Diss. ETH No. 21260

Applications of algebraic geometry in economics

A dissertation submitted to

ETH ZURICH

for the degree of

Doctor of Sciences

presented by

PHILIPP JOHANNES RENNER

Diplom Mathematiker, Technische Universität Kaiserslautern
born 9 August 1984
citizen of Germany

Accepted on the recommendation of

Prof. Dr. Robert Weismantel,	examiner
Prof. Dr. Gerhard Pfister,	co-examiner
Prof. Dr. Karl Schmedders,	co-examiner

2013

1 Acknowledgment

First and foremost I want to thank my advisor Karl Schmedders. He has helped and supported me for the past few years. Thanks to you I had the pleasure of starting to learn economics, while still being able to use my prior knowlege of mathematics. Professor Schmedders it was an educational and rewarding experience to work with you.

I also especially want to thank Robert Weismantel for helping to resolve the administrative problem we had by taking on the task of being my advisor at the ETH and all the extra work this entails.

I also am grateful to Gerhard Pfister for referring me to Karl in the first place and also for being on my evaluation committee.

My gratitude goes also to Ken Judd, who coauthored my first paper and advised me on several occasions on the ins and outs of economics. I am looking forward to my visit in Stanford and to our future collaboration. Also thanks to Diethard Klatte for helping me to get started on optimization and always patiently answering all my question about parametric optimization. I am also indebted to Eleftherios Couzoudis, who introduced me to the generalized Nash equilibria and coauthored one of the papers in this thesis.

Lastly I want to thank in no particular order Felix Kubler, Cordian Riener, Jonathan Hauenstein and Anna-Laura Wickström.

Contents

1 Acknowledgment 1

2 Abstract 5

3 Zusammenfassung 7

4 Finding all pure-strategy equilibria in games with continuous strategies 9

4.1 Introduction 9
4.2 Motivating example: Duopoly game with two equilibria 12
4.2.1 Bertrand price game 12
4.2.2 Polynomial equilibrium equations 13
4.2.3 Solution 14
4.3 All-solution homotopy methods 15
4.3.1 Mathematical background 15
4.3.2 Building intuition from the univariate case 18
4.3.3 The multivariate case 22
4.3.4 Advanced features 24
4.4 Implementation 25
4.4.1 Bertini 25
4.4.2 Alternatives 25
4.4.3 Parallelization 25
4.5 Bertrand price game continued 26
4.5.1 Solving the Bertrand price game with Bertini 26
4.5.2 Application of parameter continuation 28
4.5.3 The manifold of real positive solutions 29
4.6 Equilibrium equations for dynamic stochastic games 31
4.6.1 Dynamic stochastic games: general formulation 32
4.6.2 Equilibrium conditions 33
4.7 Learning curve 34
4.7.1 A learning-by-doing model 34
4.7.2 Solving the equilibrium equations with Bertini 35
4.8 Cost-reducing investment and depreciation 37
4.8.1 A cost-reducing investment model 37
4.8.2 Solving the equilibrium equations with Bertini 38
4.9 Conclusion 40
4.9.1 Summary 40

4.9.2 Current limitations and future work . . . 40
4.10 Appendix . . . 41
4.10.1 Homogenization . . . 41
4.10.2 m-homogeneous Bezout number . . . 45
4.10.3 Parameter continuation homotopy . . . 45
4.10.4 A splitting approach for solving larger systems . . . 47

5 A polynomial optimization approach to principal agent problems 49
5.1 Introduction . . . 49
5.2 The Principal-Agent Model . . . 52
5.2.1 The Principal-Agent Problem . . . 52
5.2.2 The First-Order Approach . . . 53
5.2.3 Existence of a Global Optimal Solution . . . 54
5.3 The Polynomial Optimization Approach for $\boldsymbol{A} \subset \mathbb{R}$ . . . 55
5.4 Derivation of the Polynomial Optimization Approach . . . 59
5.4.1 Mathematical Framework . . . 59
5.4.2 Proof of Theorem 5.1 . . . 65
5.4.3 Discussion of the Polynomial Approach's Assumptions and Limitations . . . 66
5.5 The Polynomial Optimization Approach for $\boldsymbol{A} \subset \mathbb{R}^{\boldsymbol{L}}$ . . . 68
5.5.1 Optimization of Multivariate Polynomials . . . 69
5.5.2 The Multivariate Polynomial Optimization Approach . . . 72
5.5.3 A Multivariate Example . . . 76
5.6 Conclusion . . . 78

6 Computing Generalized Nash Equilibria by Polynomial Programming 79
6.1 Introduction . . . 79
6.2 The Model . . . 81
6.3 Sum of Squares Optimization . . . 84
6.3.1 Basic Definitions and Theorems . . . 84
6.3.2 Optimization of Polynomials over Semialgebraic Sets . . . 85
6.4 Reformulating the GNEP . . . 87
6.5 Computational Results . . . 89

2 Abstract

The overall aim of this thesis was to apply techniques from algebraic geometry to problems in economics. Algebraic geometry has found many applications in various areas of mathematics and in several other fields. We have encountered three major approaches to employing these tools to economics. First, there are the symbolic methods from computer algebra (Greuel and Pfister, 2002). One possible avenue of approach here is Gröbner bases, which have already been used to great effect in integer programming (Loera et al., 2006) and also in economics (Kubler and Schmedders, 2010). Second, there is the numerical algebraic geometry route. There one uses Berstein's or Bezout's theorem, which give information on the isolated solutions of a square system of polynomial equations (Sommese and Wampler, 2005). The basic idea is to construct a homotopy and trace the paths leading to those isolated solutions. It is a very active field of research and applications range from optimal control (Rostalski et al., 2011) to biology (Hao et al., 2011). Lastly there is the real algebraic geometry route. It was recently discovered (Parrilo, 2000; Lasserre, 2001b) that representation results for positive polynomials can be used to relax polynomial optimization problems into convex optimization problems. Since then it has been shown that this is a promising approach to solving various problems, for instance in combinatorial optimization (Lasserre, 2001a) and also game theory (Laraki and Lasserre, 2012).

Over recent years I have looked at the last two of these approaches. The results have been presented in the form of several papers, two of which have already been published and the last of which is being revised at the time of writing.

The first paper is entitled "Finding all pure-strategy equilibria in games with continuous strategies" (Judd et al., 2012). Static and dynamic games are widely used tools for policy experiments and estimation studies. The multiple Nash equilibria in such models can potentially invalidate the results thus obtained. This problem of multiplicity has been well known for decades and there are several easy models in which it occurs (Fudenberg and Tirole, 1983a). However, it has been largely ignored in most publications thus far. In this paper we want to illustrate how to address this problem by means of the all solutions homotopy optimization approach (Sommese and Wampler, 2005). To apply this approach we require our problem to be polynomial with isolated optimal solutions. We then reformulate the problem by using the Karush-Kuhn-Tucker conditions to obtain a square system of polynomial equations. The basic idea of the homotopy approach is to use an easier version of the model, where all solutions are known. This easy system is then transformed via a function called homotopy to KKT conditions. The resulting paths are traced by numerical methods. The same ideas can be used in all

situations in which a version of the implicit function theorem holds. But in general this approach cannot compute all solutions. However, in the polynomial case, if we perturb the homotopy path randomly and choose an appropriate starting system, then we can reach all isolated solutions. We use the software package Bertini (Bates et al., 2005), which implements the homotopy solution approach, to solve a Bertrand price game and a stochastic dynamic model of cost-reducing investment.

My contribution to this paper was to describe the mathematics behind this approach and also to compute the various examples.

The second paper is entitled “A polynomial optimization approach to principal agent problems” (Renner and Schmedders, 2013). In it we deal with a canonical model in economics, the principal agent problem. The principal hires an agent to, for instance, manage a company. She knows the agent’s preferences but cannot observe the agent’s actions in the subsequent period. So, to maximize her own utility, she has to set the right incentives for the agent. This leads to a bi-level optimization problem in which both players optimize their expected utilities. Unless we impose restrictive assumptions on the functions used, this leads in general to a non-convex lower-level problem. Thus usual methods from bi-level optimization do not apply. We assume that the lower-level problem is polynomial with a compact feasible set. Then we use ideas developed in Lasserre, 2001b; Parrilo, 2000 to, in some cases, reformulate, and in others relax the lower level into a convex optimization problem. We solve the resulting nonlinear program with a numerical optimization routine.

My part in this work was the idea of using the Positivstellensätze to replace the lower level problem. Thus I also wrote the mathematical part of this paper and again computed the examples.

The third and final paper is entitled “Computing Generalized Nash Equilibria by Polynomial Programming” (Couzoudis and Renner, 2013). The Generalized Nash equilibrium is a solution concept that extends the classical Nash equilibrium to situations in which the opponents decision influences the player’s constraints. To compute these equilibria the literature usually assumes convexity or quasi-convexity of the player’s problems. However, in many situations it is desirable to use non-convex objective functions. We adapted the method developed in the previous paper to be able to solve this problem for the non-convex case. Our assumptions are that the functions are polynomials with compact feasible sets. We then again use real algebraic geometry to relax these problems into convex optimization problems which then can be solved using standard methods. As an example we compute a model of the New Zealand electricity spot market using a real data set.

Again I proved the relevant theorems, wrote the overview for the relaxation methods, and computed the example.

3 Zusammenfassung

Das Ziel dieser Arbeit war es die Techniken der algebraischen Geometrie und Computer Algebra auf Problem aus der Ökonomie anzuwenden. Algebraische Geometrie hat mittlerweile viele Anwendungen in verschieden Gebieten der Mathematik und anderen Forschungsrichtungen gefunden. Uns sind drei mögliche Ansätze begegnet, um dieses Ziel zu erreichen. Der erste Ansatz bedient sich der symbolischen Methoden, welche von der Computer Algebra kommen (Greuel und Pfister, 2002). Ein wichtiges Werkzeug dort sind die Gröbner Basen. Diese wurden bereits sowohl in ganzzahliger Optimierung (Loera u. a., 2006) und in den Wirtschaftswissenschaften benutzt (Kubler und Schmedders, 2010). Als weiter Möglichkeit gibt es die Methoden der numerischen algebraischen Geometrie. Mit Hilfe von den Sätzen von Bezout und Bernstein können alle isolierten Nullstellen von polynomialen Gleichungssystemen berechnet werden. Die grundlegende Idee ist eine Homotopie von einem einfachen System zu dem Ursprünglichen zu konstruieren (Sommese und Wampler, 2005). Dann folgt man mit numerischen Methoden den resultierenden Pfaden. Diese Methoden haben bereits viele Anwendungen, zum Beispiel in optimaler Steuerung (Rostalski u. a., 2011) und Biologie (Hao u. a., 2011), gefunden. Eine dritte Möglichkeit ist die reelle algebraische Geometrie (Parrilo, 2000; Lasserre, 2001b). Dabei werden die Repräsentationssätze für positive Polynome benutzt um ein polynomiales Optimierungsproblem zu einem konvexen Programm zu relaxieren. Dieser Ansatz hat sich als sehr vielversprechend erwiesen und hat bereits Anwendungen in zum Beispiel kombinatorischer Optimierung (Lasserre, 2001a) und Spieltheorie (Laraki und Lasserre, 2012) gefunden.

Meine Arbeit der letzten Jahre hat zu drei Artikel geführt, welche sich der letzten beiden Ansätzen bedienen. Zwei Papiere sind bereits veröffentlicht und das Dritte ist im Moment im Begutachtungsprozess.

Der erste Artikel ist "Finding all pure-strategy equilibria in games with continuous strategies" (Judd u. a., 2012). Statische und dynamische Spiele sind weit verbreitete Modelle für Strategie Experimente und Planspiele. Mehrere Nash Gleichgewichte in solchen Situationen können potentiell die Resultate verfälschen und sogar unbrauchbar machen. Diese Problematik ist seit Jahrzehnten bekannt und selbst in einfachen Modellen kann sie vorkommen (Fudenberg und Tirole, 1983a). Trotz den signifikanten Folgen wurde dies in der Literatur weitgehend ignoriert. In diesem Artikel wollen wir zeigen, wie, in gewissen Situationen, mehrfache Gleichgewichte gefunden werden können. Für das Optimierungsproblem setzen wir voraus, dass die Lösungen isoliert sind und die Funktionen Polynome. Die Karush-Kuhn-Tucker Bedingungen liefern dann ein quadratisches System von polynomialen Gleichungen. Dieses kann dann mit der Software Bertini (Bates

u. a., 2005) gelöst werden. Wir betrachten Bertrand-Wettbewerb und ein stochastisches dynamisches Modell mit Kosten reduzierendem Investment.

Mein Beitrag zu diesem Papier war die Beschreibung der zugrunde liegenden Mathematik und die Berechnung der Beispiele.

Das zweite Papier heisst "A polynomial optimization approach to principal agent problems" (Renner und Schmedders, 2013). Das Prinzipal-Agenten Modell ist eines der kanonischen Modelle der Wirtschaftswissenschaften. Der Prinzipal schliesst einen Vertrag mit einem Agenten ab, zum Beispiel ein Eigentümer stellt einen Manager ein. Das Spezielle an diesem Problem ist, dass der Prinzipal nicht die Aktion des Agenten in der folgenden Periode beobachten kann. Er kennt lediglich die Nutzenfunktion des Agenten, die möglichen Resultate und deren Wahrscheinlichkeiten. Dies führt zu einem zwei Ebenen Problem, wobei die optimale Aktion des Agenten teil der Restriktionen des Prinzipal sind. Beide Spieler optimieren hierbei ihren erwarteten Nutzen. Ausser unter starken Restriktionen, führt dies im Allgemeinen zu einer nicht konvexen unteren Ebene. Standardmethoden der Bilevel Optimierung greifen hier nicht mehr. Wir nehmen an, dass die untere Ebene eine Polynomiales Optimierungsproblem ist mit kompakter zulässiger Menge. Dann verwenden wir Ideen aus Lasserre, 2001b; Parrilo, 2000, um die untere Ebene im eindimensionalen Fall zu reformulieren und im Mehrdimensionalen zu relaxieren. Das resultierende nicht lineare Optimierungsproblem lösen wir mit numerischer Optimierungssoftware.

Mein Anteil war die Idee die Positivstellensätze zur Umformulierung der unteren Ebene zu benutzen. Somit habe ich auch den mathematischen Teil geschrieben und auch die Beispiele berechnet.

Der letzte Artikel hat den Titel "Computing Generalized Nash Equilibria by Polynomial Programming" (Couzoudis und Renner, 2013). Verallgemeinerte Nash Gleichgewichte sind ein Lösungskonzept, welches das klassische Konzept von Nash erweitert, indem die Entscheidung der Gegenspieler sich auch auf die eigenen Nebenbedingungen auswirkt. Der übliche Ansatz ist es Konvexität oder zumindest Quasi-Konvexität für die einzelnen Spielerprobleme anzunehmen. Es ist aber in gewissen Situationen interessant sich nicht konvexe Probleme anzusehen. Wir haben die Methodologie, welche im vorangegangenen Papier Anwendung gefunden hat, auf diese Situation angepasst. Wir nehmen an, dass die Funktionen Polynome sind und dass die zulässige Mengen der Spieler kompakt sind. Wir verwenden reelle algebraische Geometrie, um die einzelnen Spielerprobleme zu einem konvexen Optimierungsproblem zu relaxieren. Diese können dann wiederum mit Standardmethoden gelöst werden. Als ein Anwedungsbeispiel berechnen wir ein Modell des neuseeländischen Elektrizitätsmarktes.

Ich habe wieder die relevanten Sätze bewiesen, die Relaxierungsmethoden beschrieben und das Beispiel berechnet

4 Finding all pure-strategy equilibria in games with continuous strategies[1][2]

Abstract

Static and dynamic games are important tools for the analysis of strategic interactions among economic agents and have found many applications in economics. Such models are used both for policy experiments and for structural estimation studies. It is well-known that equilibrium multiplicity poses a serious threat to the validity of such analyses. This threat is particularly acute if not all equilibria of the examined model are known. Often equilibria can be described as solutions of polynomial equations (which must also perhaps satisfy some additional inequalities.) In this paper we describe state-of-the-art techniques developed in algebraic geometry for finding all solutions of polynomial systems of equations and illustrate these techniques by computing all equilibria of both static and dynamic games with continuous strategies. We compute the equilibrium manifold for a Bertrand pricing game in which the number of pure-strategy equilibria changes with the market size. Moreover, we apply these techniques to two stochastic dynamic games of industry competition and check for equilibrium uniqueness. Our examples show that the all-solution methods can be applied to a variety of static and dynamic models.

4.1 Introduction

Multiplicity of equilibria is a prevalent problem in equilibrium models with strategic interactions. This problem has long been acknowledged in the theoretical literature but until now been largely ignored in applied work even though simple examples of multiple equilibria have been known for decades, see, for example, the model of strategic investment in Fudenberg and Tirole, 1983a. Until recently this criticism was also true of one of the most prolific literatures of applied game-theoretic models, namely that

[1](Judd et al., 2012)

[2]We are indebted to five anonymous referees and co-editor Elie Tamer for very helpful comments on an earlier version of the paper. We thank Guy Arie, Paul Grieco, Felix Kubler, Andrew McLennan, Walt Pohl, Mark Satterthwaite, Andrew Sommese, Jan Verschelde, and Layne Watson for helpful discussions on the subject. We are very grateful to Jonathan Hauenstein for providing us with many details of the Bertini software package and for his patient explanations of the underlying mathematical theory. We are also grateful for comments from seminar audiences at the University of Chicago ICE workshops 2009–11, the University of Zurich, ESWC 2010 in Shanghai, the University of Fribourg, and the Zurich ICE workshop 2011. Karl Schmedders gratefully acknowledges the financial support of Swiss Finance Institute.

based on the framework for the study of industry evolution introduced by Ericson and Pakes, 1995. This framework builds the foundation for very active areas of research in industrial organization, marketing, and other fields—See the survey by Doraszelski and Pakes, 2007. Some recent work in this field is a great example of the growing interest in equilibrium multiplicity in active areas of modern applied economic analysis. Besanko et al., 2010 state that, to their knowledge, "all applications of Ericson and Pakes' (1995) framework have found a single equilibrium." They then show that multiple Markov-perfect equilibria can easily arise in a prototypical model in this framework. Borkovsky et al., 2010 and Doraszelski and Satterthwaite, 2010 present similar examples with multiple Markov-perfect equilibria. But findings of multiple equilibria are not confined to stochastic dynamic models. Bajari et al., 2010 show that multiple equilibria may arise in static games with incomplete information and discuss a possible approach to estimating such games. Clearly the difficulty of equilibrium multiplicity is not restricted to the cited papers. In fact in many other economic applications we may often suspect that there could be multiple equilibria.

In many economic models equilibria can be described as solutions to polynomial equations (which perhaps also must satisfy some additional inequalities.) Recent advances in computational algebraic geometry have led to several powerful methods and their easy-to-use computer implementations that find all solutions to polynomial systems. Two different solution approaches stand out—all-solution homotopy methods and Gröbner basis methods, both of which have their advantages and disadvantages. The methods which use Gröbner bases (Cox et al., 2007; Sturmfels, 2002) can solve only rather small systems of polynomial equations but can analyze parameterized systems. For an application of these methods to economics, see the analysis of parameterized general equilibrium models in Kubler and Schmedders, 2010. The all-solution homotopy methods (Sommese and Wampler, 2005) are purely numerical methods that cannot handle parameters but can solve much larger systems of polynomial equations. It is these homotopy methods that are the focus of the present paper.

All-solution homotopy methods for solving polynomial systems derived from economic models have been discussed previously in both the economics and mathematics literature on finite games. McKelvey and McLennan, 1996 mentions the initial work on the development of all-solution homotopy methods such as Drexler, 1977, Drexler, 1978, and Garcia and Zangwill, 1977. Herings and Peeters, 2005 outlines how to use all-solution homotopies for finding all Nash equilibria of generic finite n-person games in normal form but neither implements an algorithm nor solves any examples. Sturmfels, 2002 surveys methods for solving polynomial systems of equations and applies them to finding Nash equilibria of finite games. Datta, 2010 shows how to find all Nash equilibria of finite games by polyhedral homotopy continuation. Turocy, 2008 describes progress on a new implementation of a polyhedral continuation method using the software package PHCpack (Verschelde, 1999) in the software package Gambit (McKelvey et al., 2007). The literature on computing one, some, or all Nash equilibria in finite games remains very active—See the introduction to a recent symposium by von Stengel, 2010 and the many citations therein. For a recent application of all-solution homotopy ideas to calculating asymptotic approximations of all equilibria for static discrete games of incomplete information see Bajari et al., 2010. In the present paper, we do not consider finite games but instead analyze static and dynamic games with continuous strategies. Such games have many important economic applications. To our knowledge, the present paper is the first application of state-of-the-art all-solution homotopy methods to such games. In addition, this paper presents the first application of advanced techniques such as the parameter continuation method or the system-splitting approach to economic models.[3]

The application of homotopy methods has a long history in economics—See Eaves and Schmedders, 1999. Kalaba and Tesfatsion, 1991 proposes an adaptive homotopy method to allow the continuation parameters to take on complex values to deal with singular points along the homotopy path. Berry and Pakes, 2007 uses a homotopy approach for the estimation of demand systems. The homotopy approach was first applied to stochastic dynamic games by Besanko et al., 2010, Borkovsky et al., 2010 and Borkovsky et al., 2012. These three papers report results from the application of a classical homotopy approach to the computation of Markov-perfect equilibria in stochastic dynamic games. They show how homotopy paths can be used to find multiple equilibria. When the homotopy parameter is itself a parameter of the economic model, all points along the path represent economic equilibria (if the equilibrium equations are necessary and sufficient.) Whenever the path bends back on itself multiple equilibria exist. While this approach can detect equilibrium multiplicity it is not guaranteed to find all equilibria. Only the all-solution homotopy techniques presented in this paper allow for the computation of all equilibria. However, the classical homotopy approach has the advantage of finding (at least) one equilibrium of much larger economic models with thousands of equations which do not have to be polynomial. Currently available computational power may not allow us to solve systems with more than a few dozen equations depending on the degree of the polynomials. As we explain below, however, the all-solution homotopy methods are ideally suited to parallel computations. Our initial experience with an implementation on a computer cluster is very encouraging.

[3] In this paper we neither prove any new theorems nor present the most recent examples of frontier applications. Instead we follow the traditional approach in computational papers and describe a numerical method and apply it to examples that are familiar to most readers. This paper, as many previous computational papers have done, aims to educate the reader regarding the key ideas underlying a useful numerical method and illustrates these techniques in the context of familiar models. It does so in a way that makes it easy for readers to see how to apply these methods to their own particular problems, and points them to the appropriate software. To clarify what we mean by "traditional method" we should give a few examples. First, the paper by Kloek and Dijk, 1978 introduced Monte Carlo methods to basic econometrics using examples from the existing empirical literature and also focused on the methods as opposed to examining breakthrough applications. Second, Fair and Taylor, 1983 demonstrated how to use Gauss-Jacobi methods to solve rational expectation models. Again, the paper neither presented new theorems nor used frontier applications as examples. Instead it focused on very simple examples that made clear the mathematical structure of the algorithm and related it to the standard structure of rational expectation models. Third, Pakes and McGuire, 2001 showed how to use stochastic approximation to accelerate the Gauss-Jacobi algorithm that they had previously introduced in Pakes and McGuire, 1994 for the solution of stochastic dynamic games. Again, the paper did not analyze new applications and proved only one (convergence) theorem. Instead the paper educated the reader about stochastic ideas and illustrated their value in a well-known example. In this paper we follow the tradition of this literature.

The remainder of this paper is organized as follows. Section 4.2 describes a motivating economic example. We provide some intuition for the all-solution homotopy methods in Section 4.3. Next, Section 4.3.3 describes the theoretical foundation for the all-solution methods and Section 4.4 briefly comments on an implementation of such methods. In Section 4.5 we provide more details on the computations for the motivating example. Section 4.6 provides a description of the general set-up of dynamic stochastic games. In Section 4.7 we present an application of the all-solution methods to a stochastic dynamic learning-by-doing model. Similarly, Section 4.8 examines a stochastic dynamic model of cost-reducing investment with the all-solution homotopy. Finally, Section 4.9 concludes the paper and provides an outlook on future developments. The Appendix provides more mathematical details on four advanced features of all-solution homotopy methods.

4.2 Motivating example: Duopoly game with two equilibria

Before we describe details of all-solution homotopy methods, we motivate the application of such methods in economics by reporting results from applying such a method to a static duopoly game. Depending on the value of a parameter, this game may have no, one, or two pure-strategy equilibria. This example illustrates the various steps that are needed to find all pure-strategy Nash equilibria in a simple game with continuous strategies.

