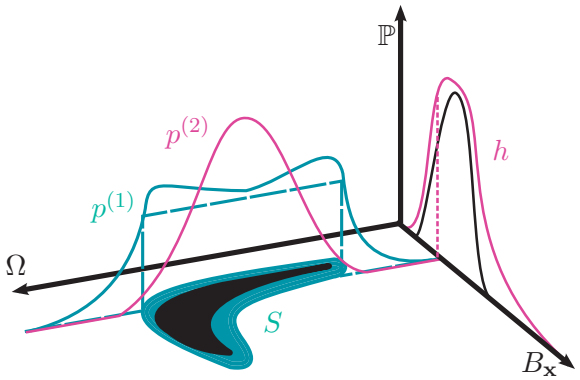

Figure 3.3: Step 2: Probability approximation. The probability is approximated by integrating $p^{(2)}$ in Ω direction for every $\mathrm{x} \in B_{\mathbf{x}}$.

3. We use the reduced set S to compute the inner approximations to the chance constraints by polynomials $h_0^*(\mathrm{x}), \ldots, h_k^*(\mathrm{x})$. To speed up convergence, we add Stokes constraints which is made possible by the availability of the polynomial $p^{(1)}$.

Now, the chance constraints in original problem (3.13c), (3.13d) can be replaced by their approximation to obtain the ACC-OPNF formulation:

$$\begin{aligned}
&\min_{\mathrm{x},\mathrm{y_x}} c(\mathrm{x},\mathrm{y_x}) \quad \text{s.t.} \\
&\quad f_i^0(\mathrm{x},\mathrm{y_x}) = 0,\ i = 1,\ldots,m, && \text{(3.21a)}\\
&\quad g_j^0(\mathrm{x},\mathrm{y_x}) \geq 0,\ j = 1,\ldots,k, && \text{(3.21b)}\\
&\quad h_0(\mathrm{x}) \geq 1 - \varepsilon_1, && \text{(3.21c)}\\
&\quad h_j(\mathrm{x}) \leq \varepsilon_2 - \varepsilon_1,\ j = 1,\ldots,k. && \text{(3.21d)}
\end{aligned}$$

Although obtaining the polynomials $h_0^*(\mathrm{x}), \ldots, h_k^*(\mathrm{x})$ might be computationally heavy, this procedure is independent of the actual solution process for the resulting ACC-OPNF. Therefore it can be considered as a pre-processing step to be executed offline. The resulting approximate CC-OPNF, despite remaining non-convex, can be solved to local optimality using a local non-linear solver. Furthermore, methods for global optimization of polynomial problems can be applied as discussed in Chapter 2.

3.2.5 Application to Chance-Constrained AC Optimal Power Flow

In this section, we present the mapping of a chance-constrained AC optimal power flow (CC-AC-OPF) problem onto the general CC-OPNF problem (3.13). Motivated by the recent increase in generation uncertainty from renewable energy sources, our CC-AC-OPF formulation attempts to minimize generation cost, subject to engineering constraints while accounting for the uncertainty in renewable power generation.

Deterministic Optimal Power Flow

We first formulate the deterministic OPF problem where we assume perfect knowledge of the system. This problem corresponds to the deterministic OPNF (3.11).

Notation We consider an electric network where $\mathcal{N}$ and $\mathcal{E}$ denote the sets of nodes and edges. Without loss of generality, we assume that there is one generator, one demand and one uncertainty source per bus. Complex power is given by $s = p + j \cdot q$, where p and q are the active and reactive power. Subscripts $_R$, $_G$ and $_D$ are for renewable energy sources, conventional generators and loads, respectively. The complex bus voltages are denoted by $v = v_{\text{real}} + v_{\text{imag}}$, and the corresponding voltage magnitudes by $|v| = (v_{\text{real}}^2 + v_{\text{imag}}^2)^{1/2}$.

Problem formulation Given the above considerations, the OPF problem is given by

$$\min_{\substack{p_{G0},\\ q_{G0},v_0}} \sum_{i\in\mathcal{G}} c_{2,i}p_{G0,i}^2 + c_{1,i}p_{G0,i} + c_0 \tag{3.22a}$$

$$\text{s.t.} \quad s_{G0,i} + s_{R,i} - s_{D,i} = \sum_{(i,j)\in\mathcal{E}} s_{0,ij}, \qquad \forall i \in \mathcal{N}, \tag{3.22b}$$

$$s_{0,ij} = \mathbf{Y}_{ij}^* v_{0,i} v_{0,i}^* - \mathbf{Y}_{ij}^* v_{0,i} v_{0,j}^*, \qquad \forall (i,j) \in \mathcal{E}, \tag{3.22c}$$

$$p_{G,i}^{\min} \leq p_{G0,i} \leq p_{G,i}^{\max}, \qquad \forall i \in \mathcal{N}, \tag{3.22d}$$

$$q_{G,i}^{\min} \leq q_{G0,i} \leq q_{G,i}^{\max}, \qquad \forall i \in \mathcal{N}, \tag{3.22e}$$

$$|v|^{\min} \leq |v_{0,j}| \leq |v|^{\max}, \qquad \forall j \in \mathcal{N}, \tag{3.22f}$$

$$|s_{0,ij}| \leq |s_{ij}|^{\max}, \qquad \forall (i,j) \in \mathcal{E}. \tag{3.22g}$$

The objective (3.22a) of the problem is to choose the generation dispatch point, given by the active and reactive power generation p_{G0}, q_{G0} and the complex voltages v_0, such that the the cost of active power generation given by the quadratic function in (3.22a) is minimized. The AC power flow equations (3.22b), (3.22c) are a set of equality constraints describing the physical laws, with the nodal power balance given by (3.22b), and transmission line flows given by Ohm's law (3.22c), where $\mathbf{Y}$ is the so-called *admittance matrix*. Note that we use the rectangular form of the power flow equations to obtain polynomial constraints. Further, we enforce a set of engineering limits (3.22d)-(3.22g). The constraints (3.22d), (3.22e) represent bounds on generation capacity, (3.22f) limits the voltage magnitudes to safe ranges and (3.22g) enforces limits on the apparent power flow. Among these constraints, (3.22b) and (3.22c) correspond to the equality constraints $f_i^0 = 0$ in the deterministic OPNF (3.11), and the remaining constraints correspond to the inequality constraints $g_j^0 \geq 0$.

Chance-Constrained Optimal Power Flow

We now extend the deterministic problem to the setting with uncertainty in the power injections.

Modelling uncertain injections We model the uncertain active power injections from renewable generators as the sum of the expected value p_R and a fluctuation ω. The expected reactive power injection is denoted by q_R. The reactive power injections are assumed to

adjust in a way that the power factor, given by $\gamma = q_R/p_R$, remains constant:

$$s_R(\omega) = (p_R + \omega) + j \cdot (q_R + \gamma\omega) \tag{3.23}$$

We assume that the probability distribution of ω is known. The active and reactive power consumption of the loads, denoted by $p_L,\ q_L$, are assumed to be constant, but could also be modelled similar to (3.23).

Power flow equations under uncertainty For non-zero uncertainty realization ω, the power flow equations (3.22b) are adapted to account for ω, i.e.

$$s_{G,i}(\omega) + s_{R,i} + \omega - s_{D,i} = \sum_{(i,j)\in\mathcal{E}} s_{ij}(\omega), \qquad \forall i \in \mathcal{N}, \tag{3.24a}$$

$$s_{ij}(\omega) = \mathbf{Y}^*_{ij} v_i(\omega) v_i^*(\omega) - \mathbf{Y}^*_{ij} v_i(\omega) v_j^*(\omega), \qquad \forall (i,j) \in \mathcal{E}. \tag{3.24b}$$

Response to Uncertainty When the power injections fluctuate, the controllable generators must adjust their generation output $s_{G,i}(\omega)$ to ensure that the power balance constraints (3.22b) are satisfied. We adopt balancing practices typical in power systems operation, which require the definition of so-called pv, pq and $v\theta$ (reference) buses.

On each node of the network there are four state variables, namely the active power injection p, the reactive power injection q, and two voltage variables corresponding to the voltage magnitude and angle $|v|,\ \theta$ (polar coordinates) or the real and imaginary voltage $v_{\text{real}},\ v_{\text{imag}}$ (rectangular coordinates). The buses are classified according to the quantities that are controllable or specified: (i) pq buses (such as loads) with specified real and reactive power, (ii) pv buses (such as generators) with controllable active power and voltage magnitude, and (iii) $v\theta$ or reference bus with the voltage angle set to zero. The sets of nodes that correspond to the three categories are denoted by subscripts $\mathcal{N}_{pq},\ \mathcal{N}_{pv}$ and $\mathcal{N}_{v\theta}$.

Given the above definitions, we assume that the active power injections from generators at $pq,\ pv$ buses remain constant throughout the fluctuations, and all fluctuations ω are balanced by the generator connected at the slack bus. Similarly, reactive power is balanced by adjusting the reactive power output of pv and $v\theta$ buses to maintain constant voltage magnitudes, while the reactive power injections at pq buses are kept constant.

Definition of x and y variables

We choose the rectangular coordinate representation in order to be able to employ the semi-algebraic methods described in this work. This gives us 4 variables per bus $p, q, v_{\text{imag}}, v_{\text{real}}$. However, as described above, the standard model for pv and $v\theta$ buses are based on polar coordinates, where we keep the voltage magnitude constant. We handle these requirements in rectangular coordinates by adding the constraints $v_{\text{imag}} = 0$ and $v_{\text{real},i}(\omega) = v_{\text{real},i}$ for $i \in \mathcal{N}_{v\theta}$, and the constraint $v_{\text{real},i}(\omega)^2 + v_{\text{imag},i}(\omega)^2 = |v|_i^2$ for $i \in \mathcal{N}_{pv}$.

This results in two independent variables per bus, which we choose to also correspond to the quantities that can be controlled by the system operator. In particular, we define the independent $\mathbf{x}$ variables as

$$p_{G0,i}, q_{G0,i}, \qquad \forall i \in \mathcal{N}_{pq},$$

$$
\begin{aligned}
&p_{G0,i}(\omega), |v|_{0,i}, && \forall i \in \mathcal{N}_{pv},\\
&v_{\text{real}0,i}, v_{\text{imag}0,i}, && \forall i \in \mathcal{N}_{v\theta}.
\end{aligned}
$$

The variables that change as a function of $\boldsymbol{\omega}$ are the elements $\mathrm{y} \in Y$ in the CC-OPNF formulation (3.13):

$$
\begin{aligned}
&v_{\text{real},i}(\omega), v_{\text{imag},i}(\omega), && \forall i \in \mathcal{N}_{pq},\\
&q_{G,i}(\omega), v_{\text{real},i}(\omega), v_{\text{imag},i}(\omega), && \forall i \in \mathcal{N}_{pv},\\
&p_{G,i}(\omega), q_{G,i}(\omega), && \forall i \in \mathcal{N}_{v\theta},\\
&s_{ij}(\omega), && \forall ij \in \mathcal{E}.
\end{aligned}
$$

Note that in the process of solving (3.13), we are not explicitly assigning a value to these dependent quantities $\mathrm{y} = \mathrm{y}(\boldsymbol{\omega})$. However, the variables $\mathrm{y}_{\mathbf{x}}$, which correspond to y at the expected operating point ($\omega = 0$), are explicitly defined.

Definition of constraints $f = 0$ and $g \geq 0$

As is evident from (3.24), both the generation outputs $p_{G,i}(\omega)$ and $q_{G,i}(\omega)$, the power flows $s_{ij}(\omega)$ and the voltage variables $v_i(\omega)$ will change depending on the realization of $\boldsymbol{\omega}$. The constraints which incorporate those quantities are therefore enforced as chance constraints.

The stochastic power flow equations (3.24) correspond to the equality constraints $f(\mathrm{x}, \mathrm{y}, \omega) = 0$. When there is no solution to this set of equations, the system is unstable and might collapse at any point leading to complete blackout of the electric grid. We hence want the probability of violating any of the equality constraints to be very low, and enforce those constraints jointly as in (3.13c) with a small acceptable violation probability ε_1. The inequality constraints $g_j(\mathrm{x}, \mathrm{y}, \omega) \geq 0$ correspond to the engineering limits

$$
\begin{aligned}
&p_{G,i}^{\min} \leq p_{G,i}(\omega) \leq p_{G,i}^{\max}, && \forall i \in \mathcal{N}_{v\theta} && \text{(3.25a)}\\
&q_{G,i}^{\min} \leq q_{G,i}(\omega) \leq q_{G,i}^{\max}, && \forall i \in \mathcal{N}_{pv}, \mathcal{N}_{v\theta} && \text{(3.25b)}\\
&|v_i|^{\min} \leq |v_i|(\omega) \leq |v_i|^{\max}, && \forall i \in \mathcal{N}_{pq} && \text{(3.25c)}\\
&v_{\text{real},i}(\omega)^2 + v_{\text{imag},i}(\omega)^2 = |v|_i^2, && \forall i \in \mathcal{N}_{pv} && \text{(3.25d)}\\
&|s_{ij}|(\omega) \leq |s_{ij}|^{\max}, && \forall (i,j) \in \mathcal{E}. && \text{(3.25e)}
\end{aligned}
$$

In contrast to a violation of the power flow equations (3.24), a violation of one of the engineering constraints (3.25) would typically have a more local impact (e.g. overloading of a component), and can often be tolerated for a certain amount of time (e.g. violations of thermal capacity limits of transmission lines). We hence enforce (3.25) as separate chance constraints, and allow for a larger violation probability $\varepsilon_2 > \varepsilon_1$.

Choosing Y

The last parameter we must determine before the mapping from the CC-AC-OPF to the generic CC-OPNF problem (3.13) is complete, is the set Y from Assumption 2. We would like to choose Y such that solutions to (3.24) are unique and have a well-defined physical meaning, which for the OPF problem implies ensuring that low voltage solutions to the

power flow equations are excluded. Therefore we define the sets Y by the inequalities

$$|v|^{\min-} \leq |v_i|(\omega), \quad \forall i \in \mathcal{N}_{pq}. \tag{3.26}$$

Here, $|v|^{\min-}$ is lower than the standard voltage bound $|v|^{\min}$, but sufficiently large to exclude low voltage solutions.

3.2.6 Case Study

We first describe the implementation and test system, before presenting the numerical results for the chance constraint approximation and the resulting approximate CC-OPNF.

Implementation

In this section, we describe our implementation to obtain the ACC-OPNF in Section 3.2.4 and evaluate its performance. To obtain the polynomials $h_0^*, \ldots, h_k^*$ in (3.21) we solve SDP relaxations to the infinite dimensional linear problems described in Section 3.2.3. We use the GloptiPoly3 Matlab toolbox [HLL09a] to model the relaxations and Mosek [MOS17] to solve the SDPs. The resulting ACC-OPNF is implemented in Julia [Bez+14] with JuMP [LD15] and PowerModels.jl [Cof+17] and then solved using the local non-linear solver Ipopt [WB06]. We also perform Monte-Carlo simulations for benchmarking which requires solving the standard power flow and the AC-OPF which are implemented using Matpower [ZMT11] and PowerModels.jl respectively.

Test System

We run our numerical experiments on a modified version of a 4-bus system in [GS94] (case4gs in the Matpower library) which is illustrated in Fig. 3.4. The system has two conventional generators at Bus 1 and Bus 4, with active and reactive power limits $p_{Gi}^{\min} = 0$,$p_{Gi}^{\max} = 500$ and $q_{Gi}^{\min} = -250$, $q_{Gi}^{\max} = 500$. Bus 1 is the reference bus, while all other buses are PQ buses. We assume that the load at Bus 2 is uncertain, with active power fluctuations $\boldsymbol{\omega}$ uniformly distributed on $[-50, 50]$. The reactive power fluctuations on Bus 2 are proportional to the active power fluctuations, with $\gamma \approx 0.62$. We assume quadratic cost for Bus 1 with $(c_{2,1}, c_{1,1}, c_{0,1}) = (0.01, 30, 200)$ and a linear cost for Bus 4 with $(c_{2,4}, c_{1,4}, c_{0,4}) = (0, 25, 400)$.

Numerical Results

We verify the approximation quality of the chance constraint approximation, and then assess the performance of the full CC-AC-OPF problem.

Approximation of Chance Constraints In the following we employ the two-step approach described in Section 3.2.3 to obtain the chance constraint approximations through the polynomials $h_0^*, \ldots, h_k^*$ given in Eq. (3.21). We investigate the accuracy of this approximation and how the accuracy improves by increasing the relaxation order d and the addition of Stokes constraints. Results for both the outer approximation and the (approximate) inner approximation described in Section 3.2.2 are shown.

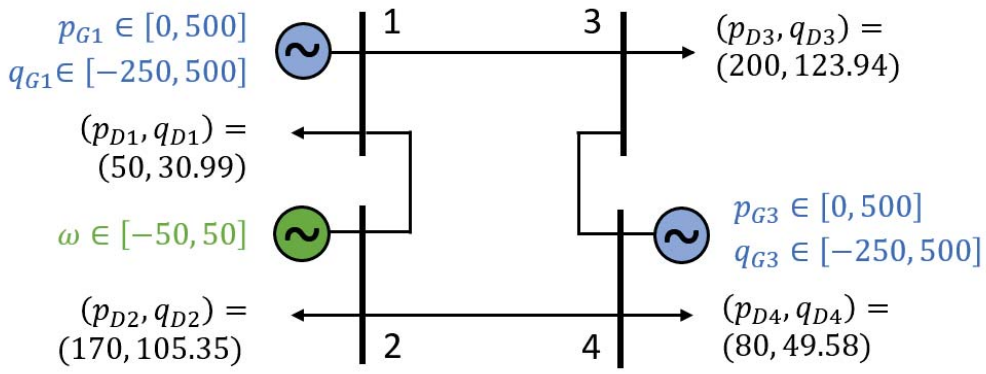

Figure 3.4: Overview of the 4-bus system. Generators marked in blue, uncertainty source in green and loads in black.

To obtain outer and inner approximations we need to compute the probability of the projections of the sets K_j defined in (3.16) and (3.18) respectively, by using the two-step method in Section 3.2.3. For the corresponding SDP relaxations, we choose the relaxation order of the first step to be $d = 2$ or 3 and for the second step to be $d + 5 = 7$ or 8. For the first step, a lower degree polynomial is sufficient to approximate the level sets of K_j, whereas the second step needs higher orders for better approximation and benefiting from Stokes constraints.

To assess how close we are to the true feasible set of the chance constraints, we created a large number of grid point to represent $B_{\mathbf{x}}$ using 100 grid points for both active and reactive power for a total of 10′000 grid points. For each grid point, we sampled 1'000 realizations of $\boldsymbol{\omega}$. For each (x, ω), we solved a standard power flow using Matpower. We then calculated the probability that a constraint holds for fixed x by dividing the number of samples for which the power flow satisfies the constraints by the total number of samples for ω.

Figure 3.5 shows the feasible region for $\varepsilon_1 = 0.01$ and $\varepsilon_2 = 0.1$. We show both the inner (green) and outer (red) approximation of the feasible region for relaxation orders $d = 2, 3$, and both with and without Stokes constraints. As a benchmark, we also show the feasible region computed through the Monte Carlo simulation (blue). The closer the approximated regions (green and red) are to the benchmark (blue), the better the approximation. We remark that both increasing the relaxation order and introducing Stokes constraints increase the quality of the solution. The improvement obtained by introducing Stokes constraints is very significant, while increasing the relaxation order only slightly increases the quality of the approximation.

To further assess the quality of approximation, we report the ratios between the volume of the approximated feasibility regions and the volume computed through the Monte Carlo simulation in Table 3.1 for $\varepsilon_1 = 0.01$ and different values of ε_2. The addition of Stokes constraints clearly offers significant improvement. Interestingly, the quality of the outer approximation does not seem to depend on the choice of ε_2, while the accuracy of the inner approximation decreases with ε_2.

We observe that the outer approximation to the chance constraints is not very tight, and

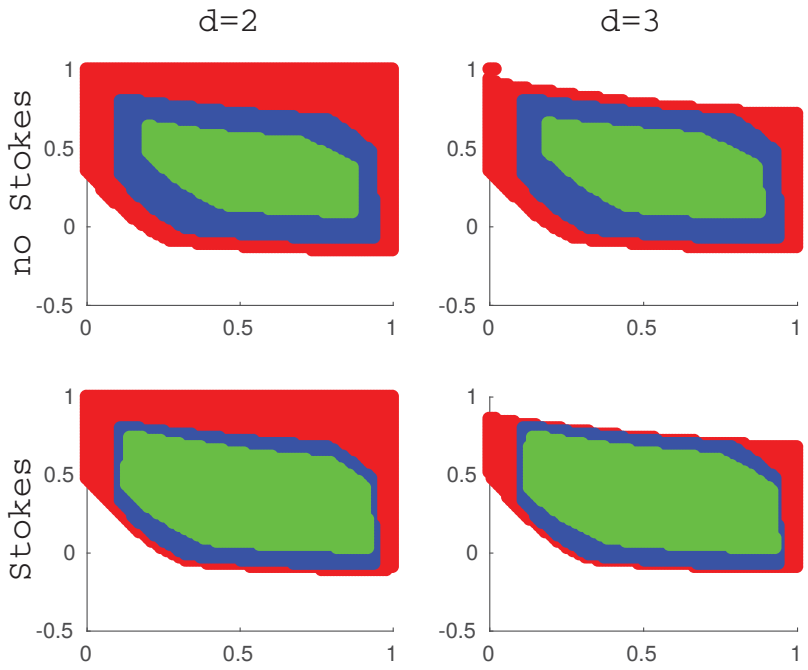

Figure 3.5: Comparison of the outer (red) and inner (green) approximation with the Monte Carlo simulation (blue) for $\varepsilon_1 = 0.01$ and $\varepsilon_2 = 0.1$.

might lead to violation probabilities significantly above the acceptable levels. The extension proposed in this paper to allow for an (approximate) inner approximation provides a significant practical advantage over the previously existing methods in terms of returning safe approximations. It is also accurate enough to provide non-empty feasible sets, even at low relaxation orders.

Solving an instance of a CC-AC-OPF We assess the performance of the ACC-OPNF formulation in (3.21) by evaluating the cost of the optimal generation dispatch, the empirical constraint violation probability and by relating it to the deterministic AC-OPF. For this experiment, we use the best inner approximation with relaxation order $d = 3$ as well as the Stokes constraints to approximate the CC-AC-OPF (3.13). We solve both the deterministic AC-OPF and the approximation of the CC-AC-OPF for different values of ε_2. We then

	outer				inner			
	$d=2$		$d=3$		$d=2$		$d=3$	
ε_2	–	Stokes	–	Stokes	–	Stokes	–	Stokes
0.20	175%	165%	141%	124%	53%	79%	56%	79%
0.15	179%	168%	143%	126%	49%	75%	53%	76%
0.10	182%	171%	144%	126%	43%	69%	47%	70%
0.05	185%	173%	144%	125%	30%	56%	36%	58%

Table 3.1: Ratio approximated vs. real volume for different values of ε_2.

	Det.	$\varepsilon_2 = 20\%$	$\varepsilon_2 = 15\%$	$\varepsilon_2 = 10\%$	$\varepsilon_2 = 5\%$
$p_{G0,1}$	8.5	30.1	36.3	44.7	58.8
$q_{G0,1}$	158.4	168.0	168.2	168.6	169.1
$p_{G0,4}$	500.0	477.6	471.2	462.4	447.9
$q_{G0,4}$	149.5	135.4	134.0	132.1	129.1
cost	13 357	13 452	13 481	13 523	13 596
ε_2^*	39.8%	18.2%	12.1%	3.7%	0.0%

Table 3.2: Optimal values and solutions to (3.11) and (3.21) for $\varepsilon_1 = 0.01$ and different values of ε_2

compare the power injections, the cost and the maximal empirical violation probability of the individual chance constraints ε_2^*, which is computed through another Monte Carlo simulation at the obtained solution point using 1'000 samples of ω.

Table 3.2 summarizes the results. In column *Det.* we show the results for the deterministic AC-OPF. The other columns are labeled by their acceptable violation probability for the individual constraint violation ε_2. The violation probability $\varepsilon_1 = 0.01$ for all experiments. The variables $p_{G0,4}$ and $q_{G0,4}$ are the independent variables $\mathbf{x}$ in our problem formulation, corresponding to the active and the reactive power of the generator at Bus 4 in the test case. The power injections at the slack bus generator $p_{G0,1}$ and $q_{G0,1}$ are among the dependent $\mathrm{y}_{\mathbf{x}}$ variables. Since these generators will adjust their values based on the realization of ω, we report their expected values in the table. Further, we list the cost of the operating point and the maximum empirical violation probabilities ε_2^* among all individual constraints. We do not show results for the empirical violation probability of the joint chance constraint ε_1^*, as it was constantly 0% for all optimal operating points. This is expected, since the engineering limits are typically more limiting than the power flow solvability conditions.

We observe that when the violation probability ε_2 decreases, more and more of the system load must be covered by the more expensive slack generator, resulting in a higher value for $p_{G0,1}$ and a higher expected cost. Considering the violation probabilities of the individual chance constraints we see that the optimal solution to the deterministic AC-OPF violates at least one of these constraints with a probability of almost 40%. For the approximations of the CC-AC-OPF the empirical violation probability ε_2^* of the individual chance constraints is always below the requested probability ε_2, reflecting the fact that we indeed obtain a true inner approximation. While the empirical violation probability is quite close to the acceptable level for $\varepsilon_2 = 20\%$ and $\varepsilon_2 = 15\%$, respectively, the approximation is significantly more conservative for lower values of ε_2. For $\varepsilon_2 = 5\%$ no violations are observed.

3.2.7 Conclusion and Directions

In this section, we developed a new approach to handle chance constrained optimization problems in non-linear physical networks. The method is based on Semidefinite Programming (SDP) techniques to compute the volume of semialgebraic sets, from which polynomial approximations of the chance constraints are obtained. To make existing results applicable in our practical setting, we (i) proposed a set reformulation in order to enable inner approx-

imations, and (ii) developed a two-step procedure to improve approximation quality at a lower computational overhead.

The method is applicable to any problem with polynomial equality and inequality constraints, when the probability distribution of the noise (or more precisely its moment sequence) is known. We next provide a framework for approximations of chance constraints when the moment information is imperfect in Section 3.3.

We have tested our approach numerically on the chance constrained AC Optimal Power Flow. In our experiments, the polynomial approximations were shown to provide sufficiently accurate representations of the feasible domain, and the resulting CC-AC-OPF was able to provide safe operating points with limited violation probability.

The method is a powerful and novel technique to handle chance constrained optimization for non-linear systems. Although, in its current form the method is applicable to systems of small dimensions only, it has the potential for several extensions and improvements. One promising future direction is to exploit the sparsity structure of networks to scale the method to instances of larger dimension.

3.3 Distributionally Robust Chance Constraints

In this section we extend the framework of chance constraint approximation presented so far. We consider the more realistic case where only some information about the probability distribution $\mu^{\boldsymbol{\omega}}$ on Ω is known. In addition we will assume that $K^{\mathrm{c}} := (B_{\mathbf{x}} \times \Omega) \setminus K$ is a basic semialgebraic set. Here, as already in the previous sections, $B_{\mathbf{x}}$ is a basic semialgebraic compact subset of $\mathbb{R}^n$ such that we can compute the moments of the normalized Lebesgue measure $\lambda_{B_{\mathbf{x}}}$. For simplicity of exposition we will only consider the case $\Omega = \mathbb{R}$ and also restrict the discussion to the case that the noise is *normal distributed with respect to unknown parameters* $\mathbf{y} = (\mathbf{m}, \boldsymbol{\sigma})$ representing the mean $\mathbf{m}$ and the standard deviation $\boldsymbol{\sigma}$. More precisely let

$$f(\mathbf{y}, \boldsymbol{\omega}) := \tfrac{1}{\sqrt{2\pi}\boldsymbol{\sigma}} \exp(-\tfrac{(\boldsymbol{\omega}-\mathbf{m})^2}{2\boldsymbol{\sigma}^2}). \tag{3.27}$$

Then we consider the family of probability measures whose density is a mixture of densities $f(\mathrm{y}, \boldsymbol{\omega})$ with y in some basic semialgebraic compact set $Y \subseteq \mathbb{R} \times \mathbb{R}_+ \setminus \{0\}$, i.e, in

$$\begin{aligned} M :=& \left\{ \mu^{\boldsymbol{\omega}} \in \mathcal{P}(\Omega) : \mathbf{d}\mu^{\boldsymbol{\omega}}(\boldsymbol{\omega}) := \left[\int_Y f \, \mathbf{d}\varphi(\mathbf{y}) \right] \mathbf{d}\boldsymbol{\omega}, \ \varphi \in \mathcal{P}(Y) \right\} \\ =& \left\{ \mu^{\boldsymbol{\omega}} \in \mathcal{P}(\Omega) : \mathbf{d}\mu^{\boldsymbol{\omega}}(\boldsymbol{\omega}) := \left[\int_Y \tfrac{1}{\sqrt{2\pi}\boldsymbol{\sigma}} \exp(-\tfrac{(\boldsymbol{\omega}-\mathbf{m})^2}{2\boldsymbol{\sigma}^2}) \, \mathbf{d}\varphi(\mathbf{m}, \boldsymbol{\sigma}) \right] \mathbf{d}\boldsymbol{\omega}, \ \varphi \in \mathcal{P}(Y) \right\}. \end{aligned} \tag{3.28}$$

In [LW18], which this section is based on, the scenario is presented in a broader set up. There, parametrised multivariate distributions are allowed as long as they satisfy some conditions naturally fulfilled by the measures described above. In addition the restriction that K^{c} has to be basic semialgebraic is relaxed to K^{c} being only a semialgebraic set. The following property of Gaussian measures is the *key condition* to compute approximations without knowledge of the reference measure.

Remark 3.3.1. Let f be as defined in (3.27). Then there exists a family of polynomials $(p_k)_{k\in\mathbb{N}} \in \mathbb{R}[\mathbf{y}]$ such that

$$\mathbf{y} \mapsto \int_\Omega \boldsymbol{\omega}^k f(\mathbf{y}, \boldsymbol{\omega})\, \mathbf{d}\boldsymbol{\omega} = p_k(\mathbf{y}).$$

Remember that by K_x we denote the set $\{\omega \in \Omega : (\mathrm{x}, \omega) \in K\}$. The aim of this section is to provide *inner approximations* to the *distributionally robust* chance constraint

$$\mathbb{P}_{\mu^{\boldsymbol{\omega}}}(K_\mathrm{x}) \geq 1 - \varepsilon, \quad \forall \mu^{\boldsymbol{\omega}} \in M. \tag{3.29}$$

The crucial difference between (3.29) and the chance constraint (3.10) is that in (3.29) the probabilistic constraint is with respect to a *whole set* of distributions, rather than the probability with respect to a single known distribution as in (3.10). This means that (3.29) is robust with respect to the family M of distributions - hence the name *distributionally robust*.

Positioning in the literature Knowledge of first and second order moments is a typical assumption when approximating distributionally robust chance constraints [DY10; EI06]. For instance Calafiore and El Ghaoui [CE06] have shown that when g is bilinear and $K = \{(\mathrm{x}, \omega) : g(\mathrm{x}, \omega) \geq 0\}$ then a tractable characterization via second order cone constraints is possible. Recently Yang and Xu [YX16] have considered non-linear optimization problems where the constraint functions are concave in the decision variables and quasi-convex in the uncertain parameters. They show that such problems are tractable if the uncertainty is characterized by its mean and variance only; in the same spirit see also Chao Duan et al. [Dua+18], Xie and Ahmed [XA18] and Zhang et al. [ZSM17] for other tractable formulations of distributionally robust chance-constrained for optimal power flow problems.

In contrast to these approaches, in our set up we do *not* assume perfect knowledge of the first order moments, and can deal with only some approximate knowledge. Notice that in this framework *no* mean, variance or higher order moments have to be estimated. However we assume knowledge of bounds on the parameters defining the family of measures M. Accordingly, our approach can be viewed as an alternative and/or a complement to those considered in e.g. [CE06; DY10; EI06; YX16] when a good estimation of such moments is not possible. Indeed in many cases, providing a box (where the mean vector can lie in) and a possible range $\underline{\delta}\,\mathbf{I} \preceq \Sigma \preceq \overline{\delta}\,\mathbf{I}$ for the covariance matrix Σ, can be more realistic than providing a single mean vector and a single covariance matrix.

Outline The strategy in order to provide approximations of distributionally chance constraints is very similar to what we described in Section 3.1.4. Note however that in (3.4) we needed *perfect knowledge* of *all moments* of the reference measure μ. In the set up of this section we do not even have precise information about the first moments. In Section 3.3.1 we state a linear problem on measures that characterizes the "worst case probability". We will explain how the polynomials p_k from Remark 3.3.1 can be used to compensate this lack of knowledge in Section 3.3.2. Then in Section 3.3.3 we explain how we can add Stokes constraint in order to accelerate convergence as we did before in Section 3.1.4. Finally we

illustrate the theoretic results with some numeric examples. In contrast to the previously cited contributions, the focus however is not on scalability but much more on the generality of the approach. In particular we show that the moment approach can approximate feasible sets X^ε that are non-convex. Even though it follows the same scheme as Section 3.1.4, the approach provided in this section is *not* a direct generalization but rather a non-trivial extension of what has been presented so far.

3.3.1 Approximations via a Moment Approach

As outlined in the introduction and motivated by applications like in Section 3.2, we are mainly interested in *inner* approximations of the chance constrained set $X^\varepsilon := \{\mathrm{x} \in B_{\mathbf{x}} : \mathbb{P}_{\mu^{\boldsymbol{\omega}}}(K_\mathrm{x}) \geq 1-\varepsilon, \forall \mu^{\boldsymbol{\omega}} \in M\}$. As the measure approach (3.1) leads to over approximations of the probability we will therefore work with the complement of K. Let $L := K^\mathrm{c} := (B_{\mathbf{x}} \times \Omega) \setminus K$ be a basic semialgebraic set and $L_{\mathbf{x}} := \{\omega \in \Omega : (\mathbf{x}, \omega) \in L\}$. Then we have that

$$X^\varepsilon = \{\mathrm{x} \in B_{\mathbf{x}} : \mathbb{P}_{\mu^{\boldsymbol{\omega}}}(L_\mathrm{x}) \leq \varepsilon, \forall \mu^{\boldsymbol{\omega}} \in M\}. \tag{3.30}$$

We want to compute overestimators of the probabilities $\mathbb{P}_{\mu^{\boldsymbol{\omega}}}(L_{\mathbf{x}})$ for every $\mu^{\boldsymbol{\omega}} \in M$ in order to approximate X^ε from the interior. Define the function

$$\rho(\mathbf{x}) := \sup_{\mu^{\boldsymbol{\omega}} \in M} \mathbb{P}_{\mu^{\boldsymbol{\omega}}}(L_{\mathbf{x}}) \tag{3.31}$$

which can be thought of as the *worst case probability*, i.e., given $\mathrm{x} \in B_{\mathbf{x}}$, $\rho(\mathrm{x})$ is the probability of L_x with respect to the measure $\mu^{\boldsymbol{\omega}} \in M$ that maximizes $\mathbb{P}_{\mu^{\boldsymbol{\omega}}}(L_\mathrm{x})$ or equivalently minimizes $\mathbb{P}_{\mu^{\boldsymbol{\omega}}}(K_\mathrm{x})$.[3] In consequence, comparing to (3.30), $X^\varepsilon = \{\mathrm{x} \in B_{\mathbf{x}} : \rho(\mathrm{x}) \leq \varepsilon\}$. In the sequel we compute polynomial over-approximations $h_d \in \mathbb{R}[\mathbf{x}]$ of ρ and establish L^1 convergence of $h_d \to \rho$ for $d \to \infty$. From these polynomials we can define basic semialgebraic sets

$$X_d^\varepsilon := \{\mathrm{x} \in B_{\mathbf{x}} : h_d(\mathrm{x}) \leq \varepsilon\} \subseteq X^\varepsilon,$$

approximating X^ε from the interior. In addition we prove $\mathrm{vol}(X^\varepsilon \setminus X_d^\varepsilon) \to 0$ when $d \to \infty$.

Note that by defining ρ as in (3.31) we commit to computing inner approximations. Unlike as in Section 3.1.4, we cannot yield outer approximations of X^ε by interchanging the roles of K and L. In order to achieve outer approximations, we would need to allow the measures $\mu^{\boldsymbol{\omega}} \in M$ to violate the probability individually and could not gather the information of all measures in one function as we do in (3.31).

Identification and Characterization of Reference Measure

As explained above, in order to define a reference measure it is sufficient to focus on the worst case probability. However, in contrast to Section 3.1.4 where the reference measure μ was defined as the product measure $\lambda_{B_{\mathbf{x}}} \otimes \mu^{\boldsymbol{\omega}}$ for a single *known* $\mu^{\boldsymbol{\omega}} \in \mathcal{P}(\Omega)$, now the

[3] For a given $\mathrm{x} \in B_{\mathbf{x}}$, the measure maximizing $\mathbb{P}_{\mu^{\boldsymbol{\omega}}}(L_\mathrm{x})$ minimizes $\mathbb{P}_{\mu^{\boldsymbol{\omega}}}(K_\mathrm{x})$ and hence is the critical measure or the "worst case" for the original chance constraint (3.29).

probability part of μ will depend on $\mathbf{x}$, i.e. $\mu = \lambda_{B_\mathbf{x}} \mu_\mathbf{x}$, where the unknown μ_x denotes the *conditional* or the *stochastic kernel* of μ given x and $\mu_\mathrm{x} \in M$ for each $\mathrm{x} \in B_\mathbf{x}$. Consequently we need to identify the measure $\mu_\mathrm{x} \in M$ such that $\mathbb{P}_{\mu_\mathrm{x}}(L_\mathrm{x}) = \rho(\mathrm{x})$ for each $\mathrm{x} \in B_\mathbf{x}$, respectively. Note that the choice of μ_x is not unique if $L_\mathrm{x} = \emptyset$.

Lemma 3.3.2. *The function ρ is measurable. In particular there exists a measurable selector $y : B_\mathbf{x} \to Y$ such that $\rho(\mathbf{x}) = \mathbb{P}_{\mu_{y(\mathbf{x})}}(L_\mathbf{x})$ where $\mu_{y(\mathbf{x})} := f(\boldsymbol{\omega}, y(\mathrm{x}))\lambda_\Omega \in M$ for every $\mathrm{x} \in B_\mathbf{x}$.*

Proof. Remark upfront that the existence of a measurable selector $y : B_\mathbf{x} \to Y$ and the equality $\rho(\mathrm{x}) = \int_{L_\mathrm{x}} f(\boldsymbol{\omega}, y(\mathrm{x}))\, \mathbf{d}\boldsymbol{\omega}$ imply that ρ is measurable. Let $\mathrm{x} \in B_\mathbf{x}$ be fixed. We show that $\rho(\mathrm{x}) := \sup_{\mu^\omega \in M} \mathbb{P}_{\mu^\omega}(L_\mathrm{x}) = \max_{\mathrm{y} \in Y} \int_{L_\mathrm{x}} f(\boldsymbol{\omega}, \mathrm{y})\, \mathbf{d}\boldsymbol{\omega}$. To see that the maximum on the right hand side is attained, it suffices to note that $\mathrm{y} \to \int_{L_\mathrm{x}} f(\boldsymbol{\omega}, \mathrm{y})\, \mathbf{d}\boldsymbol{\omega}$ is continuous (see (3.27) for the definition of f). It is clear that $\sup_{\mu^\omega \in M} \mathbb{P}_{\mu^\omega}(L_\mathrm{x}) \geq \max_{\mathrm{y} \in Y} \int_{L_\mathrm{x}} f(\boldsymbol{\omega}, \mathrm{y})\, \mathbf{d}\boldsymbol{\omega}$ as $\{\delta_\mathrm{y} : \mathrm{y} \in Y\} \subseteq \mathcal{P}(Y)$. For the converse inequality consider

$$\begin{aligned}
\sup_{\mu^\omega \in M} \mathbb{P}_{\mu^\omega}(L_\mathrm{x}) &= \sup_{\varphi \in \mathcal{P}(Y)} \int_{L_\mathrm{x}} \int_Y f(\boldsymbol{\omega}, \mathbf{y})\, \mathbf{d}\varphi(\mathbf{y})\, \mathbf{d}\boldsymbol{\omega} \\
&= \sup_{\varphi \in \mathcal{P}(Y)} \int_Y \underbrace{\int_{L_\mathrm{x}} f(\boldsymbol{\omega}, \mathbf{y})\, \mathbf{d}\boldsymbol{\omega}}_{=:F(\mathbf{y})}\, \mathbf{d}\varphi(\mathbf{y}) \\
&\leq \sup_{\varphi \in \mathcal{P}(Y)} \max_{\mathrm{y} \in Y} F(\mathrm{y}) \underbrace{\int_Y \mathbf{d}\varphi(\mathbf{y})}_{=1} \\
&= \max_{\mathrm{y} \in Y} \int_{L_\mathrm{x}} f(\boldsymbol{\omega}, \mathrm{y})\, \mathbf{d}\boldsymbol{\omega}.
\end{aligned}$$

Now, the assertion follows by [Las10a, Proposition 4.4]. □

Measure Formulation of the Worst Case Probability

Lemma 3.3.2 only states the existence of a function y such that $\mu_{y(\mathrm{x})}$ provides the worst case probability $\rho(\mathrm{x})$ for every $\mathrm{x} \in B_\mathbf{x}$. In order to use those measures to dominate the optimization variable ϕ in (3.1), we somehow need an access to this function y. Consider the linear operator $T : \mathcal{B}(B_\mathbf{x} \times \Omega) \to \mathcal{B}(B_\mathbf{x} \times Y)$ by

$$Tg(\mathrm{x}, \mathrm{y}) = \int_\Omega g(\mathrm{x}, \boldsymbol{\omega}) f(\boldsymbol{\omega}, \mathrm{y})\, \mathbf{d}\boldsymbol{\omega}.$$

For any real space $\mathcal{X}$ the bilinear form $\langle ., . \rangle : \mathcal{M}(\mathcal{X}) \times \mathcal{B}(\mathcal{X}) \to \mathbb{R}$, $\langle \mu, g \rangle := \int_\mathcal{X} g\mu$ is a dual pairing for $\mathcal{M}(\mathcal{X})$ and $\mathcal{B}(\mathcal{X})$ (compare [HL99]). Hence T induces an adjoint operator $T^* : \mathcal{M}(B_\mathbf{x} \times Y) \to \mathcal{M}(B_\mathbf{x} \times \Omega)$ by

$$\langle T^*\nu, g \rangle = \langle \nu, Tg \rangle, \quad \forall \nu \in \mathcal{M}(B_\mathbf{x} \times Y), \forall g \in \mathcal{B}(B_\mathbf{x} \times \Omega).$$

To see that T^* is well defined, let $\nu \in \mathcal{M}(B_\mathbf{x} \times Y)$, and write $\nu = \nu_\mathbf{x} \nu^\mathbf{x}$ where $\nu_\mathbf{x}$ denotes the conditional of ν on Y knowing $\mathbf{x}$ and $\nu^\mathbf{x}$ is the marginal of ν on $\mathbf{x}$. Then for all

$g \in \mathcal{B}(B_{\mathbf{x}} \times \Omega)$ it holds that

$$\begin{aligned}\langle \nu, Tg \rangle &= \int_{B_{\mathbf{x}}} \int_Y \int_\Omega g(\mathbf{x},\boldsymbol{\omega}) f(\boldsymbol{\omega},\mathbf{y})\, \mathbf{d}\boldsymbol{\omega}\, \mathbf{d}\nu_{\mathbf{x}}(\mathbf{y})\, \mathbf{d}\nu^{\mathbf{x}}(\mathbf{x}) \\ &= \int_{B_{\mathbf{x}}} \int_\Omega g(\mathbf{x},\boldsymbol{\omega}) \underbrace{\int_Y f(\boldsymbol{\omega},\mathbf{y})\, \mathbf{d}\nu_{\mathbf{x}}(\mathbf{y})}_{=\theta(\mathbf{x},\boldsymbol{\omega})}\, \mathbf{d}\boldsymbol{\omega}\, \mathbf{d}\nu^{\mathbf{x}}(\mathbf{x}) \\ &= \int_{B_{\mathbf{x}}} \int_\Omega g(\mathbf{x},\boldsymbol{\omega}) \theta(\mathbf{x},\boldsymbol{\omega})\, \mathbf{d}\boldsymbol{\omega}\, \mathbf{d}\nu^{\mathbf{x}}(\mathbf{x}) \\ &=: \langle T^*\nu, g \rangle\,.\end{aligned}$$

In particular we see that T^* provides a more direct access to y. Denote $\nu := \delta_{y(\mathbf{x})}\lambda_{B_{\mathbf{x}}}$. Then, $T^*\nu = f(\boldsymbol{\omega}, y(\mathbf{x}))\, \mathbf{d}\boldsymbol{\omega}\ \lambda_{B_{\mathbf{x}}} = \mu_{y(\mathbf{x})}\lambda_{B_{\mathbf{x}}}$. The final idea to define the reference measure μ now is very similar to the idea of polynomial optimization presented in Chapter 1: we relax $\delta_{y(\mathrm{x})}$ to an arbitrary probability measure supported on Y for each $\mathrm{x} \in B_{\mathbf{x}}$ and show that "we do not relax too much": Consider

$$\begin{aligned}&\sup_{\phi,\nu} \int_L \mathbf{d}\phi \\ &\text{s.t.} \quad \phi \le T^*\nu, \quad \nu^{\mathbf{x}} = \lambda_{B_{\mathbf{x}}}, \\ &\qquad \phi \in \mathcal{M}_+(L), \nu \in \mathcal{P}(B_{\mathbf{x}} \times Y),\end{aligned} \tag{3.32}$$

where we recall that $\nu^{\mathbf{x}}$ denotes the marginal of ν with respect to $\mathbf{x}$ and $\lambda_{B_{\mathbf{x}}}$ is the Lebesgue measure on $B_{\mathbf{x}}$ scaled to be a probability measure.

Lemma 3.3.3. *The optimal value of* (3.32) *is* $\int_L \rho\, \mathbf{dx}$ *and is attained by* $\phi^* := \mu_{y(\mathbf{x})}\lambda_{B_{\mathbf{x}}}|_L$ *and* $\nu^* := \delta_{y(\mathbf{x})}\lambda_{B_{\mathrm{x}}}$.

Proof. It is clear from the discussion above that ϕ^* and ν^* are feasible for (3.32). Further by Lemma 3.3.2

$$\int_L \mathbf{d}\phi^* = \int_L \mathbf{d}(\mu_{y(\mathbf{x})}\lambda_{B_{\mathbf{x}}}) = \int_L \rho\, \mathbf{dx}.$$

To see that ϕ^* and ν^* are optimal, let ϕ, ν be feasible. Then

$$\begin{aligned}\int_L \mathbf{d}\phi &\le \int_L \mathbf{d}T^*\nu = \int_{B_{\mathbf{x}}} \int_\Omega \mathbf{1}_L(\mathbf{x},\boldsymbol{\omega}) \int_Y f(\mathbf{x},\mathbf{y})\, \mathbf{d}\nu_{\mathbf{x}}(\mathbf{y})\, \mathbf{d}\boldsymbol{\omega}\, \mathbf{dx} \\ &\le \int_{B_{\mathbf{x}}} \int_\Omega \mathbf{1}_L(\mathbf{x},\boldsymbol{\omega}) f(\mathbf{x}, y(\mathbf{x}))\, \mathbf{d}\boldsymbol{\omega}\, \mathbf{dx} = \int_L \mathbf{d}(\mu_{y(\mathbf{x})}\lambda_{B_{\mathbf{x}}}) = \int_L \rho\, \mathbf{dx}.\end{aligned}$$

□

Before going on in the discussion let us sumarize what we have achieved in this section. First, to approximate the distributionally robust chance constraint (3.29) it suffices to approximate the worst case probability ρ defined in (3.31), which turns out to be a measurable function, due to the existence of a measurable selector $y : B_{\mathbf{x}} \to Y$. Using this selector it is possible to characterize $\int_{B_{\mathbf{x}}} \rho\, \mathbf{dx}$ as the optimal value of the linear problem (3.32). Following the strategy from Section 3.1 we next need to show that this problem has an equivalent formulation as a GMP so that we can approximate its value by the moment-SOS hierarchy. For this the polynomials p_k in Remark 3.3.1 play an important role. We expect that from

the dual formulation we can compute the polynomial h defining an inner approximation of X^ε.

3.3.2 Moment Formulation

Reformulating (3.32) as a generalized moment problem is not trivial because of i) Ω is unbounded and ii) in consequence the operator T is a priori not defined for polynomials. The goal of this section is to show that we can cope with these two issues.

Lemma 3.3.4. *The operator T extends to polynomials and $T(\mathbb{R}[\mathbf{x},\boldsymbol{\omega}]) \subseteq \mathbb{R}[\mathbf{x},\mathbf{y}]$.*

Proof. By linearity it is enough to show that $T(\mathbf{x}^\alpha\boldsymbol{\omega}^k)$ is well defined for all $\alpha \in \mathbb{N}^n$ and $k \in \mathbb{N}$ and $T(\mathbf{x}^\alpha\boldsymbol{\omega}^k) \in \mathbb{R}[\mathbf{x},\mathbf{y}]$. Let α, k be fixed and let p_k be as in Remark 3.3.1. Then

$$\left[T(\mathbf{x}^\alpha\boldsymbol{\omega}^k)\right](\mathbf{x},\mathbf{y}) = \int_\Omega \mathbf{x}^\alpha\boldsymbol{\omega}^k f(\boldsymbol{\omega},\mathbf{y})\,\mathbf{d}\boldsymbol{\omega} = \mathbf{x}^\alpha p_k(\mathbf{y}) \in \mathbb{R}[\mathbf{x},\mathbf{y}].$$

□

Note that as $B_\mathbf{x}$ is assumed to be compact, the equality $\nu^\mathbf{x} = \lambda_{B_\mathbf{x}}$ in (3.32) can be imposed by equating all moments, i.e., $\int \mathbf{x}^\alpha\,\mathbf{d}\nu^\mathbf{x} = \int \mathbf{x}^\alpha\,\mathbf{d}\lambda_{B_\mathbf{x}}$, for all $\alpha \in \mathbb{N}^n$. For the domination constraint $\phi \le T^*\nu$ we introduce a slack variable $\hat\phi \in \mathcal{M}_+(B)$ and write $\phi + \hat\phi = T^*\nu$. Then we use the following result.

Lemma 3.3.5. *Let $\mu \in \mathcal{M}_+(B_\mathbf{x}\times\Omega)$ and $\nu \in \mathcal{M}_+(B_\mathbf{x}\times Y)$ such that $\nu^\mathbf{x} = \lambda_{B_\mathbf{x}}$. Then $\int \mathbf{x}^\alpha\boldsymbol{\omega}^k\,\mathbf{d}\mu = \int \mathbf{x}^\alpha p_k\,\mathbf{d}\nu$ for all $\alpha \in \mathbb{N}^n$ and $k \in \mathbb{N}$ implies that $\mu = T^*\nu$.*

Proof. First remark that taking $k = 0$ by compactness of $B_\mathbf{x}$ we have $\mu^\mathbf{x} = \nu^\mathbf{x} = \lambda_{B_\mathbf{x}}$. Disintegrate μ and ν to $\mu_\mathbf{x}\lambda_{B_\mathbf{x}}$ and $\nu_\mathbf{x}\lambda_{B_\mathbf{x}}$, respectively and let α and k be fixed. Then by assumption

$$\int_{B_\mathbf{x}} \mathbf{x}^\alpha \int_\Omega \boldsymbol{\omega}^k\,\mathbf{d}\mu_\mathbf{x}(\boldsymbol{\omega})\,\mathbf{d}\mathbf{x} = \int_{B_\mathbf{x}} \mathbf{x}^\alpha \int_Y p_k\,\mathbf{d}\nu_\mathbf{x}(\mathbf{y})\,\mathbf{d}\mathbf{x}.$$

Then for each k fixed $\int_\Omega \boldsymbol{\omega}^k\,\mathbf{d}\mu_\mathbf{x}(\boldsymbol{\omega}) = \int_Y p_k\,\mathbf{d}\nu_\mathbf{x}(\mathbf{y})$ for almost all $\mathrm{x} \in B_\mathbf{x}$. This in turn implies by countably additivity of $\lambda_{B_\mathbf{x}}$ that the equality holds for all k for almost all $\mathrm{x} \in B_\mathbf{x}$. Recall that the moments of any Gaussian measure satisfy Carleman's Condition Definition 3.1.2 and in consequence the measures in M are moment determined. Hence

$$\int_\Omega \boldsymbol{\omega}^k\,\mathbf{d}\mu_\mathbf{x}(\boldsymbol{\omega}) = \int_Y p_k\,\mathbf{d}\nu_\mathbf{x}(\mathbf{y}) = \int_\Omega \boldsymbol{\omega}^k \underbrace{\int_Y f(\boldsymbol{\omega},\mathbf{y})\,\mathbf{d}\nu_\mathbf{x}(\mathbf{y})\,\mathbf{d}\boldsymbol{\omega}}_{\in M} \quad \text{a.-s.}$$

implies that $\mathbf{d}\mu_\mathbf{x}(\boldsymbol{\omega}) = \int_Y f(\boldsymbol{\omega},\mathbf{y})\,\mathbf{d}\nu_\mathbf{x}(\mathbf{y})\,\mathbf{d}\boldsymbol{\omega}$ for almost all $\mathrm{x} \in B_\mathbf{x}$. Putting all things together, we finally have for all $h \in C_c(B_\mathbf{x}\times\Omega)$:

$$\begin{aligned}
\langle \mu, h\rangle &= \langle \mu_\mathbf{x}\lambda_{B_\mathbf{x}}, h\rangle \\
&= \int_{B_\mathbf{x}\times\Omega} h(\mathbf{x},\boldsymbol{\omega}) \int_Y f(\boldsymbol{\omega},\mathbf{y})\,\mathbf{d}\nu_\mathbf{x}(\mathbf{y})\,\mathbf{d}\boldsymbol{\omega}\,\mathbf{d}\mathbf{x} \\
&= \int_{B_\mathbf{x}\times Y} \underbrace{\int_\Omega h(\mathbf{x},\boldsymbol{\omega}) f(\boldsymbol{\omega},\mathbf{y})\,\mathbf{d}\boldsymbol{\omega}}_{=Th(\mathbf{x},\mathbf{y})} \underbrace{\mathbf{d}\nu_\mathbf{x}(\mathbf{y})\,\mathbf{d}\mathbf{x}}_{=\mathbf{d}\nu(\mathbf{x},\mathbf{y})}
\end{aligned}$$

$$= \quad \langle Th, \nu \rangle = \langle T^*\nu, h \rangle \,.$$

□

Thanks to Lemma 3.3.5 we can reformulate (3.32) as the following GMP:

$$\begin{aligned}
\sup_{\phi,\hat{\phi},\nu} \quad & \int_L \mathbf{d}\phi \\
\text{s.t.} \quad & \left\langle \phi, \mathbf{x}^\alpha \boldsymbol{\omega}^k \right\rangle + \left\langle \hat{\phi}, \mathbf{x}^\alpha \boldsymbol{\omega}^k \right\rangle = \langle \nu, \mathbf{x}^\alpha p_k \rangle \,, \quad \forall \alpha \in \mathbb{N}^n,\ k \in \mathbb{N} && \text{(3.33a)} \\
& \langle \nu, \mathbf{x}^\alpha \rangle = \langle \lambda_{B_\mathbf{x}}, \mathbf{x}^\alpha \rangle \,, \quad \forall \alpha \in \mathbb{N}^n && \text{(3.33b)} \\
& \phi \in \mathcal{M}_+(L),\ \hat{\phi} \in \mathcal{M}_+(B_\mathbf{x} \times \Omega),\ \nu \in \mathcal{P}(B_\mathbf{x} \times Y)
\end{aligned}$$

We will see in the following that very similar as in Section 3.1.4 we can use the dual problem of (3.33) in order to define polynomial approximations h_d of ρ.

The Dual Problem

Consider the dual problem of (3.33), which reads as follows.

$$\begin{aligned}
\inf_{g,h} \quad & \int_{B_\mathbf{x}} h \, \mathbf{dx} \\
\text{s.t.} \quad & h - Tg \geq 0 \text{ on } B_\mathbf{x} \times Y && \text{(3.34a)} \\
& g \geq 0 \text{ on } B_\mathbf{x} \times \Omega && \text{(3.34b)} \\
& g \geq 1 \text{ on } L && \text{(3.34c)} \\
& h \in \mathbb{R}[\mathbf{x}], \quad g \in \mathbb{R}[\mathbf{x}, \boldsymbol{\omega}].
\end{aligned}$$

Proposition 3.3.6. *Problem (3.34) is feasible. In addition for any feasible point (h, g) and any $\mathrm{x} \in B_\mathbf{x}$ it holds that $h(\mathrm{x}) \geq \rho(\mathrm{x})$.*

Proof. Note that by Lemma 3.3.4 $T1 = 1$ and hence $(1, 1)$ is feasible for (3.34). Let (h, g) be an arbitrary feasible point and $\mathrm{x} \in B_\mathbf{x}$ fixed. Recall the definition of the function $y : B_\mathbf{x} \to Y$ from Lemma 3.3.2. Now, as $(\mathrm{x}, y(\mathrm{x})) \in B_\mathbf{x} \times Y$, constraint (3.34a) yields:

$$\begin{aligned}
h(\mathrm{x}) \geq Tg(\mathrm{x}, y(\mathrm{x})) \quad &= \quad \int_\Omega g(\mathrm{x}, \boldsymbol{\omega}) f(\boldsymbol{\omega}, y(\mathrm{x})) \, \mathbf{d}\boldsymbol{\omega} \\
&= \quad \int_{K_\mathrm{x}} \underbrace{g(\mathrm{x}, \boldsymbol{\omega})}_{\geq 0 \text{ (3.34b)}} f(\boldsymbol{\omega}, y(\mathrm{x})) \, \mathbf{d}\boldsymbol{\omega} + \int_{L_\mathrm{x}} \underbrace{g(\mathrm{x}, \boldsymbol{\omega})}_{\geq 1 \text{ (3.34c)}} f(\boldsymbol{\omega}, y(\mathrm{x})) \, \mathbf{d}\boldsymbol{\omega} \\
&\geq \quad \int_{L_\mathrm{x}} f(\boldsymbol{\omega}, y(\mathrm{x})) \, \mathbf{d}\boldsymbol{\omega} = \rho(\mathrm{x}).
\end{aligned}$$

□

Proposition 3.3.6 enables us to define inner approximations of the chance constrained set $X^\varepsilon = \{\mathrm{x} \in B_\mathbf{x} : \rho(\mathrm{x}) \leq \varepsilon\}$ by replacing ρ with any h which is feasible for (3.34). In order to compute such feasible polynomials we apply the moment-SOS hierarchy Section 1.4. More precisely, we compute a relaxation of (3.33) involving only moments up to order $2d$. An optimal solution of the d-th SOS-stregthening of (3.34) provides us with a polynomial h_d

which is feasible for (3.34). Hence, by Proposition 3.3.6, the set $X_d^\varepsilon = \{\mathrm{x} \in B_{\mathbf{x}} : h_d(\mathrm{x}) \leq \varepsilon\}$ is an inner approximation of X^ε.

Convergence of the Hierarchy

In this section we show that under Assumption 3 the moment relaxations to (3.33) are feasible and that strong duality holds on each step of the relaxation. Moreover we establish L^1-convergence of h_d to ρ, making h_d a reasonable choice for the approximation. In particular this implies that $\mathrm{vol}(X^\varepsilon \setminus X_d^\varepsilon) \to 0$ when $d \to \infty$.

Assumption 3. *Let $N \in \mathbb{N}$ be large enough such that $B_{\mathbf{x}} \subseteq \{\mathrm{x} \in \mathbb{R}^n : N - ||\mathrm{x}||^2 \geq 0\}$ and similarly $Y \subseteq \{\mathrm{y} \in \mathbb{R}^2 : N - ||\mathrm{y}||^2 \geq 0\}$. We assume that these redundant constraints are part of the semialgebraic description of $B_{\mathbf{x}}$ and Y, respectively.*

Denote by (P_d) the moment relaxation of (3.33) with moments up to order $2d$, and by (D_d) its dual program, the SOS-strengthening of (3.34) (see Section 1.4). Let d_0 be the maximal degree of the polynomials involved in the semialgebraic description of L and $B_{\mathrm{x}} \times Y$.