4.2.1 Bertrand price game

We consider a Bertrand price game between two firms. There are two products, x and y, two firms with firm x (y) producing good x (y), and three types of customers. Let p_x (p_y) be the price of good x (y). $Dx1$, $Dx2$, and $Dx3$ are the demands for product x by customer type 1, 2, and 3, respectively. Demands $Dy1$, etc. are similarly defined. Type 1 customers only want good x, and have a linear demand curve,

$$Dx1 = A - p_x; \;\; Dy1 = 0.$$

Type 3 customers only want good y and have a linear demand curve,

$$Dx3 = 0; \;\; Dy3 = A - p_y.$$

Type 2 customers want some of both. Let n be the number of type 2 customers. We assume that the two goods are imperfect substitutes for type 2 customers with a constant elasticity of substitution between the two goods and a constant elasticity of demand for a composite good. These assumption imply the demand functions

$$Dx2 = np_x^{-\sigma}\left(p_x^{1-\sigma} + p_y^{1-\sigma}\right)^{\frac{\gamma-\sigma}{-1+\sigma}}; \quad Dy2 = np_y^{-\sigma}\left(p_x^{1-\sigma} + p_y^{1-\sigma}\right)^{\frac{\gamma-\sigma}{-1+\sigma}} \; .$$

where σ is the elasticity of substitution between x and y, and γ is the elasticity of demand for the composite good $\left(q_1^{\frac{\sigma-1}{\sigma}} + q_2^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{(\sigma-1)}}$. Total demand for good x (y) is given by $Dx = Dx1 + Dx2 + Dx3$ ($Dy = Dy1 + Dy2 + Dy3$). Let m be the unit cost of production for each firm. Profit for good x is $R_x = (p_x - m)Dx$; Ry is similarly defined. Let MR_x be marginal profits for good x; similarly for MR_y. Equilibrium prices satisfy the necessary conditions $MR_x = MR_y = 0$.

Firm x (y) is a monopolist for type 1 (3) customers. The two firms only compete in the large market for type 2 customers. And so we may envision two different pricing strategies for the firms. The mass market strategy chooses a low price so that the firm can sell a large quantity to the large number of type 2 customers that would like to buy both goods but are price sensitive. Such a low price leads to small profits from the customers dedicated to the firm's product. The niche strategy is to just sell at a high price to the few customers that want only its good. Such a high price leads to small demand for its product among the price-sensitive type 2 customers.

We want to demonstrate how we can find all solutions even when there are multiple equilibria. The idea of our example is to find values for the parameters where each firm has two possible strategies. We examine a case where one firm goes for the high-price, small-sales (niche) strategy and the other firm goes after type 2 customers with a mass market strategy. Let

$$\sigma = 3,\ \gamma = 2,\ n = 2700,\ m = 1,\ A = 50.$$

The marginal profit functions are as follows.

$$MR_x = 50 - p_x + (p_x - 1)\left(\frac{2700}{p_x^6\left(p_x^{-2} + p_y^{-2}\right)^{3/2}} - \frac{8100}{p_x^4\sqrt{p_x^{-2} + p_y^{-2}}} - 1\right) + \frac{2700}{p_x^3\sqrt{p_x^{-2} + p_y^{-2}}}$$
$$MR_y = 50 - p_y + (p_y - 1)\left(\frac{2700}{p_y^6\left(p_x^{-2} + p_y^{-2}\right)^{3/2}} - \frac{8100}{p_y^4\sqrt{p_x^{-2} + p_y^{-2}}} - 1\right) + \frac{2700}{p_y^3\sqrt{p_x^{-2} + p_y^{-2}}}$$

4.2.2 Polynomial equilibrium equations

We first construct a polynomial system. The system we construct must contain all the equilibria, but it may have extraneous solutions. The extraneous solutions present no problem because we can easily identify and discard them.

We need to eliminate the radical terms. Let Z be the square root term

$$Z = \sqrt{p_x^{-2} + p_y^{-2}},$$

which implies

$$0 = Z^2 - \left(p_x^{-2} + p_y^{-2}\right).$$

This is not a polynomial. We gather all terms into one fraction and extract the numerator, which is the polynomial we include in our polynomial system to represent the variable Z,

$$0 = -p_x^2 - p_y^2 + Z^2 p_x^2 p_y^2. \tag{4.1}$$

We next use the Z definition to eliminate radicals in MR_x and MR_y. Again we gather terms into one fraction and extract the numerator. The second and third equation of our polynomial are as follows:

$$0 = -2700 + 2700p_x + 8100Z^2p_x^2 - 5400Z^2p_x^3 + 51Z^3p_x^6 - 2Z^3p_x^7, \quad (4.2)$$
$$0 = -2700 + 2700p_y + 8100Z^2p_y^2 - 5400Z^2p_y^3 + 51Z^3p_y^6 - 2Z^3p_y^7. \quad (4.3)$$

Any pure-strategy Nash equilibrium is a solution of the polynomial system (4.1,4.2,4.3).

4.2.3 Solution

Solving the above system of polynomial equations (see Section 4.5.1 for details) we find 18 real and 44 complex solutions. Nine of the 18 real solutions contain at least one variable with a negative value and are thus economically meaningless. Table 4.1 shows the remaining 9 solutions. We next check the second-order conditions of each firm. This

p_x	1.757	8.076	22.987	2.036	5.631	2.168	25.157	7.698	24.259
p_y	1.757	8.076	22.987	5.631	2.036	25.157	2.168	24.259	7.698

Table 4.1: Real, positive solutions of (4.1,4.2,4.3)

check eliminates five more real solutions and reduces the set of possible equilibria to four, namely

$$\left(p_x^1, p_y^1\right) = (1.757, 1.757), \qquad \left(p_x^2, p_y^2\right) = (22.987, 22.987),$$
$$\left(p_x^3, p_y^3\right) = (2.168, 25.157), \qquad \left(p_x^4, p_y^4\right) = (25.157, 2.168).$$

We next need to check global optimality for each player in each potential equilibrium. The key fact is that the global max must satisfy the first-order conditions given the other player's strategy. So, all we need to do is to find all solutions to a firm's first-order condition at the candidate equilibrium, and then find which one produces the highest profits. We keep the candidate equilibrium only if it is the global maximum.

First consider $\left(p_x^1, p_y^1\right)$. We first check to see if player x's choice is globally optimal given p_y. Since we take p_y as given, the equilibrium system reduces to the Z equation and the first-order condition for player x, giving us the polynomial system

$$0 = 0.32410568484991703p_x^2 + 1 - Z^2p_x^2$$
$$0 = -2700 + 2700p_x + 8100Z^2p_x^2 - 5400Z^2p_x^3 + 51Z^3p_x^6 - 2Z^3p_x^7$$

This system has 14 finite solutions, 8 complex and 6 real solutions. One of the solutions is $p_x = 25.2234$ where profits equal 607.315. Since this exceeds 504.625, firm x's profits at $\left(p_x^1,\ p_y^1\right)$, we conclude that $\left(p_x^1,\ p_y^1\right)$ is not an equilibrium. A similar approach shows that $\left(p_x^2, p_y^2\right)$ is not an equilibrium. Given $p_y^2 = 22.987$, firm x would receive a higher profit from a low price than from p_x^2. When we examine the remaining two candidate equilibria, we find that these are two asymmetric equilibria, $\left(p_x^3,\ p_y^3\right)$ and $\left(p_x^4,\ p_y^4\right)$. This may not appear to be an important multiplicity since the two equilibria are mirror images of each other. However, it is clear that if we slightly perturb the demand functions to eliminate the symmetries that there will still be two equilibria that are not mirror images.

In the equilibrium $\left(p_x^3,\ p_y^3\right) = (2.168, 25.157)$, firm x chooses a mass-market strategy and firm y a niche strategy. The low price allows firm x to capture most of the market of price-sensitive type 2 customers while it forgoes most of the possible (monopoly) profits in its niche market of type 1 customers. Firm y instead charges a high price (just below the monopoly price for the market of type 3 customers) to capture most of its niche market. In the equilibrium $\left(p_x^4, p_y^4\right) = (25.157, 2.168)$ the strategies of the two firms are reversed.

This example demonstrates that the problem of finding all Nash equilibrium reduces to solving a series of polynomial systems. The first system identifies a set of solutions for the firms' first-order conditions, which are only necessary but not sufficient. The second step is to eliminate all candidate equilibria where some firm does not satisfy the local second-order condition for optimization. The third step is to check the global optimality of each firm's reactions in each of the remaining candidate equilibria. This step reduces to finding all solutions of a set of smaller polynomial systems.

Figure 4.1 displays the manifold of a firm's equilibrium prices for values of the market size parameter n between 500 and 3400. For $500 \leq n \leq 2470$ there is a unique equilibrium. The competitive market of type 2 customers is so small that each firm chooses a niche strategy and charges a high price to focus on the few customers that only want its good. For $3318 \leq n \leq 3400$ there is again a unique equilibrium. The competitive market of type 2 customers is now sufficiently large so that each firm chooses a mass market strategy and charges a low price to sell a high quantity into the mass market of type 2 customers. For $2481 \leq n \leq 3020$ there are two equilibria. At these intermediate values of n, the two firms prefer complementary strategies, one firm chooses a (high-price) niche strategy and the other firm a (low-price) mass market strategy. And finally there are two regions with no pure-strategy equilibria, namely for $2471 \leq n \leq 2480$ and also for $3021 \leq n \leq 3317$.

4.3 All-solution homotopy methods

In this section we introduce the mathematical background of all-solution homotopy methods for polynomial systems of equations. Polynomial solution methods rely on results from complex analysis and algebraic geometry. For this purpose we first review some basic definitions.

4.3.1 Mathematical background

We define a polynomial in complex variables.

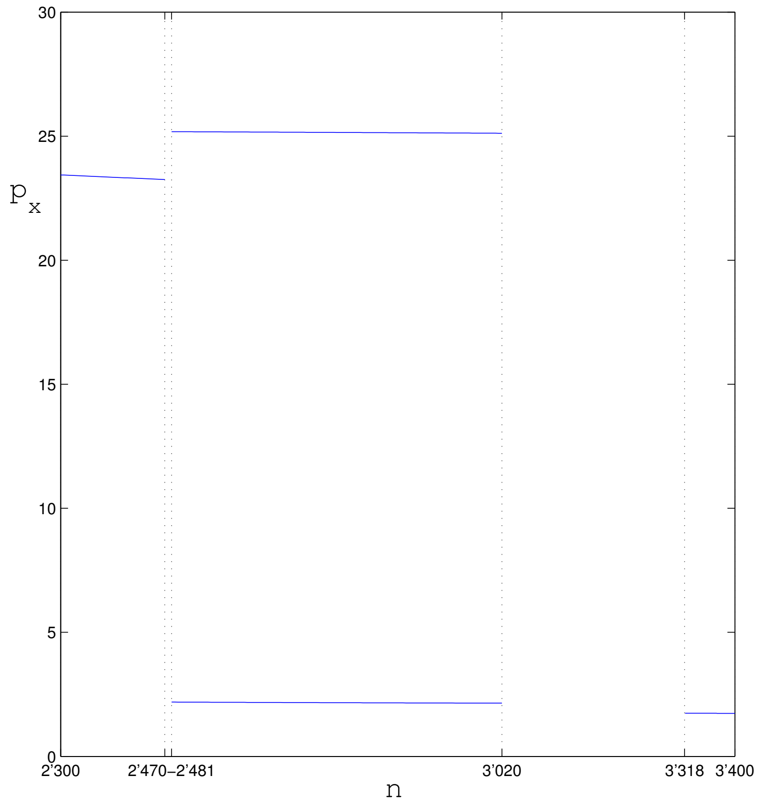

Figure 4.1: Equilibrium prices as a function of n

Definition 4.1. A polynomial f over the variables $z_1, \ldots, z_n$ is defined as

$$f(z_1, \ldots, z_n) = \sum_{j=0}^{d} \left(\sum_{d_1+\ldots+d_n=j} a_{(d_1,\ldots,d_n)} \prod_{k=1}^{n} z_k^{d_k} \right) \text{ with } a_{(d_1,\ldots,d_n)} \in \mathbb{C},\ d \in \mathbb{N}.$$

For convenience we denote $z = (z_1, \ldots, z_n)$. The expression $a_{(d_1,\ldots,d_n)} \prod_{k=1}^{n} z_k^{d_k}$ for $a_{(d_1,\ldots,d_n)} \neq 0$ is called a term of f. The degree of f is defined as $\deg f = \max_{a_{(d_1,\ldots,d_n)}\neq 0} \sum_{k=1}^{n} d_k$. The term $\sum_{d_1+\ldots+d_n=j} a_{(d_1,\ldots,d_n)} \prod_{k=1}^{n} z_k^{d_k}$ is called the homogeneous part of degree j of f and is denoted by $f^{(j)}$.

Note that $f^{(j)}$ being homogeneous of degree j means $f^{(j)}(cz) = c^j f^{(j)}(z)$ for any complex scalar $c \in \mathbb{C}$. We now regard a polynomial f in the variables $z_1, \ldots, z_n$ as a function $f : \mathbb{C}^n \to \mathbb{C}$. Then f belongs to the following class of functions.

Definition 4.2. Let $U \subset \mathbb{C}^n$ be an open subset and $f : U \to \mathbb{C}$ a function. Then we call f analytic at the point $b = (b_1, \ldots, b_n) \in U$ if and only if there exists a neighborhood V

of b such that

$$f(z) = \sum_{j=0}^{\infty} \left(\sum_{d_1+\ldots+d_n=j} a_{(d_1,\ldots,d_n)} \prod_{k=1}^{n} (z_k - b_k)^{d_k} \right), \quad \forall z \in V,$$

where $a_{(d_1,\ldots,d_n)} \in \mathbb{C}$, i.e. the above power series converges to the function f on V. It is called the Taylor series of f at b.

Obviously every function given by polynomials is analytic with one Taylor expansion on all of $\mathbb{C}^n$. However note that in general $V \subsetneq U$ and that the power series is divergent outside of V. For functions in complex space we can state the Implicit Function Theorem analogously to the case of functions in real space.

Theorem 4.1 (Implicit Function Theorem)**.** *Let*

$$H : \mathbb{C} \times \mathbb{C}^n \longrightarrow \mathbb{C}^n \quad \text{with} \quad (t, z_1, \ldots z_n) \longmapsto H(t, z_1, \ldots z_n)$$

be an analytic function. Denote by $D_z H = \left(\frac{\partial H_j}{\partial z_i} \right)_{i,j=1,\ldots n}$ *the submatrix of the Jacobian of* H *containing the partial derivatives with respect to* z_i*,* $i = 1, \ldots, n$*. Furthermore let* $(t_0, x_0) \in \mathbb{C} \times \mathbb{C}^n$ *such that* $H(t_0, x_0) = 0$ *and* $\det D_z H(t_0, x_0) \neq 0$*. Then there exist neighborhoods* T *of* t_0 *and* A *of* x_0 *and an analytic function* $x : T \to A$ *such that* $H(t, x(t)) = 0$ *for all* $t \in T$*. Furthermore the chain rule implies that*

$$\frac{\partial x}{\partial t}(t_0) = -D_z H(t_0, x_0)^{-1} \cdot \frac{\partial H}{\partial t}(t_0, x_0).$$

Next we define the notion of a path.

Definition 4.3. Let $A \subset \mathbb{C}^n$ be an open or closed subset. An analytic[4] function $x : [0, 1] \to A$ or $x : [0, 1) \to A$ is called a path in A.

Definition 4.4. Let $H(t, z) : \mathbb{C}^{n+1} \to \mathbb{C}^n$ and $x : [0, 1] \to \mathbb{C}^n$ an analytic function such that $H(t, x(t)) = 0$ for all t. Then x defines a path in $\{(t, x) \in \mathbb{C}^{n+1} \mid H(t, x) = 0\}$. We call the path *regular*, iff $\{t \in [0, 1) \mid H(t, x(t)) = 0, \det D_z H(t, x(t)) = 0\} = \emptyset$.[5]

Note that for general homotopy methods the regularity definition is less strict. One usually only wants the Jacobian to have full rank. Here we also impose which part of it has full rank. Such a definition is reasonable for polynomial homotopy methods since, as we see later, we can ensure this property for our paths.

Definition 4.5. Let $A \subset \mathbb{C}^n$. We call A *pathwise connected*, iff for all points $a_1, a_2 \in A$ there exists a continuous function $x : [0, 1] \to A$ such that $x(0) = a_1$ and $x(1) = a_2$.

Lastly we need the following notion from topology.

[4]The usual definition of a path only requires continuity, but all paths we regard are automatically given by analytic functions.

[5]We see below why we can exclude $t = 1$ from our regularity assumption.

Definition 4.6. Let $U, V \subset \mathbb{C}^n$ be open subsets and $h_0 : U \to V$, $h_1 : U \to V$ be continuous functions. Let

$$\begin{array}{rcl} H : [0,1] \times U & \longrightarrow & V \\ (t,z) & \longmapsto & H(t,z) \end{array}$$

be a continuous function such that $H(0,z) = h_0(z)$ and $H(1,z) = h_1(z)$. Then we call H a homotopy from h_0 to h_1.

4.3.2 Building intuition from the univariate case

Homotopy methods have a long history in economics, see Eaves and Schmedders, 1999, for finding one solution to a system of nonlinear equations. Recent applications of such homotopy methods in game-theoretic models include Besanko et al., 2010 and Borkovsky et al., 2010. Homotopy methods for finding all solutions of polynomial systems were first introduced by Garcia and Zangwill, 1977 and Drexler, 1977. These papers initiated an active field of research that is still advancing today, see Sommese and Wampler, 2005 for an overview. In this subsection, following Sommese and Wampler, 2005 and the many cited works therein, we provide some intuition for the theoretical foundation underlying all-solution homotopy continuation methods.

The basic idea of the homotopy approach is to find an easier system of equations and continuously transform it into our target system. We first illustrate this for univariate polynomials. Consider the univariate polynomial $f(z) = \sum_{i \leq d} a_i z^i$ with $a_d \neq 0$ and $\deg f = d$. The Fundamental Theorem of Algebra states that f has precisely d complex roots, counting multiplicities.[6] A simple polynomial of degree d with d distinctive complex roots is $g(z) = z^d - 1$, whose roots are $r_k = e^{\frac{2\pi i k}{d}}$ for $k = 0, \ldots, d-1$. (These roots are called the d-th roots of unity.) Now we can define a homotopy H from g to f by setting $H = (1-t)g + tf$. Thus H is a polynomial in t, z and therefore an analytic function. Under the assumption that $\frac{\partial H}{\partial z}(t,z) \neq 0$ for all (t,z) satisfying $H(t,z) = 0$ and $t \in [0,1]$ the Implicit Function Theorem (Theorem 4.1) states that each root r_k of g gives rise to a path that is described by an analytical function. The idea is now to start at each solution $z = r_k$ of $H(0,z) = 0$ and to follow the resulting path until a solution z of $H(1,z) = 0$ has been reached. The path-following can be done numerically using a predictor-corrector method (see, for example, Allgower and Georg, 2003). For example, Euler's method is a so-called first-order predictor and obtains a first step along the path by choosing an $\varepsilon > 0$ and calculating

$$\tilde{x}_k(0+\varepsilon) = x_k(0) + \varepsilon \frac{\partial x_k}{\partial t}(0),$$

where the $\frac{\partial x_k}{\partial t}(0)$ are implicitly given by Theorem 4.1. Then this first estimate is corrected using Newton's method with starting point $\tilde{x}_k(0+\varepsilon)$. So the method solves the equation $H(\varepsilon, z) = 0$ for z and sets $x_k(\varepsilon) = z$.

[6] Any univariate polynomial of degree d over the complex numbers can be written as $f(z) = c(z - b_1)^{r_1}(z-b_2)^{r_2} \cdots (z-b_l)^{r_l}$ with $c \in \mathbb{C} \setminus \{0\}$, $b_1, b_2, \ldots, b_l \in \mathbb{C}$, and $r_1, r_2, \ldots, r_l \in \mathbb{N}$. The exponent r_j denotes the multiplicity of the root b_j. For example, the polynomial z^3 has the single root $z = 0$ with multiplicity 3.

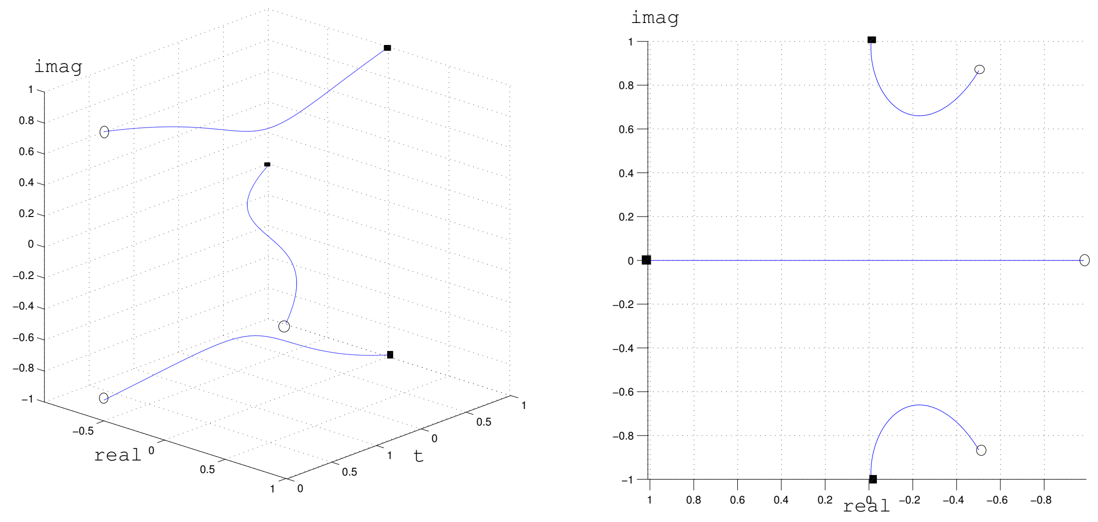

Figure 4.2: Homotopy paths in Example 4.1 and the projection to $\mathbb{C}$.

Example 4.1. As a first example we look at the polynomial $f(z) = z^3 + z^2 + z + 1$. The zeros are $\{-1, -i, i\}$. As a start polynomial we choose $g(z) = z^3 - 1$. We define a homotopy from g to f as follows,

$$H(t, z) = (1 - t)(z^3 - 1) + t(z^3 + z^2 + z + 1).$$

This homotopy generates the three solution paths shown in Figure 4.2. The starting points of the three paths, $-\frac{1}{2} - \frac{\sqrt{3}}{2}i$, $-\frac{1}{2} + \frac{\sqrt{3}}{2}i$, 1, respectively, and are indicated by circles. The respective end points, $-i$, i, and -1 are indicated by squares.

This admittedly rough outline captures the fundamental idea of the all-solution homotopy methods. This method can potentially run into difficulties. First, the paths might cross and, secondly, the paths might bend sideways and diverge. We illustrate these problems with an example and also show how to circumvent them.

Example 4.2. Let $f(z) = 5 - z^2$ and $g(z) = z^2 - 1$. Then a homotopy from g to f can be defined as

$$H(t, z) = t(5 - z^2) + (1 - t)(z^2 - 1) = (1 - 2t)z^2 + 6t - 1. \tag{4.4}$$

Now $H(\frac{1}{6}, z) = \frac{2}{3}z^2$ has the double root $z = 0$, so $\det D_z H(\frac{1}{6}, 0) = 0$. Such points are called non-regular and the assumption of the Implicit Function Theorem is not satisfied. Non-regular points are also problematic for the Newton corrector step in the path-following algorithm. But matters are even worse for this homotopy since $H(\frac{1}{2}, z) = 2$, which has no zero at all, i.e. there can be no solution path from $t = 0$ to $t = 1$. The coefficient of the leading term $(1-2t)z^2$ has become 0 and so the degree of the polynomial

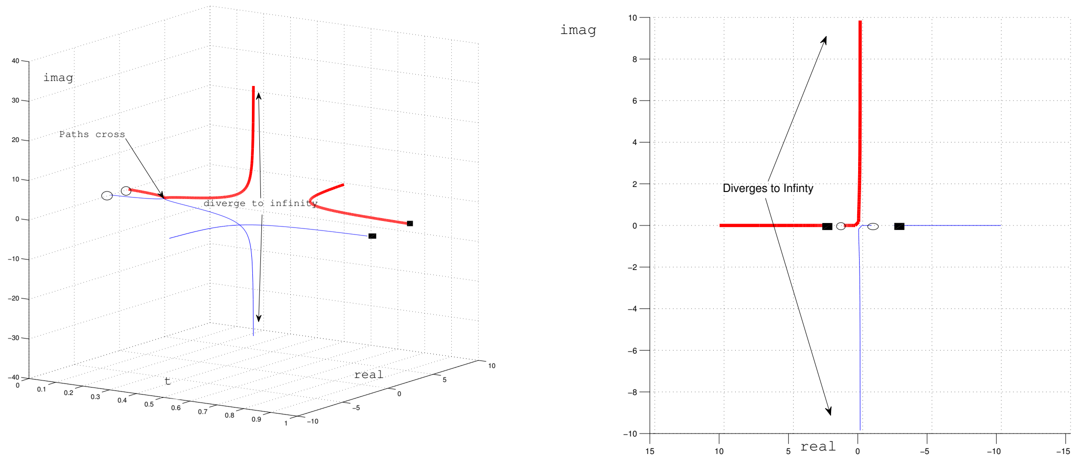

Figure 4.3: Homotopy paths in Example 4.2 and the projection to $\mathbb{C}$. One path is colored red, the other is colored blue.

H drops at $t = \frac{1}{2}$. Figure 4.3 displays the set of zeros of the homotopy. The two paths starting at $\sqrt{5}$ and $-\sqrt{5}$ diverge as $t \to \frac{1}{2}$.

The general idea to resolve the technical problems illustrated in Example 4.2 is to "walk around" the points that cause us trouble. For a description of this idea we need the following theorem which describes one of the differences between complex and real spaces.

Theorem 4.2. *Let $F = (f_1, \ldots, f_k) = 0$ be a system of polynomial equations in n variables, with $f_i \neq 0$ for some i. Then $\mathbb{C}^n \setminus \{F = 0\}$ is a pathwise connected and dense subset of $\mathbb{C}^n$.*[7]

This statement does not hold true over the reals. Take for instance $n = 2$, $k = 1$ and set $f_1(x_1, x_2) = x_1$. (Note that f_1 is not identically zero.) Now we restrict ourselves to the real numbers, $(x_1, x_2) \in \mathbb{R}^2$. If we remove the zero set $\{(x_1, x_2) \in \mathbb{R}^2 : f_1(x_1, x_2) = 0\}$, which is the vertical axis, then the resulting set $\mathbb{R}^2 \setminus \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 = 0\}$ consists of two disjoint components. Thus it is not pathwise connected.

Example 4.3. Returning to Example 4.2 we temporarily regard t also as a complex variable and thus $\{(t, z) | H(t, z) = 0\} \subset \mathbb{C}^2$. Due to theorem 4.1 we only have a path if locally the determinant is nonzero. The points that are not regular are characterized by the equations

$$\begin{aligned} (1 - 2t)z^2 + 6t - 1 &= 0 \\ \det D_z H = 2z(1 - 2t) &= 0. \end{aligned} \tag{4.5}$$

[7] This is a simpler version of the theorem that is actually needed. But for simplicity we avoid the general case.

Points at which our path is interrupted are given by

$$1 - 2t = 0. \tag{4.6}$$

In this case we can easily determine that the only solution to (4.5) is $(\frac{1}{6}, 0)$ and the solution to (4.6) is $\{t = \frac{1}{2}\}$. The union of the solution sets to the two equations is exactly the solution set of the following system of equations

$$\begin{aligned} ((1-2t)z^2 + 6t - 1)(1-2t) &= 0 \\ (2z(1-2t))(1-2t) &= 0. \end{aligned} \tag{4.7}$$

Theorem 4.2 now implies that the complement of the solution set to system (4.7) is pathwise connected. In other words, we can find a path between any two points without running into problematic points. To walk around those problematic points we define a new homotopy by multiplying the start polynomial $z^2 - 1$ by $e^{i\gamma}$ for a random $\gamma \in [0, 2\pi)$:

$$H(t,z) = t(5 - z^2) + e^{i\gamma}(1-t)(z^2-1) = (e^{i\gamma} - t - te^{i\gamma})z^2 + te^{i\gamma} - e^{i\gamma} + 5t. \tag{4.8}$$

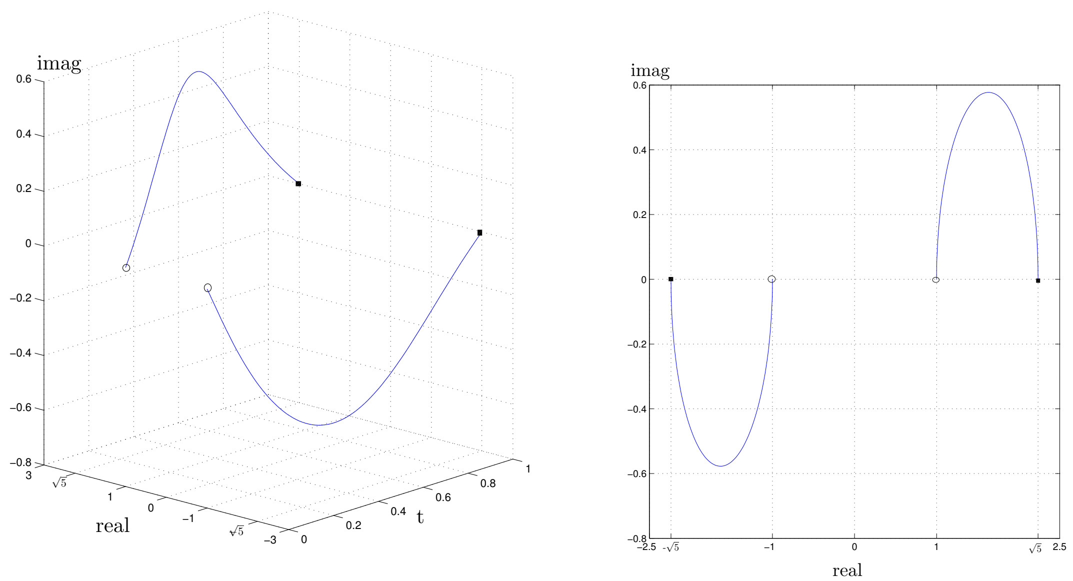

Figure 4.4: Homotopy paths in Example 4.3 after application of the gamma trick.

Now we obtain $D_z H = 2(e^{i\gamma} - t - te^{i\gamma})z$ which has $z = 0$ as its only solution if $e^{i\gamma} \notin \mathbb{R}$ and $t \in [0,1]$. Furthermore if $e^{i\gamma} \notin \mathbb{R}$ then $H(t,0) = te^{i\gamma} - e^{i\gamma} + 5t \neq 0$ for all $t \in [0,1]$. Additionally the coefficient of z^2 in (4.8) does not vanish for $t \in \mathbb{R}$ and thus $H(t,x) = 0$ has always two solutions for $t \in [0,1]$ due to the Fundamental Theorem of Algebra. Therefore this so-called gamma trick yields only paths that are not interrupted and are regular. Figure 4.4 displays the two paths; the left graph shows the paths in three dimensions, the right graph shows a projection of the paths on $\mathbb{C}$. It remains to check how strict the condition $e^{i\gamma} \notin \mathbb{R}$ is. We know $e^{i\gamma} \in \mathbb{R} \Leftrightarrow \gamma = k\pi$ for $k \in \mathbb{N}$. Since $\gamma \in [0, 2\pi)$ these are only two points. Thus for a random γ the paths exist and are regular with probability one.

This example concludes our introductory discussion of the all-solution homotopy approach. In the next subsection we describe technical details of the general multivariate homotopy approach. A reader who is mainly interested in the quick implementation of homotopies as well as economic applications may want to skip this part and continue with Section 4.4.

4.3.3 The multivariate case

When we attempt to generalize the outlined approach from the univariate to the multivariate case we encounter a significant difficulty. The Fundamental Theorem of Algebra does not generalize to multiple equations and so we do not know a priori the number of complex solutions. However, we can determine upper bounds on the number of solutions. For the sake of our discussion in this paper it suffices to introduce the simplest such bound.

Definition 4.7. Let $F = (f_1, \ldots f_n) : \mathbb{C}^n \to \mathbb{C}^n$ be a polynomial function. Then the number

$$d = \prod_i \deg f_i$$

is called the *total degree* or *Bezout number* of F.

Theorem 4.3 (Bezout's Theorem). *Let d be the Bezout number of F. Then the polynomial system $F = 0$ has at most d isolated solutions counting multiplicities.*

This bound is tight, in fact, García and Li, 1980 show that generic polynomial systems have exactly d distinct isolated solutions. But this result does not provide any guidance for specific systems, since systems arising in economics and other applications will typically be so special that the number of solutions is much smaller.

Next we address the difficulties we observed in Example 4.2 for the multivariate case. Consider a square polynomial system $F = (f_1, \ldots, f_n) = 0$ with $d_i = \deg f_i$. Construct a start system $G = (g_1, \ldots, g_n) = 0$ such that

$$g_i(z) = z_i^{d_i} - 1. \tag{4.9}$$

Note that the polynomial $g_i(z)$ only depends on the variable z_i and has the same degree as $f_i(z)$. The polynomial functions F and G have the same Bezout number. Now construct a homotopy $H = (h_1, \ldots, h_n) : \mathbb{C} \times \mathbb{C}^n \to \mathbb{C}^n$ from the square polynomial system $F(z) = 0$ and the start system $G(z) = 0$ that is linear in the homotopy parameter t. As a result $h_i(z)$ is a polynomial of degree d_i in the variables $z_1, \ldots, z_n$ and coefficients that are linear functions in t,

$$h_i(z) = \sum_{j=0}^{d_i} \left(\sum_{c_1+\ldots+c_n=j} a_{(i,c_1,\ldots,c_n)}(t) \prod_{k=1}^{n} z_k^{c_k} \right)$$

In a slight abuse of notation we denote by $a_i(t)$ the product of the coefficients of the highest-degree monomials of $h_i(z)$. As before we need to rule out non-regular points and values of the homotopy parameter for which the system $H(t,z)=0$ may have no solution. Non-regular points are solutions to the following system of equations.

$$\begin{aligned} h_i &= 0 \quad \forall i \\ \det D_z H &= 0. \end{aligned} \tag{4.10}$$

Additionally, values of the homotopy parameter for which one or more of our paths might get interrupted are all t that satisfy the following equation,

$$\prod_i a_i(t) = 0. \tag{4.11}$$

For a t' satisfying the above equation it follows that the polynomial $H(t',z)$ has a lower Bezout number than $F(z)$.[8] Analogously to example 4.3 we can cast (4.10) and (4.11) in one system of equations,

$$\begin{aligned} h_i \prod_j a_j(t) &= 0 \quad \forall i \\ \det(D_z H) \prod_i a_i(t) &= 0. \end{aligned} \tag{4.12}$$

Theorem 4.2 states that the complement of the solution set to this system of equations is a pathwise connected set. So as before we can "walk around" those points that cause difficulties for the path-following algorithm. In fact, if we choose our paths randomly just as in Example 4.3 then we do not encounter those problematic points with probability one.

Theorem 4.4 (Gamma trick). *Let $G(z):\mathbb{C}^n \to \mathbb{C}^n$ be our start system and $F(z):\mathbb{C}^n \to \mathbb{C}^n$ our target system. Then for almost all[9] choices of the constant $\gamma \in [0,2\pi)$ the homotopy*

$$H(t,z) = e^{\gamma i}(1-t)G(z) + tF(z) \tag{4.13}$$

has regular solution paths and $|\{z \mid H(t_1,z)=0\}| = |\{z \mid H(t_2,z)=0\}|$ for all $t_1,t_2 \in [0,1)$.

We say that a path *diverges to infinity* at $t=1$ if $\|z(t)\| \to \infty$ for $z(t)$ satisfying $H(t,z(t))=0$ as $t\to 1$ where $\|\cdot\|$ denotes the Euclidean norm. The Gamma trick leads to the following theorem.

Theorem 4.5. *Consider the homotopy H as in* (4.13) *with a start system as in* (4.9)*. For almost all parameters $\gamma \in [0,2\pi)$, the following properties hold.*

[8] Note that after homogenization, which we introduce in Section 4.10.1, this no longer poses any problem.

[9] Throughout this paper the terminology "almost all" means an open set of measure one. All stated results in fact hold on so-called Zariski-open sets, but for simplicity we omit a proper definition of this term.

1. *The preimage $H^{-1}(0)$ consists of d regular paths, i.e. no paths cross or bend backwards.*

2. *Each path either diverges to infinity or converges to a solution of $F(z) = 0$ as $t \to 1$.*

3. *If $\bar{z}$ is an isolated solution with multiplicity[10] m, then there are m paths converging to it.*

By construction the easy system $G(z) = 0$ has exactly d isolated solutions. Each of these solutions is the starting point of a smooth path along which the parameter t increases monotonically, that is, the Jacobian has full rank and the path does not bend backwards. To find all solutions of $F(z) = 0$ we need to follow all d paths and check whether they diverge or run into a solution of our system. In light of the aforementioned result by García and Li, 1980 that generic polynomial systems $F(z) = 0$ have d isolated solutions, Theorem 4.5 implies that the homotopy H gives rise to d distinct paths that terminate at the d isolated roots of F. So, generically the intuition of the univariate case carries over to the multivariate case.

4.3.4 Advanced features

The described method is intuitive but has two major drawbacks that make it impractical. First, the paths diverging to infinity are of no interest in economic applications. Second, the number of paths grows exponentially in the number of nonlinear equations. A practical homotopy method needs to spend as little time as possible on diverging paths. In addition, it will always be advantageous to keep the number of paths as small as possible. Advanced all-solution homotopy methods address both these problems. In the appendix we describe the underlying mathematical approaches.

The diverging paths are of no interest for finding economically meaningful solutions to systems of equations derived from an economic model. The diverging paths typically require much more computational effort than converging paths. And their potential presence requires a computer program following the paths to decide whether a path is diverging or only very long but converging. The decision when to declare that a path is diverging cannot be made without the risk of actually truncating a very long converging path. A reliable and robust computational method thus needs some feature to handle diverging paths. It is possible to "compactify" the diverging path through a homogenization of the polynomials. Appendix 4.10.1 describes this approach.

The number of paths d grows rapidly with the degree of individual equations. It also grows exponentially in the number of equations (if the equations are not linear). For many economic models we believe that there are only a few (if not unique) equilibria, that is, our systems have few real solutions and usually even fewer economically meaningful solutions. As a result we may have to follow a large number of paths that do not yield

[10] Multiplicity of a root for a system of polynomial equations is similar to multiplicity in the univariate case. We forgo any proper definition for the sake of simplicity.

useful solutions. Also, if there are only a few real and complex solutions then many paths must converge to solutions at infinity. There may even be continua of solutions at infinity which can cause numerical difficulties, see Example 4.4 in Appendix 4.10.1 below. Therefore it would be very helpful to reduce the number of paths that must be followed as much as possible. Appendices 4.10.2 and 4.10.3 describe two methods for a reduction of the number of paths.

4.4 Implementation

We briefly describe the software package Bertini and the potential computational gains from a parallel version of the software code.

4.4.1 Bertini

The software package Bertini, written in the programming language C, offers solvers for a few different types of problems in numerical algebraic geometry, see Bates et al., 2005. The most important feature for our purpose is Bertini's homotopy continuation routine for finding all isolated solutions of a square system of polynomial equations. In addition to an implementation of the advanced homotopy of Theorem 4.7 (see Appendix 4.10.1) it also allows for m-homogeneous start systems as well as parameter-continuation homotopies as in Theorem 4.8, see Appendices 4.10.2 and 4.10.3. Bertini has an intuitive interface which allows the user to quickly implement systems of polynomial equations, see Sections 4.5.1 and 4.5.2 for examples of code that a user must supply. Bertini can be downloaded free of charge under `http://www.nd.edu/~sommese/bertini/` .

All results in this paper were computed with Bertini on a laptop, namely an Intel Core 2 Duo T9550 with 2.66 GHz and 4GB RAM.

4.4.2 Alternatives

Two other all-solution homotopy software packages are PHCpack (Verschelde, 1999) written in ADA and POLSYS_PLP (Wise et al., 2000) written in FORTRAN90 and which is intended to be used in conjunction with HOMPACK90 (Watson et al., 1997), a popular homotopy path solver. Because of its versatility, stable implementation, great potential for parallelization on large computer clusters and its friendly user interface we use Bertini for all our calculations.

4.4.3 Parallelization

The overall complexity of the all-solution homotopy method is the same as for other methods used for polynomial system solving. The major advantage of this method, however, is that it is naturally parallelizable. Following each path is a distinct task, i.e. the paths can be tracked independently from each other. Moreover, the information gathered during the tracking process of a path cannot be used to help track other paths.

This advantage coincides with the recent developments in processing technology. The performance of a single processor will no longer grow as in the years before, since power consumption and the core temperature have become big issues in the production of computer chips. The new strategy of computer manufactures is to use multiple cores within a single machine to spread out the workload.

The software package Bertini is available in a parallel version. As of this writing, we have already successfully computed examples via parallelization on 200 processors at the CSCS cluster (Swiss Scientific Computing Center). In order to spread the work across many more processors a modest revision of the Bertini code is necessary. We are optimistic that we will soon be able to solve problems on clusters with thousands of processors. Such a set-up will allow us to solve problems that are orders of magnitude larger than those described below.

4.5 Bertrand price game continued

We return to the duopoly price game from Section 4.2. We now show how to solve the problem with Bertini. We also show how to use some of the advanced features from Appendices 4.10.1–4.10.3.

4.5.1 Solving the Bertrand price game with Bertini

To solve the system (4.1,4.2,4.3) in Bertini we write the following input file:

```
CONFIG
MPTYPE: 0;
END;

INPUT
variable_group px,py,z;
function f1, f2, f3;
f1 = -(px^2)-py^2+z^2*px^2*py^2;
f2 = -(2700)+2700*px+8100*z^2*px^2-5400*z^2*px^3+51*z^3*px^6-2*z^3*px^7;
f3 = -(2700)+2700*py+8100*z^2*py^2-5400*z^2*py^3+51*z^3*py^6-2*z^3*py^7;
END;
```

The option `MPTYPE:0` indicates that we are using standard path-tracking. The polynomials `f1,f2,f3` define the system of equations. The Bezout number is $6 \times 10 \times 10 = 600$. Thus, Bertini must track 600 paths. With the above code, we obtained 18 real solutions, 44 complex solutions, 270 truncated infinite paths and 268 failures.[11] In Appendix 4.10.1 we show that, if we homogenize the above equations, then we have continua of solutions at infinity as illustrated in Example 4.4. Such solutions are responsible for the large number of failures since at these solutions the rank of the Jacobian drops. Of course,

[11] In those cases the path tracker failed to converge on a solution at infinity. Note that Bertini uses random numbers to define the homotopy, so the number of failed paths varies.

such paths with convergence failures represent a serious concern. Fortunately, Bertini offers the option `MPTYPE: 2` for improved convergence. This command instructs Bertini to use adaptive precision which handles singular solutions much better but needs more computation time. We then find the same 18 real and 44 complex solutions as before. But in contrast to the previous run, we now have 538 truncated infinite paths and no failures. Bertini lists the real solution in the file `real_finite_solutions` and all finite ones in `finite_solutions`.

Next we show how to reduce the number of paths with m-homogenization (see Appendix 4.10.2). Replace `variable_group px,py,z;` by

```
variable_group px;
variable_group py;
variable_group z;
```

By separating the variables in the different groups, we indicate how to group them for the m-homogenization. As a result we have only 182 paths to track. However each new variable group adds another variable to the computations[12] and decreases numerical stability. Therefore we always have to consider the problem of reducing the number of paths versus increasing the number of variables.

A key point to note is that the number of solutions is much smaller than the Bezout number. The Bezout number of the system (4.1,4.2,4.3) is 600 but there are only 62 finite solutions. This fact may be surprising in the light of the theorem that says that systems such as (4.1,4.2,4.3) would generically have 600 finite complex solutions, see Garcia and Li (1980). However, (4.1,4.2,4.3) is not similar to the generic system since most monomials of degree 6 are missing from (4.1), and most monomials of degree 10 are missing from (4.2,4.3). The absence of so many monomials often implies a far smaller number of finite complex solutions. For many games this fact makes our strategy much more practical than we would initially think.

Another key point to note is that the all-solution methods can only be applied to polynomial systems, that is, when all variables have exponents with non-negative integer values. We cannot apply such a method to equations with irrational exponents. Such systems would occur in the Bertrand game, for example, if an elasticity were an irrational number such as π. In addition, an important prerequisite for Bertini to be able to trace all paths is that the Bezout number remains relatively small. The conversion of systems with rational exponents with large denominators to proper polynomial systems, however, leads to polynomial systems with large exponents. For example, the conversion of equations with exponents such as 54321/10000 will lead to very difficult systems that require tracing a huge number of paths. In addition, such polynomial terms with very large exponents will likely generate serious and perhaps fatal numerical difficulties for the path tracker. Therefore, we face some practical constraints on the size of the rational exponents appearing in our economic models.

[12] We repeatedly solve square systems of linear equations. Bertini performs this task with conventional methods with a complexity of roughly $\frac{1}{3}n^3$, where n is the number of variables. Thus increasing the number of variables by m adds $\frac{1}{3}(m^3 + 3m^2n + 3n^2m)$ to the complexity for each iteration of Newton's method.

4.5.2 Application of parameter continuation

To demonstrate parameter continuation, which we describe in Appendix 4.10.3, we choose n as the parameter and vary it from 2700 to 1000. Note that in Bertini the homotopy parameter goes from 1 to 0. So to do this we define a homotopy just between those two values

$$n = 2700t + (0.22334546453233 + 0.974739352i)t(1-t) + 1000(1-t).$$

Thus for $t = 1$ we have $n = 2700$ and if $t = 0$ then $n = 1000$. The complex number in the equation is the application of the gamma trick. We also have to provide the solutions for our start system. We already solved this system. We just rename Bertini's output file `finite_solutions` to `start` which now provides Bertini with the starting points for the homotopy paths. In addition, we must alter the input file as follows.

```
CONFIG
USERHOMOTOPY: 1;
MPTYPE: 0;
END;
INPUT
variable px,py,z;
function f1, f2, f3;
pathvariable t;
parameter n;
n = t*2700 + (0.22334546453233 + 0.974739352*I)*t*(1-t)+(1-t)*1000;
f1 = -(px^2)-py^2+z^2*px^2*py^2;
f2 = -(n)+n*px+3*n*z^2*px^2-2*n*z^2*px^3+51*z^3*px^6-2*z^3*px^7;
f3 = -(n)+n*py+3*n*z^2*py^2-2*n*z^2*py^3+51*z^3*py^6-2*z^3*py^7;
END;
```

If we run Bertini we obtain 14 real and 48 complex solutions. Note that the number of real solutions has dropped by 4. Thus if we had not used the gamma trick some of our paths would have failed. There are only five positive real solutions. The first three solutions in Table 4.2 fail the second-order conditions for at least one firm. The fourth solution fails the global-optimality test. Only the last solution in Table 4.2 is an equilibrium for the Bertrand game for $n = 1000$.

p_x	3.333	2.247	3.613	2.045	24.689
p_y	2.247	3.333	3.613	2.045	24.689

Table 4.2: Real, positive solutions for $n = 1000$

4.5.3 The manifold of real positive solutions

The parameter continuation approach allows us to compare solutions and thus equilibria for two different (vectors of) parameter values q_0 and q_1 of our economic model. Ideally we would like to push our analysis even further and, in fact, compute the equilibrium manifold for all convex combinations $sq_1 + (1-s)q_0$ with $s \in [0, 1]$.

Observe that Theorem 4.8 in Appendix 4.10.3 requires a path between q_0 and q_1 of the form

$$\varphi(s) = e^{i\gamma} s(s-1) + sq_1 + (1-s)q_0$$

with a random $\gamma \in [0, 2\pi)$. Note that for real values q_0 and q_1 the path $\varphi(s)$ is not real and so all solutions to $F(z, \varphi(s)) = 0$ are economically meaningless for $s \in (0, 1)$. This problem would not occur if we could drop the first term of $\varphi(s)$ and instead use the convex combination

$$\tilde{\varphi}(s) = sq_1 + (1-s)q_0$$

in the definition of the parameter continuation homotopy. Now an examination of the real solutions to $F(z, \tilde{\varphi}(s)) = 0$ would provide us with the equilibrium manifold for all $\tilde{\varphi}(s)$ with $s \in [0, 1]$. Unfortunately, such an approach does not always work. As we have seen in the previous section, while the number of isolated finite solutions remains constant with probability one, the number of real solutions may change. A parameter continuation homotopy with $\tilde{\varphi}(s)$ does not allow for this change.