Lemma 3.3.7. *Let Assumption 3 hold. Then for each $d \geq d_0$ the optimal value $(P_d)^*$ of the semidefinite relaxation (P_d) is attained. Moreover, $(P_d)^* \to (3.33)^* = \int_{B_{\mathbf{x}}} \rho \, \mathbf{dx}$ as $d \to \infty$.*

Proof. The proof follows the lines of the proof of Theorem 1.3. However there are two important differences. First, Ω is unbounded and we cannot use the strategy of the proof of Theorem 1.3 to show that the feasible set of each SDP relaxation (P_d) is compact. Second, we cannot use Putinar's Theorem to argue that the limit sequence is a moment sequence. We address both points in the following.

Denote by $\mathrm{z}^{\phi,d}$, $\mathrm{z}^{\hat{\phi},d}$, and $\mathrm{z}^{\nu,d}$ the sequences corresponding respectively to the measures ϕ, $\hat{\phi}$, and ν at relaxation (P_d).

1.) As ν is supported on a compact set and because of Assumption 3, $\mathrm{z}_\alpha^{\nu,d}$ is bounded for each $|\alpha| \leq d$ by the same argument as in the proof of Theorem 1.3. Now, $\mathrm{z}_\alpha^{\phi,d}$ and $\mathrm{z}_\alpha^{\hat{\phi},d}$ are bounded by a linear combination of $\mathrm{z}_\beta^{\nu,d}$, $\beta \in \mathbb{N}^{n+1}$ due to (3.33a). This shows that the feasible set of each (P_d) is compact and for each $d \geq d_0$ the optimal value $(P_d)^*$ of the semidefinite relaxation (P_d) is attained.

2.) Let now $\mathrm{z}^{\phi,d}$, $\mathrm{z}^{\hat{\phi},d}$, and $\mathrm{z}^{\nu,d}$ denote the optimal sequences for each relaxation (P_d). Knowing a priory bounds for $\mathrm{z}_\alpha^{\phi,d}$, $\mathrm{z}_\alpha^{\hat{\phi},d}$, and $\mathrm{z}_\alpha^{\nu,d}$ for each α and each d we can construct related sequences contained in the unit ball of ℓ_∞, for all d. By Banach-Alaoglu theorem [Bre10, Theorem 3.16] these sequences have weakly star converging subsequences. After re-normalization, we obtain sequences $\mathrm{z}^{\phi,*}$, $\mathrm{z}^{\hat{\phi},*}$, and $\mathrm{z}^{\nu,*}$ for which it holds in particular that $\mathrm{z}_\alpha^{\phi,d} \to \mathrm{z}_\alpha^{\phi,*}$, $\mathrm{z}_\alpha^{\hat{\phi},d} \to \mathrm{z}_\alpha^{\hat{\phi},*}$, and $\mathrm{z}_\alpha^{\nu,d} \to \mathrm{z}_\alpha^{\nu,*}$ for the corresponding subsequences. By linearity of the Riesz functional, these limit sequences satisfy Carleman's Condition (Definition 3.1.2) as $\mathrm{z}^{\phi,d}$, $\mathrm{z}^{\hat{\phi},d}$, and $\mathrm{z}^{\nu,d}$ satisfy Carleman's Condition. □

Theorem 3.1. *Let Assumption 3 hold and $d \geq d_0$.*

1. *If K, L, and Y have non-empty interior, then (D_d) has an optimal solution and $(D_d)^* = (P_d)^*$.*
2. *Let h_d be part of an optimal solution to (D_d). Then $\int_{B_{\mathbf{x}}} |h_d - \rho| \, \mathbf{dx} \to 0$ for $d \to \infty$.*

Proof. The first statement uses standard arguments from duality in conic optimization. In fact we only need to show that the feasible set of the SDP-relaxations (P_d) have a strictly feasible point. This is called Slater's condition and implies the assertion. The second statement is a direct consequence of the first one. Just recall that by Proposition 3.3.6 $h_d \geq \rho$ on $B_{\mathbf{x}}$ and

$$\int_{B_{\mathbf{x}}} h_d \, \mathbf{dx} = (D_d)^* = (P_d)^* \to \int_{B_{\mathbf{x}}} \rho \, \mathbf{dx}, \quad d \to \infty.$$

To see that (P_d) for $d \geq d_0$ has a strictly feasible point consider the measures $\phi, \hat{\phi}$ and ν given by

$$\nu := \lambda_Y \otimes \lambda_{\mathcal{B}_{\mathbf{x}}}, \quad \phi := \mathbf{1}_L(\mathbf{x}, \boldsymbol{\omega}) T^* \nu \quad \text{and} \quad \hat{\phi} := \mathbf{1}_K(\mathbf{x}, \boldsymbol{\omega}) T^* \nu.$$

By construction, all truncated moment sequences of these measures respect the moment constraints in the relaxation (P_d) of (3.33). As Y is assumed to have non empty interior, and ν is the uniform measure on $Y \times \mathcal{B}_{\mathbf{x}}$, the moment and localization matrices for the moments of ν are (strictly) positive definite for all $d \geq d_0$. To see that the moment and localization matrices for the moment sequences of ϕ and $\hat{\phi}$ are also positive definite, recall that $\mathbf{d}T^*\nu(\mathbf{x}, \boldsymbol{\omega}) = \int_Y f(\mathbf{y}, \boldsymbol{\omega}) \lambda_Y \lambda_{\mathcal{B}_{\mathbf{x}}}$ and f is a strictly positive density. Consequently, the sequences of moments of $\phi, \hat{\phi}$ and ν up to order d are a strictly feasible point for (P_d) for each $d \geq d_0$. □

3.3.3 Distributionally Robust Stokes Constraints

Having a look at (3.34) we face a similar problem as in Section 3.1. The polynomial g is trying to approximate the indicator function of L on $B_{\mathbf{x}} \times \Omega$. We hence expect the convergence of the hierarchy to be slow due to a Gibbs' phenomenon for g. Therefore we want to apply Stokes constraints similar to Section 3.1.2 in order to accelerate convergence. As we explain next it will be necessary to extend the measures ϕ in (3.33) from $\mathcal{M}_+(L)$ to $\mathcal{M}_+(L \times Y)$. We will explain this in the following.

As outlined in Section 3.1.4 we are only able to use Stokes constraints in the direction $\boldsymbol{\omega}$, as in the other directions the meaning of h_d would be destroyed. Let therefore $\vartheta \in \mathbb{R}[\mathbf{x}, \boldsymbol{\omega}]$ be a polynomial such that $\vartheta \mathbf{n}_{n+1}$ vanishes where $\mathbf{n}$ is the outer normal vector of L written in the standard basis of $\mathbb{R}^{n+1} \supset B_{\mathbf{x}} \times \Omega$. Recall that $\mathbf{y} = (\mathbf{m}, \boldsymbol{\sigma})$ and define $q_k := \boldsymbol{\sigma}^2 \frac{\partial}{\partial \boldsymbol{\omega}} (\boldsymbol{\omega}^k \vartheta) - \boldsymbol{\omega}^k \vartheta(\boldsymbol{\omega} - \mathbf{m}) \in \mathbb{R}[\mathbf{x}, \mathbf{y}, \boldsymbol{\omega}]$. Then as a consequence of $\int_{L_{\mathrm{x}}} \frac{\partial}{\partial \boldsymbol{\omega}} \left(\boldsymbol{\omega}^k \vartheta(\mathrm{x}, \boldsymbol{\omega}) f(\mathrm{y}, \boldsymbol{\omega}) \right) \mathbf{d}\boldsymbol{\omega} = 0$, we obtain

$$\int_{L_{\mathbf{x}}} q_k(\mathrm{x}, \mathrm{y}, \boldsymbol{\omega}) \, \mathbf{d}\mu^{\boldsymbol{\omega}}{}_{y(\mathrm{x})} = 0, \quad \forall \mathrm{x} \in B_{\mathbf{x}}, \forall \mathrm{y} \in Y. \tag{3.35}$$

The fact that $q_k \in \mathbb{R}[\mathbf{x}, \mathbf{y}, \boldsymbol{\omega}]$ explains, why we need to extend the support of the optimization measure ϕ if we want to use the additional property (3.35). Indeed, in order to integrate the polynomials q_k with respect to ϕ we need this measure to be defined over $B_{\mathbf{x}} \times \Omega \times Y$. Equation (3.35) yields a family of Stokes constraints similar as explained in Section 3.1.2. The following GMP is an extended version of (3.33).

$$\sup_{\phi, \hat{\phi}, \nu} \quad \int_{L \times Y} \mathbf{d}\phi$$

$$
\begin{aligned}
\text{s.t.} \quad & \left\langle \phi, \mathbf{x}^\alpha \boldsymbol{\omega}^k \right\rangle + \left\langle \hat{\phi}, \mathbf{x}^\alpha \boldsymbol{\omega}^k \right\rangle = \langle \nu, \mathbf{x}^\alpha p_k \rangle, \quad \forall (\alpha, k) \in \mathbb{N}^{n+1}, && \text{(3.36a)}\\
& \langle \nu, \mathbf{x}^\alpha \rangle = \langle \lambda_{B_\mathbf{x}}, \mathbf{x}^\alpha \rangle, \quad \forall \alpha \in \mathbb{N}^n, && \text{(3.36b)}\\
& \left\langle \phi, \mathbf{x}^\alpha \mathbf{y}^\beta q_k \right\rangle = 0, \quad \forall (\alpha, \beta, k) \in \mathbb{N}^{n+3}, && \text{(3.36c)}\\
& \phi \in \mathcal{M}_+(L \times Y),\ \hat{\phi} \in \mathcal{M}_+(B_\mathbf{x} \times \Omega),\ \nu \in \mathcal{P}(B_\mathbf{x} \times Y).
\end{aligned}
$$

Indeed the constraints (3.36a) and (3.36b) correspond exactly to (3.33a) and (3.33b), respectively. We show that the measures $\phi^* := \delta_{y(\mathbf{x})}\mu_{y(\mathbf{x})}\lambda_{B_\mathbf{x}}|_{L\times Y}$, $\nu^* := \delta_{y(\mathbf{x})}\lambda_{B_\mathbf{x}}$, and $\hat{\phi}^* := \delta_{y(\mathbf{x})}\mu_{y(\mathbf{x})}\lambda_{B_\mathbf{x}}|_{K\times Y}$ are optimal for (3.36) and that $(3.33)^* = (3.36)^*$: Note first that optimality of $(\phi^*, \hat{\phi}^*, \nu^*)$ and feasibility in (3.36a) and (3.36b) can be proved in a very similar way as discussed in the proof of Lemma 3.3.3. To see that ϕ^* satisfies (3.36c), note that

$$
\begin{aligned}
\left\langle \phi^*, \mathbf{x}^\alpha \mathbf{y}^\beta q_k \right\rangle &= \int_{B_\mathbf{x}} \int_{L_\mathbf{x}} \int_Y \mathbf{x}^\alpha \mathbf{y}^\beta q_k(\mathbf{x}, \mathbf{y}, \boldsymbol{\omega}) \, \mathbf{d}\delta_{y(\mathbf{x})}(\mathbf{y}) \, \mathbf{d}\mu_{y(\mathbf{x})}(\boldsymbol{\omega}) \, \mathbf{dx} \\
&= \int_{B_\mathbf{x}} \mathbf{x}^\alpha y(\mathbf{x})^\beta \underbrace{\int_{L_\mathbf{x}} q_k(\mathbf{x}, y(\mathbf{x}), \boldsymbol{\omega}) \, \mathbf{d}\mu_{y(\mathbf{x})}(\boldsymbol{\omega})}_{=0 \text{ (3.35)}} \, \mathbf{dx} = 0.
\end{aligned}
$$

Finally $(3.33)^* = (3.36)^*$ is a direct consequence of optimality of $(\phi^*, \hat{\phi}^*, \nu^*)$.

We next show that from the dual of (3.36) we obtain a polynomial h similar to the one in the case without Stokes constraints. Define the linear operator $S : \mathbb{R}[\mathbf{x}, \mathbf{y}, \mathbf{r}] \to \mathbb{R}[\mathbf{x}, \mathbf{y}, \boldsymbol{\omega}]$ via $S\mathbf{x}^\alpha \mathbf{y}^\beta \mathbf{r}^k = \mathbf{x}^\alpha \mathbf{y}^\beta q_k$ with q_k defined as in (3.35). Then the dual of (3.36) reads:

$$
\begin{aligned}
\inf_{g^{(1)}, g^{(2)}, h} \quad & \int_{B_\mathbf{x}} h \, \mathbf{dx} \\
\text{s.t.} \quad & h - Tg^{(1)} \geq 0 \text{ on } B_\mathbf{x} \times Y && \text{(3.37a)}\\
& g^{(1)} \geq 0 \text{ on } B_\mathbf{x} \times \Omega && \text{(3.37b)}\\
& g^{(1)} + Sg^{(2)} \geq 1 \text{ on } L \times Y && \text{(3.37c)}\\
& h \in \mathbb{R}[\mathbf{x}],\ g^{(1)} \in \mathbb{R}[\mathbf{x}, \boldsymbol{\omega}],\ g^{(2)} \in \mathbb{R}[\mathbf{x}, \mathbf{y}, \mathbf{r}].
\end{aligned}
$$

Proposition 3.3.8. *Let h be feasible for (3.37). Then $h(\mathrm{x}) \geq \rho(\mathrm{x})$ for every $\mathrm{x} \in B_\mathbf{x}$.*

Proof. The proof is very similar to the one of Proposition 3.3.6. Let $\mathrm{x} \in B_\mathbf{x}$. Then (3.37a) implies

$$
\begin{aligned}
h(\mathrm{x}) \geq Tg^{(1)}(\mathrm{x}, y(\mathrm{x})) &= \int_\Omega g^{(1)}(\mathrm{x}, \boldsymbol{\omega}) f(\boldsymbol{\omega}, y(\mathrm{x})) \, \mathbf{d}\boldsymbol{\omega} \\
&= \int_{K_\mathrm{x}} \underbrace{g^{(1)}(\mathrm{x}, \boldsymbol{\omega})}_{\geq 0 \text{ (3.37b)}} f(\boldsymbol{\omega}, y(\mathrm{x})) \, \mathbf{d}\boldsymbol{\omega} + \int_{L_\mathrm{x}} \underbrace{g^{(1)}(\mathrm{x}, \boldsymbol{\omega})}_{\substack{\text{(3.37c)} \\ \geq \\ 1 - Sg^{(2)}(\mathrm{x}, y(\mathrm{x}), \boldsymbol{\omega})}} f(\boldsymbol{\omega}, y(\mathrm{x})) \, \mathbf{d}\boldsymbol{\omega} \\
&\geq \int_{L_\mathrm{x}} f(\boldsymbol{\omega}, y(\mathrm{x})) \, \mathbf{d}\boldsymbol{\omega} - \underbrace{\int_{L_\mathrm{x}} Sg^{(2)}(\mathrm{x}, y(\mathrm{x}), \boldsymbol{\omega}) f(\boldsymbol{\omega}, y(\mathrm{x})) \, \mathbf{d}\boldsymbol{\omega}}_{=0 \text{ (3.35)}} = \rho(\mathrm{x}).
\end{aligned}
$$

□

It is now possible to prove similar results for the moment hierarchy associated to (3.36) and (3.37) as in Lemma 3.3.7 and Theorem 3.1.

3.3.4 Numerical Experiments

We conclude this section with some illustrative numerical experiments. To implement the semidefinite relaxations of (3.33) and (3.36) we have used the GloptiPoly software [HLL09a]. The resulting SDPs are solved using version 8.1 of Mosek [MOS17].

We discuss three examples chosen to a) illustrate the effect (and efficiency) of Stokes constraints, b) compare the approximations with the real feasible set X^ϵ in (3.30) (approximated with intensive simulations), and c) show the behavior of the approximations for different violation probabilities.

Approximations With and Without Stokes

In order to illustrate the difference in quality of the approximation of X^ϵ when using or not using Stokes constraints, consider the example where $B_{\mathbf{x}} = [-1, 1]$, $K = \{\mathrm{x} \in B_{\mathbf{x}} \times \Omega : \omega - \mathrm{x} \geq 0\}$, $Y = [-0.1, 0.1] \times [0.8, 1]$, i.e., we consider univariate Gaussian measures with mean approximately 0 and deviation slightly less than 1. For every fixed x, due to the simple definition of K we can express $\mathbb{P}_{\mu_\mathrm{y}}(L_\mathrm{x})$ as an analytic expression in y. It is hence relatively easy to obtain a good estimation of $\rho(\mathrm{x})$ by sampling over $\mathbf{y}$. In Fig. 3.6 is displayed $\mathrm{x} \mapsto \rho(\mathrm{x})$ in black and two different approximations h_d computed for relaxation orders $d = 4$ in blue and $d = 6$ in red. The dashed lines are the polynomials corresponding to problem formulations including Stokes constraints.

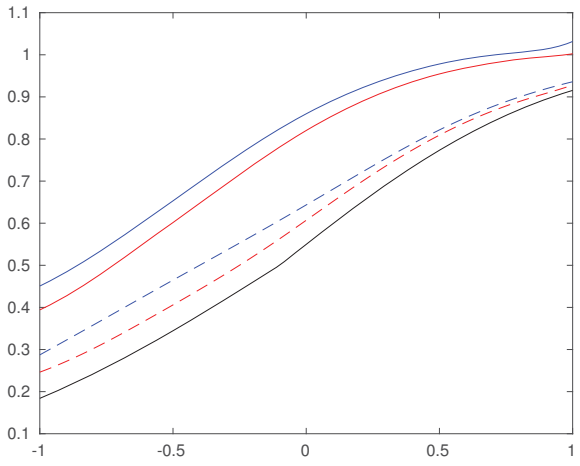

Figure 3.6: Approximation of the worst case probability ρ (black) by polynomials h_4 (blue) and h_6 (red), dashed/solid lines correspond to with/without Stokes constraints

As a first remark observe that, in accordance with the theoretic results, all approximations are overestimators of ρ. However, the approximations computed with Stokes constraints are much closer to ρ than the ones computed without. The former approximations

are particularly close to ρ for big values of the worst case probability. For lower probabilities they degrade (but are still quite good). This can be due to the non-differentiability of ρ at $\mathrm{x} = -0.1$. To visualize the sets X^ε and X_d^ε, e.g., for a value of $\varepsilon = 1/3$ one looks at the 1/3-sub-level set of the respective functions. This yields approximately that $X^{1/3} = [-1, -0.62]$ is the true feasible set. With Stokes, the approximations h_4 and h_6 yield the respective intervals $[-1, -0.96]$ and $[-1, -0.83]$ while the approximations without Stokes provide empty intervals.

Inner Approximations from Different Relaxations

As seen in the previous example, Stokes constraints are essential for the performance of our approach. In this section we therefore only report results using these additional constraints. In the second illustrative example, $B_{\mathbf{x}} = [-1, 1]^2$, $K = \{\mathrm{x} \in B_{\mathbf{x}} \times \Omega : 2\omega\,\mathrm{x}_2^2 - 2\omega\,\mathrm{x}_1^2 - 1 \geq 0\}$ and $Y = [-0.1, 0.1] \times [0.8, 1]$. In Fig. 3.7 we plot the feasible set X^ε and its approximations X_d^ε for a violation level of 10% ($\varepsilon = 0.1$).

The feasible set is approximated as follows. We discretize $B_{\mathbf{x}}$ into 200 and Y into 100 steps in each direction respectively. For each point x and each combination of parameters y we draw 1000 realizations of $\boldsymbol{\omega}$ from the normal distribution described by y. The point x is considered to be feasible whenever for each y, when for at least 900 out of the 1000 realizations ω of $\boldsymbol{\omega}$ the points (x, ω) are in K. This simulation took about 8600 seconds (without the authors claiming to be experts for Monte Carlo simulations) whereas the approximations for $d = 4, 5, 6$ take 5, 43, and 482 seconds respectively.

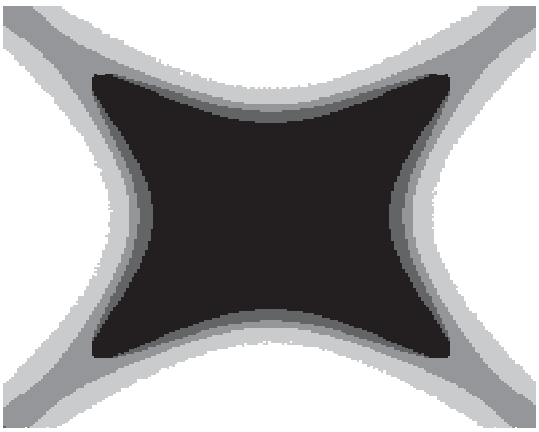

Figure 3.7: Monte Carlo simulation (light grey) of X^ε and inner approximations X_d^ε for $d = 8, 10, 12$, in decreasing intensity

Inspection of Fig. 3.7 reveals that the feasible set X^ε is non-convex. Already the lowest approximation X_4^ε (black) is able to capture this behavior. The next approximation X_5^ε (dark grey) is already a bit larger and X_6^ε (medium grey) captures a significant part of X^ε ($\approx 74\%$). Its computation time is 18 times faster than the one required for the Monte Carlo simulation of X^ε. In addition, and in contrast to the approximation via Monte Carlo, X_d^ε is guaranteed to be inside the true feasible set.

$(d, \text{time})\backslash\varepsilon$	50%	25%	12.5%	6.25%	3.125%
4 (30s)	96.94%	83.07%	69.70%	22.72%	0%
5 (107s)	99.91%	86.70%	73.21%	73.79%	2.48%
6 (633s)	100.0%	90.13%	79.94%	61.31%	27.98%

Table 3.3: Polynomial approximations vs Monte Carlo simulation.

Inner Approximations on Different Violation Levels

In the third example, $B_{\mathbf{x}} = [-1, 1]^3$, $\Omega = \mathbb{R}$, $K = \{\mathrm{x} \in B_{\mathbf{x}} \times \Omega : -2\omega\,\mathrm{x}_1^2 + 2\omega\,\mathrm{x}_2^2 - 2\omega\,\mathrm{x}_3^2 - 1 \geq 0\}$. We compute the inner approximations X_d^ε for $d = 4, 5, 6$. To compute the Monte Carlo approximation of X^ε in a reasonable time, we fix the mean of the distribution to 0 and the standard deviation σ is taken in the interval $[0.4, 0.6]$. For Monte Carlo we discretize $B_{\mathbf{x}}$ and $[0.4, 0.6]$ in 100 steps in each direction and draw again 1000 realizations of $\boldsymbol{\omega}$ for each point and each σ. This simulation takes about 2277 seconds. In the first example we have already seen that the polynomial approximations h_d are quite good for large violation probabilities. In Table 3.3 we compare the "volume" of our approximations against the Monte Carlo simulation, i.e. the ratio of the number of points admissible for our approximations over the number of points admissible in Monte Carlo. As the polynomial approximations are inner approximations, we expect the ratii to be less than one (assuming that Monte Carlo is accurate).

Again the polynomial approximations h_d are computed significantly faster than the Monte Carlo approximation ρ. As in the first example, for large ε the approximations are pretty exact. However, for all relaxation orders d the quality of approximation decreases with ε, and eventually $X_4^{0.03125} = \emptyset$. However we should not forget that good approximations with small ε are difficult to achieve in any case. Therefore it is quite interesting that we can retrieve almost 30% of $X^{0.03125}$ with $X_6^{0.03125}$ and using moments up to order 12 only.

3.4 Conclusion and Directions

In this chapter we have presented two contributions both building on the idea of approximating probabilities by solving semidefinite relaxations to a GMP. While Section 3.2 was an application of existing methods to a practical problem, Section 3.3 extended these methods to a wider theoretical framework.

A direct follow up of the presented contributions would be to apply the theory from Section 3.3 to generalize the framework in Section 3.2. A research direction that is surely necessary to apply Section 3.2 to real size OPF problems is to develop sparse versions of the volume and probability approximations, where sparsity is understood in a similar fashion as in Chapter 2. We provide first results in this direction in [Tac+18].

Chapter 4

Measure Valued Solutions to Differential Equations

In this chapter we consider solutions to non-linear (ordinary and partial) differential equations as instances of the GMP. The basic idea is quite similar to the idea of polynomial optimization, where we relaxed a point x – or the Dirac measure δ_{x} – to a general probability measure μ on space. In this chapter we relax trajectories $y(\mathbf{x})$ – or the measure $\delta_{y(\mathbf{x})}\lambda$ – to a more general measure $\mu = \mu_{\mathbf{x}}\lambda$, where $\mu_{\mathbf{x}}$ denotes the stochastic kernel or conditional of μ, knowing x, which is a probability measure for each x, and λ denotes the Lebesgue measure. The work in this chapter builds on works of Young [You69] and DiPerna [DiP85; DM87] who, among other things, considered generalized solutions to differential equations in the contexts of optimal control and conservation laws. Though these topics are quite different we will see, that in both we can use the same tools, so-called Young measures and their generalizations, in order to reformulate the problems as an instance of the GMP.

We briefly introduce the concept of *Young measures* (also called parametrized measures) before we discuss some generalizations in Section 4.2, where we use these measures to capture *limit effects in optimal control* of ordinary differential equations, when regular solutions do not converge. Then in Section 4.3 we turn to partial differential equations, more precisely to the class of *scalar hyperbolic conservation laws*. In both parts we employ the moment-SOS hierarchy in order to compute approximate solutions to the respective problems. Section 4.2 is based on [HKW18] and Section 4.3 is based on [Mar+18].

4.1 Young Measures

Let $(u_k)_{k\in\mathbb{N}} \subset L^\infty(\mathcal{X}, \mathbb{R}^m)$ be a weakly star converging sequence, i.e., for all $f \in L^1(\mathcal{X}, \mathbb{R}^m)$ it holds that

$$\lim_{k\to\infty} \int_{\mathcal{X}} f(\mathbf{x})u_k(\mathbf{x})\,\mathbf{dx} = \int_{\mathcal{X}} f(\mathbf{x})u(\mathbf{x})\,\mathbf{dx}$$

for some function $u \in L^\infty$. In particular, this implies that the local averages $\int_{\mathbf{X}} u_k(\mathbf{x})\,\mathbf{dx}$ converge to the local average $\int_{\mathbf{X}} u(\mathbf{x})\,\mathbf{dx}$, for every compact $\mathbf{X} \subseteq \mathcal{X}$. Young wanted to extract more information from this convergence and was interested in describing the limit behaviour for $\int_{\mathcal{X}} g(u_k(\mathbf{x}))\,\mathbf{dx}$ for arbitrary continuous functions $g \in C(\mathcal{X})$. To that end, he introduced the following families of measures

Definition 4.1.1 (Young measures)**.** A Young measure on a Euclidean space $\mathcal{X}$ is a map $\mu : \mathcal{X} \to \mathcal{P}(\mathbb{R}^m)$, $\mathrm{x} \mapsto \mu_{\mathrm{x}}$, such that for every $g \in C(\mathbb{R}^m)$ the function $\mathrm{x} \mapsto \int_{\mathbb{R}^m} g(\mathbf{u})\,\mathbf{d}\mu_{\mathrm{x}}(\mathbf{u})$ is measurable.

We already used a Young measure in Section 3.3 when considering the worst case probability for distributionally robust chance constraints. However we did not exploit any

properties of Young measures which is why we did not introduce the concept before. Here we use Young measures to describe the limit behaviour for continuous functions of u_k:

Theorem 4.1. *Let $(u_k)_{k\in\mathbb{N}} \subseteq L^\infty(\mathcal{X}, \mathbb{R}^m)$ be a weakly star converging sequence. Then there is a (non-relabelled) subsequence and a Young measure μ such that $g(u_k(\mathbf{x})) \overset{*}{\rightharpoonup} \int g(\mathbf{u})\,\mathbf{d}\mu_{\mathbf{x}}(\mathbf{u})$ in L^∞, i.e., for all $f \in L^1(\mathcal{X}, \mathbb{R}^m)$*

$$\lim_{k\to\infty} \int_{\mathcal{X}} f(\mathbf{x}) g(u_k(\mathbf{x}))\,\mathbf{dx} = \int_{\mathcal{X}} \int_{R^m} f(\mathbf{x}) g(\mathbf{u})\,\mathbf{d}\mu_{\mathbf{x}}(\mathbf{u})\,\mathbf{dx} \tag{4.1}$$

The result of Young has been generalized to sequences $(u_k)_{k\in\mathbb{N}} \subseteq L^p(\mathcal{X})$ for arbitrary p. Denote $C_p(\mathbb{R}^m) := \{g \in C(\mathbb{R}^m) : g(\mathrm{u}) = o(\|\mathrm{u}\|^p) \text{ for } \|\mathrm{u}\| \to \infty\}$ the set of continuous functions of less than p-th growth. Then $g(u_k(\mathbf{x})) \rightharpoonup \int g(\mathbf{u})\,\mathbf{d}\mu_{\mathbf{x}}(\mathbf{u})$ in L^1 in the weak topology for all $g \in C_p(\mathcal{X})$. In case of a compact set $\mathcal{X}$ the convergence of L^p-Young measures is weaker then the convergence mentioned above, as it only holds for continuous functions g with growth at infinity less than p and is only tested on functions $f \in L^\infty(\mathcal{X}) \subseteq L^1(\mathcal{X})$ [KR96].

Theorem 4.1 give the theoretical justification for what we explained in the introduction: in order to consider the limit behaviour of trajectories u_k we relax the limit at every point x to a measurement of the image space of u_k. In the following sections we use this concept to describe limits occurring in optimizing sequences of optimal control problems, and limits of solutions to partial differential equations. For a comprehensive reference on Young measures and their use in the control of ordinary and partial differential equations, see [Fat99, Part III].

4.2 Optimal Control Problems with Oscillations, Concentrations and Discontinuities

Optimal control problems with oscillations (chattering controls) and concentrations (impulsive controls) can have integral performance criteria such that concentration of the control signal occurs at a discontinuity of the state signal. In this section we apply a generalization to Young measures (anisotropic parametrized measures) to give a precise meaning of the integral cost and to allow for the sound application of numerical methods. We show how this can be combined with the moment-SOS hierarchy.