To illustrate the described difficulty, we examine two parameter continuation homotopies in Bertini. We vary the parameter n first from 2700 to 3400 and then from 2700 to 500. Figure 4.5 displays the positive real solutions as a function of n over the entire range from 500 to 3400. For a clear view of the different portions of the manifold we separate it into two graphs.

For the first homotopy the number of positive real, other real, and complex (within nonzero imaginary part) solutions does not change as n is increased from 2700 to 3400. Therefore, in this case, the described approach to obtain the manifold of (positive) real solutions encounters no difficulties. Things are quite different for the second homotopy when n is decreased from 2700 to 500. As n approaches 1188.6 the paths for the two largest production quantities converge and then, when n is decreased further, move into complex space. The same is true for two paths in the lower graph of Figure 4.5. Bertini reports an error message for all four paths and stops tracking them. At $n = 1188.6$ the number of real solutions changes from 18 to 14, while the number of (truly) complex solutions with nonzero imaginary part increases from 44 to 48. A similar change in the number of real and complex solutions occurs for $n = 813.8$.

To determine the equilibrium manifold, we need to check the second-order and global optimality conditions for all positive real solutions. Doing so yields the equilibrium manifold in Figure 4.1 in Section 4.2.

In sum, we observe that a complete characterization of the equilibrium manifold is not a simple exercise. When we employ the parameter continuation approach with a path of parameters in real space then we have to allow for the possibility of path-tracking failures whenever the number of real and complex solution changes. The determination of the

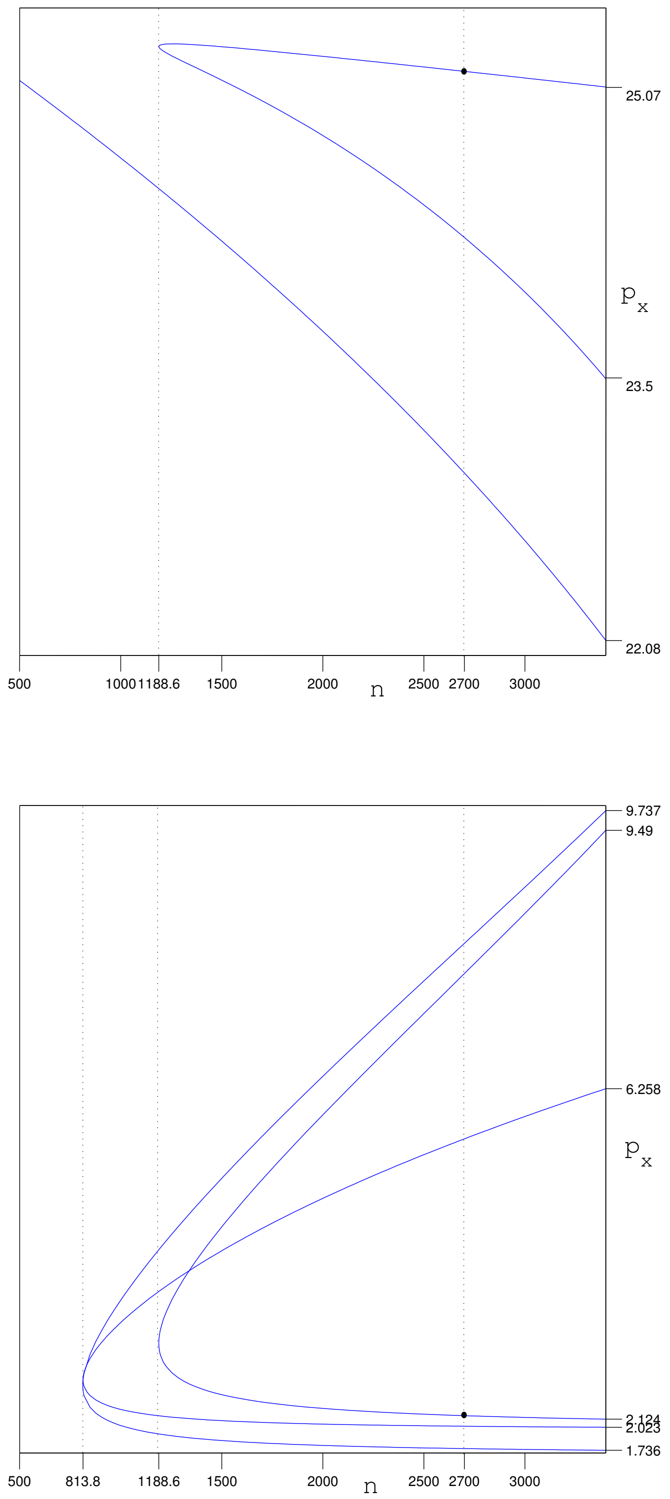

Figure 4.5: Real positive solutions as a function of n

entire manifold of positive real solutions may, therefore, require numerous homotopy runs. Despite these difficulties we believe that the parameter continuation approach is a very helpful tool for the examination of equilibrium manifolds.

We can continue our analysis for larger values of the market size n. Figure 4.6 shows the unique equilibrium price $p_x = p_y$ for $3400 \le n \le 10000$. The market of type 2 customers is so large that both firms choose a mass market strategy and charge a low price. While the number of equilibria remains constant for large values of n, the number

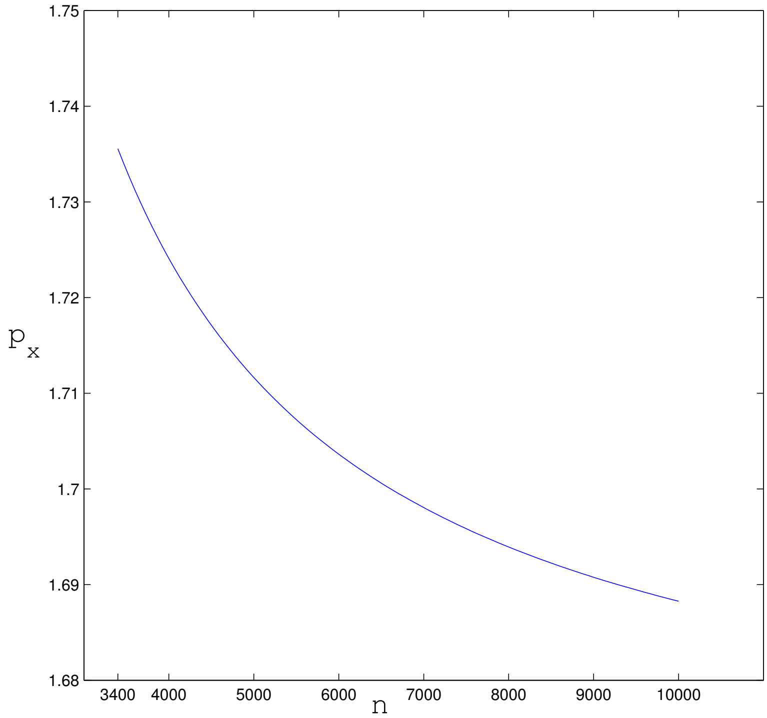

Figure 4.6: Unique equilibrium for large values of n

of real solutions changes twice in the examined region. Recall that there are 18 real solutions for $n = 3400$. This number decreases to 16 at about $n = 5104.5$ and further to 14 at about $n = 5140.8$.

4.6 Equilibrium equations for dynamic stochastic games

In this section we first briefly describe a general set-up of dynamic stochastic games. Such games date back to Shapley, 1953, for a textbook treatment see Filar and Vrieze, 1997. Subsequently we explain how Markov-perfect equilibria (MPE) in these games can be characterized by nonlinear systems of equations.

4.6.1 Dynamic stochastic games: general formulation

We consider discrete-time infinite-horizon dynamic stochastic games of complete information with N players. In period $t = 0, 1, 2, \ldots$, player $i \in \{1, 2, \ldots, N\}$ is characterized by its state $\omega_{i,t} \in \Omega_i$. The set of possible states, Ω_i, is finite and without loss of generality we thus define $\Omega_i = \{1, 2, \ldots, \hat{\omega}_i\}$ for some number $\hat{\omega}_i \in \mathbb{N}$. The product $\Omega = \prod_{i=1}^{N} \Omega_i$ is the state space of the game; the vector $\omega_t = (\omega_{1,t}, \omega_{2,t}, \ldots, \omega_{N,t}) \in \Omega$ denotes the state of the game in period t.

Players choose actions simultaneously. Player i's action in period t is $a_{i,t} \in \mathcal{A}_i(\omega_t)$, where $\mathcal{A}_i(\omega_t)$ is the set of feasible actions for player i in state ω_t. In many economic applications of dynamic stochastic games $\mathcal{A}_i(\omega_t)$ is a convex subset of $\mathbb{R}^M$, $M \in \mathbb{N}$, and we adopt this assumption here to employ standard first-order conditions in the analysis. We denote the collection of all players' actions in period t by $a_t = (a_{1,t}, a_{2,t}, \ldots, a_{N,t})$ and the collection of all but player i's actions by $a_{-i,t} = (a_{1,t}, \ldots, a_{i-1,t}, a_{i+1,t}, \ldots, a_{N,t})$.

Players' actions affect the probabilities of state-to-state transitions. If the state in period t is ω_t and the players choose actions a_t, then the probability that the state in period $t+1$ is ω^+ is $\Pr(\omega^+ | a_t; \omega_t)$. In many applications the transition probabilities for player i's state are assumed to depend on player i's actions only and to be independent of other players' actions and transitions in their states. We follow this custom and make the same assumption. Denoting the transition probability for player i's state by $\Pr_i\left(\left(\omega^+\right)_i | a_{i,t}; \omega_{i,t}\right)$, the transition probability for the state of the game therefore satisfies

$$\Pr\left(\omega^+ | a_t; \omega_t\right) = \prod_{i=1}^{N} \Pr_i\left(\left(\omega^+\right)_i | a_{i,t}; \omega_{i,t}\right).$$

If the state of the game is ω_t in period t and the players choose actions a_t then player i receives a payoff $\pi_i(a_t, \omega_t)$. Players discount future payoffs using a discount factor $\beta \in (0, 1)$. The objective of player i is to maximize the expected net present value of all its future cash flows,

$$\mathrm{E}\left\{\sum_{t=0}^{\infty} \beta^t \pi_i(a_t; \omega_t)\right\}.$$

Economic applications of dynamic stochastic games typically rely on the equilibrium notion of a pure strategy Markov-perfect equilibrium (MPE). That is, attention is restricted to pure equilibrium strategies that depend only on the current state and are independent of the history of the game. We can thus drop the time subscript. Player i's strategy A_i maps each state $\omega \in \Omega$ into its set of feasible actions $\mathcal{A}_i(\omega)$. The actions of all other players in state ω prescribed by their respective strategies are denoted $A_{-i}(\omega) = (A_1(\omega), \ldots, A_{i-1}(\omega), A_{i+1}(\omega), \ldots, A_N(\omega))$. Finally, we denote by $V_i(\omega)$ the expected net present value of future cash flows to player i if the current state is ω. The mapping $V_i : \Omega \to \mathbb{R}$ is player i's value function.

For given Markovian strategies A_{-i} of all other players, player i faces a discounted infinite-horizon dynamic programming problem. As Doraszelski and Judd, 2012 point out, Bellman's principle of optimality implies that the optimal solution for this dynamic programming problem is again a Markovian strategy A_i. That is, a Markov-perfect equilibrium remains subgame perfect even without the restriction to Markovian strategies. The Bellman equation for player i's dynamic programming problem is

$$V_i(\omega) = \max_{a \in A_i(\omega)} \left\{ \pi_i(a, A_{-i}(\omega); \omega) + \beta \mathrm{E}\left[V_i(\omega^+) | a, A_{-i}(\omega); \omega \right] \right\} \tag{4.14}$$

where the expectation operator $\mathrm{E}[\ \cdot\ |\ \cdot\ ]$ determines the conditional expectation of the player's continuation values $V_i(\omega^+)$ which are a function of next period's state ω^+, which in turn depends on the players current action a, the other players' actions $A_{-i}(\omega)$, and the current state ω. We denote by

$$h_i(a, A_{-i}(\omega); \omega; V_i) = \pi_i(a, A_{-i}(\omega); \omega) + \beta \mathrm{E}\left[V_i(\omega^+) | a, A_{-i}(\omega); \omega \right]$$

the maximand in the Bellman equation. Player i's optimal action $A_i(\omega) \in \mathcal{A}_i(\omega) \subset \mathbb{R}^M$ in state ω is given by

$$A_i(\omega) = \arg \max_{a \in \mathcal{A}_i(\omega)} h_i(a, A_{-i}(\omega); \omega; V_i). \tag{4.15}$$

For each player $i = 1, 2, \ldots, N$, equations (4.14) and (4.15) yield optimality conditions on the unknowns $V_i(\omega)$ and $A_i(\omega)$ in each state $\omega \in \Omega$. A Markov-perfect equilibrium (in pure strategies) is now a simultaneous solution to equations (4.14) and (4.15) for all players and states.

4.6.2 Equilibrium conditions

Doraszelski and Satterthwaite, 2010 develop sufficient conditions for the existence of a Markov-perfect equilibrium for a class of dynamic stochastic games. A slightly modified version of the existence result in their Proposition 2 holds in the described model under the assumptions that both actions and payoffs are bounded and the maximand function $h_i(\cdot, A_{-i}(\omega); \omega; V_i)$ is strictly concave for all $\omega \in \Omega$, other players ' strategies A_{-i}, and value functions V_i satisfying the Bellman equation. Under these assumptions the maximand $h_i(\cdot, A_{-i}(\omega); \omega; V_i)$ has a unique maximizer $A_i(\omega)$. This unique maximizer could lie on the boundary of or be an interior solution of the set of feasible actions $\mathcal{A}_i(\omega)$. (As V_i changes so will the maximizer and there could be several consistent solutions and thus equilibria.)

For the purpose of this paper we restrict attention to models that satisfy two further assumptions which are frequently made in economic applications. First, the function $h_i(\cdot, A_{-i}(\omega); \omega; V_i)$ is continuously differentiable. Second, we assume that the maximizer in equation (4.15) is always an interior solution. Under these assumptions we can equivalently characterize players' optimality conditions (4.14) and (4.15) by a set of necessary and sufficient first-order conditions.

$$0 = \frac{\partial}{\partial a} \left\{ \pi_i(a, A_{-i}(\omega); \omega) + \beta \mathrm{E}\left[V_i(\omega_+) | a, A_{-i}(\omega); \omega \right] \right\} \Big|_{a = A_i(\omega)} \tag{4.16}$$

$$V_i(\omega) = \pi_i(a, A_{-i}(\omega); \omega) + \beta \mathrm{E}\left[V_i(\omega_+) | a, A_{-i}(\omega); \omega \right] \Big|_{a = A_i(\omega)} \tag{4.17}$$

Thus we have $M + 1$ equations for each state $\omega \in \Omega$ and for each player $i = 1, 2, \ldots, N$. Any simultaneous solution of pure strategies $A_1(\omega), \ldots, A_N(\omega)$ and values $V_1(\omega), \ldots, V_N(\omega)$ for all states $\omega \in \Omega$ yields an MPE.

If the payoff functions π_i and the probability functions $Pr(\omega^+|\cdot;\omega)$ are rational functions then the nonlinear equilibrium equations can be transformed into a polynomial system of equations. In the next two sections we examine two economic models that satisfy these assumptions.

4.7 Learning curve

In many industries the marginal cost of production decreases with the cumulative output, this effect is often called learning-by-doing. The impact of learning-by-doing on market equilibrium has been studied in the industrial organization literature for decades. Early work in this area includes Spence, 1981 and Fudenberg and Tirole, 1983b. Besanko et al., 2010 analyze learning-by-doing and organizational forgetting within the framework of Ericson and Pakes, 1995.

In this section we examine a basic learning-by-doing model in the Ericson and Pakes, 1995 framework. Although the functional forms for the price functions and transition probabilities are not polynomial we can derive a system of polynomial equations such that all positive real solutions of this system are Markov-perfect equilibria.

4.7.1 A learning-by-doing model

There are $N = 2$ firms and two goods. Firm i produces good i, $i = 1, 2$. The output of firm i is denoted by q_i which is the firm's only action. (In the language of our general formulation, $a_i = q_i$.) The state variable ω_i for firm i is a parameter in the firm's production cost function $c_i(q_i; \omega_i)$. In our numerical example we assume $c_i(q_i; \omega_i) = \omega_i q_i$ implying that the state ω_i is firm i's unit cost of production. For simplicity we assume w.l.o.g. that $\omega_i \in \Omega_i = \{1, 2, \ldots, \hat{\omega}_i\}$.

In each period the two firms engage in Cournot competition. Customers' utility function over the two goods (and money M) is

$$u(q_1, q_2) = w \frac{\gamma}{\gamma - 1} \left(q_1^{\frac{\sigma-1}{\sigma}} + q_2^{\frac{\sigma-1}{\sigma}} \right)^{\frac{(\gamma-1)\sigma}{\gamma(\sigma-1)}} + M$$

where σ is the elasticity of substitution between goods 1 and 2, γ is the elasticity of demand for the composite good $\left(q_1^{\frac{\sigma-1}{\sigma}} + q_2^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{(\sigma-1)}}$, and w is a weighting factor. The resulting market clearing prices for the two goods are then

$$P_1(q_1, q_2) = w q_1^{-\frac{1}{\sigma}} \left(q_1^{\frac{\sigma-1}{\sigma}} + q_2^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\gamma-\sigma}{\gamma(\sigma-1)}}, \quad P_2(q_1, q_2) = w q_2^{-\frac{1}{\sigma}} \left(q_1^{\frac{\sigma-1}{\sigma}} + q_2^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\gamma-\sigma}{\gamma(\sigma-1)}},$$

where $P_i(q_1, q_2) = \frac{\partial}{\partial q_i} u(q_1, q_2)$ denotes the price of good i if sales of the two goods are (q_1, q_2). And so, if the two firms produce the quantities (q_1, q_2) in state $\omega = (\omega_1, \omega_2)$, their resulting payoffs are

$$\pi_i(q_i, q_{-i}; \omega) = P_i(q_1, q_2)q_i - c_i(q_i; \omega_i). \tag{4.18}$$

Note that in this model firm i's payoff does not explicitly depend on the other firm's state but only implicitly via the other firm's production quantity.

The dynamic aspect of the model arises from changes in the unit cost ω_i. Through learning-by-doing the firms can reduce their unit cost. In our numerical example we use the popular functional form (see Pakes and McGuire, 1994, Borkovsky et al., 2012, and many other papers) for the transition probabilities

$$\Pr_i[\omega_i - 1 | q_i; \omega_i] = \frac{Fq_i}{1 + Fq_i}, \quad \Pr_i[\omega_i | q_i; \omega_i] = \frac{1}{1 + Fq_i}, \quad 0 \text{ otherwise} \tag{4.19}$$

with some constant $F > 0$ for $\omega_i \geq 2$. The lowest-cost state $\omega_i = 1$ is an absorbing state. Note that outside the absorbing state the higher a firm's production quantity the higher its probability to move to the next lower cost state. We assume that the transition probability functions are independent across firms.

Substituting the expressions (4.18) and (4.19) into the equilibrium equations (4.16) and (4.17) yields a system of equilibrium equations for the learning-by-doing model. This system has 4 equations for each state $\omega = (\omega_1, \omega_2)$ and thus a total of $4|\hat{\omega}_1||\hat{\omega}_2|$ equations and unknowns.

Solving the system of equations is greatly simplified by the observation that the nature of the transitions in this model induces a partial order on the state space Ω. The unit cost ω_i can only decrease but never increase during the course of the game. Instead of solving one large system of equations we can successively solve systems of 4 equations state by state. For the lowest-cost state $(1, 1)$ we only need to find the static Cournot equilibrium and calculate the values $V_i(1, 1)$. Next we can successively solve the systems for the states $(\omega_1, 1)$ with $\omega_1 = 2, 3, \ldots, \hat{\omega}_1$ and for the states $(1, \omega_2)$ with $\omega_2 = 2, 3, \ldots, \hat{\omega}_2$. Next we can do the same for all $(\omega_1, 2)$ with $\omega_1 = 2, 3, \ldots, \hat{\omega}_1$, for all nodes $(2, \omega_2)$ with $\omega_2 = 3, \ldots, \hat{\omega}_1$ and so on. For symmetric games we can further reduce the workload. We only need to solve system of equations for the states (ω_1, ω_2) with $\omega_2 \leq \omega_1$, that is, for $(1, 1)$, $(\omega_1, 2)$ for $\omega_1 = 2, 3, \ldots, \hat{\omega}_1$, $(\omega_1, 3)$ for $\omega_1 = 3, \ldots, \hat{\omega}_1$, and so on.

4.7.2 Solving the equilibrium equations with Bertini

We compute Markov-perfect equilibria for the learning-by-doing game for the following parameter values. We consider a utility function with $\sigma = 2$, $\gamma = 3/2$, and $w = 100/3$. The parameter for the transition probability function is $F = 1/5$. The firms use the discount factor $\beta = 0.95$. We only examine symmetric cases with $\Omega_1 = \Omega_2$.

Similar to the static game in Section 4.5, the equilibrium equations in this model contain fractions and radical terms. The transformation of the equations into polynomial form forces us to introduce auxiliary variables Q_1, Q_2, Q_3 that are defined as follows,

$$Q_1^2 = q_1, \quad Q_2^2 = q_2, \quad Q_3^2 = Q_1 + Q_2.$$

The introduction of these new variables enables us to eliminate the value function terms $V_i(q_1, q_2)$ of both firms. For each state (ω_1, ω_2) we obtain a system of five equations in the five unknowns q_1, q_2, Q_1, Q_2, Q_3. There is a multiple root at 0. To remove it we add another variable t and a normalization equation $tQ_1 - 1 = 0$, thereby obtaining a system with six variables and six equations.

We solve four different types of polynomial systems. First, we solve the system of the absorbing state $(1, 1)$. The monomials with the highest degrees of the six equations are

$$tQ_1,\ Q_3^3,\ Q_1^2,\ Q_2^2,\ -Q_1Q_3(Q_1+Q_2),\ -Q_2Q_3(Q_1+Q_2),$$

respectively, resulting in a Bezout number of $2^3 \cdot 3^3 = 216$. Using m-homogeneity the number of paths to track reduces to 44. Bertini tracks these 44 paths in just under 4 seconds.

Next we solve the equations for the states $(1, \omega_2)$ for $\omega_2 \geq 2$. The highest degree terms of the six equations are

$$tQ_1,\ Q_3^3,\ Q_1^2,\ Q_2^2,\ -Q_1Q_3(Q_1+Q_2),\ (9F^2\omega_2)Q_1Q_2Q_3q_2^2 + (9F^2\omega_2)Q_2^2Q_3q_2^2,$$

respectively, resulting in a Bezout number of $2^3 \cdot 3^2 \cdot 5 = 360$. Thanks to m-homogeneity we need to track 140 paths and this takes us with Bertini about 1 minute for each ω_2.

Then we solve the equations for state $(2, 2)$, where the highest-degree terms are

$$tQ_1,\ Q_3^3,\ Q_1^2,\ Q_2^2,\ (9F^4\omega_1)Q_1^2Q_3q_1^2q_2^2 + (9F^4\omega_1)Q_1Q_2Q_3q_1^2q_2^2,$$
$$(9F^4\omega_2)Q_1Q_2Q_3q_1^2q_2^2 + (9F^4\omega_2)Q_2^2Q_3q_1^2q_2^2.$$

So the Bezout number is $2^3 \cdot 3 \cdot 7^2 = 1176$. Exploiting m-homogeneity we have to track 364 paths which takes about 5 minutes. There are 152 real and complex (finite) solutions.

For the remaining states we can now use parameter continuation since the degree structure of the systems is identical to that of the equations for state $(2, 2)$. The Bezout number remains the same as for state $(2, 2)$, but now we only have to track 152 paths since that was the number of solutions to the system at $(2, 2)$. (To check whether 152 is indeed the maximal number k of isolated finite solutions as in Theorem 4.8 we solve a few systems with randomly chosen coefficients but the same degree structure. In all cases there are 152 isolated finite solutions.) Tracking these 152 paths takes about 25 seconds for each state. Again we observe that tracking paths ending at finite solutions takes much less time than tracking paths that end at points at infinity. The reason is again that some of the solutions at infinity lie within continua of solutions and thus cause numerical difficulties.

We solved instances of the described learning-by-doing model with many states for each firm. We wrote a C++ script that solved the problem by backwards induction by calling Bertini at each state.[13] To keep the presentation of the results manageable we report here the results for a symmetric game with $\hat{\omega}_1 = 5$. In all our systems there was a unique real positive solution for all variables. Therefore, we found a unique Markov-perfect equilibrium for the learning-by-doing model. Table 4.3 reports the production

[13] The script is available on http://www.business.uzh.ch/professorships/qba/publications/Software.html.

quantities q_1 and the values of the value function V_1 of firm 1. For example, in state $(\omega_1, \omega_2) = (3, 4)$ firm 1 produces $q_1 = 11.385$ and the game has a value of $V_1 = 982$ for the firm. By symmetry the corresponding values for firm 2 are $(q_2, V_2) = (8.620, 913)$.

$\omega_1 \setminus \omega_2$	5		4		3		2		1	
5	7.202	874	7.108	861	7.009	851	6.889	843	6.626	838
4	8.850	939	8.748	925	8.620	913	8.464	905	8.137	899
3	11.475	996	11.385	982	11.233	969	11.016	959	10.573	953
2	16.921	1042	16.840	1027	16.699	1014	16.401	1003	15.714	997
1	38.228	1072	38.171	1057	38.056	1043	37.773	1032	36.600	1025

Table 4.3: Production quantities q_1 and value function V_1 of firm 1

Table 4.4 reports running times on a laptop (Intel Core 2 Duo T9550 with 2.66 GHz and 4GB RAM) for the learning-by-doing model. The running times grow approximately

$\hat{\omega}_1 = \hat{\omega}_2$	3	5	7	10
time (sec)	477	745	1359	2852

Table 4.4: Running Times

linearly in the number of states $\hat{\omega}_1 \times \hat{\omega}_2$ and so we could easily solve games with many more states per firm.

4.8 Cost-reducing investment and depreciation

In models of cost-reducing investment, spending on investment reduces future production cost, see, for example, Flaherty, 1980 and Spence, 1984. In models of irreversible investment, current investment spending increases future production capacity, see Fudenberg and Tirole, 1983a. Besanko and Doraszelski, 2004 presents a model with both capacity investments and depreciation within the Ericson and Pakes, 1995 framework. Depreciation tends to offset investment. In this section we describe a stochastic dynamic game model in which the marginal cost of production may decrease through investment or increase through depreciation.

4.8.1 A cost-reducing investment model

The model of Cournot competition is the same as in the learning-by-doing model with the only exception that a firm's production quantity does not affect its unit cost. The dynamic aspect of the model arises again from changes in the unit cost ω_i. Both increases and decreases of the unit cost are possible. Firms may be hit by a depreciation shock resulting in a cost increase but they can also make a cost-reducing investment. A depreciation shock increases the unit cost from ω_i to $\omega_i + 1$ and has probability $\delta > 0$. If firm i makes a cost-reducing investment y_i at a cost $cr_i(y_i)$ then it achieves a probabilistic reduction of its cost state. In our numerical examples we assume a quadratic investment cost function, $cr_i(y) = D_i y^2$. Total per-period payoff is then the difference of the Cournot profit and the investment cost,

$$\pi_i(q_i, y_i, q_{-i}, y_{-i}; \omega) \;=\; \pi_i^C(q_i, q_{-i}; \omega) - cr_i(y_i) \;=\; P_i(q_1, q_2)q_i - c_i(q_i; \omega_i) - D_i y_i^2.$$

We assume a transition function of the form (4.19) with the investment level y_i replacing the Cournot quantity. Assuming independence of the depreciation probabilities and the investment transition function then results in the transition probabilities (see also Besanko and Doraszelski, 2004)

$$\Pr_i[\omega_i - 1|y_i; \omega_i] = \frac{F y_i}{1 + F y_i}(1 - \delta) \qquad 2 \le \omega_i \le \hat{\omega}_i \quad (4.20)$$

$$\Pr_i[\omega_i + 1|y_i; \omega_i] = \frac{1}{1 + F y_i}\delta \qquad 1 \le \omega_i \le \hat{\omega}_i - 1 \quad (4.21)$$

$$\Pr_i[\omega_i|y_i; \omega_i] = 1 - \Pr_i[\omega_i - 1|y_i; \omega_i] - \Pr_i[\omega_i + 1|y_i; \omega_i] \qquad 2 \le \omega_i \le \hat{\omega}_i - 1 \quad (4.22)$$

The remaining transition probabilities are

$$\Pr_i[1|y_i; 1] \;=\; 1 - \Pr_i[2|y_i; 1] \quad (4.23)$$

$$\Pr_i[\hat{\omega}_i|y_i; \hat{\omega}_i] \;=\; 1 - \Pr_i[\hat{\omega}_i - 1|y_i; \hat{\omega}_i] \quad (4.24)$$

Substituting the expressions for payoffs and transition probabilities into the equilibrium equations (4.16) and (4.17) yields a system of equilibrium equations for the model. The static Cournot game played in each period does not affect the transition probabilities and so we can solve the two equations at each state that are derived from differentiating with respect to the production quantities q_1 and q_2 independently from the remaining equations. The remaining system consists of 4 equations for each state $\omega = (\omega_1, \omega_2)$ and thus has a total of $4|\hat{\omega}_1||\hat{\omega}_2|$ equations and unknowns. The degree of each equation is 4.

4.8.2 Solving the equilibrium equations with Bertini

Since the unit cost ω_i may increase or decrease we cannot solve the equations state by state as in the learning-by-doing model. Instead we need to solve a single system of equations[14].

Two states for each firm

We describe the solution of the cost-reducing investment game with depreciation for the following parameter values, $\beta = 0.95, D_1 = D_2 = 1, F = 0.2, \delta = 0.1$. The parameters

[14]We perform all calculations and derive the final system in Mathematica. The Mathematica file is available on http://www.business.uzh.ch/professorships/qba/publications/Software.html.

for the utility functions are again $\sigma = 2$, $\gamma = 3/2$, and $w = 100/3$. Each firm can be in one of two states. We set $\Omega_1 = \Omega_2 = \{1, 5\}$ (in a slight abuse of previous notation).

We first solve the Cournot game for each state. The production quantities of firm 1 are

$$q_1(5,5) = 3.2736,\ q_1(5,1) = 2.4664,\ q_1(1,5) = 38.224,\ q_1(1,1) = 36.600.$$

For this model with $2 \times 2 = 4$ states there are 16 equations and variables. The resulting Bezout number is $4^{16} = 4,294,967,296$. By utilizing symmetry we simplified our problem to 8 equations and variables with a total Bezout number of $4^8 = 65,536$. Utilizing m-homogeneity we reduce the number of paths to 3328. It took us 1 hour 40 minutes to solve this problem. We found a total of 589 finite, i.e. complex and real, solutions that lie in affine space, 44 of which are real. We had no path failures, when using adaptive precision.[15] Only one of those real solutions is economically relevant. The investment levels of firm 1 are

$$y_1(5,5) = 3.306,\ y_1(5,1) = 3.223,\ y_1(1,5) = 0.763,\ y_1(1,1) = 0.736,$$

resulting in the following values of the value function,

$$V_1(5,5) = 816.313,\ V_1(5,1) = 794.329,\ V_1(1,5) = 926.059,\ V_1(1,1) = 895.570.$$

Three states for each firm

We choose $\Omega_1 = \Omega_2 = \{1, 5, 10\}$ and our other parameters as in the two-state case. The production quantities of firm 1 in the additional high-cost states are $q_1(10, 10) = 1.1574$ and

$$q_1(10,5) = 1.0648,\ q_1(10,1) = 0.70015,\ q_1(5,10) = 3.3975,\ q_1(1,10) = 37.915.$$

Solving the system of equilibrium equations for the three-state model now poses significantly more problems than the two-state case. The initial system has 36 equations and unknowns. The Bezout number is $4^{36} \approx 4.72 \cdot 10^{21}$. After exploiting symmetry and using some algebraic operations to simplify some equations we obtain a system that has 21 equations and unknowns. Its Bezout number is $1,528,823,808$. This system, however, is still unsolvable on a single laptop if we use the standard homotopy approach. For this reason we now apply the splitting approach from Appendix 4.10.4. We split the system into two subsystems which are both small enough to be solvable. In our example the first system has $M_1 = 358$ nonsingular solutions. The second system has $M_2 = 4510$ nonsingular solutions. Therefore, if we focus only on the nonsingular solutions we have $358 \times 4510 = 1,614,580$ paths to track when we combine the two subsystems via a parameter continuation homotopy. Note that this is an order of magnitude smaller than

[15] If we do not use adaptive precision we can finish computations in just under 3 minutes. However, then 396 paths fail to converge. Nevertheless we still obtain all finite solutions. Clearly, if we could prove that all equilibria are regular solutions to the polynomial system of equilibrium equations then we could relax the precision parameters in Bertini and thus significantly reduce both the computational effort and running times.

taken the system as a whole. We obtain a unique nonsingular equilibrium, see Table 4.5. The time to solve this on a single core is over a week.[16]

$\omega_1 \setminus \omega_2$	10		5		1	
10	3.42	705.00	3.31	680.31	3.24	663.01
5	3.78	820.36	3.70	789.73	3.62	765.53
1	0.86	945.48	0.83	911.76	0.80	878.89

Table 4.5: Equilibrium investment levels y_1 and value function V_1

4.9 Conclusion

We summarize the paper and discuss the current limitations of all-solution methods.

4.9.1 Summary

This paper describes state-of-the-art techniques for finding all solutions of polynomial systems of equations and illustrates these techniques by computing all equilibria of both static and dynamic games with continuous strategies. The requirement of polynomial equations may, at first, appear very restrictive. In our first application, a static Bertrand pricing game, we show how certain types of non-polynomial equilibrium conditions can be transformed into polynomial equations. We also show how with repeated application of the polynomial techniques we can deal with first-order conditions that are necessary but not sufficient. Finally, this example also depicts the power of the parameter-continuation homotopy approach. This approach greatly reduces the number of homotopy paths that need to be traced and, therefore, increases the size of models that we can analyze. When handled carefully, it even allows us to trace out the equilibrium manifold.

We also apply the all-solution techniques to two stochastic dynamic games of industry competition and check for equilibrium uniqueness. In the first application, a learning-by-doing model of industry competition, the equilibrium system separates into many small systems of equations which can be solved sequentially. As a result we can solve specifications of this model with many states. In our second application, a model with cost-reducing investment and cost-increasing depreciation, such a separation of the equilibrium system is impossible. Solving the resulting equilibrium system requires the tracing of a huge number of paths. On a single laptop we can solve specifications of the model with only a small number of states.

4.9.2 Current limitations and future work

For stochastic dynamic games, the number of equations grows exponentially in the number N of players and polynomially (with degree N) in the number of states. In turn, the

[16] The files are available on http://www.business.uzh.ch/professorships/qba/publications/Software.html.

Bezout number grows exponentially in the number of nonlinear equations. Additionally the degree of the polynomials is essential which limits the parameter choice for the exponents in the utility functions. As a result, the number of paths that an all-solution method must trace grows extremely fast in the size of the economic model. This growth clearly limits the size of problems we can hope to solve.

Modern policy-relevant models quickly generate systems of polynomial equations with thousands of equations. For example, the model in Besanko et al., 2010 has up to 900 states and 1800 equations. Finding all equilibria of models of this size is impossible with the computer power available as of the writing of this paper and it will remain out of reach for the foreseeable future. However, we will likely be able to solve smaller models such as the dynamic model of capacity accumulation of Besanko and Doraszelski, 2004 with at most 100 states within a few years. Progress will come on at least three frontiers. First, computer scientists have yet to optimize the performance of software packages such as Bertini. Second, the all-solution homotopy methods are ideally suited for parallel computations. Our initial experience has been very promising. And so, as soon as the existing software will have been adapted to large parallel computing systems, we will see great progress in the size of the models we can analyze with the methods described in this paper. And third, methodological advances such as the equation splitting approach will also help us to solve larger systems.

4.10 Appendix

4.10.1 Homogenization

The all-solution homotopy method presented in Section 4.3.3 has the unattractive feature that it must follow diverging paths. Homogenization of the polynomials greatly reduces the computational effort to track such paths.

Definition 4.8. The homogenization $\hat{f}_i(z_0, z_1, \ldots, z_n)$ of the polynomial $f_i(z_1, \ldots, z_n)$ of degree d_i is defined by

$$\hat{f}_i(z_0, z_1, \ldots, z_n) = z_0^{d_i} f_i\left(\frac{z_1}{z_0}, \ldots, \frac{z_n}{z_0}\right).$$

Effectively, each term of $\hat{f}_i$ is obtained from multiplying the corresponding term of f_i by the power of z_0 that leads to a new degree of that term of d_i. So, if the term originally had degree d_{ij} then it is multiplied by $z_0^{d_i - d_{ij}}$. Performing this homogenization for each polynomial f_i in the system

$$F(z_1, \ldots, z_n) = 0 \tag{4.25}$$

leads to the transformed system

$$\hat{F}(z_0, z_1, \ldots, z_n) = 0. \tag{4.26}$$

For convenience we use the notation $\hat{z} = (z_0, z_1, \ldots, z_n)$ and write $\hat{F}(\hat{z}) = 0$. By construction all polynomials $\hat{f}_i$, $i = 1, \ldots, n$, are homogeneous and so for any solution $\hat{b}$ of

$\hat{F}(\hat{z}) = 0$ it holds that $\hat{F}(\lambda\hat{b}) = 0$ for any complex scalar $\lambda \in \mathbb{C}$. So, the solutions to system (4.26) are complex lines through the origin in $\mathbb{C}^{n+1}$.

Definition 4.9. The n-dimensional complex projective space CP^n is the set of lines in $\mathbb{C}^{n+1}$ that go through the origin. The space $\mathbb{C}^{n+1}$ is called the affine space.

A point in projective space CP^n corresponds to a line through the origin of the affine space $\mathbb{C}^{n+1}$. Let $[\hat{b}] \in CP^n$ denote a point in CP^n then there is a point $\hat{b} = (\hat{b}_0, \hat{b}_1, \ldots, \hat{b}_n) \in \mathbb{C}^{n+1} \setminus \{0\}$ that determines this line. We denote the line $[\hat{b}]$ by $(\hat{b}_0 : \hat{b}_1 : \ldots : \hat{b}_n)$ to distinguish it from a single point. The notation $(z_0 : z_1 : \ldots : z_n)$ is called the *homogeneous coordinates* of CP^n. Note however that this notation is not unique, we can take any $\lambda\hat{b}$ with $\lambda \in \mathbb{C} \setminus \{0\}$ as a representative. Furthermore $(0 : 0 : \ldots : 0)$ is not a valid point in projective space. Thus for any point $(\hat{b}_0 : \ldots : \hat{b}_n)$ there exists at least one element $\hat{b}_i \neq 0$.

There is a one-to-one relationship between the solutions of system (4.25) in $\mathbb{C}^n$ and the solutions of system (4.26) in $\mathbb{C}^{n+1}$ with $\hat{b}_0 \neq 0$. If b is a solution to (4.25) then the line through $\hat{b} = (1, b)$, that is, $[\hat{b}] \in CP^n$, is a solution to (4.26). For the converse, if $(\hat{b}_0 : \hat{b}_1 : \ldots : \hat{b}_n)$ with $\hat{b}_0 \neq 0$ is a solution to (4.26) then the point $(\frac{\hat{b}_1}{\hat{b}_0}, \ldots, \frac{\hat{b}_n}{\hat{b}_0})$ is a solution of (4.25).

One of the advantages of the homogenized system (4.26) is that it can model "infinite" solutions. If we have a line $\{(\lambda b) \mid \lambda \in \mathbb{C}\} \subset \mathbb{C}^n$, $b \in \mathbb{C}^n \setminus \{0\}$ and look at the corresponding line $\{(1 : \lambda b_1 : \ldots, \lambda b_n) \mid \lambda \in \mathbb{C}\}$ in projective space then for any λ, $(\frac{1}{\lambda} : b_1 : \ldots : b_n)$ is also a valid representation of that point on the projective line. So if $\|\lambda\| \to \infty$ then $\|\frac{1}{\lambda}\| \to 0$ and we are left with the point $(0 : b_1 : \ldots : b_n)$. Note that $\|\lambda\| \to \infty$ in the affine space means $\|\lambda b\| \to \infty$. Thus we traverse the line to "infinity". This observation leads to the following definition.

Definition 4.10. Consider the natural embedding of $\mathbb{C}^n$ with coordinates $(z_1, \ldots, z_n)$ in the projective space CP^n with homogeneous coordinates $(z_0 : \ldots : z_n)$. Then we call points $(0 : b_1 : \ldots : b_n) \in CP^n$ *points at infinity.*

The value $\hat{b}_0 = 0$ for a solution $\hat{b}$ to $\hat{F}$ implies $\hat{f}_i(\hat{b}_0 : \hat{b}_1 : \ldots : \hat{b}_n) = f_i^{(d_i)}(\hat{b}_1, \ldots, \hat{b}_n) = 0$. Therefore the solutions at infinity of $\hat{F}(\hat{z}) = 0$ correspond to the solutions to the system $(f_1^{(d_1)}, \ldots, f_n^{(d_n)}) = 0$. The fact that we now have a representation of solutions at infinity leads to a new version of Bezout's theorem for projective space.

Theorem 4.6 (Bezout's theorem in projective space CP^n)**.** *If system* (4.26) *has only a finite number of solutions in CP^n and if d is the Bezout number of F, then it has exactly d solutions (counting multiplicities) in CP^n.*

If we view the system of equation (4.26) in affine space $\mathbb{C}^{n+1}$ instead of in complex projective space CP^n then it is actually underdetermined because it consists of n equations in $n + 1$ unknowns. For a computer implementation of a homotopy method, however, we need a determinate system of equations. For this purpose we add a simple normalization. Using the described relationship between solutions of the two systems (4.25) and (4.26) we can now introduce a third system to find the solutions of system (4.25). Define a new linear function

$$u(z_0, z_1, \ldots, z_n) = \xi_0 z_0 + \xi_1 z_1 + \ldots + \xi_n z_n$$

with random coefficients $\xi_i \in \mathbb{C}$. (The nongeneric cases are where the normalization line is parallel to a solution "line".) Now define

$$\begin{aligned} \tilde{f}_i(z_0, z_1, \ldots, z_n) &:= \hat{f}_i(z_0, z_1, \ldots, z_n), \qquad\qquad i = 1, \ldots, n, \\ \tilde{f}_0(z_0, z_1, \ldots, z_n) &:= u(z_0, z_1, \ldots, z_n) - 1. \end{aligned} \tag{4.27}$$

The resulting system of equations

$$\tilde{F} = (\tilde{f}_0, \tilde{f}_1, \ldots, \tilde{f}_n) = 0 \tag{4.28}$$

has $n+1$ equations in $n+1$ variables. Note that the system $\tilde{F}(\hat{z})$ has the same total degree d as the system $F(z)$ in the original system of equations (4.25). As a start system we choose

$$\begin{aligned} G_i(z_0, z_1, \ldots, z_n) &= z_i^{d_i} - z_0^{d_i}, \qquad\qquad i = 1, \ldots, n, \\ G_0(z_0, z_1, \ldots, z_n) &= u(z_0, z_1, \ldots, z_n) - 1. \end{aligned} \tag{4.29}$$

We write the resulting system as $G(\hat{z}) = 0$ and define the homotopy

$$H(t, \hat{z}) = t\tilde{F}(\hat{z}) + e^{\gamma i}(1-t)G(\hat{z}) \tag{4.30}$$

for a $\gamma \in [0, 2\pi)$. To illustrate a possible difficulty with this approach we examine the system of equations (4.1,4.2,4.3) that we derived for the Bertrand price game in Section 4.2.2.

Example 4.4. After homogenization of the equilibrium system (4.1,4.2,4.3) in the variables p_x, p_y, and Z with the variable x_0 we obtain the following polynomial equations.

$$\begin{aligned} 0 &= -p_x^2 x_0^4 - p_y^2 x_0^4 + Z^2 p_x^2 p_y^2 \\ 0 &= -2700 x_0^{10} + 2700 p_x x_0^9 + 8100 Z^2 p_x^2 x_0^6 - 5400 Z^2 p_x^3 x_0^5 + 51 Z^3 p_x^6 x_0^1 - 2Z^3 p_x^7 \\ 0 &= -2700 x_0^{10} + 2700 p_y x_0^9 + 8100 Z^2 p_y^2 x_0^6 - 5400 Z^2 p_y^3 x_0^5 + 51 Z^3 p_y^6 x_0^1 - 2Z^3 p_y^7 \end{aligned}$$

The solutions at infinity are those for which $x_0 = 0$. In this case the system simplifies as follows

$$Z^2 p_x^2 p_y^2 = 0, \quad -2Z^3 p_x^7 = 0, \quad -2Z^3 p_y^7 = 0.$$

After setting $Z = 0$ all equations hold for any values of p_x and p_y. There is a continuum of solutions at infinity. Such continua can cause numerical difficulties for the path-following procedure.

The following theorem now states that in spite of the previous example our paths converge to the relevant isolated solutions.

Theorem 4.7. *Let the homotopy H be as in* (4.30) *with Bezout number d. Then the following statements hold for almost all $\gamma \in [0, 2\pi)$:*

1. *The homotopy has d continuous solution paths.*

2. *Each path will either converge to an isolated nonsingular or to a singular[17] solution, i.e. one where the rank of the Jacobian drops.*

3. *If b is an isolated solution with multiplicity m, then there are m paths converging to it.*

4. *Paths are monotonically increasing in t, i.e. the paths do not bend backwards.*

Now we can apply the homotopy H as defined in equation (4.30) and find all solutions of the system (4.28). There will be no diverging paths. From the solutions of (4.28) we easily obtain the solutions of the original system (4.25).

An additional advantage of the above approach lies in the possibility to scale our solutions via u. If a solution component z_i becomes too large, then this will cause numerical problems, e.g. the evaluation of polynomials at such a point becomes rather difficult. Thus if something like this happens we pick a new set of ξ_i. Furthermore we eliminated the special case of infinite paths and we do not have to check whether the length of the path grows too large. Instead every diverging path has become a converging one. So while tracking a path we do not need to check whether the length of the path exceeds a certain bound.

Theoretically we have eliminated the problem of solutions at infinity. Note that the problem of diverging paths still remains in practice. A solution b belongs to a diverging path if $b_0 = 0$. We still need to decide when b_0 becomes zero numerically. Thus there is no absolute certainty if a path converges to a solution at infinity or if the solution is extremely large. However, we are in the convergence zone of Newton's method and can quickly sharpen our solutions to an arbitrary precision.

Remark. *Here we attempt to give some intuition for the problem of infinite paths. Take two lines $L_1 = \{(x_1, x_2) | x_1 + a_{12}x_2 + b_1 = 0\}$ and $L_2 = \{(x_1, x_2) | x_1 + a_{22}x_2 + b_2 = 0\}$ with $a_{12}, a_{22} \in \mathbb{R}$. Then there are three possibilities for $L_1 \cap L_2$. First $L_1 \cap L_2 = L_1$ so $a_{12} = a_{22}$ and $b_1 = b_2$. Secondly $L_1 \cap L_2 = \{p\}$ for some point $p \in \mathbb{R}^2$. Lastly we have $L_1 \cap L_2 = \emptyset$, i.e. the lines are parallel and so $a_{12} = a_{22}$ but $b_1 \neq b_2$. By using projective space we eliminate the last possibility by adding infinity where the two lines can meet. So in projective space the lines are given by the zero sets of the two polynomials $x_1 + a_{12}x_2 + b_1x_0$ and $x_1 + a_{22}x_2 + b_2x_0$. Clearly $(0 : -a_{12} : 1)$ is a common zero for these polynomials if $a_{12} = a_{22}$. So in projective space CP^n, n linear homogeneous polynomials which are not pairwise identical intersect at exactly one point.*

Bezout's theorem generalizes this idea to n polynomials. However the theorem implicitly embeds the system of polynomials in projective space. Therefore we have to consider

[17] This might be an isolated root with multiplicity higher than one, e.g. a double root of the system F, or a non-isolated solution component as in Example 4.4.

the possibility that solutions are at infinity and thus the paths that belong to those diverge. The case that one of those intersection points lies at infinity is equivalent to demanding that $z_0 = 0$. This is clearly a non-generic case. But the systems that interest us are highly non-generic, the reason being that they are sparse. That means for a degree d polynomial in n variables there are $\binom{n+d}{d}$ monomials of degree equal or smaller than d but most of their coefficients are zero which is a non-generic condition. Thus those systems tend to have many solutions at infinity.

4.10.2 m-homogeneous Bezout number

The number of paths d grows rapidly with the degree of individual equations. For many economic models we believe that there are only a few (if not unique) equilibria, that is, our systems have few real solutions and usually even fewer economically meaningful solutions. As a result we may have to follow a large number of paths that do not yield useful solutions. As we have seen in Example 4.4, there may be continua of solutions at infinity which can cause numerical difficulties. Therefore it would be very helpful to reduce the number of paths that must be followed as much as possible.