4.2.1 Introduction

As a consequence of optimality, various limit behaviours can be observed in optimal control: minimizing control law sequences may feature increasingly fast variations, called oscillations (chattering controls [You69]), or increasingly large values, called concentrations (impulsive controls [Lue69]). The simultaneous presence of oscillations and concentrations in optimal control needs careful analysis and specific mathematical tools, so that the numerical methods behave correctly. Previous work of two of the authors [CHK17] combined tools from partial differential equation analysis (DiPerna-Majda measures [DM87]) and semidefinite programming relaxations (the moment-sums-of-squares or Lasserre hierarchy [Las+08b]) to describe a sound numerical approach to optimal control in the simultaneous presence of oscillations and concentrations. To overcome difficulties in the analysis, a certain number

of technical assumptions were made, see [CHK17, Assumption 1, Section 2.2], so as to avoid the simultaneous presence of concentrations (in the control signals) and discontinuities (in the system trajectories).

In the present contribution we would like to remove these technical assumptions and accommodate the simultaneous presence of concentrations and discontinuities, while allowing oscillations as well. For this, we exploit a recent extension of the notion of DiPerna-Majda measures called anisotropic parametrized measures [KKK17], so that it makes sense mathematically while allowing for an efficient numerical implementation with semidefinite programming relaxations.

To motivate further our work, let us use an elementary example to illustrate the difficulties that may be faced in the presence of discontinuities and concentrations. Consider the optimal control problem

$$\begin{aligned} &\inf_u \int_0^1 (\mathbf{t}+y(\mathbf{t}))u(\mathbf{t})\,\mathbf{dt} \\ &\text{s.t.} \quad \dot{y}(\mathrm{t}) = u(t), \qquad \text{a.e. on } [0,1], \\ &\qquad 1 \geq y(\mathrm{t}) \geq 0,\ u(\mathrm{t}) \geq 0, \quad \text{a.e. on } [0,1], \\ &\qquad y(0)=0, \quad y(1)=1, \end{aligned} \tag{4.2}$$

where the infimum is with respect to measurable controls of time. The trajectory y should move the state from zero at initial time to one at final time, yet for the non-negative integrand to be as small as possible, the control u should be zero all the time, except maybe at time zero. We can design a sequence of increasingly large controls u that drive y from zero to one increasingly fast. We observe that this sequence has no limit in the space of measurable functions but it tends (in a suitable weak sense) to the Dirac measure at time zero. We speak of control signal concentration or impulsive control. The integrand contains the product yu of a function whose limit becomes discontinuous at a point where the other function has no limit, hence requiring careful analysis. Here however, this product can be written $y\dot{y} = \frac{\mathbf{d}}{\mathbf{dt}}\frac{y^2}{2}$ and hence the integral term is well defined since $\int_0^1 y\dot{y}\,\mathbf{dt} = \frac{y(1)^2-y(0)^2}{2} = \frac{1}{2}$. Consequently the cost in (4.2) is equal to $\int_0^1 \mathbf{t}u(\mathbf{t})\,\mathbf{dt} + \frac{1}{2}$ and independent of the actual trajectory.

This reasoning is valid because $\dot{y} = u$ in problem (4.2), but this integration trick cannot be carried out for more general differential equations. For example we cannot solve analytically the following modified optimal control problem

$$\begin{aligned} &\inf_u \int_0^1 (\mathbf{t}+y(\mathbf{t}))u(\mathbf{t})\,\mathbf{dt} \\ &\text{s.t.} \quad \dot{y}(\mathrm{t}) = \sqrt{\varepsilon^2+u^2(\mathrm{t})}, \quad \text{a.e. on } [0,1], \\ &\qquad 1 \geq y(\mathrm{t}) \geq 0, \quad u(\mathrm{t}) \geq 0, \quad \text{a.e. on } [0,1], \\ &\qquad y(0)=0, \quad y(1)=1, \end{aligned} \tag{4.3}$$

where ε is a given real number. Providing a mathematically sound framework for the analysis of this kind of phenomenon combining concentration and discontinuity, and possibly also oscillation (not illustrated by the simple example above), is precisely the purpose of our section.

4.2.2 Relaxing Optimal Control

Let $L : [0,1] \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and $F : [0,1] \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ be continuous functions. For initial and final conditions $\mathrm{y}_0, \mathrm{y}_1 \in \mathbb{R}^n$ and some integer $1 \leq p \leq \infty$, the formulation of the classical optimal control problem is

$$\begin{aligned} \inf_u \int_0^1 & L(\mathbf{t}, y(\mathbf{t}), u(\mathbf{t}))\,\mathbf{dt} \\ \text{s.t.} \quad & \dot{y}(\mathrm{t}) = F(\mathrm{t}, y(\mathrm{t}), u(\mathrm{t})), \quad \text{a.e. on } [0,1], \\ & y(0) = \mathrm{y}_0, \;\; y(1) = \mathrm{y}_1, \\ & y \in W^{1,1}(0,1;\mathbb{R}^n),\; u \in L^p(0,1;\mathbb{R}^m), \end{aligned} \tag{OCP}$$

where $W^{1,p}(0,1;\mathcal{X})$ is the space of functions from $(0,1)$ to $\mathcal{X}$ whose weak derivative belongs to $L^p(0,1;\mathcal{X})$, the space of functions from $(0,1)$ to $\mathcal{X}$ whose p-th power is Lebesgue integrable. We suppress $\mathcal{X}$ in the notation if $\mathcal{X} = \mathbb{R}$.

Consider a minimizing sequence of controls $(u_k)_{k\in\mathbb{N}} \subseteq L^p(0,1;\mathbb{R}^m)$ for problem (OCP) and the corresponding sequence of trajectories $(y_k)_{k\in\mathbb{N}} \subseteq W^{1,1}(0,1;\mathbb{R}^n)$, the space of absolutely continuous functions. Then the infimum in (OCP) might not be attained because $(u_k)_{k\in\mathbb{N}}$ might not converge in $L^p(0,1;\mathbb{R}^m)$ and $(y_k)_{k\in\mathbb{N}}$ might not converge in $W^{1,1}(0,1;\mathbb{R}^n)$. To overcome this issue, it has been proposed to relax the regularity assumptions on u. In the following we discuss some of the approaches in detail.

Oscillations

The limit of a minimizing sequence for (OCP) might fall out of the feasible space because of oscillation effects of $(u_k)_{k\in\mathbb{N}}$. Consider for example the optimal control problem

$$\begin{aligned} \inf_u \int_0^1 & (u(\mathbf{t})^2 - 1)^2 + y(\mathbf{t})^2\,\mathbf{dt} \\ \text{s.t.} \quad & \dot{y}(\mathrm{t}) = u(\mathrm{t}), \quad \text{a.e. on } [0,1], \\ & y(0) = 0,\; y(1) = 0, \\ & y \in W^{1,4}(0,1), \quad u \in L^4(0,1). \end{aligned} \tag{4.4}$$

As the integrand in the cost is a sum of squares, the value is at least zero. To see that actually it is equal to zero, consider the sequence of controls $(u_k)_{k\in\mathbb{N}} \subseteq L^4(0,1)$ defined by

$$u_k(\mathrm{t}) := \begin{cases} 1, & \text{if } \mathrm{t} \in \left[\frac{2l+1}{2^k}, \frac{l+1}{2^{k-1}}\right],\; 0 \leq l \leq k-1 \\ -1, & \text{otherwise} \end{cases} \tag{4.5}$$

for $k > 1$ and $u_1 := 0$. For the corresponding sequence of trajectories $(y_k)_{k\in\mathbb{N}}$ defined by $y_k(\mathrm{t}) := \int_0^{\mathrm{t}} u_k(\mathbf{s})\,\mathbf{ds}$ it holds that $y_k \in W^{1,4}(0,1)$ and $y_k(1) = 0$ as desired. Hence, $(u_k)_{k\in\mathbb{N}}$ is a sequence of feasible controls. A short calculation shows that using this sequence the cost in (4.4) converges to zero. While the limit $y_\infty := 0$ of $(y_k)_{k\in\mathbb{N}}$ stays in $W^{1,4}(0,1)$, the sequence of controls $(u_k)_{k\in\mathbb{N}}$ however does not converge in $L^4(0,1)$.

In contrast to that, the sequence of measures defined by $\mu_k = \delta_{u_k(\mathbf{t})}\lambda_{[0,1]}$ converges weakly to $\mu := \frac{1}{2}(\delta_{-1} + \delta_1)\lambda_{[0,1]}$ in the sense that for all $f \in C([0,1])$ and $g \in C_p(\mathbb{R})$:

$$\lim_{k\to\infty} \int_0^1 \int_{\mathbb{R}} f(\mathbf{t})g(\mathbf{u})\,\mathbf{d}\mu_k = \int_0^1 \int_{\mathbb{R}} f(\mathbf{t})g(\mathbf{u})\,\mathbf{d}\mu \tag{4.6}$$

where $C_p(\mathbb{R}) := \{g \in C(\mathbb{R}) : g(\mathrm{u}) = o(\|\mathrm{u}\|^p) \text{ for } \|\mathrm{u}\| \to \infty\}$ is the set of continuous functions of less than p-th growth. Integration then yields

$$y_\infty(1) = \int_0^1 \int_{\mathbb{R}} \mathbf{u}\,\mathbf{d}\mu = \int_0^1 \int_{\mathbb{R}} \mathbf{u}\,\mathbf{d}\tfrac{1}{2}(\delta_{-1} + \delta_1)(\mathbf{u})\,\mathbf{dt} = 0.$$

A similar reasoning shows that the cost with respect to μ is zero.

More generally, this observation motivates to relax the regularity assumptions on the control u in (OCP) and also allow for limits $\mu = \omega_{\mathbf{t}}\lambda_{[0,1]}$ of control sequences $(u_k)_{k\in\mathbb{N}} \subseteq L^p(0,1;\mathbb{R}^m)$. In general the measure ω depends on time, i.e., we have a family of probability measures $(\omega_\mathrm{t})_{\mathrm{t}\in[0,1]} \subseteq \mathcal{P}(\mathbb{R}^m)$, where $\mathcal{P}(\mathcal{X})$ denotes the set of probability measures on $\mathcal{X}$, i.e. non-negative Borel regular measures with unit mass. Such parametrized measures obtained as limits of a sequence of functions $(u_k)_{k\in\mathbb{N}} \subseteq L^p(0,1;\mathbb{R}^m)$ have been called L^p-Young measures (compare 4.1).

The relaxed version of (OCP) that now takes into account oscillating control sequences can be written as

$$\begin{aligned} \inf_{\omega} \int_0^1 \int_{\mathbb{R}^m} & L(\mathbf{t}, y(\mathbf{t}), \mathbf{u})\,\mathbf{d}\omega_\mathrm{t}(\mathbf{u})\,\mathbf{dt} \\ \text{s.t.} \int_0^1 \int_{\mathbb{R}^m} & F(\mathbf{t}, y(\mathbf{t}), \mathbf{u})\,\mathbf{d}\omega_\mathrm{t}(\mathbf{u})\,\mathbf{dt} = \mathrm{y}_1 - \mathrm{y}_0 \\ & y \in W^{1,1}(0,1;\mathbb{R}^n),\ \omega_\mathrm{t} \in \mathcal{P}(\mathbb{R}^m) \forall \mathrm{t} \in [0,1] \end{aligned} \tag{4.7}$$

where the constraint is a reformulation of the differential equation

$$\dot{y}(\mathrm{t}) = \int_{\mathbb{R}^m} F(\mathrm{t}, y(\mathrm{t}), \mathbf{u})\,\mathbf{d}\omega_\mathrm{t}(\mathbf{u}),\ \text{a.e. on } [0,1]$$

with the boundary conditions $y(0) = \mathrm{y}_0$ and $y(1) = \mathrm{y}_1$.

Concentrations

Oscillation of the control sequence is not the only reason that prevents the infimum in (OCP) from being attained. As a second example consider the following problem of optimal control:

$$\begin{aligned} \inf_{u} \int_0^1 & (\mathbf{t} - \tfrac{1}{2})^2 u(\mathbf{t})\,\mathbf{dt} \\ \text{s.t.} \quad & \dot{y}(\mathrm{t}) = u(\mathrm{t}) \geq 0, \quad \text{a.e. on } [0,1], \\ & y(0) = 0,\ y(1) = 1, \\ & y \in W^{1,1}(0,1), \quad u \in L^1(0,1). \end{aligned} \tag{4.8}$$

Note that the control enters into the problem linearly. The value is zero as the integrand is positive and using the sequence of controls

$$u_k(\mathrm{t}) := \begin{cases} k, & \text{if } \mathrm{t} \in \left[\frac{k-1}{2k}, \frac{k+1}{2k}\right] \\ 0, & \text{else} \end{cases} \tag{4.9}$$

the cost converges to zero. As in the previous subsection neither $(u_k)_{k\in\mathbb{N}}$ nor any subsequence converges in $L^1(0,1)$. In contrast to the previous example this time $(y_k)_{k\in\mathbb{N}}$ does not converge in $W^{1,1}(0,1)$ neither. We hence use the extension $BV(0,1)$, the space of functions with bounded variation, as a relaxed space for the trajectory. Following the same approach as before we consider the control as a measure $\mu_k := \delta_{u_k(\mathbf{t})}\lambda_{[0,1]}$. As u appears linearly in (4.8) we can directly integrate over $\mathbf{u}$ and define a sequence of probability measures $(\tau_k)_{k\in\mathbb{N}} \subseteq \mathcal{P}([0,1])$ by $\mathbf{d}\tau_k(\mathbf{t}) := \int_{\mathbb{R}} \mathbf{u}\,\mathbf{d}\mu_k(\mathbf{t},\mathbf{u})$. A short calculation shows that this sequence has the weak limit $\tau := \delta_{\frac{1}{2}}$, i.e. for all $f \in C([0,1])$:

$$\lim_{k\to\infty} \int_0^1 f\,\mathbf{d}\tau_k = \int_0^1 f\,\mathbf{d}\tau.$$

Note that by integrating before passing to the limit we transfer the unboundedness of the control into the measurement of time and only keep the direction (i.e. $+1$ in this example) of the control. Whereas we observed a superposition of two different controls in the previous example, here we see a concentration of the control in time. For optimal control problems with linear growth in the control:

$$\begin{aligned} \inf_u \int_0^1 & L(\mathbf{t}, y(\mathbf{t}))u(\mathbf{t})\,\mathbf{dt} \\ \text{s.t.} \quad & \dot{y}(\mathrm{t}) = F(\mathrm{t}, y(\mathrm{t}))u(\mathrm{t}), \quad \text{a.e. on } [0,1] \\ & y(0) = \mathrm{y}_0, \quad y(1) = \mathrm{y}_1, \\ & y \in W^{1,1}(0,1;\mathbb{R}^n),\ u \in L^1(0,1;\mathbb{R}^m) \end{aligned}$$

we can therefore build the following relaxation that can take into account concentration effects of the control:

$$\begin{aligned} \inf_\tau \int_0^1 & L(\mathbf{t}, y(\mathbf{t}))\,\mathbf{d}\tau \\ \text{s.t.} \quad & \int_0^1 F(\mathbf{t}, y(\mathbf{t}))\,\mathbf{d}\tau = \mathrm{y}_1 - \mathrm{y}_0, \\ & y \in BV(0,1;\mathbb{R}^n), \quad \tau \in \mathcal{P}([0,1]). \end{aligned} \tag{4.10}$$

See [Cla+14] for an application of the moment-SOS hierarchy for solving numerically nonlinear control problems in the presence of concentration.

Oscillation and Concentration

The relaxations proposed so far allow to consider controls that are either oscillating in value or concentrating in time. However it is possible that both effects appear in the same

problem. Consider for example

$$\begin{aligned} \inf_u \int_0^1 \frac{u(\mathbf{t})^2}{1+u(\mathbf{t})^4} + (y(\mathbf{t})-\mathbf{t})^2 \, \mathbf{dt} \\ \text{s.t.} \quad \dot{y}(\mathrm{t}) = u(\mathrm{t}) \geq 0, \quad \text{a.e. on } [0,1], \\ y(0)=0,\ y(1)=1, \\ y \in W^{1,1}(0,1),\ u \in L^1(0,1). \end{aligned} \tag{4.11}$$

The infimum value zero of (4.11) can be approached arbitrarily close by a sequence of controls $(u_k)_{k\in\mathbb{N}}$ defined by

$$u_k(\mathrm{t}) := \begin{cases} k, & \text{if } \mathrm{t} \in \left[\frac{l}{k} - \frac{1}{2k^2}, \frac{l}{k} + \frac{1}{2k^2}\right],\ 1 \leq l < k \\ 0, & \text{else} \end{cases} \tag{4.12}$$

for $k > 1$ and $u_1 := 1$. The idea to capture the limit behaviour of this sequence is to combine a Young measure on the control and replacing the uniform measure on time by a more general measure on time. Note that due to linearity it was possible in Section 4.2.2 to transfer the limit behaviour of the control into the measurement of time. In the present example the control enters non-linearly in the cost, which is why we will need to allow the control to take values at infinity. We consider a metrizable compactification $\beta_{\mathcal{U}}\mathbb{R}$ of the control space corresponding to a complete and separable sub ring $\mathcal{U}$ of the space of continuous bounded functions from $\mathbb{R}^m$ to $\mathbb{R}$ (see Section 4.2.3 for more details). Elements of $\mathcal{U}$ we mark with an index b to emphasize that they are *bounded* functions. The sequence of measures $\mu_k := \delta_{u_k(\mathbf{t})}\lambda_{[0,1]}$ converges to $\mu := \omega\tau$ with $\omega := \frac{1}{2}(\delta_0 + \delta_\infty)$ and $\tau := 2\lambda_{[0,1]}$ understood in the following weak sense: for all $f \in C([0,1])$ and $g_b \in \mathcal{U}$:

$$\lim_{k\to\infty} \int_0^1 \int_{\mathbb{R}} f(\mathbf{t}) g_b(\mathbf{u})(1+\|\mathbf{u}\|^p)\, \mathbf{d}\mu_k(\mathbf{t},\mathbf{u}) = \int_0^1 \int_{\beta_{\mathcal{U}}\mathbb{R}} f(\mathbf{t}) g_b(\mathbf{u})\, \mathbf{d}\mu(\mathbf{t},\mathbf{u}). \tag{4.13}$$

Measures $\mu \in \mathcal{P}([0,1]\times\beta_{\mathcal{U}}\mathbb{R}^m)$ obtained as limits of sequences $(u_k)_{k\in\mathbb{N}} \subseteq L^p(0,1;\mathbb{R}^m)$ in the sense of (4.13) have been called DiPerna-Majda measures. They will be discussed in more detail in Section 4.2.3. Let $L \in C([0,1]\times\mathbb{R}^n\times\mathbb{R}^m)$ such that $L(\mathrm{t},\mathrm{y},\mathbf{u}) \in C_p(\mathbb{R}^m)$ for all $(\mathrm{t},\mathrm{y}) \in [0,1]\times\mathbb{R}^n$ and $F \in C([0,1]\times\mathbb{R}^n\times\mathbb{R}^m;\mathbb{R}^n)$ such that $L(\mathrm{t},\mathrm{y},\mathbf{u}) \in C_p(\mathbb{R}^m;\mathbb{R}^n)$ for all $(\mathrm{t},\mathrm{y}) \in [0,1]\times\mathbb{R}^n$. A relaxed version of (OCP) taking into account both oscillation and concentration effects can hence be stated as

$$\begin{aligned} \inf_\mu \int L_b(\mathbf{t}, y(\mathbf{t}), \mathbf{u})\, \mathbf{d}\mu(\mathbf{t},\mathbf{u}) \\ \text{s.t.} \int F_b(\mathrm{t}, y(\mathrm{t}), \mathrm{u})\, \mathbf{d}\mu(\mathbf{t},\mathbf{u}) = \mathrm{y}_1 - \mathrm{y}_0, \\ \mu \in \mathcal{P}([0,1]\times\beta_{\mathcal{U}}\mathbb{R}^m) \end{aligned} \tag{4.14}$$

where

$$L_b(\mathbf{t},\mathbf{y},\mathbf{u}) := \frac{L(\mathbf{t},\mathbf{y},\mathbf{u})}{1+\|\mathbf{u}\|^p}, \quad F_b(\mathbf{t},\mathbf{y},\mathbf{u}) := \frac{F(\mathbf{t},\mathbf{y},\mathbf{u})}{1+\|\mathbf{u}\|^p}. \tag{4.15}$$

In [CHK17], the moment-SOS hierarchy is adapted to compute numerically DiPerna-Majda measures and solve optimal control problem featuring oscillations and concentrations. How-

ever, the approach is valid under a certain number of technical assumptions on the data L and F, see [CHK17, Assumption 1, Section 2.2]. These assumptions are enforced to prevent the simultaneous presence of concentration and discontinuity.

Oscillations, Concentrations and Discontinuities

As mentioned in the introduction, the integrals in (OCP) might not be well defined, as concentration effects of the control are likely to cause discontinuities in the trajectory occurring at the same time. In view of the previous examples we propose to generalize the DiPerna-Majda measures, which themselves are a generalization of Young measures, even further and now also relax the trajectory to a measure valued function depending on time and control. In the sequel we describe accordingly the set of anisotropic parametrized measures. Then we provide a linear formulation of optimal control problem (OCP) that can cope with oscillations, concentrations and discontinuities in a unified fashion.

4.2.3 Anisotropic Parametrized Measures

In the following we describe a generalization of DiPerna-Majda measures. For this it will be instructive to review first the classical DiPerna-Majda measures.

DiPerna-Majda measures

Let $\mathcal{U}$ be a complete[1] and separable subring of continuous bounded functions from $\mathbb{R}^m$ to $\mathbb{R}$. It is known [Eng89, Sect. 3.12.22] that there is a one-to-one correspondence between such rings and metrizable compactifications of $\mathbb{R}^m$. By a compactification we mean a compact set, denoted by $\beta_{\mathcal{U}}\mathbb{R}^m$, into which $\mathbb{R}^m$ is embedded homeomorphically and densely. For simplicity, we will not distinguish between $\mathbb{R}^m$ and its image in $\beta_{\mathcal{U}}\mathbb{R}^m$. Similarly, we will not distinguish between elements of $\mathcal{U}$ and their unique continuous extensions defined on $\beta_{\mathcal{U}}\mathbb{R}^m$.

DiPerna and Majda [DM87], see also [Rou97], have shown that every bounded sequence $(u_k)_{k\in\mathbb{N}}$ in $L^p(0,1;\mathbb{R}^m)$ with $1 \le p < \infty$ has a subsequence (denoted by the same indices) such that there exists a probability measure $\tau \in \mathcal{P}([0,1])$ and an L^p-Young measure $\omega := \omega_{\mathbf{t}} : [0,1] \to \mathcal{P}(\beta_{\mathcal{U}}\mathbb{R}^m)$ satisfying for all $f \in C(0,1)$ and $g_b \in \mathcal{U}$:

$$\lim_{k\to\infty} \int_0^1 f(\mathbf{t}) g_b(u_k(\mathbf{t}))(1 + \|u_k(\mathbf{t})\|^p)\,\mathbf{dt} = \int_0^1 \int_{\beta_{\mathcal{U}}\mathbb{R}^m} f(\mathbf{t}) g_b(\mathbf{u})\,\mathbf{d}\omega_{\mathbf{t}}(\mathbf{u})\,\mathbf{d}\tau(\mathbf{t}) \tag{4.16}$$

compare with (4.13). The limit measure $\mu := \omega_{\mathbf{t}}\tau$ of such a sequence, or sometimes the pair (τ, ω), is called a DiPerna-Majda measure.

Note that, letting $g_b \equiv 1 \in \mathcal{U}$ in (4.16), the measure on time τ can be computed as the weak star limit of the sequence $(1 + \|u_k\|^p)_{k\in\mathbb{N}}$, i.e. for all $f \in C(0,1)$:

$$\lim_{k\to\infty} \int_0^1 (1 + \|u_k\|^p)\,\mathbf{dt} = \int_0^1 \int_{\beta_{\mathcal{U}}\mathbb{R}^m} f(\mathbf{t})\,\mathbf{d}\omega_{\mathbf{t}}(\mathbf{u})\,\mathbf{d}\tau(\mathbf{t}) = \int_0^1 f(\mathbf{t})\,\mathbf{d}\tau(\mathbf{t}) \tag{4.17}$$

[1] A ring of functions is complete if it contains all constant functions, it separates points from closed subsets and it is closed with respect to the supremum norm.

where the last equality follows from the fact that a Young measure is a probability measure i.e. $\int_{\beta_{\mathcal{U}}\mathbb{R}^m} \omega_{\mathrm{t}}(\mathbf{du}) = 1$ for each $\mathrm{t} \in [0,1]$.

As a second remark, consider any $f \in C(0,1) \subseteq L^\infty(0,1)$ and $g_b \in \mathcal{U} \cap C_0(\mathbb{R}^m)$. Then, as $g_b(\mathbf{u})(1+\|\mathbf{u}\|^p) \in C_p(\mathbb{R}^m)$, the limit in (4.16) is already given by (4.6). This means that the restriction of a DiPerna-Majda measure (τ, ω) to $[0,1] \times \mathbb{R}^m$ is $(\lambda_{[0,1]}, \tilde{\omega})$, where $\tilde{\omega}_{\mathbf{t}} : [0,1] \to \mathcal{P}(\mathbb{R}^m)$ is the L^p-Young measure generated by $(u_k)_{k\in\mathbb{N}}$. Hence the right side of (4.16) can – now again in full generality – be written as

$$\int_0^1 \int_{\mathbb{R}^m} f(\mathbf{t}) g_b(\mathbf{u})(1+\|\mathbf{u}\|^p)\,\mathbf{d}\tilde{\omega}_{\mathbf{t}}(\mathbf{u})\,\mathbf{dt} + \int_0^1 \int_{\beta_{\mathcal{U}}\mathbb{R}^m \setminus \mathbb{R}^m} f(\mathbf{t}) g_b(\mathbf{u})\,\mathbf{d}\omega_{\mathbf{t}}(\mathbf{u})\,\mathbf{d}\tau(\mathbf{t}). \tag{4.18}$$

This illustrates clearly that Young measures can only capture oscillations of the sequence but not concentrations.

Generalization

The drawback of DiPerna-Majda measures is that f in (4.16) must be a continuous function. This does not fit to our aim to study interactions of discontinuities and concentrations, as we cannot consider f as a function $f(y(\mathbf{t}))$ in the case that y is discontinuous. We therefore need a notion of convergence which generalizes the measures of DiPerna and Majda.

To cope with the simultaneous presence of oscillations, concentrations and discontinuities, a new tool was recently introduced in [KKK17], namely anisotropic parametrized measures generated by pairs $(y_k, u_k)_{k\in\mathbb{N}}$ where u_k is the control and y_k the corresponding state trajectory.

Proposition 4.2.1. *Any admissible trajectory of optimal control problem* (OCP) *is such that $y \in L^\infty(0,1;Y)$ for some compact set $Y \subseteq \mathbb{R}^n$, e.g. a ball of sufficiently large radius.*

Proof. The function $\mathrm{t} \mapsto y(\mathrm{t})$ is the integral of a Lebesgue integrable function and hence it is bounded on a bounded time interval. □

Now, the following result is a special case of [KKK17, Theorem 4]:

Theorem 4.2. *Let $1 \le p < +\infty$. Let $(u_k)_{k\in\mathbb{N}}$ be a bounded sequence in $L^p(0,1;\mathbb{R}^m)$ and $(y_k)_{k\in\mathbb{N}}$ a bounded sequence in $W^{1,1}(0,1;Y)$ for some compact set $Y \subseteq \mathbb{R}^n$. Then there is a (non-relabelled) subsequence $(u_k, y_k)_{k\in\mathbb{N}}$, a measure $\tau \in \mathcal{P}([0,1])$, a measure $\omega_{\mathrm{t}} \in \mathcal{P}(\beta_{\mathcal{U}}\mathbb{R}^m)$ parametrized in $\mathrm{t} \in [0,1]$ and a measure $\upsilon_{\mathrm{t,u}} \in \mathcal{P}(Y)$ parametrized in $\mathrm{t} \in [0,1]$ and $u \in \beta_{\mathcal{U}}\mathbb{R}^m$ such that for every $f \in C([0,1])$, $g_b \in \mathcal{U}$, $h \in C(Y)$, it holds*

$$\begin{aligned} &\lim_{k\to\infty} \int_0^1 f(\mathbf{t}) g_b(u_k(\mathbf{t}))(1+\|u_k(\mathbf{t})\|^p) h(y_k(\mathbf{t}))\,\mathbf{dt} \\ &\qquad = \int_0^1 \int_{\beta_{\mathcal{U}}\mathbb{R}^m} \int_Y f(\mathbf{t}) g_b(\mathbf{u}) h(\mathbf{y})\,\mathbf{d}\upsilon_{\mathbf{t},\mathbf{u}}(\mathbf{y})\,\mathbf{d}\omega_{\mathbf{t}}(\mathbf{u})\,\mathbf{d}\tau(\mathbf{t}) \end{aligned} \tag{4.19}$$

Moreover, (ω, τ) is the DiPerna-Majda measure generated by $(u_k)_{k\in\mathbb{N}}$. The measure $\mu := \upsilon_{\mathbf{t},\mathbf{u}}\omega_{\mathbf{t}}\tau$, or sometimes the triplet (τ, ω, υ), is called an anisotropic parametrized measure *or* generalized DiPerna-Majda measure *and the subsequence $(u_k, y_k)_{k\in\mathbb{N}}$ its* generating sequence.

We revisit and slightly modify an example from [KKK17] to give an intuition about these newly introduced measures.

Example 4.2.2. Consider the following sequence of trajectories and its weak derivatives $u_k := \dot{y}_k$, illustrated in Fig. 4.1.

$$y_k(\mathrm{t}) := \begin{cases} 0 & \text{if } 0 \le \mathrm{t} \le \frac{1}{2} - \frac{1}{k}, \\ k(\mathrm{t} - \frac{1}{2} + \frac{1}{k}) & \text{if } \frac{1}{2} - \frac{1}{k} \le \mathrm{t} \le \frac{1}{2}, \\ -2k(\mathrm{t} - \frac{1}{2} - \frac{1}{2k}) & \text{if } \frac{1}{2} \le \mathrm{t} \le \frac{1}{2} + \frac{1}{k}, \\ -1 & \text{if } \frac{1}{2} + \frac{1}{k} \le \mathrm{t} \le 1, \end{cases} \quad u_k(\mathrm{t}) := \begin{cases} 0 & \text{if } 0 \le \mathrm{t} \le \frac{1}{2} - \frac{1}{k}, \\ k & \text{if } \frac{1}{2} - \frac{1}{k} \le \mathrm{t} \le \frac{1}{2}, \\ -2k & \text{if } \frac{1}{2} \le \mathrm{t} \le \frac{1}{2} + \frac{1}{k}, \\ 0 & \text{if } \frac{1}{2} + \frac{1}{k} \le \mathrm{t} \le 1. \end{cases}$$

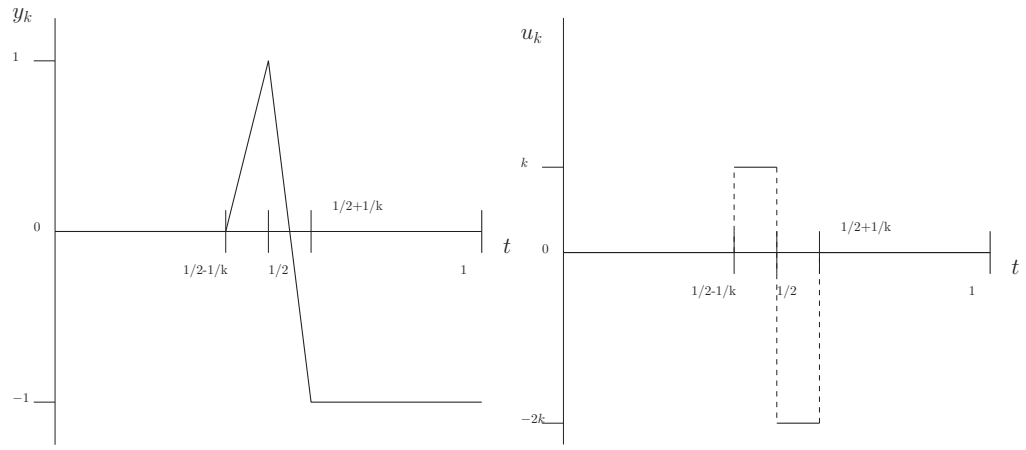

Figure 4.1: Sequences $(y_k, u_k)_{k\in\mathbb{N}}$ from Example 4.2.2.

Let $f \in C([0,1])$, and $h \in C_b(\mathbb{R})$ with primitive denoted by H. Further let $\mathcal{U}$ correspond to the two-point compactification of $\mathbb{R}$, which is $\beta_{\mathcal{U}}\mathbb{R}^m = \mathbb{R} \cup \{-\infty, +\infty\}$, and let $g(\mathbf{u}) = (1 + \|\mathbf{u}\|)g_b(\mathbf{u})$ where $g_b \in \mathcal{U}$, i.e. $g_b(\pm\infty) := \lim_{\mathrm{u}\to\pm\infty} g_b(u)$, respectively. Then it holds

$$\begin{aligned}
&\lim_{k\to\infty} \int_0^1 f(\mathbf{t})g(u_k(\mathbf{t}))h(y_k(\mathbf{t}))\,\mathbf{dt} \\
&= \lim_{k\to\infty} \int_0^{\frac{1}{2}-\frac{1}{k}} f(\mathbf{t})g(0)h(0)\,\mathbf{dt} + \lim_{k\to\infty} \int_{\frac{1}{2}-\frac{1}{k}}^{\frac{1}{2}} f(\mathbf{t})g(k)h(k(\mathbf{t} - \tfrac{1}{2} + \tfrac{1}{k}))\,\mathbf{dt} \\
&\quad + \lim_{k\to\infty} \int_{\frac{1}{2}}^{\frac{1}{2}+\frac{1}{k}} f(\mathbf{t})g(-2k)h(-2k(\mathbf{t} - \tfrac{1}{2} - \tfrac{1}{2k}))\,\mathbf{dt} + \lim_{k\to\infty} \int_{\frac{1}{2}+\frac{1}{k}}^1 f(\mathbf{t})g(0)h(-1)\,\mathbf{dt} \\
&= \int_0^{\frac{1}{2}} f(\mathbf{t})g(0)h(0)\,\mathbf{dt} + \int_{\frac{1}{2}}^1 f(\mathbf{t})g(0)h(-1)\,\mathbf{dt} + \lim_{k\to\infty} \int_{\frac{1}{2}-\frac{1}{k}}^{\frac{1}{2}} f(\mathbf{t})g_b(k)\dot{H}(k(\mathbf{t} - \tfrac{1}{2} + \tfrac{1}{k}))\tfrac{1+k}{k}\,\mathbf{dt} \\
&\quad + \lim_{k\to\infty} \int_{\frac{1}{2}}^{\frac{1}{2}+\frac{1}{k}} f(\mathbf{t})g_b(-2k)\dot{H}(-2k(\mathbf{t} - \tfrac{1}{2} - \tfrac{1}{2k}))\tfrac{1+2k}{-2k}\,\mathbf{dt} \\
&= \int_0^{\frac{1}{2}} f(\mathbf{t})g(0)h(0)\,\mathbf{dt} + \int_{\frac{1}{2}}^1 f(\mathbf{t})g(0)h(-1)\,\mathbf{dt} + f(\tfrac{1}{2})g_b(+\infty)(H(1) - H(0)) \\
&\quad + f(\tfrac{1}{2})g_b(-\infty)(H(1) - H(-1))
\end{aligned}$$

$$= \int_0^1 \int_{\beta_{\mathcal{U}}\mathbb{R}^m} \int_Y f(\mathbf{t}) g_b(\mathbf{u}) h(\mathbf{y}) \, \mathbf{d}\upsilon_{\mathbf{t},\mathbf{u}}(\mathbf{y}) \, \mathbf{d}\omega(\mathbf{u}) \, \mathbf{d}\tau(\mathbf{t})$$

where $Y = [-1,1]$, $\tau = \lambda_{[0,1]} + 3\delta_{\frac{1}{2}}$, $\lambda_{\mathcal{X}}$ denotes the Lebesgue measure on $\mathcal{X}$ scaled to be a probability measure and

$$\omega_{\mathrm{t}} = \begin{cases} \frac{1}{2}\delta_{+\infty} + \frac{1}{2}\delta_{-\infty} & \text{if } \mathrm{t} = \frac{1}{2}, \\ \delta_0 & \text{otherwise,} \end{cases} \qquad \upsilon_{\mathrm{t,u}} = \begin{cases} \delta_0 & \text{if } \mathrm{t} \in [0, \frac{1}{2}), \\ \lambda_{[0,1]} & \text{if } \mathrm{t} = \frac{1}{2},\ \mathrm{u} = +\infty, \\ \lambda_{[-1,1]} & \text{if } \mathrm{t} = \frac{1}{2},\ \mathrm{u} = -\infty, \\ \delta_{-1} & \text{if } \mathrm{t} \in (\frac{1}{2}, 1]. \end{cases}$$

4.2.4 Relaxed Optimal Control with Oscillations, Concentrations and Discontinuities

In the following we propose a relaxation of (OCP) based on the generalized DiPerna-Majda measures introduced in the previous section. Therefore we first discuss an equivalent formulation on the space of measures which we relax to a linear problem on measures in a second step. To prove equivalence of the first problem and (OCP) we need to make the following assumption on the regularity of the data.

Assumption 4 (Regularity of the data)**.** *Assume that $L \in C([0,1] \times \mathbb{R}^n \times \mathbb{R}^m)$ is such that $L_b(\mathbf{t},\mathbf{y},\mathbf{u}) := L(\mathbf{t},\mathbf{y},\mathbf{u})(1+\|\mathbf{u}\|^{\mathrm{p}})^{-1} \in C_b([0,1]\times\mathbb{R}^n\times\mathbb{R}^m;\mathbb{R})$ and $F \in C([0,1]\times\mathbb{R}^n\times\mathbb{R}^m;\mathbb{R}^n)$ is such that $F_b(\mathbf{t},\mathbf{y},\mathbf{u}) := F(\mathbf{t},\mathbf{y},\mathbf{u})(1+\|\mathbf{u}\|^{\mathrm{p}})^{-1} \in C_b([0,1]\times\mathbb{R}^n\times\mathbb{R}^m;\mathbb{R}^n)$. Moreover, assume that there is a constant $c_L > 0$ such that*

$$L(\mathrm{t},\mathrm{u},\mathrm{y}) \geq c_L \|\mathrm{u}\|^p \tag{4.20}$$

for all t, u, y *and there is a constant $c_F > 0$ such that*

$$\|F(\mathrm{t},\mathrm{u},\mathrm{y}_1) - F(\mathrm{t},\mathrm{u},\mathrm{y}_2)\| \leq c_F(1+\|\mathrm{u}\|^p)\|\mathrm{y}_1 - \mathrm{y}_2\| \tag{4.21}$$

for all t, u, y_1, y_2.

The coercivity condition (4.20) is to guarantee that optimizing sequences $(u_k)_{k\in\mathbb{N}}$ for (OCP) are bounded in L^p. The second condition (4.21) is a Lipschitz condition on F with respect to the variable $\mathbf{y}$, which typically is used to establish uniqueness of solutions.

In the subsequent we show that the classical optimal control problem (OCP) is equivalent to the following problem on the generalized DiPerna-Majda measures.

$$\begin{aligned} \inf_{\mu} & \int L_b \, \mathbf{d}\mu \\ \text{s.t.} & \int F_b \, \mathbf{d}\mu = \mathrm{y}_1 - \mathrm{y}_0, \\ & \exists (y_k, u_k)_{k\in\mathbb{N}} \subseteq W^{1,1}(0,1;Y) \times L^1(0,1,\mathbb{R}^m) \text{ solving} \\ & \quad \dot{y_k}(\mathrm{t}) = F(\mathrm{t}, u_k(\mathrm{t}), y_k(\mathrm{t})) \quad \text{a.e. on } [0,1], \\ & \quad y_k(0) = \mathrm{y}_0,\ y_k(1) = \mathrm{y}_1, \text{ and generating} \\ & \quad \mu \in \mathcal{P}([0,1] \times \beta_{\mathcal{U}}\mathbb{R}^m \times Y). \end{aligned} \tag{GOCP}$$

It is immediate to see that (OCP)$^* =$ (GOCP)*. However we cannot optimize over measures that possess generating sequences. Therefore, to the reformulation (GOCP) of classical optimal control problem (OCP) we associate the relaxed optimal control problem

$$\begin{aligned} &\inf_{\mu} \int L_b \,\mathbf{d}\mu \\ &\quad \text{s.t.} \int F_b \,\mathbf{d}\mu = \mathrm{y}_T - \mathrm{y}_0, \\ &\qquad \mu \in \mathcal{P}([0,1] \times \beta_{\mathcal{U}}\mathbb{R}^m \times Y). \end{aligned} \tag{ROCP}$$

Note that (ROCP) is *linear* in the unknown measure μ. In contrast, the classical problem (OCP) is non-linear in the unknown trajectory y and control u.

Since optimal control problem (ROCP) is a relaxation of (GOCP), it may happen that the infimum in (ROCP) is strictly less than the infimum in (GOCP), i.e. (ROCP)$^* <$ (GOCP)$^* =$ (OCP)*. Formulating necessary and sufficient conditions on the problem data F and L such that (GOCP)$^* =$ (ROCP)*, i.e., that there is no relaxation gap, is an open problem. However, if we can verify a posteriori that the minimizing measure in (ROCP) is generated by limits of functions, there is no relaxation gap.

4.2.5 Relaxed Optimal Control with Occupation Measures

In the previous subsection, we proposed a linear relaxation of the non-linear optimal control, based on anisotropic parametrized measures. In the current subsection, we describe another linear reformulation proposed in [Las+08b] and relying on the notion of occupation measure. The relation between this linear reformulation and the classical DiPerna-Majda measures was investigated in [CHK17], with the help of a graph completion argument. In the sequel we show that the generalized DiPerna-Majda measures also fit naturally this framework.

Let $\varphi \in C^1([0,1] \times Y)$. Then for any continuous function y, it holds that

$$\int_0^1 \mathbf{d}\varphi(\mathbf{t}, y(\mathbf{t}))\,\mathbf{dt} = \varphi(1, \mathrm{y}(1)) - \varphi(0, \mathrm{y}(0)) = \int_0^1 \left(\frac{\partial \varphi}{\partial \mathbf{t}}(\mathbf{t}, y(\mathbf{t})) + \frac{\partial \varphi}{\partial \mathbf{y}}(\mathbf{t}, y(\mathbf{t})) \cdot \dot{y}(\mathbf{t}) \right) \mathbf{dt}.$$

Consequently optimal control problem (OCP) can be rewritten as

$$\begin{aligned} &\inf_{u} \int_0^1 L(\mathbf{t}, u(\mathbf{t}), y(\mathbf{t}))\,\mathbf{dt} \\ &\text{s.t.} \int_0^1 \left(\frac{\partial \varphi}{\partial \mathbf{t}} + \frac{\partial \varphi}{\partial \mathbf{y}} \cdot F \right)(\mathbf{t}, u(\mathbf{t}), y(\mathbf{t}))\,\mathbf{dt} = \varphi(1, \mathrm{y}_1) - \varphi(0, \mathrm{y}_0),\ \forall \varphi \in C^1([0,1] \times \mathbb{R}^n) \\ &\quad y \in W^{1,1}(0,1;\mathbb{R}^n), \quad u \in L^p(0,1;\mathbb{R}^m). \end{aligned} \tag{4.22}$$

Definition 4.2.3 (Occupation measure)**.** For given control u and trajectory y solving the differential equation in (OCP), we define the occupation measure of $\mu_{u,y} \in \mathcal{P}([0,1] \times \mathbb{R}^n \times \mathbb{R}^m)$ by

$$\mu_{u,y} := \delta_{y(\mathbf{t})} \delta_{u(\mathbf{t})} \lambda_{[0,1]}.$$

Geometrically $\mu_{u,y}(A \times B \times C)$ is the time spent by the trajectory $(\mathbf{t}, u(\mathbf{t}), y(\mathbf{t}))$ in any Borel subset $A \times B \times C$ of $[0,1] \times Y \times \mathbb{R}^m$. Analytically, integration with respect to $\mu_{u,y}$ is the same as integration along $(u(\mathbf{t}), y(\mathbf{t}))$ with respect to time. In particular

$$\int_0^1 L(\mathbf{t}, u(\mathbf{t}), y(\mathbf{t}))\, \mathbf{dt} = \int_0^1 \int_{\mathbb{R}^m} \int_Y L(\mathbf{t}, \mathbf{u}, \mathbf{y})\, \mathbf{d}\mu_{u,y}$$

and for all test functions $\varphi \in C^1([0,1] \times Y)$, it holds that

$$\int_0^1 \left(\frac{\partial \varphi}{\partial \mathbf{t}} + \frac{\partial \varphi}{\partial \mathbf{y}} \cdot F\right)(\mathbf{t}, u(\mathbf{t}), y(\mathbf{t}))\, \mathbf{dt} = \int_0^1 \int_{\mathbb{R}^m} \int_Y \left(\frac{\partial \varphi}{\partial \mathbf{t}} + \frac{\partial \varphi}{\partial \mathbf{y}} \cdot F\right)(\mathbf{t}, \mathbf{u}, \mathbf{y})\, \mathbf{d}\mu_{u,y}$$

Using the same arguments as in [CHK17, Proposition 4], we can reformulate optimal control problem (4.22) as a linear problem on measures, leading to the following relaxed formulation:

$$\begin{aligned} &\inf_{\mu} \int L_b\, \mathbf{d}\mu \\ &\text{s.t.} \int \left(\frac{\partial \varphi}{\partial \mathbf{t}}(1 + \|\mathbf{u}\|^p)^{-1} + \frac{\partial \varphi}{\partial \mathbf{y}} \cdot F_b\right) \mathbf{d}\mu = \varphi(1, \mathrm{y}_1) - \varphi(0, \mathrm{y}_0), \\ &\qquad \forall \varphi \in C^1([0,1] \times Y), \quad \mu \in \mathcal{P}([0,1] \times \beta_{\mathcal{U}}\mathbb{R}^m \times Y). \end{aligned} \tag{MOCP}$$

Note that μ in the above problem is not necessarily an occupation measure in the sense of Definition 4.2.3, but a general probability measure measure in $\mathcal{P}([0,1] \times \beta_{\mathcal{U}}\mathbb{R}^m \times Y)$. For this reason, the infimum in relaxed problem (MOCP) can be strictly less than the infimum in classical problem (OCP), i.e. $(\text{MOCP})^* < (\text{OCP})^*$.

Proposition 4.2.4 (No relaxation gap)**.** *It holds* $(\text{ROCP})^* \leq (\text{MOCP})^* \leq (\text{OCP})^*$ *and hence if there is no relaxation gap in relaxed problem* (ROCP) *then there is no relaxation gap in relaxed problem* (MOCP)*.*

Proof. Just observe that problem (ROCP) corresponds to the particular choice of test functions $\varphi(t, y) := y_k$, $k = 1, \ldots, n$ in problem (MOCP). Hence the infimum in (ROCP) is smaller than the infimum in (MOCP), which is in turn smaller than the infimum in (OCP), i.e. $(\text{ROCP})^* \leq (\text{MOCP})^*$. Now if $(\text{ROCP})^* = (\text{OCP})^*$ then obviously $(\text{MOCP})^* = (\text{OCP})^*$. □

4.2.6 Numerical Example

Once we get to the problem (MOCP), we follow the same strategy as in [CHK17, Section 4], to compute the measures numerically. The procedure is summarized as follows:

1. apply a change of variables to u such that $\beta\mathbb{R}^m$ is mapped into the unit ball;
2. introduce lifting variables to express all data as polynomials;
3. construct an equivalent moment problem where the unknown are moments of the occupation measure supported on a compact semialgebraic set;

4. use the moment-SOS hierarchy to obtain a sequence of approximate moments at the price of solving numerically semidefinite programming problems;

5. from the approximate moments, construct an approximate solution to the optimal control problem.

We illustrate this strategy on our introductory example (4.3). The trajectory y should move the state from zero at initial time to one at final time, yet for the non-negative integrand to be as small as possible, the control u should be zero all the time, except maybe at time zero. If $\varepsilon = 1$ this problem has a trivial optimal solution $u(\mathbf{t}) = 0$. For $\varepsilon = 0$ as explained already we can solve the problem by integration by parts because $\dot{y} = u$. The integration trick cannot be carried out in the case of $\varepsilon \in (0, 1)$.

We use the relaxation (MOCP) to formulate problem (4.3) as a linear problem on measures:

$$\begin{aligned} &\inf_{\mu} \int (\mathbf{t} + \mathbf{y}) \frac{\mathbf{u}}{1 + \mathbf{u}} \, \mathbf{d}\mu \\ &\text{s.t.} \int \frac{\partial \varphi}{\partial \mathbf{t}} \frac{1}{1 + \mathbf{u}} + \frac{\partial \varphi}{\partial \mathbf{y}} \frac{\mathbf{u}}{1 + \mathbf{u}} \, \mathbf{d}\mu = \varphi(1, 1) - \varphi(0, 0), \text{ for all } \varphi \in C^1([0, 1]^2) \\ &\quad \mu \in \mathcal{P}([0, 1] \times \beta_{\mathcal{U}} \mathbb{R}_+ \times [0, 1]). \end{aligned} \tag{4.23}$$

Note that we can omit the absolute value in the denominator, as u is constrained to be non-negative. We expect the control to concentrate. Therefore let $u(\mathbf{t}) := \frac{r(\mathbf{t})}{1 - r(\mathbf{t})}$ with $r(\mathrm{t}) \in [0, 1]$ for all $\mathrm{t} \in [0, 1]$. Then the dynamics of y reads

$$\dot{y}(\mathrm{t}) = \sqrt{\left(\tfrac{r(\mathrm{t})}{1 - r(\mathrm{t})}\right)^2 + \varepsilon^2} = \frac{\sqrt{r(\mathrm{t})^2 + \varepsilon^2 (1 - r(\mathrm{t}))^2}}{1 - r(\mathrm{t})}.$$

Introduce the auxiliary variable $w(\mathbf{t})$ such that $w(\mathbf{t})^2 = r(\mathbf{t})^2 + \varepsilon^2 (1 - r(\mathbf{t}))^2$. By knowledge of bounds for ε and $r(\mathbf{t})$ we can conclude that $0 \leq w(t) \leq 1$ for all $\mathrm{t} \in [0, 1]$. The linear problem on moments than reads

$$\begin{aligned} &\inf_{\gamma} \int (\mathbf{t} + \mathbf{y}) \, \mathbf{r} \, \mathbf{d}\gamma \\ &\text{s.t.} \int \frac{\partial \mathbf{t}^i \mathbf{y}^j}{\partial \mathbf{t}} (1 - \mathbf{r}) + \frac{\partial \mathbf{t}^i \mathbf{y}^j}{\partial \mathbf{y}} w(\mathbf{t}) \, \mathbf{d}\gamma = \varphi(1, 1) - \varphi(0, 0), \text{ for all }, (i, j) \in \mathbb{N}^2 \\ &\quad \gamma \in \mathcal{P}([0, 1]^3). \end{aligned} \tag{4.24}$$

Here we used the notation $w(\mathbf{t})$ for ease of readability. The actual implementation of the moment-SOS relaxation using GloptiPoly optimizes over measures depending on four variables $\mathbf{t}, \mathbf{y}, \mathbf{r}, \mathbf{w}$ and supported on $[0, 1]^4 \cap \{(\mathrm{t}, \mathrm{y}, \mathrm{r}, \mathrm{w}) \| \mathrm{w}^2 = \mathrm{r}^2 + \varepsilon^2 (1 - \mathrm{r})^2\}$ (see Fig. 4.2). With this implementation we could solve the problem numerically for different values of the parameter ε and could guess the analytic optimal solution.

The measure $\mu = \tau \omega_{\mathbf{t}} \upsilon_{\mathbf{t},\mathbf{u}}$ with

$$\tau = \lambda_{[0,1]} + (1 - \varepsilon)\delta_0, \quad \omega_{\mathrm{t}} = \begin{cases} \delta_\infty, & t = 0 \\ \delta_0, & t > 0 \end{cases}, \quad \upsilon_{\mathrm{t,u}} = \begin{cases} \frac{1}{1-\varepsilon} \lambda_{[0, 1-\varepsilon]}, & t = 0 \\ \delta_{1 - \varepsilon + \varepsilon t}, & t > 0 \end{cases}$$

```
%% Parameters
d = 6;      % relaxation order (moments of size up to 2*d)
e = 0.2;    % epsilon

%% GMP
% occupation measure
mpol t y r w
gamma = meas([t y r w]);

% test function
v = mmon([t y],2*d);

% initial condition
% assignment of r and w is not taken into when evaluating v(t,y)
assign([t y r w],[0 0 0 0]);
v0 = double(v);

% final condition
assign([t y r w],[1 1 0 0]);
vT = double(v);

% moment problem
objective = min((t+y)*r);
support = [t-t^2>=0, y-y^2>=0, r-r^2>=0, w-w^2>=0, w^2==r^2+e^2*(1-r)^2];
dynamic = vT-v0 == mom(diff(v,t)*(1-r) + diff(v,y)*w);

P = msdp(objective, support, dynamic);

% solve
[stat,obj] = msol(P);

% display moments of solution
if stat>=0
fprintf('computed optimum: %6.4f\n', obj)
fprintf('int t^k int y^k int r^k int w^k\n')
fprintf('%6.4f\t %6.4f\t %6.4f\t %6.4f\n',[double(mom(mmon(t,2*d))),...
double(mom(mmon(y,2*d))),double(mom(mmon(r,2*d))),double(mom(mmon(w,2*d)))])
end
```

Figure 4.2: GloptiPoly code to solve the 6th relaxation with $\varepsilon = 0.2$.

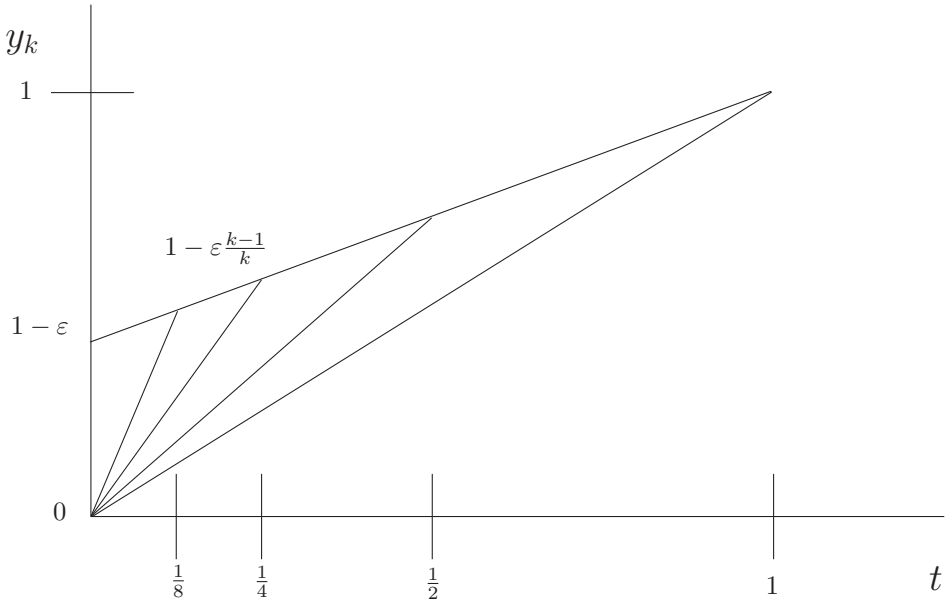

Figure 4.3: Sequence $(y_k)_{k=1,2,4,8}$ from Eq. (4.3).

is optimal for (4.3) and yields the value $\frac{(1-\varepsilon)^2}{2}$. It is attained by the following sequences (see Fig. 4.3):

$$u_k(t) = \begin{cases} \sqrt{(k(1-\varepsilon)+\varepsilon)^2-\varepsilon^2}, & t \in [0, \frac{1}{k}] \\ 0, & t > \frac{1}{k} \end{cases}, \quad y_k(t) = \begin{cases} (k(1-\varepsilon)+\varepsilon)t, & t \in [0, \frac{1}{k}] \\ \varepsilon t + 1 - \varepsilon, & t > \frac{1}{k} \end{cases}$$

The numerical moments for the 6th relaxation (i.e. moments of degree up to 12) obtained with GloptiPoly and the SeDuMi semidefinite solver are reported in Table 4.1. They match to 4 significant digits with the analytic moments reported in Table 4.2.

4.2.7 Conclusion

In this section we considered critical limit behaviour of controls and trajectories in problems of optimal control. In particular we proposed an approach giving meaning to the integral cost when the trajectory is discontinuous at a point where the control concentrates, using anisotropic measures generated by pairs of sequences. This work is a consequent continuation of [CHK17] where the authors already were able to consider discontinuities of the trajectory and concentration of the control, but under the severe condition that they do not occur at the same time.

We further proposed to approximate solutions by the moment-SOS hierarchy. Therefore we formulated a relaxation based on occupation measures, which is stronger than the relaxation based on generalized DiPerna-Majda measures, but still weaker than the original problem. We propose to compute solutions to the relaxation with the moment-SOS hierarchy.

k	$\int \mathbf{t}^k \, \mathbf{d}\mu$	$\int \mathbf{y}^k \, \mathbf{d}\mu$	$\int \mathbf{r}^k \, \mathbf{d}\mu$	$\int \mathbf{w}^k \, \mathbf{d}\mu$
0	1.8000	1.8000	1.8000	1.8000
1	0.5000	1.2200	0.8000	1.0000
2	0.3333	0.9840	0.8000	0.8400
3	0.2500	0.8404	0.8000	0.8080
4	0.2000	0.7379	0.8000	0.8016
5	0.1667	0.6586	0.8000	0.8003
6	0.1429	0.5944	0.8000	0.8001
7	0.1250	0.5411	0.8000	0.8000
8	0.1111	0.4959	0.8000	0.8000
9	0.1000	0.4571	0.8000	0.8000
10	0.0909	0.4233	0.8000	0.8000
11	0.0833	0.3938	0.8000	0.8000
12	0.0769	0.3677	0.8000	0.8000

Table 4.1: Approximate moments for $\varepsilon = 0.2$, computed with Gloptipoly and SeDuMi.

$\int \mathbf{t}^k \, \mathbf{d}\mu$	$\frac{1}{k+1} + (1-\varepsilon)0^k$
$\int \mathbf{y}^k \, \mathbf{d}\mu$	$\frac{(1-\varepsilon)^{k+1}}{(k+1)} - \frac{(1-\varepsilon)^{k+1}-1}{\varepsilon(k+1)}$
$\int \mathbf{r}^k \, \mathbf{d}\mu$	$0^k + (1-\varepsilon)$
$\int \mathbf{w}^k \, \mathbf{d}\mu$	$\varepsilon^k + (1-\varepsilon)$

Table 4.2: Analytic expressions of the moments.

4.3 A Moment Approach for Entropy Solutions to Non-linear Hyperbolic PDEs

This section is concerned with the numerical study of scalar non-linear hyperbolic conservation laws, which model numerous physical phenomena such as fluid mechanics, traffic flow or non-linear acoustics [Daf00], [Whi11]. The existence and uniqueness of solutions to the associated Cauchy problem crucially depends on the flux and the initial condition [Kru70]. Even if the solution is unique, its numerical computation is a challenge – in particular in the case when the solution has a shock, i.e., a discontinuity, locating the position of the shock at a given time is a challenge. Existing numerical schemes based on discretization such as [God59] suffer from numerical dissipation: the shock is smoothened in the numerical solution and it cannot be represented accurately. In fact, sometimes the exact location of the shock is of crucial interest for applications that need to deal with conservation laws. Note however that some existing numerical schemes are able to capture shock in the case where the conservation laws under consideration are linear, see e.g. [DL01].

In contrast to existing methods, a distinguishing feature of the numerical scheme we present in this section is to *not* rely on discretization; it computes the solution in a given time-space-window globally. From such a solution, the location of the shock at a given time can potentially be computed up to the limits of machine precision.

Measure-Valued Solutions While partial differential equations are usually understood in a weak sense, DiPerna proposed an even weaker notion, so-called *measure-valued solutions* (mv solutions) [DiP85], which are based on Young measures, see Section 4.1. DiPerna introduces mv solutions to conservation laws as measures on the state space, depending on time and space.