Two approaches for a reduction in the number of paths exist. The first approach sets the homogenized polynomial system not into CP^n but in a product of m projective spaces $CP^{n_1} \times \ldots \times CP^{n_m}$. For this purpose the set of variables is split into m groups. In the homogenization of the original polynomial F each group of variables receives a separate additional variable, thus this process is called m-homogenization. The resulting bound on the number of solutions, called the m-homogeneous Bezout number, is often much smaller than the original bound and thus leads to the elimination of paths tending to solutions at infinity. In this paper we do not provide details on this approach but only show its impact in our computational examples. We refer the interested reader to Sommese and Wampler, 2005 and the citations therein. The first paper to introduce m-homogeneity appears to be Morgan and Sommese, 1987.

The second approach to reduce the number of paths is the use of parameter continuation homotopies. We believe that this approach is perfectly suited for economic applications.

4.10.3 Parameter continuation homotopy

Economic models typically make use of exogenous parameters such as risk aversion coefficients, price elasticities, cost coefficients, or many other pre-specified constants. Often we do not know the exact values of those parameters and so would like to solve the model for a variety of different parameter values. Clearly solving the model each time "from scratch" will prove impractical whenever the number of solution paths is very large. The parameter continuation homotopy approach enables us to greatly accelerate the repeated solution of an economic model for different parameter values. After solving one instance of the economic model we can construct a homotopy that alters the parameters from their previous to their new values and allows us to track solutions paths from the previous solutions to new solutions. Therefore, the number of paths we need to follow is greatly reduced.

The parameter continuation approach rests on the following theorem which is a special case of a more general result, see (Sommese and Wampler, 2005, Theorem 7.1.1).

Theorem 4.8 (Parameter Continuation)**.** *Let* $F(z,q) = (f_1(z,q), \dots, f_n(z,q))$ *be a system of polynomials in the variables* $z \in \mathbb{C}^n$ *with parameters* $q \in \mathbb{C}^m$*,*

$$F(z,q) : \mathbb{C}^n \times \mathbb{C}^m \to \mathbb{C}^n.$$

Additionally let $q_0 \in \mathbb{C}^m$ *be a point in the parameter space, where* $k = \max_q |\{z \mid F(z,q) = 0;\ \det\left(\frac{\partial F}{\partial z}(z,q_0)\right) \neq 0\}|$ *is the number of nonsingular isolated solutions. For any other set of parameters* q_1 *and a random* $\gamma \in [0, 2\pi)$ *define*

$$\varphi(s) = e^{i\gamma} s(s-1) + sq_1 + (1-s)q_0$$

Then the following statements hold.

1. $k = |\{z \mid F(z,q) = 0;\ \det\left(\frac{\partial F}{\partial z}(z,q)\right) \neq 0\}|$ *for almost all* $q \in \mathbb{C}^m$*.*

2. *The homotopy* $F(z, \varphi(s)) = 0$ *has* k *nonsingular solution paths for almost all* $\gamma \in [0, 2\pi)$*.*

3. *All solution paths converge to all isolated nonsingular solutions of* $F(z, \varphi(1)) = 0$ *for almost all* $\gamma \in [0, 2\pi)$*.*

The theorem has an immediate practical implication. Suppose we already solved the system $F(z, q_0) = 0$ for some parameter vector q_0. Under the assumption that this system has the maximal number k of locally isolated solutions across all parameter values, we can use this system as a start system for solving the system $F(z, q_1) = 0$ for another parameter vector q_1. The number of paths that need to be tracked is k instead of the Bezout number d or some m-homogeneous Bezout number. In our applications k is much smaller (sometimes orders of magnitude smaller) than these upper bounds. As a result the parameter continuation homotopy drastically reduces the number of paths that we must track. More importantly, no path ends at a solution at infinity for almost all $q_1 \in \mathbb{C}^n$. As we observe in our examples, exactly these solutions often create numerical problems for the path-tracking software, in particular if there are continua of solutions at infinity as in Example 4.4. And due to those numerical difficulties the running times for tracking these paths is often significantly larger than for tracking paths that end at finite solutions. In sum, we believe that the parameter continuation homotopy approach is of great importance for finding all equilibria of economic models.

A statement similar to that of Theorem 4.8 holds if we regard isolated solutions of some fixed multiplicity. But we then have to track paths which have the same multiplicity. Tracking such paths requires a lot more computational effort than non-singular paths. The homotopy continuation software Bertini enables the user to track such paths since it allows for user-defined parameter continuation homotopies.

4.10.4 A splitting approach for solving larger systems

In our application of the all-solutions methods to dynamic stochastic games we quickly run into problems that are too large to be solved on a single computer. We now briefly describe an approach that enables us to increase the size of problems we can solve.

A splitting approach[18] breaks the square system

$$F(z_1, z_2, \ldots, z_n) = (f_1, f_2, \ldots, f_n)\,(z_1, z_2, \ldots, z_n) = 0$$

of polynomial equations into two sub-systems $F_1 = (f_1, \ldots, f_p)$ and $F_2 = (f_{p+1}, \ldots, f_n)$. Similarly, the variables are grouped

$$(z_1, z_2, \ldots, z_n) = (x, y) = (x_1, \ldots, x_p, y_1, \ldots y_{n-p}).$$

Thus, we can write the entire system as follows,

$$\begin{aligned} F_1(x_1, \ldots, x_p, y_1, \ldots y_{n-p}) &= (f_1, \ldots f_p)(x_1, \ldots, x_n, y_1, \ldots y_{n-p}) = 0 \\ F_2(y_1, \ldots, y_{n-p}, x_1, \ldots x_p) &= (f_{p+1}, \ldots f_n)(y_1, \ldots, y_{n-p}, x_1, \ldots x_p) = 0. \end{aligned}$$

Clearly, F_1 and F_2 are not square systems of polynomial equations. We now solve the systems

$$\begin{aligned} F_1(x_1, \ldots, x_p, y_1, \ldots y_{n-p}) &= 0 \\ y_i &= a_i, \quad i = 1, \ldots, n-p, \end{aligned}$$

and

$$\begin{aligned} F_2(y_1, \ldots, y_{n-p}, x_1, \ldots x_p) &= 0 \\ x_j &= b_j, \quad j = 1, \ldots, p, \end{aligned}$$

where $a \in \mathbb{C}^{n-p}$ and $b \in \mathbb{C}^p$ are random complex numbers. Each of these two new square systems has a smaller (m-homogeneous) Bezout number than the original system.

Now suppose that we obtain finite solution sets M_1 and M_2 for each of the two systems, respectively. Any pair $(x^*, a, y^*, b) \in M_1 \times M_2$ is a solution to the following square system of polynomial equations in the unknowns $x_1, \ldots, x_p$, $y_1, \ldots, y_{n-p}$, $r_1, \ldots, r_{n-p}$, and $s_1, \ldots, s_p$,

$$\begin{aligned} F_1(x_1, \ldots, x_p, r_1, \ldots r_{n-p}) &= 0 \\ r_i - a_i &= 0, \quad i = 1, \ldots, n-p, \\ F_2(y_1, \ldots, y_{n-p}, s_1, \ldots s_p) &= 0 \\ s_j - b_j &= 0, \quad j = 1, \ldots, p. \end{aligned}$$

[18] We thank Jonathan Hauenstein for suggesting this method to us.

This system is now the start system for the following parameter continuation homotopy, where r and s are the parameters,

$$\begin{aligned} F_1(x_1, \ldots, x_p, r_1, \ldots r_{n-p}) &= 0 \\ (1-t)(r_i - a_i) + t(r_i - y_i) + (1-t)te^{i\gamma} &= 0, \quad i = 1, \ldots, n-p, \\ F_2(y_1, \ldots, y_{n-p}, s_1, \ldots s_p) &= 0 \\ (1-t)(s_j - b_j) + t(s_j - x_j) + (1-t)te^{i\gamma} &= 0, \quad j = 1, \ldots, p, \end{aligned}$$

with all elements in $M_1 \times M_2$ being start points. Thus there are $|M_1| \cdot |M_2|$ paths to track. Observe that for $t = 1$ we obtain a system that is equivalent to the original system $F(z) = 0$.

To see why this approach works, note that our parameters r and s have been chosen randomly. Statement (1) of Theorem 4.8 states that for almost all choices of those parameters we have the maximal number of isolated roots. Thus all the requirements of the theorem are met and our homotopy converges to all isolated solutions.

A judicious separation of the original equations produces two subsystems with respective Bezout numbers that are roughly equal to the square root of the Bezout number of the original system. This significant reduction in the number of paths to be tracked may make it feasible to solve the subsystems even if the complete system cannot be solved in reasonable time. And if the number of finite solutions of the subsystems is also not too large, then the parameter continuation homotopy will generate all finite solutions of the original system of equations.

In Section 4.8.2 this splitting approach enables us to solve a system of polynomial equations that otherwise would have been too large to be solvable on a single laptop.

5 A polynomial optimization approach to principal agent problems [1][2]

Abstract

This paper presents a new method for the analysis of moral hazard principal-agent problems. The new approach avoids the stringent assumptions regarding the distribution of outcomes made by the classical first-order approach and instead only requires the agent's expected utility to be a rational function of the action. This assumption allows for a reformulation of the agent's utility maximization problem as an equivalent system of equations and inequalities. This reformulation in turn transforms the principal's utility maximization problem into a nonlinear program. Under the additional assumptions that the principal's expected utility is a polynomial and the agent's expected utility is a rational function, the final nonlinear program can be solved to global optimality. The paper also shows that the polynomial optimization approach, unlike the classical approach, extends to principal-agent models with multi-dimensional action sets.

5.1 Introduction

In moral hazard principal-agent problems, the principal maximizes her utility subject to two constraints involving the agent's utility function, a participation constraint and an incentive-compatibility constraint. While the participation constraint is rather straightforward, just imposing a lower bound on the agent's expected utility, the incentive constraint involves an expected utility maximization problem of the agent. As a consequence, principal-agent problems are a type of bi-level optimization problem,[3] a class of optimization problem that is notoriously difficult. The most popular solution approach

[1] (Renner and Schmedders, 2013)

[2] We are indebted to Eleftherios Couzoudis, Johannes Horner, Ken Judd, Diethard Klatte, Felix Kubler, Rida Laraki, George Mailath, Steve Matthews, Walt Pohl, Andy Postlewaite, Gregor Reich, Che-Lin Su, and Rakesh Vohra for helpful discussions on the subject. We thank seminar audiences at the University of Zurich, the 2012 Cowles Summer Conference on Economic Theory, and the 2012 ICE Conference at the Becker Friedman Institute for comments. We are very grateful to Janos Mayer for detailed comments on an earlier version. Karl Schmedders gratefully acknowledges financial support from the Swiss Finance Institute.

[3] The major feature of bi-level optimization problems is that they include two mathematical programs in a single optimization problem. One of the mathematical programs is part of the constraints of the other. This hierarchical relationship is expressed by calling the two programs the lower-level- and the upper-level problem, respectively. In the principal-agent problem, the agent's problem is the lower-level- and the principal's problem is the upper-level problem.

to principal-agent problems is the first-order approach, which replaces the agent's maximization problem by the corresponding first-order condition and leads to an optimization problem for the principal that is more tractable. Unfortunately, this approach requires very restrictive assumptions regarding the probability distribution of outcomes; assumptions which fail to hold in many economic applications.[4] A more widely applicable solution approach for principal-agent problems is obviously desirable.

In this paper, we present a new method for the analysis of moral hazard principal-agent problems. The new approach avoids the stringent assumptions regarding the distribution of outcomes made by the classical first-order approach and instead only requires the agent's expected utility to be a rational function of the action. This assumption enables us to apply ideas from polynomial optimization and to relax the principal-agent problem to a system of polynomial equations and inequalities. So, similar to the first-order approach, we transform the principal's utility maximization problem from a bi-level optimization problem into a nonlinear program. For the special case of univariate effort, we obtain an equivalent reformulation. In the multidimensional case we show that our relaxation converges to the optimal value and that the solution converges to an optimal solution of the original problem.

For principal-agent problems with a one-dimensional effort set for the agent, our assumption that the agent's expected utility function is rational in effort allows us to employ the global optimization approach for rational functions of Jibetean and Klerk, 2006. We transform the agent's expected utility maximization approach into an equivalent semidefinite programming (SDP) problem via a sum of squares representation of the agent's utility function. Semidefinite programs are a special class of convex programming problems which can be solved efficiently both in theory and in practice—See Vandenberghe and Boyd, 1996 and Boyd and Vandenberghe, 2004. We can further reformulate the SDP into a set of inequalities and equations, thereby transforming the principal's bi-level optimization problem into a "normal" nonlinear program. Under the additional assumptions that all objective functions and constraints are rational, the action set is an interval and, if the set of wages is compact, then the resulting problem is a polynomial optimization problem, which is globally solvable. We can then use the methods implemented in GloptiPoly, see Henrion et al., 2009, to find a globally optimal solution to the principal-agent problem. That is, we can obtain a numerical certificate of global optimality.

The first-order approach, a widely used solution method for principal-agent problems, replaces the incentive-compatibility constraint that the agent chooses a utility-maximizing action, by the first-order condition for the agent's utility maximization problem. Mirrlees, 1999 (originally circulated in 1975) was the first to show that this approach is invalid in general (even though it had frequently been applied in the literature.) Under two conditions regarding the probability function of outcomes, the monotone likelihood-ratio condition (MLRC) and the convexity of distribution function

[4]In economic applications, the first-order approach is, then, often just assumed to be applicable. In this case, of course, the resulting conclusions may or may not be valid. Needless to say, this custom is rather unsatisfactory.

condition (CDFC), Rogerson, 1985 proved the validity of the first-order approach. Mirrlees, 1979 had previously surmised that these two assumptions would be sufficient for a valid first-order approach and so these conditions are also known as the Mirrlees-Rogerson conditions. The CDFC is a rather unattractive restriction. Rogerson, 1985 pointed out that the CDFC generally does not hold in the economically intuitive case of a stochastic production function with diminishing returns to scale generating the output. In addition, Jewitt, 1988 observed that most of the standard textbook probability distributions do not satisfy the CFDC.[5] Jewitt, 1988 provided a set of sufficient technical conditions avoiding the CDFC and two sets of conditions for principal-agent models with multiple signals on the agent's effort. Sinclair-Desgagné, 1994 introduced a generalization of the CDFC for an extension of the Mirrless-Rogerson conditions to a first-order approach for multi-signal principal-agent problems. Finally, Conlon, 2009 clarified the relationship between the different sets of sufficient conditions and presented multi-signal generalizations of both the Mirrlees-Rogerson and the Jewitt sufficient conditions for the first-order approach. Despite this progress,[6] all of these sufficient sets of conditions are regarded as highly restrictive—see Conlon, 2009 and Kadan et al., 2011.

Principal-agent models in which the agent's action set is one-dimensional dominate both the literature on the first-order approach and the applied and computational literature—see, for example, (Araujo and Moreira, 2001; Judd and Su, 2005; Armstrong et al., 2010). However, the analysis of linear, multi-task principal-agent models in Holmstrom and Milgrom, 1991 demonstrates that multivariate agent problems exhibit some fundamental differences in comparison to the common one-dimensional models. The theoretical literature that allows the set of actions to be multi-dimensional—for example, Grossman and Hart, 1983, Kadan et al., 2011, and Kadan and Swinkels, 2012—focuses on the existence and properties of equilibria. To the best of our knowledge, the first-order approach has not been extended to models with multi-dimensional action sets.

We show how to extend our polynomial optimization approach to principal-agent models in which the agent has more than one decision variable. When we apply the multivariate optimization approach of Jibetean and Klerk, 2006 we encounter a theoretical difficulty. Un- like univariate nonnegative polynomials, multivariate nonnegative polynomials are not necessarily sums of squares of fixed degree. This fact has the consequence that we can no longer provide an exact reformulation of the agent's utility maximization problem but only a relaxation depending on the degree of the involved polynomials. The relaxed problem yields an upper bound on the agent's maximal utility. We then use this relaxation to replace the agent's optimization problems by equations and inequalities including a constraint that requires the upper utility bound not to deviate from the true maximal utility by more than some pre-specified tolerance level. We then prove that as the tolerance level converges to zero, the optimal solutions of the se-

[5] (LiCalzi and Spaeter, 2003) described two special classes of distributions that satisfy the CDFC.

[6] (Araujo and Moreira, 2001) introduced a Lagrangian approach different from Mirrlees, 1999. Instead of imposing conditions on the outcome distribution, they included more information in the Lagrangian, namely a second-order condition as well as the behavior of the utility function on the boundary in order to account for possible non-concave objective functions. A number of additional technical assumptions considerably limits the applicability of this approach as well.

quence of nonlinear programs involving the relaxation converge, and—in fact—the limit points yield optimal solutions to the original principal-agent problem.

Although our main results are of a theoretical nature, our paper also contributes to the computational literature on principal-agent problems. Due to the strong assumptions of the first-order approach, the computational literature has shied away from it. Prescott, 1999 and Prescott, 2004 approximated the action and compensation sets by finite grids and then allowed for action and compensation lotteries. The resulting optimization problem is linear and thus can be solved with efficient large-scale linear programming algorithms. Judd and Su, 2005 avoided the compensation lotteries and only approximated the action set by a finite grid. This approximation results in a mathematical program with equilibrium constraints (MPEC). Contrary to the LP approach, the MPEC approach may face difficulties finding global solutions, since the standard MPEC algorithms only search for locally optimal solutions. Despite this shortcoming, MPEC approaches have recently received a lot of attention in economics—see, for instance, Su and Judd, 2012 and Dubé et al., 2012. Our polynomial optimization approach does not need lotteries and instead allows us to solve principal-agent problems with continuous action and compensation sets.

The remainder of this paper is organized as follows. Section 5.2 describes the principal-agent model and the classical first-order approach. In Section 5.3 we introduce our main result for the polynomial optimization approach. Section 5.4 summarizes the mathematical background for our analysis and provides a proof of the main result. We extend the polynomial approach to models with multi-dimensional action sets in Section 5.5. Section 5.6 concludes.

5.2 The Principal-Agent Model

In this section, we briefly describe the principal-agent model under consideration. Next we review the first-order approach. We complete our initial discussion of principal-agent problems by proving the existence of a global optimal solution.

5.2.1 The Principal-Agent Problem

The agent chooses an action ("effort level") $\boldsymbol{a}$ from a set $A \subset \mathbb{R}^L$. The outcome ("output" or "gross profit") received by the principal from an action $\boldsymbol{a}$ taken by the agent can be one of N possible values, $y_1 < y_2 < \ldots < y_N$, with $y_i \in \mathbb{R}$. Let $\mu(\bullet|\boldsymbol{a})$ be a parameterized probability measure on the set of outcomes $Y = \{y_1, y_2, \ldots, y_N\}$. Then for any y_i, $\mu(y_i|\bullet)$ is a function mapping A into $[0,1]$. Of course, $\sum_{i=1}^{N} \mu(y_i|\boldsymbol{a}) = 1$ for all $\boldsymbol{a} \in A$.

The principal cannot monitor the agent's action but only the outcome. Thus, the principal will pay the agent conditional on the observed outcome. Let $w_i \in \mathcal{W} \subset \mathbb{R}$ denote the wage paid to the agent if outcome y_i occurs. A contract ("compensation scheme") between the principal and the agent is then a vector $\boldsymbol{w} = (w_1, w_2, \ldots, w_N) \in W \equiv \mathcal{W}^N$. The principal has a Bernoulli utility function over income, $u : I \to \mathbb{R}$, with domain $I = (\underline{I}, \infty) \subset \mathbb{R}$ for some $\underline{I} \in \mathbb{R} \cup \{-\infty\}$. For example, if the principal receives the outcome y_i and pays the wage w_i, then she receives utility $u(y_i - w_i)$. The agent has a Bernoulli utility function over income and actions given by $v : J \times A \to \mathbb{R}$, with $J = (\underline{J}, \infty) \subset \mathbb{R}$ for some $\underline{J} \in \mathbb{R} \cup \{-\infty\}$. Both the principal and the agent have von Neumann-Morgenstern utility functions. The expected utility functions of the principal and agent are

$$U(\boldsymbol{w}, \boldsymbol{a}) = \sum_{i=1}^{N} u(y_i - w_i)\mu(y_i|\boldsymbol{a}) \quad \text{and} \quad V(\boldsymbol{w}, \boldsymbol{a}) = \sum_{i=1}^{N} v(w_i, \boldsymbol{a})\mu(y_i|\boldsymbol{a}),$$

respectively. We are now in the position to state the principal-agent problem.

$$\begin{aligned} \max_{\boldsymbol{w} \in W, \boldsymbol{a} \in A} \quad & U(\boldsymbol{w}, \boldsymbol{a}) \\ \text{s.t.} \quad & \boldsymbol{a} \in \arg\max_{\boldsymbol{b} \in A} V(\boldsymbol{w}, \boldsymbol{b}) \\ & V(\boldsymbol{w}, \boldsymbol{a}) \geq \underline{V} \end{aligned} \tag{5.1}$$

The objective of this optimization problem is to maximize the principal's expected utility. The first constraint,

$$\boldsymbol{a} \in \arg\max_{\boldsymbol{b} \in A} V(\boldsymbol{w}, \boldsymbol{b}) \tag{5.2}$$

is the incentive-compatibility constraint for the agent; he will only take actions that maximize his own expected utility. We assume implicitly that the agent does not work against the principal, that is, if he is indifferent between several different actions then he will choose the action most beneficial to the principal. The second constraint is the participation constraint for the agent. He has an outside option and will accept a contract only if he receives at least the expected utility $\underline{V}$ of that outside opportunity.

The principal cannot observe the agent's actions but knows his utility function. Thus, the described principal-agent model exhibits pure moral hazard and no hidden information. The first-order approach for models of this type has been examined by Mirrlees, 1999, Rogerson, 1985, Jewitt, 1988, Sinclair-Desgagné, 1994, Alvi, 1997, Jewitt et al., 2008, Conlon, 2009, and others.

5.2.2 The First-Order Approach

In general it is very difficult to find a global optimal solution to the principal-agent problem (5.1). For the model with a one-dimensional action set, $A = [\underline{a}, \bar{a}]$ with $\bar{a} \in \mathbb{R} \cup \{\infty\}$, the popular first-order approach replaces the incentive-compatibility constraint (5.2) by a stationarity condition. If the set A is sufficiently large so that the optimal solution to the agent's expected utility maximization problem has an interior solution, the necessary first-order condition is

$$\frac{\partial}{\partial a} V(\boldsymbol{w}, a) = \sum_{i=1}^{N} \left(\frac{\partial}{\partial a} v(w_i, a)\ \mu(y_i|a) + v(w_i, a)\ \frac{\partial}{\partial a} \mu(y_i|a) \right) = 0. \tag{5.3}$$

For an application of the first-order approach, standard monotonicity, curvature, and differentiability assumptions are imposed. Rogerson, 1985 introduces the following assumptions (in addition to some other minor technical conditions).

(1) The function $\mu(y|\bullet) : A \to [0, 1]$ is twice continuously differentiable for all $y \in Y$.

(2) The principal's Bernoulli utility function $u : I \to \mathbb{R}$ is strictly increasing, concave, and twice continuously differentiable.

(3) The agent's Bernoulli utility function $v : J \times A \to \mathbb{R}$ satisfies $v(w, a) = \psi(w) - a$. The function $\psi : J \to \mathbb{R}$ is strictly increasing, concave and twice continuously differentiable.

These three assumptions alone are not sufficient for the first-order approach to be valid, since the probabilities $\mu(y_i|a)$ depend on the action a and thus affect the monotonicity and curvature of the expected utility functions. Rogerson, 1985 proved the validity of the first-order approach under two additional assumptions on the probability function, see also Mirrlees, 1979. We define the following function $F_j(a) = \sum_{i=1}^{j} \mu(y_i|a)$. For $\mu(y_i|a) >> 0$ for all $a \in A$ and all i , the conditions of Mirrlees, 1979 and Rogerson, 1985 are as follows.

(MLRC) (monotone likelihood-ratio condition[7]) The measure μ has the property that for $a_1 \leq a_2$ the ratio $\frac{\mu(y_i|a_1)}{\mu(y_i|a_2)}$ is decreasing in i.

(CDFC) (convexity of the distribution function condition) The function F has the property that $F_i''(a) \geq 0$ for all $i = 1, 2, \ldots, N$ and $a \in A$.

According to Conlon, 2009, these assumptions are the most popular conditions in economics, even though other sufficient conditions exist, see Jewitt, 1988. Sinclair-Desgagné, 1994 generalized the conditions of Mirrlees, 1979 and Rogerson, 1985 for the multi-signal principal-agent problem. Conlon, 2009 in turn presented multi-signal generalizations of both the Mirrlees-Rogerson and the Jewitt sufficient conditions for the first-order approach. Despite this progress, all of these conditions are regarded as highly restrictive, see Conlon, 2009 and Kadan et al., 2011.

5.2.3 Existence of a Global Optimal Solution

For the sake of completeness, we show the existence of a global optimal solution to the principal-agent problem (5.1) without assumptions on the differentiability, monotonicity, and curvature of the utility and probability functions. For this purpose we introduce the following three assumptions.

Assumption 5.1 (Feasibility). *There exists a contract $\boldsymbol{w} \in W$ such that the agent is willing to participate, that is, $V(\boldsymbol{w}, \boldsymbol{a}) \geq \underline{V}$ for some $\boldsymbol{a} \in A$.*

[7] The MLRC implies a stochastic dominance condition, $F_i'(a) \leq 0$ for all $i = 1, 2, \ldots, N$ and $a \in A$.

Assumption 5.2 (Compactness)**.** *Both decision variables are chosen from compact domains.*

(1) The set A of actions is a non-empty, compact subset of a finite-dimensional Euclidean space, $A \subset \mathbb{R}^L$.

(2) The set $\mathcal{W}$ of possible wages is a nonempty, compact interval $[\underline{w}, \overline{w}] \subset \mathbb{R}$.

Assumption 5.3 (Continuity)**.** *All functions in the model are continuous.*

(1) The function $\mu(y|\bullet) : A \to [0, 1]$ is continuous for all $y \in Y$.

(2) The principal's Bernoulli utility function $u : I \to \mathbb{R}$ is continuous on I.

(3) The agent's Bernoulli utility function $v : J \times A \to \mathbb{R}$ is continuous on $J \times A$.

For simplicity we also assume that the expected utility functions U and V are well-defined on their domain $W \times A$. (Sufficient conditions for this innocuous assumption are $\underline{J} < \underline{w}$ and $\underline{I} < y_1 - \overline{w}$). Under the stated assumptions, a global optimal solution to the optimization problem (5.1) exists.

Proposition 5.1. *If Assumptions 5.1 – 5.3 hold, then the principal-agent problem (5.1) has a global optimal solution.*

Proof. Consider the optimal value function $\Psi : W \to \mathbb{R}$ for the agent defined by $\Psi(\boldsymbol{w}) = \max\{V(\boldsymbol{w}, \boldsymbol{a}) \mid a \in A\}$. By Assumptions 5.2 and 5.3, the expected utility function V is continuous on the compact domain $W \times A$. Thus, (a special case of) Berge's Maximum Theorem (Berge, 1963) implies that Ψ is continuous on its domain W. Using the function Ψ, we can state the feasible region F of the principal-agent problem (5.1),

$$F = \{(\boldsymbol{w}, \boldsymbol{a}) \in W \times A \mid V(\boldsymbol{w}, \boldsymbol{a}) = \Psi(\boldsymbol{w}), V(\boldsymbol{w}, \boldsymbol{a}) \geq \underline{V}\}\,.$$

The feasible region F is nonempty by Assumption 5.1. As a subset of $W \times A$ it is clearly bounded. Since both V and Ψ are continuous functions and the constraints involve only an equation and a weak inequality, the set F is also closed. And so the optimization problem (5.1) requires the maximization of the continuous function U on the nonempty, compact feasible region F. Now the proposition follows from the extreme value theorem of Weierstrass. □

5.3 The Polynomial Optimization Approach for $A \subset \mathbb{R}$

The purpose of this section is to state our main result, Theorem 5.1, and illustrate it by an example.

Recall that a symmetric matrix $M \in \mathbb{R}^{n \times n}$ is called positive semidefinite if and only if $\boldsymbol{v}^T M \boldsymbol{v} \geq 0$ for all $\boldsymbol{v} \in \mathbb{R}^n$. We denote such matrices by $M \succcurlyeq 0$. The set of all symmetric positive semidefinite $n \times n$ matrices is a closed convex cone.

Next we introduce an assumption on the agent's expected utility function.

Assumption 5.4 (Rational Expected Utility Function)**.** *The parameterized probability distribution functions $\mu(y|\bullet) : A \to [0,1]$ and the agent's Bernoulli utility function $v : J \times A \to \mathbb{R}$ are such that the agent's expected utility function is a rational function of the form*

$$-V(\boldsymbol{w}, a) = -\sum_{j=1}^{N} v(w_j, a)\mu(y_j|a) = \frac{\sum_{i=0}^{d} c_i(\boldsymbol{w})a^i}{\sum_{i=0}^{d} f_i(\boldsymbol{w})a^i}$$

for functions $c_i, f_i : W \to \mathbb{R}$ with $\sum_{i=0}^{d} f_i(\boldsymbol{w})a^i > 0$ for all $(\boldsymbol{w}, a) \in W \times A$.[8] *Moreover, the two polynomials in the variable a, $\sum_{i=0}^{d} c_i(\boldsymbol{w})a^i$ and $\sum_{i=0}^{d} f_i(\boldsymbol{w})a^i$, have no common factors and $d \in \mathbb{N}$ is maximal such that $c_d(\boldsymbol{w}) \neq 0$ or $f_d(\boldsymbol{w}) \neq 0$.*

Recall the notation $\lceil x \rceil$ for the smallest integer not less than x. Using this notation, we define the number $D = \left\lceil \frac{d}{2} \right\rceil$.

Without loss of generality we assume for the set of actions, $A = [-1,1] = \{a \in \mathbb{R} \mid 1 - a^2 \geq 0\}$. The following theorem[9] provides us with an equivalent problem to the principal-agent problem (5.1).

Theorem 5.1. *Let $A = [-1,1]$ and suppose Assumption 5.4 holds. Then $(\boldsymbol{w}^*, a^*)$ solves the principal-agent problem (5.1) if and only if there exist $\rho^* \in \mathbb{R}$ as well as matrices $Q^{(0)*} \in \mathbb{R}^{(D+1)\times(D+1)}$ and $Q^{(1)*} \in \mathbb{R}^{D\times D}$ such that $(\boldsymbol{w}^*, a^*, \rho^*, Q^{(0)*}, Q^{(1)*})$ solves the following optimization problem:*

$$\max_{\boldsymbol{w},a,\rho,Q^{(0)},Q^{(1)}} \quad U(\boldsymbol{w}, a) \qquad \textit{subject to} \tag{5.4}$$

$$\begin{aligned}
c_0(\boldsymbol{w}) - \rho f_0(\boldsymbol{w}) &= Q^{(0)}_{0,0} + Q^{(1)}_{0,0} && (5.4a)\\
c_l(\boldsymbol{w}) - \rho f_l(\boldsymbol{w}) &= \sum_{i+j=l} Q^{(0)}_{ij} + \sum_{i+j=l} Q^{(1)}_{ij} - \sum_{i+j=l-2} Q^{(1)}_{ij}, \quad l = 1,\dots,d && (5.4b)\\
Q^{(0)}, Q^{(1)} &\succcurlyeq 0 && (5.4c)\\
\rho\left(\sum_{i=0}^{d} f_i(\boldsymbol{w})a^i\right) &= \sum_{i=0}^{d} c_i(\boldsymbol{w})a^i && (5.4d)\\
\sum_{i=0}^{d} c_i(\boldsymbol{w})a^i &\leq -\underline{V}\left(\sum_{i=0}^{d} f_i(\boldsymbol{w})a^i\right) && (5.4e)\\
-a^2 + 1 &\geq 0 && (5.4f)\\
\boldsymbol{w} &\in W && (5.4g)
\end{aligned}$$

The new optimization problem (5.4) has the same objective function as the original principal-agent problem (5.1). Unlike the original problem, the new problem (5.4) is not

[8]The positivity condition for the denominator is necessary, since a change in sign would lead to division by zero.

[9]Note that the row and column indexing of the two matrices in the theorem starts at 0. The reason for this convention becomes clear in the theoretical arguments presented Section 5.4.1.

a bilevel optimization problem. Instead the constraint involving the agent's expected utility maximization problem has been replaced by inequalities and equations. Problem (5.4) has the additional decision variables $\rho \in \mathbb{R}$, $Q^{(0)} \in \mathbb{R}^{(D+1)\times(D+1)}$, and $Q^{(1)} \in \mathbb{R}^{D\times D}$. The optimal value ρ^* of the variable ρ in problem (5.4) will be $-V(\boldsymbol{w}^*, a^*)$, the negative of the agent's maximal expected utility. Constraints (5.4a)–(5.4c) use a sum of squares representation of nonnegative polynomials to ensure that for a contract $\boldsymbol{w}$ chosen by the principal, $-V(\boldsymbol{w}, a) \geq \rho$ for all $a \in A$. That is, $-\rho$ is an upper bound on all possible utility levels for the agent. Note that equations (5.4a) and (5.4b) are linear in ρ and the elements of the matrices $Q^{(0)} \in \mathbb{R}^{(D+1)\times(D+1)}$ and $Q^{(1)} \in \mathbb{R}^{D\times D}$. Constraint (5.4c) requires that these two matrices are symmetric positive semi-definite. (Later on we summarize properties of positive semi-definite matrices, which show that constraint (5.4c) can be written as a set of polynomial inequalities.) Next, constraint (5.4d) ensures that the variable $-\rho$ is actually equal to the agent's utility for effort a and contract $\boldsymbol{w}$. Therefore, this constraint together with the constraints (5.4a)–(5.4c) forces any value of a satisfying the equation to be the agent's optimal effort choice as well as the value of ρ to be the corresponding maximal expected utility value. Put differently, for a given contract $\boldsymbol{w}$ the first four constraints ensure an optimal effort choice by the agent. The last three constraints are straightforward. Constraint (5.4e) is the transformed participation constraint for the agent's rational expected utility function. Constraint (5.4f) is a polynomial representation of the feasible action set and constraint (5.4g) is just the constraint on the compensation scheme from the original principal-agent problem (5.1).

We illustrate the statement of the theorem by a simple example.

Example 5.1. Let $A = [0,1]$ and $\mathcal{W} = \mathbb{R}_+$. There are $N = 3$ possible outcomes $y_1 < y_2 < y_3$ which occur with the probabilities

$$\mu(y_1|a) = \binom{2}{0} a^0(1-a)^2,\ \mu(y_2|a) = \binom{2}{1} a(1-a),\ \mu(y_3|a) = \binom{2}{2} a^2(1-a)^0.$$

The principal is risk-neutral with Bernoulli utility $u(y-w) = y-w$. The agent is risk-averse and has utility

$$v(w,a) = \frac{w^{1-\eta}-1}{1-\eta} - \kappa a^2,$$

where $\eta \neq 1, \eta \geq 0$ and $\kappa > 0$. The agent's expected utility is

$$V(w_1,w_2,w_3,a) = (1-a)^2 \frac{w_1^{1-\eta}-1}{1-\eta} + 2(1-a)a \frac{w_2^{1-\eta}-1}{1-\eta} + a^2 \frac{w_3^{1-\eta}-1}{1-\eta} - \kappa a^2.$$

The second-order derivative of V with respect to a,

$$\frac{\partial^2 V}{\partial a^2} = \frac{2w_1^{1-\eta}}{1-\eta} - \frac{4w_2^{1-\eta}}{1-\eta} + \frac{2w_3^{1-\eta}}{1-\eta} - 2\kappa,$$

changes sign on $W \times A$. Thus, this function is not concave and so the classical first-order approach does not apply. We apply Theorem 5.1 to solve this principal-agent problem. For simplicity, we consider a specific problem with $\eta = \frac{1}{2}$, $\underline{V} = 0$, $\kappa = 2$, and $(y_1, y_2, y_3) = (0, 2, 4)$.

First we transform the set of actions $A = [0, 1]$ into the interval $A = [-1, 1]$ via the variable transformation $a \mapsto \frac{a+1}{2}$. The resulting expected utility functions are

$$\begin{aligned} U(\boldsymbol{w}, a) =& 2 + 2a - \frac{w_1}{4} + \frac{aw_1}{2} - \frac{a^2 w_1}{4} - \frac{w_2}{2} + \frac{a^2 w_2}{2} - \frac{w_3}{4} - \frac{aw_3}{2} - \frac{a^2 w_3}{4} \\ V(\boldsymbol{w}, a) =& -\frac{5}{2} + \frac{\sqrt{w_1}}{2} + \sqrt{w_2} + \frac{\sqrt{w_3}}{2} - a - a\sqrt{w_1} + a\sqrt{w_3} - a^2\sqrt{w_2} + \frac{a^2\sqrt{w_1}}{2} - \frac{a^2}{2} \\ & + \frac{a^2\sqrt{w_3}}{2} \end{aligned}$$

We observe that $V(\boldsymbol{w}, a)$ is a quadratic polynomial in a. The representation of $-V(\boldsymbol{w}, a)$ according to Assumption 5.4 has the nonzero coefficients $f_0(\boldsymbol{w}) = 1$ and $c_0(\boldsymbol{w}) = \frac{5}{2} - \frac{\sqrt{w_1}}{2} - \sqrt{w_2} - \frac{\sqrt{w_3}}{2}$, $c_1(\boldsymbol{w}) = 1 + \sqrt{w_1} - \sqrt{w_3}$, and $c_2(\boldsymbol{w}) = \frac{1}{2} - \frac{\sqrt{w_1}}{2} + \sqrt{w_2} - \frac{\sqrt{w_3}}{2}$. According to Theorem 5.1, the matrix $Q^{(0)}$ is a 2×2 matrix and $Q^{(1)}$ is just a single number. With

$$Q^{(0)} = \begin{pmatrix} n_{00} & n_{01} \\ n_{01} & n_{11} \end{pmatrix} \quad \text{and} \quad Q^{(1)} = m_{00}$$

we can rewrite the principal-agent problem following the theorem.

$$\begin{aligned} \max_{w_1, w_2, w_3, a, \rho, n_{00}, n_{01}, n_{11}, m} \quad & U(w_1, w_2, w_3, a) \\ \text{s.t.} \quad & \frac{5}{2} - \frac{\sqrt{w_1}}{2} - \sqrt{w_2} - \frac{\sqrt{w_3}}{2} - \rho = n_{00} + m_{00} \\ & 1 + \sqrt{w_1} - \sqrt{w_3} = 2n_{01} \\ & \frac{1}{2} - \frac{\sqrt{w_1}}{2} + \sqrt{w_2} - \frac{\sqrt{w_3}}{2} = n_{11} - m_{00} \\ & \rho = -V(w_1, w_2, w_3, a) \\ & n_{00} \geq 0, n_{11} \geq 0, n_{00}n_{11} - n_{01}^2 \geq 0, m_{00} \geq 0 \\ & V(w_1, w_2, w_3, a) \geq 0 \\ & -a^2 + 1 \geq 0 \\ & w_1, w_2, w_3 \geq 0 \end{aligned}$$

We can solve this nonlinear optimization problem with Gloptipoly, see Henrion et al., 2009, and obtain the globally optimal contract $\boldsymbol{w}^* = (0.3417, 1.511, 3.511)$ and the resulting optimal effort $a^* = 0.6446$. Table 5.1 reports solutions for different levels of the agent's risk aversion η. For completion the table also reports the corresponding first-best solutions[10] indexed by FB. For $\eta = 0$, when the agent is risk-neutral, a continuum of contracts exists. However, the intervals of values for w_1 and w_2 are economically irrelevant since for $w_3 = 1$ the optimal effort of $a^* = 1$ results in zero probability of outcomes 1 and 2 and the first-best solution.

[10]Omitting the incentive compatibility constraint and maximizing the principal's expected utility only subject to the participation constraint leads to the first-best solution.

η	$U(w_1^*, w_2^*, w_3^*, a^*)$	a^*	w_1^*	w_2^*	w_3^*	U_{FB}	a_{FB}	w_{FB}
0	1	1	$[0,1)$	$[0,1]$	3	1	1	3
$\frac{1}{4}$	0.6760	0.8260	0.2777	1.177	3.344	0.7471	0.7993	2.450
$\frac{1}{3}$	0.5723	0.7637	0.2879	1.273	3.441	0.6850	0.7541	2.332
$\frac{1}{2}$	0.3844	0.6446	0.3417	1.511	3.511	0.5814	0.6823	2.148
$\frac{4}{5}$	0.1292	0.4881	0.5314	1.798	3.296	0.4410	0.5918	1.926
2	-0.3444	0.2413	0.8749	1.817	2.416	0.1349	0.4196	1.544
4	-0.6102	0.1277	0.9657	1.597	1.866	-0.09165	0.3117	1.338

Table 5.1: Numerical solutions to the principal-agent problem as a function of η

5.4 Derivation of the Polynomial Optimization Approach

In this section we first review the mathematical foundation of Theorem 5.1 and then prove the theorem. We also discuss the assumptions of the theorem as well as the limitations of the polynomial optimization approach.

5.4.1 Mathematical Framework

First we introduce semidefinite programs, a class of convex optimization problems that is relevant for our analysis. Next we define sums of squared polynomials and state representation theorems for such polynomials. Then we describe how the representation results allow us to simplify constrained polynomial optimization problems. And finally we describe the extension to rational objective functions.

Semidefinite Programming

For a matrix $M = (m_{ij}) \in \mathbb{R}^{n\times n}$ the sum of its diagonal elements,

$$\mathrm{tr}(M) = \sum_{i=1}^{n} m_{ii},$$

is called the trace of M. Note that

$$\mathrm{tr}(CX) = \sum_{i,j=1}^{n} C_{ij}X_{ij}$$

for matrices $C, X \in S^n$ is a linear function on the set S^n of symmetric $n \times n$ matrices. A semidefinite optimization problem (in standard form) is defined as follows.

Definition 5.1. Let $C, A_j \in \mathbb{R}^{n\times n}$ for all $j = 1, \ldots, m$ be symmetric matrices and $b \in \mathbb{R}^m$. We then call the following convex optimization problem a *semidefinite program*

(*SDP*).

$$\begin{aligned} \sup_X \ & \operatorname{tr}(CX) \\ \text{s.t. } & \operatorname{tr}(A_jX) = b_j \quad j = 1, \ldots, m \\ & X \succcurlyeq 0 \end{aligned} \tag{5.5}$$

Note that the (*SDP*) has a linear objective function and a closed convex feasible region. Thus, semi-definite programs are a special class of convex optimization problems. In fact, semidefinite programs can be solved efficiently both in theory and in practice, see Vandenberghe and Boyd, 1996 and Boyd and Vandenberghe, 2004.

We need to reformulate this into an NLP and so we first look at the following definition.

Definition 5.2. Let $M = (m_{ij})_{i=1,\ldots,n,j=1,\ldots,n} \in \mathbb{R}^{n \times n}$ a matrix and let $I \subset \{1, \ldots, n\}$. Then $\det((m_{ij})_{i,j \in I \times I})$ is called a *principal minor*. If $I = \{1, \ldots, k\}$ then $\det((m_{ij})_{i,j \in I \times I})$ is called the *leading principal minor*.

Proposition 5.2. *Let $Q \in \mathbb{R}^{n \times n}$ be a symmetric matrix with rank m. Then the following statements are equivalent.*

(a) Q is positive semidefinite.

(b) All principal minors of Q are nonnegative.

(c) There exists a matrix $V \in \mathbb{R}^{n \times m}$ with $Q = VV^T$ and $m \leq n$.

(d) There exists a lower triangular matrix $L \in \mathbb{R}^{n \times n}$ with nonnegative diagonal such that $Q = LL^T$.

(e) All eigenvalues are nonnegative.

Note here that the equivalent statements for positive semidefiniteness can be expressed by polynomial equations and inequalities. Statement (b) gives a set of polynomial inequalities. Statement (c) involves a system of polynomial equations. Statements (d) and (e) are given by a system of equations and inequalities.

Polynomials and Sums of Squares

For the study of polynomial optimization it is necessary to first review a few fundamental concepts from the study of polynomials in real algebraic geometry. Our brief review is based upon the survey by Laurent, 2009 and the book by Lasserre, 2010b.

The expression $\mathbb{R}[x_1, \ldots, x_n]$ denotes the ring of polynomials in n variables over the real numbers. Whenever possible we use the abbreviation $\mathbb{R}[\boldsymbol{x}]$ with $\boldsymbol{x} = (x_1, \ldots, x_n)$. For clarification, we denote the set of nonnegative integers by $\mathbb{N}$. For a vector $\boldsymbol{\alpha} \in \mathbb{N}^n$, we denote the monomial $x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ by $\boldsymbol{x}^{\boldsymbol{\alpha}}$. The degree of this monomial is $|\boldsymbol{\alpha}| = \sum_{i=1}^n \alpha_i$. A polynomial $p \in \mathbb{R}[\boldsymbol{x}]$, $p = \sum_{\boldsymbol{\alpha}} a_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\alpha}}$ is a sum of terms $a_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\alpha}}$ with finitely many nonzero $a_{\boldsymbol{\alpha}} \in \mathbb{R}$. The degree of p is $\deg(p) = \max_{\{\boldsymbol{\alpha} | a_{\boldsymbol{\alpha}} \neq 0\}} |\boldsymbol{\alpha}|$.

Let $g_1, \ldots, g_m \in \mathbb{R}[\boldsymbol{x}]$. Then the set

$$K = \{\boldsymbol{x} \in \mathbb{R}^n \mid g_i(\boldsymbol{x}) \geq 0, \ \forall i = 1, \ldots, m\}$$

is called a basic semi-algebraic set.

A central concept of polynomial optimization is the notion of a sum of squares.

Definition 5.3. A polynomial $\sigma \in \mathbb{R}[\boldsymbol{x}]$ is called a *sum of squares* if there exists finitely many polynomials $p_1, \ldots, p_m \in \mathbb{R}[\boldsymbol{x}]$ such that $\sigma = \sum_{i=1}^m p_i^2$. The expression $\Sigma[\boldsymbol{x}] \subset \mathbb{R}[\boldsymbol{x}]$ denotes the set of sums of squares. And $\Sigma_d[\boldsymbol{x}] \subset \mathbb{R}[\boldsymbol{x}]$ denotes the set of sums of squares up to degree d.

A sum of squares σ is always a nonnegative function. The converse however is not always true, i.e. not every non negative polynomial is a sum of squares. Also it is clear that a polynomial can only be a sum of squares if it has even degree. Moreover, the degree of each polynomial p_i in the sum is bounded above by half the degree of σ. To see the link to positive semi-definite matrices, we consider the vector

$$\boldsymbol{v}_d(\boldsymbol{x}) = (\boldsymbol{x}^{\boldsymbol{\alpha}})_{|\boldsymbol{\alpha}| \leq d} = \left(1, x_1, \ldots, x_n, x_1^2, x_1 x_2, \ldots, x_{n-1}x_n, x_n^2, \ldots, x_n^d\right)^T$$

of all monomials $\boldsymbol{x}^{\boldsymbol{\alpha}}$ of degree at most d. This vector is of dimension $\binom{n+d}{d}$. There is a strong connection between sums of squares, the vector $\boldsymbol{v}_d(\boldsymbol{x})$ and positive semi-definite matrices.

Lemma 5.1. **[(Lasserre, 2010b, Proposition 2.1)]** *A polynomial $\sigma \in \mathbb{R}[\boldsymbol{x}]$ of degree $2d$ is a sum of squares if and only if there exists a symmetric positive semidefinite $\binom{n+d}{d} \times \binom{n+d}{d}$ matrix Q such that $\sigma = \boldsymbol{v}_d(\boldsymbol{x})^T Q \boldsymbol{v}_d(\boldsymbol{x})$, where $\boldsymbol{v}_d(\boldsymbol{x})$ is the vector of monomials in $\boldsymbol{x}$ of degree at most d.*

Sum of Squares and SDP in $\mathbb{R}$

We illustrate the relationship between finding sum of squares representations and SDPs for the univariate case. For $n = 1$,

$$\boldsymbol{v}_d(x) = \left(1, x, x^2, \ldots, x^d\right)^T.$$

We can identify a polynomial $p_i(x) = \sum_{j=0}^d a_{ij} x^j$ with its vector of coefficients $\boldsymbol{a}_i = (a_{i0}, a_{i1}, \ldots, a_{id})$ and write $p_i(x) = \boldsymbol{a}_i \boldsymbol{v}_d(x)$. Next we aggregate m such polynomials in a matrix-vector product

$$\begin{bmatrix} p_1(x) \\ p_2(x) \\ \vdots \\ p_m(x) \end{bmatrix} = \begin{bmatrix} a_{10} & a_{11} & \ldots & a_{1d} \\ a_{20} & a_{21} & \ldots & a_{2d} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m0} & a_{m1} & \ldots & a_{md} \end{bmatrix} \begin{bmatrix} 1 \\ x \\ \vdots \\ x^d \end{bmatrix}$$

Denoting the $(m \times (d+1))$ coefficient matrix on the right-hand side by V, we can write a sum of squares as

$$\sigma(x) = \sum_{i=1}^{m} p_i^2(x) = (V\boldsymbol{v}_d(x))^T (V\boldsymbol{v}_d(x)) = \boldsymbol{v}_d(x)^T Q \boldsymbol{v}_d(x)$$

for $Q = V^T V$. By construction the matrix Q is symmetric, positive semi-definite and has at most rank m. Note that if we start indexing Q with 0 then Q_{ij} with $i + j = h$ contributes to the term of σ with degree h.