Naturally, every weak solution gives rise to a mv solution when identifying a solution $y(\mathbf{t}, \mathbf{x})$ with the Young measure $\delta_{y(\mathbf{t},\mathbf{x})}$. In such a situation, we say that the mv solution is concentrated on (the graph of) the solutions. In the following, we focus on a set up where both weak and mv solutions are unique. In this case, both solutions coincide via the identification just mentioned. Note however that our approach also applies without any change to situations where the mv solution is not concentrated, e.g., because of an initial condition that is not concentrated either.

Recently there has been an increasing interest in numerical schemes to compute mv solutions for hyperbolic conversation laws with non-concentrated initial condition [Fjo+17; FLM18]. These schemes apply standard discretization methods to compute sufficiently many trajectories according to the distribution of the initial condition and recover the moments of the mv solution by considering limits of the trajectories. In contrast to this our approach *directly computes the moments* of the mv solution. Therefore in some sense the work presented in this section is in the opposite direction. We compute moments to recover trajectories in the case where the solution is concentrated.

To ensure uniqueness of the solution and concentration of the mv solution, we rely on the notion of *entropy solutions* which has been extended to entropy mv solutions. Entropy is a concept from thermodynamics and makes reference to the fact that differences in physical systems, e.g., the densities of particles in a room, tend to adjust to each other. It is well known that the entropy solution of a scalar non-linear hyperbolic conservation law is unique. For the generalized situation things are more involved. However under suitable assumptions on the initial condition, uniqueness of entropy mv solutions can be proved.

Generalized Moment Problem The key idea underlying the approach is to consider mv solutions as solutions to a particular instance of the Generalized Moment Problem. This is in line with Section 4.2 and many other applications proposed, e.g, in [Las10b], i.e., (i) mv solutions are viewed (or formulated) as solutions of a particular instance of the GMP, and (ii) the moments of mv solutions are approximated as closely as desired by solving an appropriate hierarchy of semidefinite programs of increasing size (seeChapter 1). Any optimal solution of each moment-SOS relaxation at step d in the hierarchy provides information about the mv solution in the form of a sequence of its (approximated) moments up to degree $2d$; the higher is d the better is the approximation of its moments. As we restrict to measures with compact support, they are fully characterized from knowledge of the complete sequence of their moments. It is worth noting that in [Fjo+17] it was already pointed out that the statistical moments of mv solutions are precisely the quantities of interest.

Outline This section is based on [Mar+18] and organized as follows. Section 4.3.1 introduces different notions of solutions for scalar conservation laws and provides some links between these notions. In Section 4.3.2 we show that the mv solution framework can be

written as an instance of the GMP. Finally in Section 4.3.3 we provide a numerical study of our approach on the Burgers equation for the Riemann problem.

4.3.1 Notions of solutions

We start with a brief overview of different notions of solutions to scalar polynomial partial differential equations. For details, we refer to [Daf00] for weak solutions and [Mál+96] for mv solutions. The aim of this section is to give a clear link between these two concepts of solutions.

Weak and entropy solutions

In order to study mv solutions, it is instructive to revise the classical concept of weak solutions first. Consider therefore the Cauchy problem of finding a function $y \in L^\infty(\mathbb{R}_+ \times \mathbb{R})$ such that

$$\frac{\partial}{\partial \mathbf{t}} y(\mathbf{t}, \mathbf{x}) + \frac{\partial}{\partial \mathbf{x}} f(y(\mathbf{t}, \mathbf{x})) = 0, \tag{4.25a}$$
$$y(0, \mathrm{x}) = y_0(\mathrm{x}), \quad \text{a.e. on } \mathbb{R}, \tag{4.25b}$$

where (4.25a) is a scalar hyperbolic conservation law with $f \in C^1(\mathbb{R})$ and (4.25b) provides an initial condition $y_0 \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$. We will restrict the discussion to $f \in \mathbb{R}[\mathbf{y}]$ in order to be able to apply the moment-SOS approach later on. Our standard example for such a conservation law is Burgers equation, which corresponds to (4.25a) by setting $f(y) = \frac{1}{2}y^2$. Even if the initial condition y_0 is smooth, solutions to (4.25) might be discontinuous [Eva98, p. 143]. Solutions to this problem are hence usually understood in the following *weak sense*:

Definition 4.3.1 (Weak solution)**.** A function $y \in L^\infty(\mathbb{R}_+ \times \mathbb{R})$ is a weak solution to (4.25) if for all test functions $\varphi \in C_c^1(\mathbb{R}_+ \times \mathbb{R})$

$$\int_{\mathbb{R}_+} \int_{\mathbb{R}} \left\{ \frac{\partial}{\partial t} \varphi(\mathbf{t}, \mathbf{x}) y(\mathbf{t}, \mathbf{x}) + \frac{\partial}{\partial \mathbf{x}} \varphi(\mathbf{t}, \mathbf{x}) f(y(\mathbf{t}, \mathbf{x})) \right\} \mathbf{dx}\, \mathbf{dt} + \int_{\mathbb{R}} \varphi(0, \mathbf{x}) y_0(\mathbf{x})\, \mathbf{dx} = 0. \tag{4.26}$$

In general, weak solutions to (4.25) are not unique. However it can be shown (see e.g. [Kru70]) that among all possible weak solutions, only one has a physical meaning. This solution is called the *entropy solution* and can be characterized as follows.

Definition 4.3.2 (Entropy pair/entropy solution)**.** (i) A pair of functions $\eta, q \in C^1(\mathbb{R})$ is called an *entropy pair* for (4.25a) if η is strictly convex and $q' = f'\eta'$.

(ii) A weak solution $y \in L^\infty(\mathbb{R}_+ \times \mathbb{R})$ of (4.25) is said to be an *entropy solution* if *for all entropy pairs* and all non-negative test functions $\psi \in C_c^1(\mathbb{R}_+ \times \mathbb{R})$,

$$\int_{\mathbb{R}_+} \int_{\mathbb{R}} \left\{ \frac{\partial}{\partial t} \psi(\mathbf{t}, \mathbf{x}) \eta(y(\mathbf{t}, \mathbf{x})) + \frac{\partial}{\partial \mathbf{x}} \psi(\mathbf{t}, \mathbf{x}) q(y(\mathbf{t}, \mathbf{x})) \right\} \mathbf{dx}\, \mathbf{dt} + \int_{\mathbb{R}} \psi(0, \mathbf{x}) \eta(y_0(\mathbf{x}))\, \mathbf{dx} \geq 0. \tag{4.27}$$

Measure-valued solutions

In general, regularity results for conservation laws are obtained from regularized conservation laws

$$\frac{\partial}{\partial \mathbf{t}} y^{(\varepsilon)}(\mathbf{t}, \mathbf{x}) + \frac{\partial}{\partial \mathbf{x}} f(y^{(\varepsilon)}(\mathbf{t}, \mathbf{x})) - \varepsilon \frac{\partial^2}{\partial \mathbf{x}^2} y^{(\varepsilon)}(\mathbf{t}, \mathbf{x}) = 0,$$

where $\varepsilon > 0$ [Daf00, Section 6.3]. Then one studies the limit of solutions $y^{(\varepsilon)}$ as ε goes to 0 and tries to preserve some regularity properties of the regularized equation for the original conservation law. Due to the lack of reflexivity of L^∞, regularized solution $y^{(\varepsilon)}$ do not necessarily converge to a weak solution y of (4.25). However, as seen in Section 4.1 such sequences posses subsequences that converge in the sense of Young to a measure valued measurable function. When allowing for measure valued functions as solutions to differential equations we can consider problems more general than (4.25), where the initial condition (4.25b) is replaced by a Young measure parametrized in space (see e.g. [Fjo+17] and the references therein). The generalized problem now is to find a Young measure $\mu : \mathbb{R}_+ \times \mathbb{R} \to \mathcal{P}(\mathbb{R})$, which satisfies the following Cauchy problem:

$$\partial_{\mathbf{t}} \left\langle \mu_{(\mathbf{t},\mathbf{x})}, \mathbf{y} \right\rangle + \partial_{\mathbf{x}} \left\langle \mu_{(\mathbf{t},\mathbf{x})}, f(\mathbf{y}) \right\rangle = 0, \tag{4.28a}$$

$$\mu_{(0,\mathrm{x})} = \sigma_0, \quad \text{a.e. on } \mathbb{R}, \tag{4.28b}$$

where $\langle \mu, f \rangle$ denotes as usual integration of a function f with respect to a measure μ, and the measure σ_0 is a given Young measure on $\mathbb{R}$. The conservation law (4.28a) has to be understood in the sense of distributions, i.e.:

Definition 4.3.3 (Measure-valued solution)**.** A Young measure μ is said to be a measure-valued solution to (4.28) if, for all test functions $\varphi \in C_c^1(\mathbb{R}_+ \times \mathbb{R})$, it satisfies,

$$\int_{\mathbb{R}_+} \int_{\mathbb{R}} \left\{ \tfrac{\partial}{\partial \mathbf{t}} \varphi(\mathbf{t},\mathbf{x}) \left\langle \mu_{(\mathbf{t},\mathbf{x})}, \mathbf{y} \right\rangle + \tfrac{\partial}{\partial \mathbf{x}} \varphi(\mathbf{t},\mathbf{x}) \left\langle \mu_{(\mathbf{t},\mathbf{x})}, f(\mathbf{y}) \right\rangle \right\} \mathbf{dx}\,\mathbf{dt} + \int_{\mathbb{R}} \varphi(0,\mathbf{x}) \left\langle \sigma_0, \mathbf{y} \right\rangle \mathbf{dx} = 0, \tag{4.29}$$

Note that the weak solution y has been replaced by a time-space parametrized probability measure μ supported on the range of y. Whereas a weak solution y is requested to satisfy (4.26), only averages of the mv solution (i.e. moments up to order $\deg(f)$) are constrained in (4.29). In this sense the concept of mv solutions is much weaker than the classical weak sense (compare to Section 4.2).

It is easy to see that every weak solution induces an mv solution via the canonical embedding $y(\mathbf{t},\mathbf{x}) \mapsto \delta_{y(\mathbf{t},\mathbf{x})}$. As in the case of weak solutions, an entropy condition is needed in order to select solutions with a physical meaning. Quite in analogy to entropy solutions, entropy mv solutions are defined as follows.

Definition 4.3.4 (Entropy measure-valued solution)**.** A measure-valued solution μ is said to be an entropy mv solution to (4.28) if it satisfies, for all entropy pairs (η, q) and all *non-negative* test functions $\psi \in C_c^1(\mathbb{R}_+ \times \mathbb{R})$,

$$\begin{aligned} \int_{\mathbb{R}_+} \int_{\mathbb{R}} \left\{ \tfrac{\partial}{\partial \mathbf{t}} \psi(\mathbf{t},\mathbf{x}) \left\langle \mu_{(\mathbf{t},\mathbf{x})}, \eta(\mathbf{y}) \right\rangle + \tfrac{\partial}{\partial \mathbf{x}} \psi(\mathbf{t},\mathbf{x}) \left\langle \mu_{(\mathbf{t},\mathbf{x})}, q(\mathbf{y}) \right\rangle \right\} \mathbf{dx}\,\mathbf{dt} \\ + \int_{\mathbb{R}} \psi(0,\mathbf{x}) \left\langle \sigma_0, \eta(\mathrm{y}) \right\rangle \mathbf{dx} \geq 0. \end{aligned} \tag{4.30}$$

Again it is straightforward to see that entropy solutions are entropy mv solutions via the canonical embedding $y(\mathbf{t},\mathbf{x}) \mapsto \delta_{y(\mathbf{t},\mathbf{x})}$. However, as demonstrated on an example in [Fjo+17, p. 775], in contrast to entropy solutions, entropy mv solutions are not necessarily unique.

We have seen that the concept of mv solutions is weaker than the classical concept of weak solutions. Hence mv solutions are a relaxation of weak solutions. However the following result states that considering mv solutions is not relaxing too much.

Theorem 4.3 (Concentration of the entropy mv solution)**.** *Let y be an entropy solution to (4.25) and μ be an entropy measure-valued solution to (4.28). If $\sigma_0 = \delta_{y_0(\mathbf{x})}$, then $\mu_{(\mathbf{t},\mathbf{x})} = \delta_{y(\mathbf{t},\mathbf{x})}$.*

This states that when the initial measure in (4.28b) is concentrated on the support of the initial condition in (4.25b), the entropy mv solution to (4.28) is unique and concentrated on the graph of the (unique) entropy solution to (4.25). A proof of the theorem can be found in [DiP85; Fjo+17]. It is based on the doubling variable strategy and considers the following family of entropy pairs:

$$\eta_{\mathrm{v}}(\mathbf{y}) = |\mathbf{y} - \mathrm{v}|, \quad q_{\mathrm{v}}(\mathbf{y}) = \operatorname{sgn}(\mathbf{y} - \mathrm{v})(f(\mathbf{y}) - f(\mathrm{v})), \quad \forall \mathrm{v} \in \mathbb{R}. \tag{4.31}$$

In [Lax71], it has been proved that a linear combination of this special entropy pairs, together with the convex hull of linear functions, generate all entropy pairs. In [Mar+18] we adapt the proof of DiPerna to a compact set up.

An emphasis on compact sets

In practice, one computes solution on compact subsets of $\mathbb{R}_+ \times \mathbb{R}$. Therefore let $(L, R, T) \in \mathbb{R}^2 \times \mathbb{R}_+$ be fixed and define

$$\mathbf{X} := [L, R], \ \mathbf{T} := [0, T].$$

After scaling, we assume without loss of generality that $T = R - L = 1$. Note that the entropy condition (4.27) induces a stability property:

$$\|y(\mathrm{t}, \mathbf{x})\|_{L^\infty(\mathbf{X})} \leq \|y_0(\mathbf{x})\|_{L^\infty(\mathbb{R})}, \ \forall \mathrm{t} \geq 0 \tag{4.32}$$

see e.g. [Daf00, Theorem 6.2.4]. Since y_0 is bounded in L^∞, it follows from the maximum principle [Daf00, Theorem 6.3.2] that $y(\mathrm{t}, \mathbf{x})$ is bounded in L^∞ for all $\mathrm{t} \geq 0$. Hence, we can consider that y takes values in a compact set

$$\mathbf{Y} := [\underline{\mathrm{y}}, \overline{\mathrm{y}}], \tag{4.33}$$

where the bounds $\underline{\mathrm{y}} := \operatorname{ess\,inf}_{\mathrm{x} \in \mathbb{R}} y_0(\mathrm{x})$ and $\overline{\mathrm{y}} := \operatorname{ess\,sup}_{\mathrm{x} \in \mathbb{R}} y_0(\mathrm{x})$ depend on the initial condition.

On $\mathbf{T} \times \mathbf{X}$, an entropy solution to (4.25) is defined as follows:

Definition 4.3.5 (Entropy solution on compact sets)**.** A function $y \in L^\infty(\mathbf{T} \times \mathbf{X})$ is said to be an *entropy solution* to (4.25) *on* $\mathbf{T} \times \mathbf{X}$ if y satisfies, for all test functions $\varphi \in C^1(\mathbf{T} \times \mathbf{X})$,

$$\begin{aligned}\int_{\mathbf{T}}\int_{\mathbf{X}}\left\{\tfrac{\partial}{\partial \mathbf{t}}\varphi(\mathbf{t},\mathbf{x})y(\mathbf{t},\mathbf{x})+\tfrac{\partial}{\partial \mathbf{x}}\varphi(\mathbf{t},\mathbf{x})f(y(\mathbf{t},\mathbf{x}))\right\}\mathbf{dx}\,\mathbf{dt}\\ +\int_{\mathbf{X}}\{\varphi(0,\mathbf{x})y_0(\mathbf{x})-\varphi(T,\mathbf{x})y(T,\mathbf{x})\}\,\mathbf{dx}\\ +\int_{\mathbf{T}}\{\varphi(\mathbf{t},L)f(y(\mathbf{t},L))-\varphi(\mathbf{t},R)f(y(\mathbf{t},R))\}\,\mathbf{dt}=0\end{aligned}\tag{4.34}$$

and for all convex pairs (η,q) and all non-negative test function $\psi\in C^1(\mathbb{R}_+\times\mathbb{R})$,

$$\begin{aligned}\int_{\mathbf{T}}\int_{\mathbf{X}}\left\{\tfrac{\partial}{\partial \mathbf{t}}\varphi(\mathbf{t},\mathbf{x})\eta(y(\mathbf{t},\mathbf{x}))+\tfrac{\partial}{\partial \mathbf{x}}\varphi(\mathbf{t},\mathbf{x})q(y(\mathbf{t},\mathbf{x}))\right\}\mathbf{dx}\,\mathbf{dt}\\ +\int_{\mathbf{X}}\{\varphi(0,\mathbf{x})\eta(y_0(\mathbf{x}))-\varphi(T,\mathbf{x})\eta(y(T,\mathbf{x}))\}\,\mathbf{dx}\\ +\int_{\mathbf{T}}\{\varphi(\mathbf{t},L)q(y(\mathbf{t},L))-\varphi(\mathbf{t},R)q(y(\mathbf{t},R))\}\,\mathbf{dt}\geq 0.\end{aligned}\tag{4.35}$$

As we work on compact sets, the test functions do not need to vanish at infinity. However new terms $y(\mathbf{t},R)$, $y(\mathbf{t},L)$, and $y(T,\mathbf{x})$ now appear. Similar to this notion of solutions on compact sets, we can consider mv entropy solution on compact sets.

Definition 4.3.6 (Measure-valued entropy solution on compact sets)**.** A Young measure $\mu:\mathbf{T}\times\mathbf{X}\mapsto\mathcal{P}(\mathbf{Y})$ is said to be an *entropy measure-valued solution* to (4.28) *on* $\mathbf{T}\times\mathbf{X}$ if it satisfies for all test function $\varphi\in C^1(\mathbf{T}\times\mathbf{X})$,

$$\begin{aligned}\int_{\mathbf{T}}\int_{\mathbf{X}}\left\{\tfrac{\partial}{\partial \mathbf{t}}\varphi(\mathbf{t},\mathbf{x})\left\langle\mu_{(\mathbf{t},\mathbf{x})},\mathbf{y}\right\rangle+\frac{\partial}{\partial \mathbf{x}}\varphi(\mathbf{t},\mathbf{x})\left\langle\mu_{(\mathbf{t},\mathbf{x})},f(\mathbf{y})\right\rangle\right\}\mathbf{dx}\,\mathbf{dt}\\ +\int_{\mathbf{X}}\{\varphi(0,\mathbf{x})\langle\sigma_0,\mathbf{y}\rangle-\varphi(T,\mathrm{x})\langle\sigma_T,\mathbf{y}\rangle\}\,\mathbf{dx}\\ +\int_{\mathbf{T}}\{\varphi(\mathbf{t},L)\langle\sigma_L,f(\mathbf{y})\rangle-\varphi(\mathbf{t},R)\langle\sigma_R,f(\mathbf{y})\rangle\}\,\mathbf{dt}=0\end{aligned}\tag{4.36}$$

and for all convex pairs (η,q) and all non-negative test functions $\psi\in C^1(\mathbf{T}\times\mathbf{X})$,

$$\begin{aligned}\int_{\mathbf{T}}\int_{\mathbf{X}}\left\{\tfrac{\partial}{\partial \mathbf{t}}\psi(\mathbf{t},\mathbf{x})\left\langle\mu_{(\mathbf{t},\mathbf{x})},\eta(\mathbf{y})\right\rangle+\frac{\partial}{\partial \mathbf{x}}\psi(\mathbf{t},\mathbf{x})\left\langle\mu_{(\mathbf{t},\mathbf{x})},q(\mathbf{y})\right\rangle\right\}\mathbf{dx}\,\mathbf{dt}\\ +\int_{\mathbf{X}}\{\psi(0,\mathbf{x})\langle\sigma_0,\eta(\mathbf{y})\rangle-\psi(T,\mathrm{x})\langle\sigma_T,\eta(\mathbf{y})\rangle\}\,\mathbf{dx}\\ +\int_{\mathbf{T}}\{\psi(\mathbf{t},L)\langle\sigma_L,q(\mathbf{y})\rangle-\psi(\mathbf{t},R)\langle\sigma_R,q(\mathbf{y})\rangle\}\,\mathbf{dt}\geq 0\end{aligned}\tag{4.37}$$

where σ_L, σ_R and σ_T are Young measures supported on $\mathbf{T}$ and $\mathbf{X}$, respectively.

Remark 4.3.7 (Imposing constraints on the boundary). To ensure concentration of $\mu_{(\mathbf{t},\mathbf{x})}$ on the graph of the solution to (4.34)-(4.35), in addition to the condition $\sigma_0=\delta_{y_0}$, one may impose conditions on the boundary measures σ_L and/or σ_R. In practice, one knows the initial condition in an interval larger than $\mathbf{X}$ and so one is able to impose σ_L and/or σ_R. The width of this interval depends on the Lipschitz constant of the flux, T, L and R. As an illustrative example, consider the case where the initial condition is positive and the flux is strictly convex. By the classical method of characteristics, if the initial condition

y_0 is positive then so is the solution y for all $t \geq 0$. In particular if f is strictly convex we only need to impose knowledge at the left of the box $\mathbf{X}$ to guarantee unique feasibility. Therefore σ_L has to be known for all $t \in \mathbf{T}$, and σ_R is unconstrained. We refer to [LeV92] for a more precise discussion on the choice of the boundary constraint.

4.3.2 Mv Solutions as Solutions of the GMP

In the previous section, we introduced measure-valued solution for scalar hyperbolic equations. The aim of this section is to express formulations (4.36)-(4.37) as constraints on the moments in order to apply the moment-SOS hierarchy to approach the solutions. We also discuss how one can interpret the sequence of moments to recover information about the graph of the solution.

Moment constraints for the entropy mv solution

Let $\phi \in \mathcal{M}(\mathbf{T} \times \mathbf{X} \times \mathbf{Y})_+$. In the following, we derive moment constraints that will imply that ϕ can be disintegrated as follows

$$\phi = \lambda_{\mathbf{T}\times\mathbf{X}} \mu_{(\mathbf{t},\mathbf{x})}, \tag{4.38}$$

where μ is an entropy measure-valued solution satisfying in particular (4.36) and (4.37) and $\lambda_{\mathcal{X}}$ denotes the Lebesgue measure on $\mathcal{X}$ scaled to be a probability measure. Similar to ϕ we introduce the "boundary" measures $\phi_T, \phi_0, \phi_L, \phi_L \in \mathcal{M}_+(\mathbf{T} \times \mathbf{X} \times \mathbf{Y})$ that respect

$$\begin{aligned} \phi_T &= (\delta_T \otimes \lambda_{\mathbf{X}})\sigma_T \\ \phi_0 &= (\delta_0 \otimes \lambda_{\mathbf{X}})\delta_{y_0} \\ \phi_L &= (\lambda_{\mathbf{T}} \otimes \delta_L)\sigma_L \\ \phi_L &= (\lambda_{\mathbf{T}} \otimes \delta_R)\sigma_R \end{aligned} \tag{4.39}$$

where σ_T is an unknown Young measure depending on space and σ_L and σ_R are Young measures depending on time and are known or unknown, depending on the problem formulation (compare Remark 4.3.7).

Remark 4.3.8. We highlight that the framework presented here could be easily modified to consider the *inverse problem* (e.g., [GZ17]), where one is interested in finding initial conditions that yield some terminal criterion at $t = T$. One only needs to replace the given initial measure ϕ_0 by an unknown measure $\sigma_T(\delta_{\mathbf{t}=0} \otimes \lambda_{\mathbf{X}})$ and prescribe the measure ϕ_T instead. How this framework can be made sound for the inverse problem is part of ongoing research.

Similar as mentioned in the Chapter 3 and by compactness of $\mathbf{T}$, $\mathbf{X}$, and $\mathbf{Y}$, we can impose the marginal properties in (4.38) and (4.39) as moment constraints. Here, we focus on reformulating the properties (4.36) and (4.37) as moment constraints. We split the exposition into two steps: the first one deals with (4.36), while the second one deals with (4.37).

Enforcing (4.36) by moment constraints is rather easy. It suffices to show that restricting the test functions φ to monomials is sufficient to enforce (4.36).

Lemma 4.3.9. *Let $\phi, \phi_T, \phi_0, \phi_L, \phi_R$ be as in (4.38) and (4.39). Then (4.36) is equivalent to*

$$\begin{aligned}\left\langle \phi, \tfrac{\partial}{\partial \mathbf{t}}(\mathbf{t}^{\alpha_1}\mathbf{x}^{\alpha_2}\mathbf{y}) + \tfrac{\partial}{\partial \mathbf{x}}(\mathbf{t}^{\alpha_1}\mathbf{x}^{\alpha_2} f(\mathbf{y}))\right\rangle = \\ \left\langle \phi_T - \phi_0, \tfrac{\partial}{\partial \mathbf{t}}(\mathbf{t}^{\alpha_1}\mathbf{x}^{\alpha_2}\mathbf{y})\right\rangle + \left\langle \phi_R - \phi_L, \tfrac{\partial}{\partial \mathbf{x}}(\mathbf{t}^{\alpha_1}\mathbf{x}^{\alpha_2} f(\mathbf{y}))\right\rangle, \quad \forall \alpha \in \mathbb{N}^2.\end{aligned} \tag{4.40}$$

Proof. We note first that (4.36) implies (4.40) as $(\mathrm{t}, \mathrm{x}) \mapsto \mathrm{t}^{\alpha_1}\mathrm{x}^{\alpha_2} \in C^1(\mathbf{T}\times\mathbf{X})$ for all $\alpha \in \mathbb{N}^2$. For the converse implication note that it is sufficient to restrict the set of test functions in (4.36) to $\varphi \in \mathbb{R}[\mathbf{t}, \mathbf{x}]$ since $\mathbf{T} \times \mathbf{X}$ is compact. Therefore let, $\varphi := \sum_{\alpha\in\mathbb{N}^2} c_\alpha \mathbf{t}^{\alpha_1}\mathbf{x}^{\alpha_2}$ with finitely many non-zero coefficients $c_\alpha \in \mathbb{R}$. Then, (4.40) implies by linearity of integration and the differentiation rules for polynomials that

$$\begin{aligned}0 = \sum_{\alpha\in\mathbb{N}^2} c_\alpha \Big(&\left\langle \phi, \tfrac{\partial}{\partial \mathbf{t}}(\mathbf{t}^{\alpha_1}\mathbf{x}^{\alpha_2}\mathbf{y}) + \tfrac{\partial}{\partial \mathbf{x}}(\mathbf{t}^{\alpha_1}\mathbf{x}^{\alpha_2} f(\mathbf{y}))\right\rangle \\ &- \left\langle \phi_T - \phi_0, \tfrac{\partial}{\partial \mathbf{t}}(\mathbf{t}^{\alpha_1}\mathbf{x}^{\alpha_2}\mathbf{y})\right\rangle - \left\langle \phi_R - \phi_L, \tfrac{\partial}{\partial \mathbf{x}}(\mathbf{t}^{\alpha_1}\mathbf{x}^{\alpha_2} f(\mathbf{y}))\right\rangle\Big) \\ = &\left\langle \phi, \tfrac{\partial}{\partial \mathbf{t}}(\varphi\mathbf{y}) + \tfrac{\partial}{\partial \mathbf{x}}(\varphi f(\mathbf{y}))\right\rangle - \left\langle \phi_T - \phi_0, \tfrac{\partial}{\partial \mathbf{t}}(\varphi\mathbf{y})\right\rangle - \left\langle \phi_R - \phi_L, \tfrac{\partial}{\partial \mathbf{x}}(\varphi f(\mathbf{y}))\right\rangle.\end{aligned}$$

which shows that (4.36) holds for all $\varphi \in \mathbb{R}[\mathbf{t}, \mathbf{x}]$ and hence concludes the proof. □

Enforcing (4.37) by moment constraints is more involved and relies on the entropies of Kruzkhov. As already mentioned it suffices to state the entropy inequality (4.37) for all Kruzkhov entropies given in (4.31) [Lax71]. However, in order to express (4.37) as moment constraints, we are faced with three issues: first, the entropies of Kruzkhov induce an uncountable family of constraints parametrized by $\mathrm{v} \in \mathbf{Y}$, second, the absolute value cannot be expressed as linear combination of monomials, and third, in contrast to Lemma 4.3.9, here we need to consider *non-negative* test functions.

To deal with the first two issues, we introduce a new variable $\mathbf{v}$ to the problem and double the number of measures. More precisely, Let $\mathbf{\Lambda} := \mathbf{Y}$, $\mathbf{W}^- := \{(\mathrm{y}, \mathrm{v}) \in \mathbf{Y} \times \mathbf{\Lambda} : \mathrm{y} \le \mathrm{v}\}$, and $\mathbf{W}^+ := \{(\mathrm{y}, \mathrm{v}) \in \mathbf{Y} \times \mathbf{\Lambda} : \mathrm{y} \ge \mathrm{v}\}$. We define Borel measures $\phi^- \in \mathcal{M}_+(\mathbf{T} \times \mathbf{X} \times \mathbf{W}^-)$ and $\phi^+ \in \mathcal{M}_+(\mathbf{T} \times \mathbf{X} \times \mathbf{W}^-)$ by

$$\phi^- + \phi^+ = \phi \otimes \lambda_{\mathbf{\Lambda}}. \tag{4.41}$$

Note, that $\lambda_{\mathbf{T}\times\mathbf{X}\times\mathbf{Y}\times\mathbf{V}}\left(\mathrm{supp}(\phi^-) \cap \mathrm{supp}(\phi^+)\right) = 0$ and the marginals of both measures with respect to time-space $(\phi^-)^{\mathbf{tx}}$ and $(\phi^+)^{\mathbf{tx}}$ are $\lambda_{\mathbf{T}\times\mathbf{X}}$, respectively. Moreover, and importantly, for the conditionals $\phi^\pm_{\mathrm{v}}$ on $\mathbf{T} \times \mathbf{X} \times \mathbf{Y}$ given $\mathrm{v} \in \mathbf{\Lambda}$ we have

$$\left\langle (\phi^-_{\mathrm{v}}, (\mathrm{v} - \mathbf{y})\right\rangle + \left\langle \phi^+_{\mathrm{v}}, (\mathbf{y} - \mathrm{v})\right\rangle = \left\langle \phi, |\mathbf{y} - \mathrm{v}|\right\rangle. \tag{4.42}$$

Similarly, we double the "boundary" measures ϕ_0, ϕ_T, ϕ_L, ϕ_R and extend their doubled versions to $\mathbf{T} \times \mathbf{X} \times \mathbf{W}^-$ and $\mathbf{T} \times \mathbf{X} \times \mathbf{W}^+$, respectively. With this notation at hand, we can formulate the following lemma.

Lemma 4.3.10 (Kruzkhov entropies)**.** *Let ϕ^-, ϕ^+, ϕ^-_T, ϕ^+_T, ϕ^-_0, ϕ^+_0, ϕ^-_L, ϕ^+_L, ϕ^-_R, and ϕ^+_R*

be as described above. Then, (4.37) *is equivalent to*

$$\begin{aligned}
&\int \chi(\mathbf{v}) \left(\tfrac{\partial}{\partial \mathbf{t}} \psi(\mathbf{t},\mathbf{x})(\mathbf{y}-\mathbf{v}) + \tfrac{\partial}{\partial \mathbf{x}} \psi(\mathbf{t},\mathbf{x})(f(\mathbf{y})-f(\mathbf{v})) \right) \mathbf{d}\phi^+ \\
&+ \int \chi(\mathbf{v}) \left(\tfrac{\partial}{\partial \mathbf{t}} \psi(\mathbf{t},\mathbf{x})(\mathbf{v}-\mathbf{y}) + \tfrac{\partial}{\partial \mathbf{x}} \psi(\mathbf{t},\mathbf{x})(f(\mathbf{v})-f(\mathbf{y})) \right) \mathbf{d}\phi^- \\
&+ \int \chi(\mathbf{v})\psi(\mathbf{t},\mathbf{x})(\mathbf{y}-\mathbf{v})\, \mathbf{d}(\phi_0^+ - \phi_T^+) + \int \chi(\mathbf{v})\psi(\mathbf{t},\mathbf{x}))(\mathbf{v}-\mathbf{y})\, \mathbf{d}(\phi_0^- - \phi_T^-) \\
&+ \int \chi(\mathbf{v})\psi(\mathbf{t},\mathbf{x})(f(\mathbf{y})-f(\mathbf{v}))\, \mathbf{d}(\phi_L^+ - \phi_R^+) + \int \chi(\mathbf{v})\psi(\mathbf{t},\mathbf{x}))(f(\mathbf{v})-f(\mathbf{y}))\, \mathbf{d}(\phi_L^- - \phi_R^-) \geq 0
\end{aligned} \tag{4.43}$$

for all $\chi \in C(\mathbf{\Lambda})_+$.

Proof. Note that for all $\chi \in C(\mathbf{\Lambda})_+$, by (4.41) and (4.42), (4.43) can be rewritten as

$$\begin{aligned}
&\int_{\mathbf{\Lambda}} \chi(\mathbf{v}) \left[\int \left\{ \tfrac{\partial}{\partial t} \psi(\mathbf{t},\mathbf{x})|\mathbf{y}-\mathbf{v}| + \tfrac{\partial}{\partial \mathbf{x}} \psi(\mathbf{t},\mathbf{x})(\mathrm{sgn}(\mathbf{y}-\mathbf{v})(f(\mathbf{y})-f(\mathbf{v})) \right\} \mathbf{d}\phi \right. \\
&\left. + \int \{\psi(\mathbf{t},\mathbf{x})|\mathbf{y}-\mathbf{v}|\}\, \mathbf{d}(\phi_0 - \phi_T) + \int \{\psi(\mathbf{t},\mathbf{x})\mathrm{sgn}(\mathbf{y}-\mathbf{v})(f(\mathbf{y})-f(\mathbf{v})\}\, \mathbf{d}(\phi_R - \phi_L) \right] \mathbf{dv} \geq 0
\end{aligned}$$

Note that the term in brackets is a measurable function, say $\theta(\mathbf{v})$. With this notation we have $\int \chi(\mathbf{v})\theta(\mathbf{v})\, \mathbf{d}\mathrm{v} \geq 0$ for all $\chi \in C(\mathbf{\Lambda})_+$. This implies that θ is non-negative almost everywhere on $\mathbf{\Lambda}$. Using the identities (4.38) and (4.39), θ is exactly (4.37) for all Kruzkhov entropies. □

The expression (4.43) does still involve non-monomial functions and is hence not a generalized moment constraint, yet. We next show how the non-negative test functions ψ and χ can be replaced by monomial constraints. Of course we can argue, that the set of polynomials is dense in the set of test functions on $\mathbf{T} \times \mathbf{X}$ and $\mathbf{\Lambda}$ respectively. However, as we have an inequality constraint in (4.43) we cannot argue as in the proof of Lemma 4.3.9 to conclude from monomials to polynomials, because multiplying the monomial constraints by a coefficient $c_\alpha \in \mathbb{R}$, i.e., a coefficient which is potentially negative, might inverse the inequality sign. Indeed we can only allow for non-negative coefficients in order to conserve the inequality.

Lemma 4.3.11. *For* $\alpha \in \mathbb{N}^6$ *define* $h_\alpha := \mathbf{t}^{\alpha_1}(T-\mathbf{t})^{\alpha_2}(\mathbf{x}-L)^{\alpha_3}(R-\mathbf{x})^{\alpha_4}(\mathbf{v}-\underline{\mathrm{y}})^{\alpha_5}(\overline{\mathrm{y}}-\mathbf{v})^{\alpha_6}$. *Then,* (4.37) *is equivalent to*

$$\begin{aligned}
&\left\langle \phi^+, \tfrac{\partial}{\partial \mathbf{t}} h_\alpha(\mathbf{y}-\mathbf{v}) + \tfrac{\partial}{\partial \mathbf{x}} h_\alpha(f(\mathbf{y})-f(\mathbf{v})) \right\rangle \\
&+ \left\langle \phi^-, \tfrac{\partial}{\partial \mathbf{t}} h_\alpha(\mathbf{v}-\mathbf{y}) + \tfrac{\partial}{\partial \mathbf{x}} h_\alpha(f(\mathbf{v})-f(\mathbf{y})) \right\rangle \\
&+ \left\langle \phi_0^+ - \phi_T^+, h_\alpha(\mathbf{y}-\mathbf{v}) \right\rangle + \left\langle \phi_0^- - \phi_T^-, h_\alpha(\mathbf{v}-\mathbf{y}) \right\rangle \\
&+ \left\langle \phi_L^+ - \phi_R^+, h_\alpha(f(\mathbf{y})-f(\mathbf{v})) \right\rangle + \left\langle \phi_L^- - \phi_R^-, h_\alpha(f(\mathbf{v})-f(\mathbf{y})) \right\rangle \geq 0
\end{aligned} \tag{4.44}$$

for all $\alpha \in \mathbb{N}^6$.

Proof. As $\mathbf{T} \times \mathbf{X}$ and $\mathbf{\Lambda}$ are compact, by Stone-Weierstrass Theorem, we can restrict to $\psi \in \mathbb{R}[\mathbf{t},\mathbf{x}]$ and $\chi \in \mathbb{R}[\mathbf{v}]$ in (4.43). Let $\psi \in \mathbb{R}[\mathbf{t},\mathbf{x}]$ and $\chi \in \mathbb{R}[\mathbf{v}]$ be non-negative on $\mathbf{T} \times \mathbf{X}$

and $\mathbf{\Lambda}$ respectively and $\varepsilon > 0$. Then, Krivine's Positivstellensatz Theorem 2.5 implies that

$$\psi(\mathbf{t},\mathbf{x})\chi(\mathbf{v}) + \varepsilon = \sum_{\alpha\in\mathbb{N}^6} c_\alpha^\varepsilon h_\alpha, \tag{4.45}$$

for finitely many non-zero coefficients $c_\alpha^\varepsilon \in \mathbb{R}_+$. Note that by (4.45) we have $\chi\frac{\partial}{\partial \mathbf{t}}\psi = \sum_{\alpha\in\mathbb{N}^6} c_\alpha^\varepsilon \frac{\partial}{\partial \mathbf{t}} h_\alpha$ and the respective equation for $\frac{\partial}{\partial \mathbf{x}}$. Now, after some calculation, (4.43) can be written as

$$\begin{aligned}
&\sum_{\alpha\in\mathbb{N}^6} c_\alpha \Big(\Big\langle \phi^+, \tfrac{\partial}{\partial \mathbf{t}} h_\alpha(\mathbf{y}-\mathbf{v}) + \tfrac{\partial}{\partial \mathbf{x}} h_\alpha(f(\mathbf{y})-f(\mathbf{v})) \Big\rangle \\
&+ \Big\langle \phi^-, \tfrac{\partial}{\partial \mathbf{t}} h_\alpha(\mathbf{v}-\mathbf{y}) + \tfrac{\partial}{\partial \mathbf{x}} h_\alpha(f(\mathbf{v})-f(\mathbf{y})) \Big\rangle \\
&+ \Big\langle \phi_0^+ - \phi_T^+, h_\alpha(\mathbf{y}-\mathbf{v}) \Big\rangle + \Big\langle \phi_0^- - \phi_T^-, h_\alpha(\mathbf{v}-\mathbf{y}) \Big\rangle \qquad (4.46) \\
&+ \Big\langle \phi_L^+ - \phi_R^+, h_\alpha(f(\mathbf{y})-f(\mathbf{v})) \Big\rangle + \Big\langle \phi_L^- - \phi_R^-, h_\alpha(f(\mathbf{v})-f(\mathbf{y})) \Big\rangle \Big) \\
&- \varepsilon \underbrace{\Big(\Big\langle \phi_0^+ - \phi_T^+ - \phi_0^- + \phi_T^-, \mathbf{y}-\mathbf{v} \Big\rangle + \Big\langle \phi_L^+ - \phi_R^+ - \phi_L^- + \phi_R^-, f(\mathbf{y})-f(\mathbf{v}) \Big\rangle \Big)}_{\leq C} \qquad (4.47)
\end{aligned}$$

where $C \in \mathbb{R}$ is a fixed constant by definition of the involved measures. Furthermore, (4.44) and the fact that $c_\alpha \geq 0$ imply that part (4.46) of the expression above is non-negative. This yields (4.43) $\geq -\varepsilon C$, for all $\varepsilon > 0$, and implies (4.43) ≥ 0. Finally Lemma 4.3.10 concludes the proof. □

Enforcing Marginal Properties: As mentioned earlier due to compactness of $\mathbf{T} \times \mathbf{X} \times \mathbf{Y}$ we can enforce the marginal properties by moment constraints. More precisely, to ensure that the marginals with respect to $\mathbf{t}$ and $\mathbf{x}$ and $\mathbf{v}$ are the Lebesgue measure on $\mathbf{T}$, $\mathbf{X}$ and $\mathbf{\Lambda}$, it suffices to impose for all $\alpha \in \mathbb{N}^3$,

$$\Big\langle \phi^- + \phi^+, \mathbf{t}^{\alpha_1}\mathbf{x}^{\alpha_2}\mathbf{v}^{\alpha_3} \Big\rangle = \langle \lambda_{\mathbf{T}\times\mathbf{Y}\times\mathbf{\Lambda}}, \mathbf{t}^{\alpha_1}\mathbf{x}^{\alpha_2}\mathbf{v}^{\alpha_3} \rangle \tag{4.48}$$

$$\Big\langle \phi_T^- + \phi_T^+, \mathbf{t}^{\alpha_1}\mathbf{x}^{\alpha_2}\mathbf{v}^{\alpha_3} \Big\rangle = \langle \delta_T \lambda_{\mathbf{X}\times\mathbf{\Lambda}}, \mathbf{t}^{\alpha_1}\mathbf{x}^{\alpha_2}\mathbf{v}^{\alpha_3} \rangle \tag{4.49}$$

$$\Big\langle \phi_{L/R}^- + \phi_{L/R}^+, \mathbf{t}^{\alpha_1}\mathbf{x}^{\alpha_2}\mathbf{v}^{\alpha_3} \Big\rangle = \Big\langle \lambda_{\mathbf{T}} \delta_{L/R} \lambda_{\mathbf{\Lambda}}, \mathbf{t}^{\alpha_1}\mathbf{x}^{\alpha_2}\mathbf{v}^{\alpha_3} \Big\rangle \tag{4.50}$$

Generalized Moment Problem and its Moment Relaxation

Entropy mv solution as a GMP. We are finally in the position to state an entropy mv solution to (4.28) as solution to a generalized moment problem. Note that the previous sections have shown, that imposing (4.38), (4.39) (marginal and boundary conditions), (4.40) (dynamic), (4.41) (auxiliary measures), and (4.44) (Kruzkhov entropies), in our set up reduces the set of feasible measures to a single point. Hence, *optimizing* a functional over the set of feasible measures, as suggested in (1), reduces to its evaluation. When passing to the moment-SOS relaxation, and hence cutting the number of constraints, the feasible set of the relaxation is likely to have more than one single point. A suitable objective functional might therefore help to converge to a useful solution more quickly. We discuss the choice of an objective functional for the Riemann problem of the Burgers equation in Section 4.3.3. Let $\mathcal{L}$ be any expression obtained by integrating polynomials against measures $\phi^\pm, \phi_T^\pm, \phi_{L/R}^\pm$

and consider

$$\inf_{\phi^\pm,\phi_T^\pm,\phi_{L/R}^\pm} \mathcal{L} \tag{4.51a}$$

$$(4.40) \qquad \text{(conservation law)} \tag{4.51b}$$

$$(4.44) \qquad \text{(entropy condition)} \tag{4.51c}$$

$$(4.48) - (4.50) \quad \text{(marginal constraints)} \tag{4.51d}$$

$$\phi^\pm, \phi_T^\pm, \phi_{L/R}^\pm \in \mathcal{M}_+(\mathbf{T} \times \mathbf{X} \times \mathbf{Y} \times \mathbf{\Lambda}), \tag{4.51e}$$

where the constraints (4.48)-(4.50) are quantified over all $\alpha \in \mathbb{N}^3$, (4.51b) is quantified over $\alpha \in \mathbb{N}^2$ and (4.51c) over $\alpha \in \mathbb{N}^6$. The constraints (4.48)-(4.50) encode the marginal informations given in (4.38), (4.39), and (4.41). Note that the measures $\phi_0^\pm$ and at least one of the measures $\phi_L^\pm$ and $\phi_R^\pm$ must be given in order to guarantee a unique solution to (4.51). Then via (4.38) the optimal measure for (4.51) satisfies $\phi = \mu_{(\mathbf{t},\mathbf{x})}\lambda_{\mathbf{T}\times\mathbf{X}}$ for the entropy mv solution $\mu_{(\mathbf{t},\mathbf{x})}$ to (4.28). As the supports of the unknown measures in (4.51) are basic semialgebraic compact sets, we can apply the moment-SOS hierarchy to approximate $\mu_{(\mathbf{t},\mathbf{x})}$.

Convergence of the relaxations of (4.51) Observe that the mass of all measures appearing in (4.51) is bounded, because their marginals with respect to time and/or space are Lebesgue. Adding redundant constraint to the semialgebraic descriptions of the supports of the unknown measures to ensure that the associated quadratic modules are archimedean, we can establish optimal feasibility of the moment relaxations (P_d) to (4.51) for each $d \geq d_0$, where d_0 is the maximal degree appearing in the polynomial data of (4.51). Finally, the sequences $(\mathrm{z}_{\phi^-}^d + \mathrm{z}_{\phi^+}^d)_{d\in\mathbb{N}}$ converge to the moment sequence of $\lambda_{\mathbf{T}\times\mathbf{X}}\mu_{(\mathbf{t},\mathbf{x})}$, when $d \to \infty$, see Chapter 1.

Interpretation of the moment solutions

An optimal solution z^d at step d of the hierarchy of SDP-relaxations (P_d) to (4.51), consists of finite sequences of moments, one for each unknown measure in (4.51). If one is interested in statistical properties of the mv solution such as its mean or its variance, the moments provide the perfect information, at least for for sufficiently large d. However, if one is rather interested in properties of the graph of the entropy solution, a post processing step is required.

An inverse problem Recovering the graph of a function $\{(\mathrm{t},\mathrm{x},y(\mathrm{t},\mathrm{x})) : \mathrm{t} \in \mathbf{T}, \mathrm{x} \in \mathbf{X}\}$ from the moments of the measure $\phi = \lambda_{\mathbf{T}\times\mathbf{X}}\delta_{y(\mathbf{t},\mathbf{x})}$ is an inverse problem which we cannot study in detail here. The *maximum entropy* approach is to approximate y by a function from a parametrized family of functions, the approximation $\tilde{y}$ is chosen in a way, that $\int \mathbf{t}^{\alpha_1}\mathbf{x}^{\alpha_2}\tilde{y}(\mathbf{t},\mathbf{x})\,\mathbf{dt}\,\mathbf{dx} = \int \mathbf{t}^{\alpha_1}\mathbf{x}^{\alpha_2}\mathbf{y}\,\mathbf{d}\phi$ for a finite number of $\alpha \in \mathbb{N}^2$, and $\tilde{y}$ maximizes some criterion (see [Las10b, Section 12.3] and the references therein). A second approach is an approximation via a polynomial minimizing the difference to the y in L^2 norm [HLM14].

This approximation can be obtained in closed form, when choosing the polynomial bases of Legendre or Chebychev. While this method works quite well in the case when the solution y is smooth, due to the Gibbs phenomenon it behaves rather bad as soon as y has a discontinuity (compare Section 3.1.2).

In the following, we briefly outline a heuristic strategy to recover the graph from the sequence of moments of a measure $\lambda_{\mathbf{T}\times\mathbf{X}}\delta_{y(\mathbf{t},\mathbf{x})}$. It turns out that it works surprising well in both our examples.

Let d be a fixed degree and let $\mathrm{z}_\alpha := \int_{\mathbf{T}\times\mathbf{X}\times\mathbf{Y}} \mathbf{t}^{\alpha_1}\mathbf{x}^{\alpha_2}\mathbf{y}^{\alpha_3}\,\mathbf{d}\phi$ for $|\alpha| \leq d$. For any polynomial, $p \in \mathbb{R}[\mathbf{t},\mathbf{x},\mathbf{y}]_d$ we have

$$\mathrm{p}^\top M_d(\mathrm{z})\mathrm{p} = \int_{\mathbf{T}\times\mathbf{X}\times\mathbf{Y}} p^2\,\mathbf{d}\phi.$$

where p denotes the vector of coefficients of p and $M_d(\mathrm{z})$ is the moment matrix associated to z (see Chapter 1). Suppose for the moment, that $\ker(M_d(\mathrm{z})) \neq \emptyset$. Then if p is in the kernel of $M_d(\mathrm{z})$, we have that

$$\int_{\mathbf{T}\times\mathbf{X}\times\mathbf{Y}} p^2\,\mathbf{d}\phi = 0.$$

In other words, the support of ϕ is contained in the zero level set of every polynomial whose vector of coefficients is in the kernel of $L_{\mathrm{z}}(\mathrm{v}_d\mathrm{v}_d^\top)$. In practice we do not compute the kernel exactly but consider all eigenvectors $\mathrm{p}_1,\ldots,\mathrm{p}_\ell$ such that the sum of the corresponding eigenvalues is less than some threshold ε. For $(t_0,x_0) \in \mathbf{T}\times\mathbf{X}$ fixed, we then consider the optimization problem

$$\inf_{\mathrm{y}\in\mathbf{Y}} \sum_{k=1}^{\ell} p_k(t_0,x_0,\mathrm{y})^2 \tag{4.52}$$

If we consider only polynomials corresponding to eigenvalues 0 the optimal value in (4.52) is 0 and attained at $y(\mathrm{t}_0,\mathrm{x}_0) \in \mathbf{Y}$. However $y(\mathrm{t}_0,\mathrm{x}_0)$ might not be the unique minimizer. Indeed the support of ϕ, i.e., the graph of y is contained in the variety of the polynomials corresponding to the eigenvectors of the kernel of the moment matrix, but the converse inclusion does not hold in general, as the graph of a function is not necessarily the variety of some set of polynomials. However, we do not only consider the polynomials "in the kernel" but allow for a certain threshold ε when choosing the eigenvectors defining $p_1,\ldots,p_\ell$. The optimal value in (4.52) is hence slightly positive. In general, for each $(\mathrm{t}_0,\mathrm{x}_0)$ there exist finitely many minimizers y_0 of (4.52). However we might be lucky and find that (4.52) has a unique solution for each $(\mathrm{t}_0,\mathrm{x}_0) \in \mathbf{T}\times\mathbf{X}$. In this case we can conclude that $\mathrm{y}_0 = y(\mathrm{t}_0,\mathrm{x}_0)$ and recover the graph of the entropy solution from the sequence of moments of the entropy mv solution.

Note that if one is interested in properties of the graph for only one instant in time t_1 or only an interval of space $[\underline{\mathrm{x}},\overline{\mathrm{x}}]$, one can solve (4.52) only for $(t_0,x_0) \in \{t_1\}\times[\underline{\mathrm{x}},\overline{\mathrm{x}}]$. In particular, since the problem is already solved globally, there is no need to discretize time instances before or after t_1 or the space outside $[\underline{\mathrm{x}},\overline{\mathrm{x}}]$ as it would be necessary in other numerical schemes.

To be fair, we need to say, that the moment matrix $M_d(\mathrm{z})$ computed by the SDP solver is only exact up to precision of the SDP solver. A second source of numerical errors is the method to solve the optimization problem (4.52). We are doing all this by simple Matlab

functions for illustration purpose only. To obtain reliable results from this extraction procedure one definitely needs to consider all numeric parameters in detail.

4.3.3 The Riemann Problem for the Burgers Equation

For a numerical demonstration of our result, we consider the classical Riemann problem (see e.g., [Eva98]) for a scaled version of the Burgers equation. In particular, we choose the flux

$$f(\mathbf{y}) = \frac{1}{4}\mathbf{y}^2.$$

The Riemann problem to this conservation law is a Cauchy problem with the following initial condition, piecewise constant with one point of discontinuity:

$$y_0(\mathrm{x}) = \begin{cases} l & \text{if } \mathrm{x} < 0, \\ r & \text{if } \mathrm{x} > 0, \end{cases}$$

where $l, r \in \mathbb{R}$. The solution to the Riemann problem depends strongly on the values of l and r. In particular:

1. If $l > r$, the solution is unique. The shock of the initial condition spreads along the characteristics.

2. If $l < r$, the solution is not necessarily unique. The entropy condition allows to select the meaningful solution, which is known as a rarefaction wave. For the specific case of Burgers equation, it is shown in [DOW04] and [Pan94] that the single entropy $\eta = \mathbf{y}^2$ is sufficient to guarantee uniqueness of the solution. To the best of our knowledge, there is no similar result for the uniqueness of entropy mv solutions for Burgers equation with concentrated initial data, except for classical solutions [DST12].

Both cases are interesting form a numerical point of view for their own reasons. In general, the first case is difficult to address because of the discontinuity. Numerical schemes based on discretization tend to smooth out the shock. Indeed, recovering numerically the exact point of discontinuitiy is a challenge for these schemes.

We present numerical results for both cases of the Riemann problem. We are going to consider $l, r \in \{0, 1\}$. Following the discussion in Remark 4.3.7, we can assume that the solution takes values only in $\mathbf{Y} = [0, 1]$. Note that, from the initial condition, we can derive that $y(\mathrm{t}, L) = l, \quad \forall \mathrm{t} \in \mathbf{T}$. Moreover, due to positivity of y, the solution on $\mathbf{T} \times \mathbf{X}$ does not depend on the initial condition for $x > \frac{1}{2}$. Hence the value of y on the right boundary has no influence on the solution. Finally, the time-space-window on which we consider the solution is $\mathbf{T} = [0, 1]$ and $\mathbf{X} = [-\frac{1}{2}, \frac{1}{2}]$. Note that we have chosen the data in such a way, that all variables are less than 1 in absolute value and all involved measures are probability measures. This is important for numerical stability.

Remark on the significance of the numeric results upfront We need to emphasize that these experiments are by no means conclusive. Our implementation is based on the Matlab interface Gloptipoly3 [HLL09a] and the SDP solver Mosek [MOS17]. The purpose of the numerical examples is to show that our framework actually works in practice and

with a proper implementation might actually provide an alternative to schemes based on discretization to compute entropy solutions to scalar hyperbolic conservation laws.

Rarefaction wave

We consider the Riemann problem with $l = 0$ and $r = 1$. As has been noticed before, with such an initial condition, entropy conditions are crucial to select the right solution, i.e., the solution with a good physical meaning. The analytical entropy solution corresponding to this example is

$$y(\mathrm{t},\mathrm{x}) = \begin{cases} 0 & x \leq 0, \\ \frac{2x}{t} & 0 \leq x \leq \frac{t}{2}, \\ 1 & x \geq \frac{t}{2}. \end{cases} \tag{4.53}$$

Numerically implementing all entropy pairs of Kruzkhov is possible (as seen in Section 4.3.2), but heavy. As already mentioned for the entropy solution the single entropy $\eta(\mathbf{y}) = \mathbf{y}^2$ provides all necessary information to select a unique solution [DOW04]. For the entropy mv solution we are not aware of such a result. Still, instead of using the Kruzkhov pairs, we impose the following family of entropies in this example:

$$\eta_k(y) = y^k, \qquad \forall k \in \mathbb{N} \tag{4.54}$$

and the corresponding polynomial functions q_k. Note that η is strictly convex on $\mathbf{Y} = [0, 1]$. Not implementing Kruzkhov makes the implementation more robust. In particular, we do not have to divide the measures in (4.37) into two measures, since there is no absolute value appearing in (4.54). It is neither necessary to introduce another variable $\mathbf{v}$ as it is discussed in Section 4.3.2. Finally, we define the sum over all entropy constraints as objective functional and maximize this quantity. In doing so at relaxation $d = 6$, we obtain the following moments for the marginal on $\mathbf{y}$

$$(\mathrm{z}_{0,0,k})_{k=0,1,\ldots} = [1.0000, 0.3750, 0.3333, 0.3125, 0.3000, 0.2917, 0.2857, 0.2812, \ldots]$$

which coincide with the moments of the actual analytic entropy solution up to precision 10^{-5}. Applying the technique from Section 4.3.2, leads to the graph presented in Fig. 4.4 which is a quite accurate presentation.

Shock wave

In the second example we inverse the roles of l and r. As it has been noticed before, with such an initial condition the solution is discontinuous for all $\mathrm{t} > 0$. The unique analytical solution corresponding to this initial condition is

$$y^*(\mathrm{t},\mathrm{x}) = \begin{cases} 1 & \mathrm{x} > \frac{\mathrm{t}}{4}, \\ 0 & \mathrm{x} < \frac{\mathrm{t}}{4}. \end{cases} \tag{4.55}$$

As objective function, we choose the standard objective function provided by Gloptipoly, which minimizes the trace of the moment matrix. Since the trace is the convex hull of the rank on the unit ball of matrices, this is likely to cause early convergence of the semidefinite convergence because low rank solutions correspond to measures supported on points only.

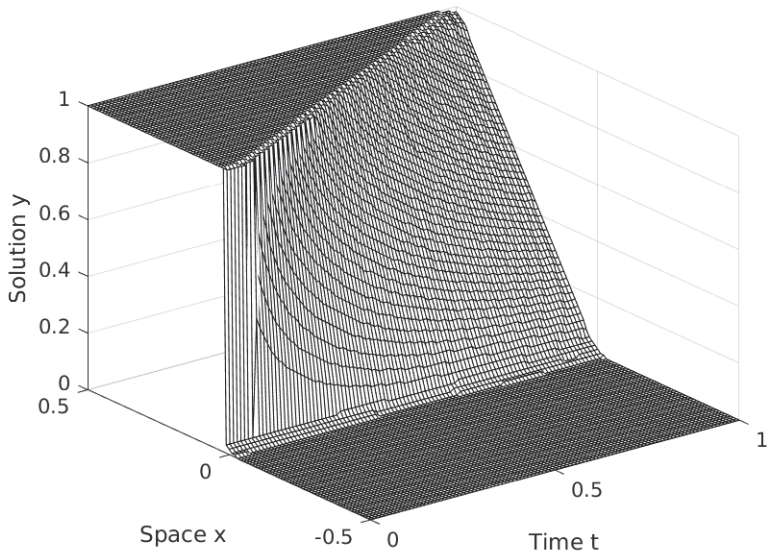

Figure 4.4: Approximation of the solution to the Riemann problem with $l = 0$ and $r = 1$.

As in this case the marginal of ϕ with respect to y is supported on $\{0, 1\}$ we expect this criterion to be appropriate to accelerate convergence. Indeed, for $d = 6$ we end up with the following moments for y

$$(\mathrm{z}_{0,0,k})_{k=0,1,\ldots} = [1.0000, 0.6250, 0.6250, 0.6250, 0.6250, \ldots]$$

which correspond up to numeric precision exactly with the moments of the analytic solution $\delta_{y^*(\mathrm{t},\mathrm{x})}\lambda_{\mathbf{T}\times\mathbf{X}}$. Applying the procedure form Section 4.3.2 leads to Fig. 4.5, where we display the difference between the analytic solution and our approximation of the graph, based on the computed moments.

Localizing the shock As already mentioned the computed moments can be used in order to approximate the location of the shock at some given time t_0. Here we will take $\mathrm{t}_0 = 0.75$. Consequently the shock is located at exactly $\mathrm{x} = 0.1875$.

We used a standard Godunov scheme (we refer to [LeV92] for more details) to compute the solution up to this time. For space discretization, we took $\Delta\mathrm{x} = 0.0005$ and an according discretization in time such that the scheme stays stable. In Table 4.3, we display the obtained values from this approach on an interval around the shock. We can see the typical behaviour that the shock is smoothed out.

We use the approach from Section 4.3.2 to approximate the solution on $\{0.75\} \times [0.1850, 0.1885]$. We expect the numerical data to be defective at this point. We therefore consider all eigen vectors of the moment matrix such that the sum of the corresponding eigenvalues is less than 10^{-8} in absolute value as belonging to its kernel. Let $p_1, \ldots, p_\ell \in \mathbb{R}[\mathbf{t}, \mathbf{x}, \mathbf{y}]$ be the polynomials in the kernel of the moment matrix. Then for each x_i from the $\mathbf{X}$ discretized in $\Delta\mathrm{x} = 0.0005$ we solve the problem $\mathrm{y}_i := \mathrm{argmin}\{\sum_{j=1}^{\ell} p_j(0.75, \mathrm{x}_i, \mathrm{y}) \,:\, \mathrm{y} \in \mathbf{Y}\}$ where $\mathbf{Y}$ is discretized in $\Delta\mathrm{y} = 0.0001$. All optimal values of these problems have been less

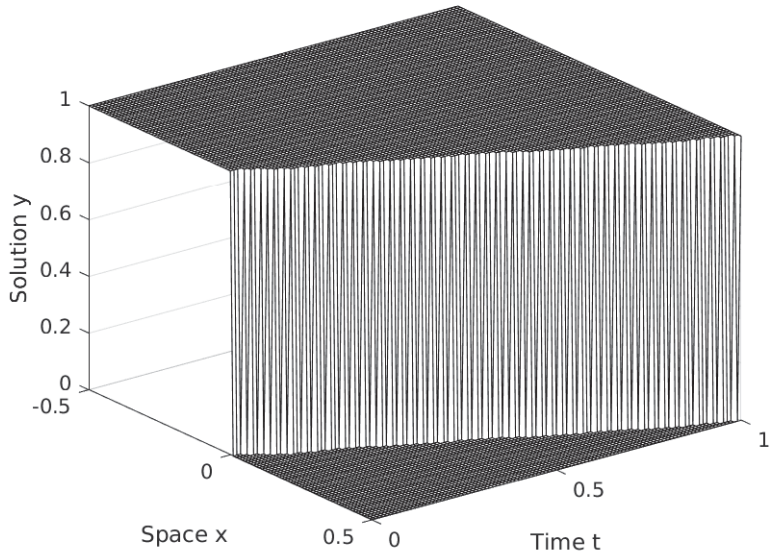

Figure 4.5: Approximation of the solution to the Riemann problem with $l = 1$ and $r = 0$.

x	0.1850	0.1855	0.1860	0.1865	0.1870	0.1875	0.1880	0.1885
Godunov	0.9999	0.9991	0.9936	0.9580	0.7647	0.2724	0.0123	0.0000
GMP	1.0000	1.0000	1.0000	1.0000	1.0000	0.0000	0.0000	0.0000

Table 4.3: Location of the shock at $t_0 = 0.75$ approximated in $\Delta x = 0.0005$ with Godunov and GMP approach.

than 10^{-9}. Hence, we consider to have found for any x_i a corresponding y_i in the support of the mv solution.

As can be seen in Table 4.3 the values obtained in this way exactly represent the position of the shock. Actually we even obtained closer bounds on the location of the shock by reducing the size of Δx when processing the already computed moments. To obtain better results with Godunov one would need to restart the simulation from the beginning.

4.3.4 Conclusion

In this section, we have discussed a new method to solve scalar polynomial hyperbolic equations based on the moment-SOS hierarchy. More precisely, we have proved that the truncated moments associated to the measure-valued solution formulation converge to the Dirac measure concentrated on entropy solution to the scalar polynomial solution. This foundational work opens many direction for future research:

- The formulation of mv solutions as solutions to an GMP leaves space for control of PDEs. A typical control in this set up would be to prescribe the measures on the boundaries $x = L, R$. Our framework opens the way for research in this direction.
- The Burgers equation is irreversible. Roughly speaking, given a terminal condition

$y_T(\mathbf{x})$ with $T > 0$, there exists a continuum of initial conditions that leads to $y(\mathbf{t}, \mathbf{x})$ (see e.g., [GZ17]). Such a continuum cannot be described by functions but it can be described with measures. Hence, our linear formulation might be useful to solve such inverse problems.

- It might also be interesting to focus on another type of equations, such as parabolic ones. One of the interest of these equations is that they regularize the solution, whatever is the initial condition. Therefore, as it is done for ODE in [Las+08a], it might be possible to define test functions depending on the solution to the parabolic equation and then define an occupation measure associated to the latter. This together with the relaxed control theory surveyed in [Fat99] might be instrumental to solve optimal control problem for non-linear parabolic equations.

Conclusion and Perspectives

In this thesis we have treated a wide spectrum of topics related to the GMP. On one side we considered alternative strategies to approximate solutions to the GMP by programs that require less computational power, in particular in the case when some sparsity structure is apparent in the original problem. The performance of such programs depends strongly on implementation details. One direction of further research would be the development of appropriate modelling software to efficiently compute and compare existing and upcoming new certificates. In particular so far there is no software package dedicated to sparse or symmetric GMPs. Another software package could be dedicated to use sparsity in order to optimize non-linear (polynomial) mixed integer problems.

On the other side we contributed to the application of the GMP in new fields, such as chance constraint approximation and measure valued solutions to scalar conservation laws. Both topics are pioneering for many further applications. The approximation of chance constraints in its version presented in this thesis is limited to small or medium size problems. A direct extensions of this work would be, e.g., the use of structured sparsity to consider real world large scale problems. Another continuation would be to approximate distributionally robust chance constraints in physical network flow problems.

Using the framework of moment-SOS relaxations to compute measure valued solutions to partial differential equations is a novelty and this thesis is amongst the first documents explaining this method for scalar non-linear hyperbolic conservation laws. Naturally, the method could be extended to non-scalar conservation laws or be used to compute controls for such problems as well as treating the inverse problem of finding initial conditions that yield a given solution at terminal time.

Solutions obtained from the moment-SOS hierarchy are sequences of moments in general. This information is often not in a format people can understand easily. In this thesis we proposed interpretations of such solutions, e.g., the extraction of a minimizer in the case of polynomial optimization for the sparse BSOS hierarchy, and the polynomials defining approximations of chance constraints. In the course of the discussion of measure valued solutions we also proposed a new approach to extract knowledge information about the support of a measure from its moments. These interpretations of solutions are of crucial interest for the development of GMP approaches as they permit to illustrate its power and utility to the engineering other mathematics community.

Bibliography

[AM14] Amir Ali Ahmadi and Anirudha Majumdar. "DSOS and SDSOS optimization: LP and SOCP-based alternatives to sum of squares optimization". In: *2014 48th Annu. Conf. Inf. Sci. Syst. CISS 2014*. 2014. ISBN: 978-1-4799-3001-2. DOI: 10.1109/CISS.2014.6814141.

[AS86] Eugene L. Allgower and Phillip H. Schmidt. "Computing Volumes of Polyhedra". In: *Math. Comput.* 46.173 (Jan. 1986), pp. 171–174. ISSN: 0025-5718. DOI: 10.2307/2008221. URL: http://dx.doi.org/10.2307/2008221.

[Bez+14] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah. "Julia: A Fresh Approach to Numerical Computing". In: *arXiv 1411.1607* (2014).

[BF08] Karl Bringmann and Tobias Friedrich. "Approximating the Volume of Unions and Intersections of High-Dimensional Geometric Objects". In: *Algorithms Comput.* Ed. by Seok-Hee Hong, Hiroshi Nagamochi, and Takuro Fukunaga. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 436–447. ISBN: 978-3-540-92182-0.

[BF87] Imre Barany and Zoltan Furedi. "Computing the volume is difficult". In: *Discret. & Comput. Geom.* 2.4 (Dec. 1987), pp. 319–326. ISSN: 1432-0444. DOI: 10.1007/BF02187886. URL: https://doi.org/10.1007/BF02187886.

[BGN09] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust Optimization*. Princeton University Press, 2009.

[Bre10] H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Springer, 2010.

[Car25] H. S. Carslaw. "A historical note on Gibbs' phenomenon in Fourier's series and integrals". In: *Bull. Amer. Math. Soc.* 31.8 (Oct. 1925), pp. 420–424. URL: https://projecteuclid.org:443/euclid.bams/1183486614.

[CC06] G. Calafiore and M. C. Campi. "The scenario approach to robust control design". In: *IEEE Trans. on Automatic Control* 51 (2006), pp. 742–753.

[CE06] G. C. Calafiore and L. El Ghaoui. "On Distributionally Robust Chance-Constrained Linear Programs". In: *J. Optim. Theory Appl.* 130.1 (2006), pp. 1–22. DOI: 10.1007/s10957-006-9084-x.

[CHK17] Mathieu Claeys, Didier Henrion, and Martin Kruzik. "Semi-definite relaxations for optimal control problems with oscillation and concentration effects". In: *ESAIM Control. Optim. Calc. Var.* 23 (2017), pp. 95–117. ISSN: 1292-8119. DOI: 10.1051/cocv/2015041. arXiv: 1412.2278. URL: http://arxiv.org/abs/1412.2278.

[Cla+14] M. Claeys, D. Arzelier, D. Henrion, and J.-B. Lasserre. "Measures and LMI for impulsive optimal control with applications to space rendezvous problems". In: *IEEE Trans. Automat. Contr.* 59.5 (2014), pp. 1374–1379. ISSN: 07431619. arXiv: arXiv:1110.3674v1.

[Cof+17] Carleton Coffrin et al. *PowerModels.jl: An Open-Source Framework for Exploring Power Flow Formulations*. 2017. eprint: arXiv:1711.01728. URL: http://arxiv.org/abs/1711.01728.

[Daf00] C. M. Dafermos. *Hyperbolic Conservation Laws in Continuum Physics*. Vol. 325. Springer, 2000.

[DBS17] E. Dall'Anese, K. Baker, and T. Summers. "Chance-Constrained AC Optimal Power Flow for Distribution Systems With Renewables". In: *IEEE Trans. Pwr. Sys.* 32.5 (Sept. 2017), pp. 3427–3438. ISSN: 0885-8950. DOI: 10.1109/TPWRS.2017.2656080.

[DF88] M Dyer and A Frieze. "On the Complexity of Computing the Volume of a Polyhedron". In: *SIAM J. Comput.* 17.5 (1988), pp. 967–974. DOI: 10.1137/0217060. URL: https://doi.org/10.1137/0217060.

[DFK91] Martin Dyer, Alan Frieze, and Ravi Kannan. "A Random Polynomial-time Algorithm for Approximating the Volume of Convex Bodies". In: *J. ACM* 38.1 (Jan. 1991), pp. 1–17. ISSN: 0004-5411. DOI: 10.1145/102782.102783. URL: http://doi.acm.org/10.1145/102782.102783.

[DiP85] R.J. DiPerna. "Measure-valued solutions to conservation laws". In: *Archive for Rational Mechanics and Analysis* 88.3 (1985), pp. 223–270.

[DL01] B. Després and F. Lagoutière. "Contact Discontinuity Capturing Schemes for Linear Advection and Compressible Gas Dynamics". In: *Journal of Scientific Computing* 16.4 (Dec. 1, 2001), pp. 479–524.

[DM87] Ronald J. DiPerna and Andrew J. Majda. "Oscillations and concentrations in weak solutions of the incompressible fluid equations". In: *Commun. Math. Phys.* 108.4 (1987), pp. 667–689. ISSN: 00103616. DOI: 10.1007/BF01214424.

[DOW04] C. De Lellis, F. Otto, and M. Westdickenberg. "Minimal entropy conditions for Burgers equation". In: *Quarterly of applied mathematics* 62.4 (2004), pp. 687–700.

[DST12] S. Demoulini, D. MA Stuart, and A. E Tzavaras. "Weak–strong uniqueness of dissipative measure-valued solutions for polyconvex elastodynamics". In: *Archive for Rational Mechanics and Analysis* 205.3 (2012), pp. 927–961.

[Dua+18] Chao Duan et al. "Distributionally Robust Chance-Constrained Approximate AC-OPF with Wasserstein Metric". In: *IEEE Trans. Power Syst.* PP (2018), p. 1.

[DY10] Erick Delage and Yinyu Ye. "Distributionally Robust Optimization Under Moment Uncertainty with Application to Data-Driven Problems". In: *Oper. Res.* 58.3 (2010), pp. 595–612. DOI: 10.1287/opre.1090.0741. URL: https://doi.org/10.1287/opre.1090.0741.

[EG92] L. C. Evans and R. F. Gariepy. *Measure Theory and Fine Properties of Functions*. Boca Raton: CRC Press, 1992.

[EI06] E Erdougan and G Iyengar. "Ambiguous chance constrained problems and robust optimization". In: *Math. Program.* 107.1 (June 2006), pp. 37–61. ISSN: 1436-4646. DOI: 10.1007/s10107-005-0678-0. URL: https://doi.org/10.1007/s10107-005-0678-0.

[Eng89] Ryszard Engelking. *General Topology*. Berlin: Heldermann, 1989.

[Eva98] L. C. Evans. *Partial Differential Equations*. Providence: American Mathemtatical Society, 1998.

[Fat99] H. O. Fattorini. *Infinite Dimensional Optimization and Control Theory*. Cambridge: Cambridge University Press, 1999.

[Fjo+17] U.S. Fjordholm, R. Käppeli, S. Mishra, and E. Tadmor. "Construction of approximate entropy measure-valued solutions for hyperbolic systems of conservation laws". In: *Foundations of Computational Mathematics* 17.3 (2017), pp. 763–827.

[FLM18] E. Feireisl, M. Lukacova-Medvidova, and H. Mizerova. "Convergence of finite volume schemes for the Euler equations via dissipative measure–valued solutions". 2018. URL: arXiv:1803.08401.

[God59] S. K. Godunov. "A difference method for numerical calculation of discontinuous solutions of the equations of hydrodynamics". In: *Matematicheskii Sbornik* 89.3 (1959), pp. 271–306.

[GP04] Karin Gatermann and Pablo A Parrilo. "Symmetry groups , semidefinite programs , and sums of squares". In: *J. Pure Appl. Algebr.* 192 (2004), pp. 95–128. DOI: 10.1016/j.jpaa.2003.12.011.

[GS94] J. Grainger and W. Stevenson. *Power System Analysis*. McGraw-Hill, 1994.

[GZ17] L. Gosse and E. Zuazua. "Filtered gradient algorithms for inverse design problems of one-dimensional Burgers equation". In: *Innovative algorithms and analysis*. Springer, 2017, pp. 197–227.

[Han88] David Handelman. "Representing polynomials by positive linear functions on compact convex polyhedra". In: *Pacific J. Math.* 132.1 (1988), pp. 35–62.

[HB07] John H. Hubbard and Barbara Burke Hubbard. *Vector Calculus, Linear Algebra, and Differential Forms: A Unified Approach*. 3rd ed. Ithaca, NY: Matrix Editions, 2007, p. 802. ISBN: 9780971576636.

[Hil88] David Hilbert. "Ueber die Darstellung definiter Formen als Summe von Formenquadraten". In: *Math. Ann.* 32 (1888), pp. 342–350.

[HKW18] Didier Henrion, Martin Kruzik, and Tillmann Weisser. "Optimal control problems with oscillations, concentrations and discontinuities". In: *preprint* (2018).

[HL99] Onesimo; Hernandez-Lerma and Jean Bernard Lasserre. *Further Topics on Discrete-Time Markov Control Processes*. Ed. by I. Karatzas and M. Yor. Springer, 1999, pp. viii, 277.

[HLL09a] D. Henrion, J.-B. Lasserre, and J. Löfberg. "GloptiPoly 3: Moments, Optimization and Semidefinite Programming". In: *Optim. Methods Software* 24 (2009), pp. 761–779.

[HLL09b] Didier Henrion, Jean Bernard Lasserre, and Johan Löfberg. "GloptiPoly 3: moments, optimization and semidefinite programming". In: *Optim. Methods Softw.* 24.4-5 (2009), pp. 761–779. ISSN: 1055-6788. DOI: 10.1080/10556780802699201. arXiv: 0709.2559. URL: http://www.tandfonline.com/doi/abs/10.1080/10556780802699201?journalCode=goms20#.VQLUd-HMIug.

[HLM14] Didier Henrion, Jean Bernard Lasserre, and Martin Mevissen. "Mean squared error minimization for inverse moment problems". In: *Appl. Math. Optim.* 70.1 (2014), pp. 83–110. ISSN: 14320606. DOI: 10.1007/s00245-013-9235-z. arXiv: 1208.6398.

[HLS09] D. Henrion, J. B. Lasserre, and C. Savorgnan. "Approximate Volume and Integration for Basic Semialgebraic Sets". In: *SIAM Rev.* 51.4 (2009), pp. 722–743. ISSN: 0036-1445. DOI: 10.1137/080730287. arXiv: arXiv:0807.2505v2. URL: http://epubs.siam.org/doi/abs/10.1137/080730287.

[HN10] J. William Helton and Jiawang Nie. "Semidefinite representation of convex sets". In: *Mathematical Programming* 122.1 (2010), pp. 21–64. DOI: 10.1007/s10107-008-0240-y. URL: https://doi.org/10.1007/s10107-008-0240-y.

[JM18] Cedric Josz and Daniel K. Molzahn. *Multi-ordered Lasserre hierarchy for large scale polynomial optimization in real and complex variables*. Version 2. 2018. URL: https://arxiv.org/abs/1709.04376v2.

[KKK17] A. Kalamajska, S. Kroemer, and M. Kruzik. "Weak lower semicontinuity by means of anisotropic parametrized measures". In: *Proc. STAMM 2016*. Preprint arXiv:1704.00368: Springer, 2017.

[KP02] Etienne de Klerk and D. V. Pasechnik. "Approximation of the stability number of a graph". In: *SIAM J. Optim.* 12.4 (2002), pp. 875–892.

[KP07] Salma Kuhlmann and Mihai Putinar. "Positive Polynomials on Fibre Products". In: *Comptes Rendus l'Academie des Sci.* 1344.681–684 (2007).

[KP09] Salma Kuhlmann and Mihai Putinar. "Positive Polynomials on Projective Limits on Real Algebraic Varieties". In: *Bull. des Sci. Math.* 133 (2009), pp. 92–111.

[KR96] Martin Kruzik and Tomas Roubicek. "Explicit Characterization of L^p -Young Measures". In: *J. Math. Anal. Appl.* 198 (1996), pp. 830–843.

[Kri64] J.L. Krivine. "Anneaux préordonnés". In: *J. d'analyse mathématique* 12 (1964), pp. 307–326.

[Kru70] S.N. Kruzkov. "First order quasilinear equations in several independent variables". In: *Mathematics of the USSR-Sbornik* 10.2 (1970), p. 217.

[Lan87] Henry J. Landau. "Moments in Mathematics". In: *Proc. Symp. Appl. Math.* 1987.

[Las+08a] J.-B Lasserre, D. Henrion, C. Prieur, and E. Trélat. "Nonlinear optimal control via occupation measures and LMI-relaxations". In: *SIAM Journal on Control and Optimization* 47.4 (2008), pp. 1643–1666.

[Las+08b] Jean Bernard Lasserre, Didier Henrion, Christophe Prieur, and Emmanuel Trélat. "Nonlinear optimal control via occupation measures and LMI-relaxations". In: *SIAM J. Control Optim.* 47.4 (2008), pp. 1643–1666. ISSN: 03630129. DOI: 10.1137/070685051. arXiv: 0703377 [math]. URL: http://arxiv.org/abs/math/0703377.

[Las01] Jean Bernard Lasserre. "Global Optimization with Polynomials and the Problem of Moments". In: *SIAM J. Optim.* 11.3 (2001), pp. 796–817. ISSN: 1052-6234. DOI: 10.1137/S1052623400366802.

[Las06] Jean Bernard Lasserre. "Convergent SDP-relaxations in polynomial optimization with sparsity". In: *SIAM J. OPTIM.* 17.3 (2006), pp. 822–843. ISSN: 1052-6234. DOI: 10.1137/05064504X.

[Las08] Jean Bernard Lasserre. "Representation of nonnegative convex polynomials". In: *Arch. der Math.* 91 (2008), pp. 126–130. DOI: 10.1007/s00013-008-2687-8.

[Las10a] Jean Bernard Lasserre. "A Joint+Marginal Approach to Parametric Polynomial Optimization". In: *SIAM J. Optim.* 20.4 (2010), pp. 1995–2022. ISSN: 0022-3999. DOI: 10.1137/090750688. arXiv: arXiv:1302.5877.

[Las10b] Jean Bernard Lasserre. *Moments, Positive Polynomials and their Applications*. Imperial College Press, 2010.

[Las13] Jean Bernard Lasserre. "The K-moment problem for continuous linear functionals". In: *Trans. Am. Math. Soc.* 365 (2013), pp. 2489–2504.

[Las15] Jean Bernard Lasserre. *An Introduction to Polynomial and Semi-Algebraic Optimization*. Cambridge Texts in Applied Mathematics. Cambridge University Press, 2015. DOI: 10.1017/CBO9781107447226.

[Las17a] Jean Bernard Lasserre. "Computing gaussian and exponential measures of semi-algebraic sets". In: *Adv. Appl. Math.* 91 (2017), pp. 137–163.

[Las17b] Jean Bernard Lasserre. "Representation of Chance-Constraints With Strong Asymptotic Guarantees". In: *IEEE Control Syst. Lett.* 1.1 (2017), pp. 50–55.

[Lau09] Monique Laurent. "Sums of Squares, Moment Matrices and Optimization Over Polynomials". In: *Emerg. Appl. Algebr. Geom.* Ed. by Mihai Putinar and Seth Sullivant. New York, NY: Springer New York, 2009, pp. 157–270. ISBN: 978-0-387-09686-5. DOI: 10.1007/978-0-387-09686-5_7. URL: https://doi.org/10.1007/978-0-387-09686-5_7.

[Lax71] P. Lax. "Shock waves and entropy". In: *Contributions to nonlinear functional analysis*. Elsevier, 1971, pp. 603–634.

[LD15] M. Lubin and I. Dunning. "Computing in Operations Research Using Julia". In: *INFORMS J. on Computing* 27.2 (2015), pp. 238–248.

[LeV92] Randall J. LeVeque. "Numerical methods for conservation laws". In: *Lectures in Mathematics ETH Zürich* (1992).

[Löf04] Johan Löfberg. "YALMIP : a toolbox for modeling and optimization in MATLAB". In: *2004 IEEE Int. Conf. Comput. Aided Control Syst. Des.* (2004), pp. 284–289. ISSN: 03014215. DOI: 10.1109/CACSD.2004.1393890.

[LS17] A. Lorca and X. A. Sun. "The Adaptive Robust Multi-Period Alternating Current Optimal Power Flow Problem". In: to appear in *IEEE Trans. Power Syst.* (2017).

[LTS17] Jean Bernard Lasserre, Kim Chuan Toh, and Yang Shouguang. "A bounded degree SOS hierarchy for polynomial optimization". In: *EURO J. Comput. Optim.* 5.1–2 (2017), pp. 87–17. ISSN: 2192-4406. DOI: 10.1007/s13675-015-0050-y. arXiv: 1501.06126. URL: http://arxiv.org/abs/1501.06126.

[Lue69] David G. Luenberger. *Optimization by vector space methods*. New York: Wiley, 1969.

[LW18] Jean Bernard Lasserre and Tillmann Weisser. "Representation of distributionally robust chance-constraints". In: *preprint* (2018). URL: https://arxiv.org/abs/1803.11500.

[Mag+15] Victor Magron, Xavier Allamigeon, Stéphane Gaubert, and Benjamin Werner. "Formal proofs for Nonlinear Optimization". In: *Journal of Formalized Reasoning* 8.1 (2015), pp. 1–24.

[Mál+96] J. Málek, J. Necas, M. Rokyta, and M. Ruzicka. *Weak and measure-valued solutions to evolutionary PDEs*. Vol. 13. CRC Press, 1996.

[Mar+18] Swann Marx, Tillmann Weisser, Jean Bernard Lasserre, and Didier Henrion. "A moment approach to solving Burgers equation". 2018.

[MH15] D. K. Molzahn and I. A. Hiskens. "Sparsity-Exploiting Moment-Based Relaxations of the Optimal Power Flow Problem". In: *IEEE Transactions on Power Systems* 30.6 (2015), pp. 3168–3180. ISSN: 0885-8950. DOI: 10.1109/TPWRS.2014.2372478.

[MHL15] V. Magron, D. Henrion, and J.-B. Lasserre. "Semidefinite approximations of projections and polynomial images of semialgebraic sets". In: *SIAM Journal on Optimization* 25-4 (2015), pp. 2143–2164.

[MOS17] MOSEK ApS. *Mosek Matab Toolbox, Release 8.0.0.59*. 2017.

[Nas+16] A. Nasri, S. J. Kazempour, A. J. Conejo, and M. Ghandhari. "Network-Constrained AC Unit Commitment Under Uncertainty: A Benders' Decomposition Approach". In: *IEEE Trans. Power Syst.* 31.1 (Jan. 2016), pp. 412–422.

[Pan94] E. Y. Panov. "Uniqueness of the solution of the Cauchy problem for a first order quasilinear equation with one admissible strictly convex entropy". In: *Mathematical Notes* 55.5 (1994), pp. 517–525.

[Put93] Mihai Putinar. "Positive polynomials on compact semi-algebraic sets". In: *Indiana Univ. Math. J.* 42.3 (1993), pp. 969–984. ISSN: 07644442. DOI: 10.1016/S0764-4442(99)80251-1.

[RA17] L. A. Roald and G. Andersson. "Chance-Constrained AC Optimal Power Flow: Reformulations and Efficient Algorithms". In: to appear in *IEEE Trans. Power Syst.* (2017).

[Rou97] Tomas Roubicek. *Relaxation in Optimization Theory and Variational Calculus*. Berlin: Gruyter, 1997.

[Sch91] Konrad Schmüdgen. "The K-moment problem for compact semi-algebraic sets." In: *Mathematische Annalen* 289.2 (1991), pp. 203–206. URL: http://eudml.org/doc/164777.

[Sho98] Naum Z. Shor. *Nondifferentiable Optimization and Polynomial Problems*. Vol. 24. Dordrecht: Kluver Academic Publishers, 1998, pp. XVII, 396. DOI: 10.1007/978-1-4757-6015-6.

[Ste74] Gilbert Stengle. "A nullstellensatz and a positivstellensatz in semialgebraic geometry". In: *Math. Ann.* 207.2 (1974), pp. 87–97. ISSN: 00255831. DOI: 10.1007/BF01362149.

[Stu99] Jos F. Sturm. "Using SeDuMi 1.02, A Matlab toolbox for optimization over symmetric cones". In: *Optim. Methods Softw.* 11.1-4 (1999), pp. 625–653. ISSN: 1055-6788. DOI: 10.1080/10556789908805766. URL: http://www.tandfonline.com/doi/abs/10.1080/10556789908805766.

[Tac+18] Matteo Tacchi, Tillmann Weisser, Jean Bernard Lasserre, and Didier Henrion. "Exploiting sparsity in volume computation". In: *preprint* (2018).

[Trn05] Maria Trnovska. "Strong Duality Conditions in Semidefinite Programming". In: *J. Electr. Eng.* 56.12 (2005), pp. 2–4.

[TTT12] Kim-Chuan Toh, Michael J Todd, and Reha H Tütüncü. "On the Implementation and Usage of SDPT3 – A Matlab Software Package for Semidefinite-Quadratic-Linear Programming, Version 4.0". In: *Handb. Semidefinite, Conic Polynomial Optim.* Ed. by Miguel F Anjos and Jean B Lasserre. Boston, MA: Springer US, 2012, pp. 715–754. ISBN: 978-1-4614-0769-0. DOI: 10.1007/978-1-4614-0769-0_25. URL: https://doi.org/10.1007/978-1-4614-0769-0_25.

[VA14] Lieven Vandenberghe and Martin S. Andersen. "Chordal Graphs and Semidefinite Optimization". In: *Found. Trends Optim.* 1.4 (2014), pp. 241–433. ISSN: 2167-3888. DOI: 10.1561/2400000006. URL: http://www.nowpublishers.com/article/Details/OPT-006.

[Vas03] F.-H. Vasilescu. "Spectral measures and moment problems". In: *Spectr. Theory Its Appl.* Theta 2003 (2003), pp. 173–215. URL: http://math.univ-lille1.fr/$%5Csim$fhvasil/articles/SMMP.pdf.

[Vra+13] M. Vrakopoulou et al. "Probabilistic security-constrained AC optimal power flow". In: *PowerTech*. Grenoble, France, June 2013.

[Wak+06] Hayato Waki, Sunyoung Kim, Masakazu Kojima, and Masakazu Muramatsu. "Sums of squares and semidefinite program relaxations for polynomial optimization problems with structured sparsity". In: *SIAM J. Optim.* 17.1 (2006), pp. 218–242.

[WB06] A. Wächter and L. T. Biegler. "On the Implementation of a Primal-Dual Interior Point Filter Line Search Algorithm for Large-Scale Nonlinear Programming". In: *Mathematical Programming* 1.106 (2006), pp. 25–57.

[Wei17] Tillmann Weisser. *Sparse BSOS Implementation*. 2017. URL: https://github.com/tweisser/Sparse_BSOS.

[Whi11] G. B. Whitham. *Linear and nonlinear waves*. Vol. 42. John Wiley & Sons, 2011.

[WLT17] Tillmann Weisser, Jean Bernard Lasserre, and Kim-Chuan Toh. "Sparse-BSOS: a bounded degree SOS hierarchy for large scale polynomial optimization with sparsity". In: *Math. Program. Comput.* (May 2017). ISSN: 1867-2957. DOI: 10.1007/s12532-017-0121-6. URL: https://doi.org/10.1007/s12532-017-0121-6.

[WRM18] Tillmann Weisser, Line A Roald, and Sidhant Misra. "Chance-Constrained Optimizaton for Non-Linear Network Flow Problems". In: *submitted* (2018). URL: https://arxiv.org/abs/1803.02696.

[XA18] W Xie and S Ahmed. "Distributionally Robust Chance Constrained Optimal Power Flow with Renewables: A Conic Reformulation". In: *IEEE Trans. Power Syst.* 33.2 (Mar. 2018), pp. 1860–1867. ISSN: 0885-8950. DOI: 10.1109/TPWRS.2017.2725581.

[You69] L. C. Young. *Lectures on the calculus of variations and optimal control theory.* Philadelphia: W. B. Saunders Co., 1969.

[YX16] Wenzhuo Yang and Huan Xu. "Distributionally robust chance constraints for non-linear uncertainties". In: *Math. Program.* 155.1 (Jan. 2016), pp. 231–265. ISSN: 1436-4646. DOI: 10.1007/s10107-014-0842-5. URL: https://doi.org/10.1007/s10107-014-0842-5.

[ZMT11] R. D. Zimmermann, C. E. Murillo-Sanchez, and R. J. Thomas. "MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education". In: *IEEE Trans. Power Systems* 23.1 (2011), pp. 12–19.

[ZSM17] Y Zhang, S Shen, and J L Mathieu. "Distributionally Robust Chance-Constrained Optimal Power Flow With Uncertain Renewables and Uncertain Reserves Provided by Loads". In: *IEEE Trans. Power Syst.* 32.2 (Mar. 2017), pp. 1378–1388. ISSN: 0885-8950. DOI: 10.1109/TPWRS.2016.2572104.