Observe that finding a sum of squares representation for the polynomial $\sigma(x)$ requires finding a symmetric positive semi-definite matrix Q such that the polynomials on the left-hand and right-hand side are identical. But that condition just requires the polynomials to have identical coefficients for all monomials. If σ has degree $2d$, then the coefficient conditions are $2d + 1$ linear equations in the $(d+1)^2/2 + d + 1$ unknown elements of Q. This set of linear equations together with the requirement that Q is symmetric positive semi-definite are just the constraints of an SDP. And so finding a sum of squares representation of a univariate polynomial σ is equivalent to an SDP feasibility problem.

Sum of Squares Representation in $\mathbb{R}$

For polynomials in a single variable x, the set of nonnegative polynomials and the set $\Sigma[x]$ of sums of squares are identical.

Lemma 5.2. [(Laurent, 2009, Lemma 3.5)] *Any nonnegative univariate polynomial is a sum of (at most) two squares.*

We next consider nonnegative univariate polynomials on closed intervals. For a general treatment it suffices to examine two cases, $[-1, 1]$ and $[0, \infty)$. The next proposition states that nonnegative polynomials on these intervals can be expressed via two sums of squares and a polynomial that describes the respective interval via a semi-algebraic set. Note that $[-1, 1] = \{x \in \mathbb{R} \mid 1 - x^2 \geq 0\}$ and $[0, \infty) = \{x \in \mathbb{R} \mid x \geq 0\}$.

Proposition 5.3. [(Lasserre, 2010b, Theorems 2.6, 2.7), (Laurent, 2009, Theorems 3.21, 3.23)] *Let $p \in \mathbb{R}[x]$ be of degree d.*

(a) $p \geq 0$ on $[-1, 1]$ if and only if

$$p = \sigma_0 + \sigma_1 \cdot (1 - x^2) \quad \sigma_0, \sigma_1 \in \Sigma[x]$$

with $\deg(\sigma_0), \deg(\sigma_1 \cdot (1 - x^2)) \leq d$ if d is even and $\deg(\sigma_0), \deg(\sigma_1 \cdot (1 - x^2)) \leq d + 1$ if d is odd.

(b) $p \geq 0$ on $[0, \infty)$ if and only if

$$p = \sigma_0 + \sigma_1 x \quad \sigma_0, \sigma_1 \in \Sigma[x]$$

with $\deg(\sigma_0), \deg(x\sigma_1) \leq d$.

These results depend critically on the specific description of the intervals via the polynomials $1 - x^2$ and x, respectively. Other descriptions lead to weaker results with representations involving higher degree sum of squares polynomials.

Proposition 5.3 can also be used to show more general cases. The univariate polynomial $f(x)$ is nonnegative on $K = [a, \infty), K = (-\infty, b]$ and $K = [a, b]$ if and only if

$$\begin{aligned}
&p(x) = f(x + a) \geq 0 \quad \forall x \in [0, \infty),\\
&p(x) = f(b - x) \geq 0 \quad \forall x \in [0, \infty),\\
&p(x) = f((x(b - a) + (a + b))/2) \geq 0 \quad \forall x \in [-1, 1],
\end{aligned}$$

respectively.

Next we describe the application of the representation results for nonnegative univariate polynomials to polynomial optimization.

Polynomial Optimization in $\mathbb{R}$

For a polynomial $p \in \mathbb{R}[x]$ and a nonempty semi-algebraic set $K \subset \mathbb{R}$ consider the constrained polynomial optimization problem,

$$p_{\min} = \inf_{x \in K} p(x). \tag{5.6}$$

We can rewrite Problem (5.6) as follows,

$$\begin{aligned}
&\sup_{\rho} \ \rho\\
&\text{s.t. } p(x) - \rho \geq 0 \ \forall x \in K.
\end{aligned} \tag{5.7}$$

For any feasible $\rho \in \mathbb{R}$ the following inequality holds,

$$\rho \leq p_{\min}. \tag{5.8}$$

Note that the constraints of the rewritten problem state that the polynomial $p - \rho$ must be nonnegative on the set K. Now consider the domain $K = [-1, 1] = \{x \mid 1 - x^2 \geq 0\}$. In this case applying part (a) of Proposition 5.3 enables us to rewrite the infinitely many constraints of Problem (5.7). With the polynomial g defined by $g(x) = 1 - x^2$ we obtain the following optimization problem,

$$\begin{aligned}
\sup_{\rho, \sigma_0, \sigma_1} \ & \rho\\
\text{s.t. } & p - \rho = \sigma_0 + \sigma_1 g\\
& \sigma_0, \sigma_0 \in \Sigma[x]
\end{aligned} \tag{5.9}$$

Note that the equality constraint here signifies equality as polynomials. Lemma 5.1 enables us to rewrite the optimization problem once more by replacing the unknown sums of squares σ_0 and σ_1 by positive semi-definite matrices. We define the number

$d_p = \left\lceil \frac{\deg(p)}{2} \right\rceil$ for a polynomial $p \in \mathbb{R}[\boldsymbol{x}]$. According to Proposition 5.3 the number d_p is an upper bound for the degrees of σ_0 and σ_1. And so we can rewrite the optimization problem.

$$\begin{aligned} \sup_{\rho, Q^{(0)}, Q^{(1)}} \quad & \rho \\ \text{s.t. } & p - \rho = v_{d_p}^T Q^{(0)} v_{d_p} + g v_{d_p - 1}^T Q^{(1)} v_{d_p - 1} \\ & Q^{(0)}, Q^{(1)} \succcurlyeq 0 \\ & Q^{(0)} \in \mathbb{R}^{(d_p+1)\times(d_p+1)}, \; Q^{(1)} \in \mathbb{R}^{d_p \times d_p} \\ & v_{d_p} = (1, x, \dots, x^{d_p})^T, \; v_{d_p - 1} = (1, x, \dots, x^{d_p - 1})^T \end{aligned} \tag{5.10}$$

Note that the first functional constraint holds if and only if all coefficients (of identical monomials on the left- and right-hand side) are identical. Thus this functional constraint reduces to a set of linear constraints which only involve the coefficients of the terms. Let $p = \sum_{l=0}^{\deg(p)} c_l x^l$ and write $Q_{ij}^{(0)}$, $i, j = 0, 1, \dots, d_p$, for the (i, j)-th entry of the matrix $Q^{(0)}$ (similarly for $Q^{(1)}$). Then we can rewrite the first constraint of Problem (5.10),

$$\begin{aligned} c_0 - \rho &= Q_{0,0}^{(0)} + Q_{0,0}^{(1)}, \\ c_l &= \sum_{i+j=l} Q_{ij}^{(0)} + \sum_{i+j=l} Q_{ij}^{(1)} - \sum_{i+j=l-2} Q_{ij}^{(1)} \quad l = 1, \dots, d. \end{aligned} \tag{5.11}$$

This set of constraints is just a set of linear equations in the unknowns ρ and $Q_{ij}^{(m)}$. In particular we observe that the final optimization problem is an SDP. Note that the positive semi-definite constraint for the matrices $Q^{(0)}$ and $Q^{(1)}$ can be interpreted as polynomial inequality constraints. This fact follows from Proposition 5.2.

The following proposition summarizes the relationship between the original problem and the reformulation.

Proposition 5.4. [**(Lasserre, 2010b, Theorem 5.8)**] *If* $p(x) = \sum_i c_i x^i$ *and* $K = \{x \in \mathbb{R} \mid 1 - x^2 \geq 0\} = [-1, 1]$ *then problem* (5.10) *is equivalent to* $\inf_{x \in [-1,1]} p(x)$ *and both problems have an optimal solution.*

The optimal solutions satisfy $\rho = p_{min}$. In sum, the constrained optimization problem of minimizing a univariate polynomial on an interval of $\mathbb{R}$ reduces to an SDP, a convex optimization problem. In particular we can write the optimization problem in theorem 5.1 as a maximization and not a supremum.

Rational Objective Function

Jibetean and Klerk, 2006 prove an analogous result for the case of rational objective functions. Let $p(\boldsymbol{x}), q(\boldsymbol{x})$ be two polynomials defined on a set $K \subset \mathbb{R}^n$. Consider the following optimization problem,

$$p_{\min} = \inf_{\boldsymbol{x} \in K} \frac{p(\boldsymbol{x})}{q(\boldsymbol{x})}. \tag{5.12}$$

We can rewrite this problem in polynomial form.

Proposition 5.5. [**(Jibetean and Klerk, 2006, Theorem 2)**] *If p and q have no common factor and K is an open connected set or a (partial) closure of such a set then*

(a) If q changes sign on K, then $p_{min} = -\infty$.

(b) If q is nonnegative on K, problem (5.12) *is equivalent to*

$$p_{min} = \sup\{\rho \mid p(\boldsymbol{x}) - \rho q(\boldsymbol{x}) \geq 0,\ \forall \boldsymbol{x} \in K\}.$$

Now consider the univariate case, so let $p, q \in \mathbb{R}[x]$ and set $d = \max(d_p, d_q)$. For $K = [-1, 1]$ and $g(x) = 1 - x^2$, we can again use Proposition 5.3 and reformulate problem (5.12),

$$\begin{aligned} \sup_{\rho,\sigma_0,\sigma_1} \ & \rho \\ \text{s.t. } & p - \rho q = \sigma_0 + g\sigma_1 \\ & \sigma_0 \in \Sigma_{2d},\ \sigma_1 \in \Sigma_{2(d-1)} \end{aligned} \tag{5.13}$$

And so we can solve the constrained optimization problem (5.12) also as an SDP.

5.4.2 Proof of Theorem 5.1

Now we are in the position to prove Theorem 5.1.

Proof. Note that the upper level problem has not been altered. In particular we still maximize over U. Thus to show that these problems are indeed equivalent it suffices to see that any feasible point for (5.4) corresponds to a feasible point for (5.1) and vice versa.

Let $(\hat{\boldsymbol{w}}, \hat{a}, \hat{\rho}, \hat{Q}^{(0)}, \hat{Q}^{(1)})$ be a feasible point for problem (5.4). Then by inequality (5.8) we have that $\rho \leq \min_{a\in[-1,1]} -V(\hat{\boldsymbol{w}}, a) = -\max_{a\in[-1,1]} V(\hat{\boldsymbol{w}}, a) \leq -V(\hat{\boldsymbol{w}}, a)$ for any $a \in [-1, 1]$. Thus by the equality condition $-V(\hat{\boldsymbol{w}}, \hat{a}) = \rho$ we have that $V(\hat{\boldsymbol{w}}, \hat{a}) = \max_{a\in[-1,1]} V(\hat{\boldsymbol{w}}, a)$. Therefore $\hat{a} \in \arg\max_{a\in[-1,1]} V(\hat{\boldsymbol{w}}, a)$ and $V(\hat{\boldsymbol{w}}, \hat{a}) \geq \underline{V}$. Hence $(\hat{\boldsymbol{w}}, \hat{a})$ is a feasible point for (5.1).

Now let $(\hat{\boldsymbol{w}}, \hat{a})$ be a feasible point for (5.1). So $\hat{a} \in \arg\max_{a\in[-1,1]} V(\hat{\boldsymbol{w}}, a)$. By Proposition 5.5 there exist $\hat{Q}^{(0)}, \hat{Q}^{(1)} \succcurlyeq 0$ and a maximal $\hat{\rho}$ such that the following system of equations is satisfied

$$\begin{aligned} c_0(\hat{\boldsymbol{w}}) - \hat{\rho} f_0(\hat{\boldsymbol{w}}) &= \hat{Q}^{(0)}_{0,0} + \hat{Q}^{(1)}_{0,0} \\ c_l(\hat{\boldsymbol{w}}) - \hat{\rho} f_l(\hat{\boldsymbol{w}}) &= \sum_{i+j=l} \hat{Q}^{(0)}_{ij} + \sum_{i+j=l} \hat{Q}^{(1)}_{ij} - \sum_{i+j=l-2} \hat{Q}^{(1)}_{ij}, \quad l = 1, \dots, d. \end{aligned}$$

Then $\hat{\rho} = \min_{a\in[-1,1]} -V(\hat{\boldsymbol{w}}, a) = -V(\hat{\boldsymbol{w}}, \hat{a})$ and therefore $(\hat{\boldsymbol{w}}, \hat{a}, \hat{\rho}, \hat{Q}^{(0)}, \hat{Q}^{(1)})$ is feasible for (5.4). □

The proof establishes that the feasible region of the original principal-agent problem (5.1) is a projection of the feasible region of the optimization problem (5.4). The first four constraints of problem (5.4) capture the agent's expected utility maximization problem. The constraints (5.4a)–(5.4d) force any value of a in a feasible solution to be the agent's optimal effort choice as well as the value of ρ to be the corresponding maximal expected utility value. Put differently, for a given contract $\boldsymbol{w}$ the first four constraints ensure an optimal effort choice by the agent.

With some additional assumptions, we can solve the optimization problem (5.4) to global optimality.

Corollary 5.1. *Suppose Assumption 5.4 holds and that the functions* $c_i, f_i : W \to \mathbb{R}$ *(in Assumption 5.4) are polynomials in* $\boldsymbol{w} \in W$*. Moreover, assume that* U *is a polynomial,* $A = [-1, 1]$*, and* W *is a basic semi-algebraic set. Then* (5.4) *is a polynomial optimization problem over a basic semi-algebraic set.*

Proof. The only problematic constraints are the semi-definiteness constraints for the matrix. However, the positive definiteness condition on the $Q^{(i)}$ is equivalent to the condition that the principal minors, that are themselves polynomials, are nonnegative. Thus the set of constraints defines a semi-algebraic set. □

If the conditions of the corollary are satisfied, we can use the methods employed in GloptiPoly, see Henrion et al., 2009, to find a globally optimal solution to the principal agent problem. That is, we can obtain a numerical certificate of global optimality. We use such an approach in Example 5.1 to ensure global uniqueness.

5.4.3 Discussion of the Polynomial Approach's Assumptions and Limitations

Theorem 5.1 rests on two key assumptions, namely that the agent's choice set is a compact interval and his expected utility function is rational in effort. The review of the mathematical background and the derivation of the theorem show that we can easily dispense with the compactness assumption and replace it by an unbounded interval such as $[0, \infty)$. While the second assumption limits the generality of the theorem, it does include the special case of agents' utility functions that are separable in wage and effort and feature a linear cost of effort (together with a rational probability distribution of outcomes).

Corollary 5.1 imposes additional assumptions on the utility functions and the set of wages; the principal's expected utility is polynomial and the agent's expected utility is rational in wages; the set of wages is a basic semi-algebraic set. The assumption on the set of wages appears to be innocuous. The assumptions on the utility functions rule out many standard utility functions such as exponential or logarithmic utility functions. Moreover, the principal's utility cannot exhibit constant risk aversion. Although the assumption on the principal's utility function is rather strong, it includes the special case of a risk-neutral principal and a polynomial probability distribution. Note that the agent's utility can be of the CRRA type. If the assumptions of Corollary 5.1 do not hold, we can still attempt to solve the final NLP with standard nonlinear optimization routines. Moreover, by invoking the Weierstrass approximation theorem that every continuous function can be uniformly approximated as closely as desired on a compact interval by a polynomial, we can argue that, at least from a theoretical viewpoint, even the assumptions on the expected utility functions in both the theorem and its corollary are not as limiting as they may appear at first.

The most serious limitation of our polynomial optimization approach is that it is not suited for a subsequent traditional theoretical analysis of the principal-agent model. A central topic of the economic literature on moral hazard problems has been the study of the nature of the optimal contract and its comparative statics properties. Studies invoking the first-order approach rely on the KKT conditions for the relaxed principal's problem to perform such an analysis. For example, Rogerson, 1985 considers the case of a separable utility function with linear cost of effort; using our notation, we can write (slightly abusing notation) $v(w_i, a) = v(w_i) + a$. Rogerson, 1985 states the KKT conditions for the relaxed principal's problem, part of which are the equations

$$\frac{u'(y_i - w_i)}{v'(w_i)} = \lambda + \delta \frac{\mu'(y_i|a)}{\mu(y_i|a)} \tag{5.14}$$

for $i = 1, 2, \ldots, N$ with Lagrange multipliers λ and δ. Rogerson, 1985 then uses these equations not only to prove the validity of the first-order approach but also to show that the optimal wage contract is increasing in the output. An analogous approach to the analysis of the optimal contract has been used in many studies, see, for example, Hölmstrom, 1979, Jewitt, 1988 and Jewitt et al., 2008. The KKT conditions for the relaxed principal's problem are rather simple since that problem has only two constraints, the participation constraint and the first-order condition for the agent's problem. The optimization problem (5.4) stated in Theorem 5.1, however, has many more constraints. In addition, the constraints characterizing the agent's optimal effort choice are not intuitive. As a result, we cannot follow the traditional approach for analyzing the principal's problem based on the new optimization problem (5.10).

Since we cannot follow the traditional theoretical route, we would instead have to rely on numerical solutions of many instances of problem (5.4) for a further analysis of the properties of the optimal contract. While at first such a numerical analysis may look rather unattractive compared to the theoretical analysis based on the first-order approach, it also offers some advantages. The first-order approach requires very strong assumptions and so applies only to a small set of principal-agent problems. A numerical analysis based on our polynomial optimization approach can examine many other problems that fall outside the classical first-order approach.

Economic theorists often make strong assumptions that allow them to prove theorems. They will generally acknowledge that their assumptions limit their analysis to a small, often measure zero, subset of economically interesting specifications of some more general and realistic theory. The only way they can justify this focus is if they believe that the results of these special cases are representative of the results in more general cases, even ones that fall far outside the set of cases their theorems examine. They believe that the assumptions are mainly necessary for the theoretical analysis leading to theoretical results and not for the theoretical results themselves. And, of course, this point certainly has some logical validity, the failure of sufficient conditions does not imply the failure of the conclusion. If there are no methods for examining the more general cases, then this approach is the only option an economist has. This paper allows us to examine the described belief in the context of principal-agent problems. Our polynomial optimization approach enables us to examine model properties for much larger classes of models than previously possible. In particular, a numerical examination of models based on the polynomial approach offers great advantages over an analysis based on the relaxed principal's problem.

The relaxed principal's problem will generally be a rather difficult nonlinear program (NLP) for many models. For example, it will have a nonlinear equation as a constraint (if the optimal effort level is interior), unless the agent's first-order condition is linear in both $\boldsymbol{w}$ and a. As a consequence, the principal's new problem will be a non-convex NLP for any utility function of the principal. The analysis of non-convex NLPs faces many theoretical and numerical difficulties. For example, the Karush-Kuhn-Tucker (KKT) conditions are often only necessary and not sufficient. Among the KKT solutions may be local maxima that are not solutions of the NLP. NLP solvers, therefore, cannot guarantee convergence to a global maximum. Furthermore, it is often rather difficult to prove that a constraint qualification holds, which is an important sufficient condition for the KKT conditions to even be necessary. However, as far as we can tell, this difficulty has been largely ignored in the literature on moral hazard problems.[11] Our approach following the corollary and using polynomial methods circumvents these problems. In fact, the approach guarantees a globally optimal solution.

5.5 The Polynomial Optimization Approach for $\boldsymbol{A} \subset \mathbb{R}^{\boldsymbol{L}}$

Principal-agent models in which the agent's action set is one-dimensional dominate not only the literature on the first-order approach but also the applied and computational literature, see for example, Araujo and Moreira, 2001, Judd and Su, 2005, Armstrong et al., 2010. However, the analysis of linear multi-task principal-agent models in Holmstrom and Milgrom, 1991 demonstrates that multivariate agent problems exhibit some fundamental differences in comparison to the common one-dimensional models. For example, the compensation paid to the agent does not only serve the dual purpose of incentive for hard work and risk-sharing but, in addition, influences the agent's attention among his various tasks. The theoretical literature that allows the set of actions to be multi-dimensional, for example, Grossman and Hart, 1983, Kadan et al., 2011, and Kadan and Swinkels, 2012, focuses on the existence and properties of equilibria. To the best of our knowledge, the first-order approach has not been extended to models with multi-dimensional action sets.

[11] For example, Rogerson, 1985 makes no reference to a constraint qualification in his derivation of (5.14). The same is true for Hölmstrom, 1979, Jewitt, 1988, Conlon, 2009, Sinclair-Desgagné, 1994, and Jewitt et al., 2008 when they state the same or an analogous first-order condition for the relaxed principal's problem.

We now extend our polynomial optimization approach to principal-agent models in which the agent has more than one decision variable, so $\boldsymbol{a} \in A \subset \mathbb{R}^L$. For this purpose, we first describe multivariate polynomial optimization. Subsequently we state and prove a generalization of Theorem 5.1. We complete our discussion with an illustration of the multi-dimensional approach by a numerical example.

5.5.1 Optimization of Multivariate Polynomials

We observed in the previous section that the reformulation of univariate polynomial optimization problems involves two steps. First, we need to rewrite the optimization problem such that the optimal value is characterized by a(n infinite) set of nonnegativity constraints. In the second step, we use a sum of squares representation of nonnegative polynomials to replace the nonnegativity constraints by finitely many convex (SDP-style) constraints in order to obtain an equivalent optimization problem. Our method for multivariate optimization follows the same general two-step reformulation approach. However, we encounter an important difficulty. While the two sets of nonnegative and positive polynomials are identical for univariate polynomials, this identity does not hold true for multivariate polynomials. A classical result of Hilbert, 1888 states that this identity holds only for quadratic multivariate polynomials and for degree 4 polynomials in two variables; or, equivalently, it holds for degree 4 homogeneous polynomials in three variables. The general lack of the identity of the sets of nonnegative and positive multivariate polynomials forces us to work directly with positive polynomials. As a result, our final optimization problem is not equivalent to the original principal-agent problem. Instead, it delivers (only) an upper bound on the optimal objective function value. Nevertheless this approach also proves very useful.

We again rely on Laurent, 2009 and Lasserre, 2010b for a review of mathematical results.

Multivariate Representation and Optimization

Putinar's Positivstellensatz is the analogue of the univariate sum of squares representation result from Proposition 5.3 for the multivariate case.

Proposition 5.6. [Putinar's Positivstellensatz, (Lasserre, 2010b, Theorem 2.14)]
Let $f, g_1 \ldots, g_m \in \mathbb{R}[\boldsymbol{x}]$ be polynomials and $K = \{\boldsymbol{x} \in \mathbb{R}^n \mid g_1(\boldsymbol{x}) \geq 0, \ldots, g_m(\boldsymbol{x}) \geq 0\} \subset \mathbb{R}^n$ a basic semi-algebraic set such that at least one of the following conditions holds,

(1) $g_1, \ldots, g_m$ are affine and K is bounded; or

(2) for some j the set $\{\boldsymbol{x} \in \mathbb{R}^n \mid g_j(\boldsymbol{x}) \geq 0\}$ is compact.

If f is strictly positive on K then

$$f = \sigma_0 + \sum_{i=1}^{m} \sigma_i g_i \tag{5.15}$$

for some $\sigma_0, \dots, \sigma_m \in \Sigma[\boldsymbol{x}]$.

The assumptions of Putinar's Positivstellensatz are not as restrictive as they may appear at first glance. For example, if we know an upper bound B such that $\|\boldsymbol{x}\|_2 \leq B$ for all $\boldsymbol{x} \in K$, then we can add the redundant ball constraint $B^2 - \sum_i x_i^2 \geq 0$. Note that in contrast to Proposition 5.3 for univariate polynomials, Putinar's Positivstellensatz does not provide any bounds on the degree of the sums of squares σ_j.

For a multivariate polynomial $p \in \mathbb{R}[x_1, x_2, \dots, x_n]$ and a nonempty semi-algebraic set $K = \{\boldsymbol{x} \in \mathbb{R}^n \mid g_1(\boldsymbol{x}) \geq 0, \dots, g_m(\boldsymbol{x}) \geq 0\}$ consider the constrained polynomial optimization problem,

$$p_{\min} = \inf_{\boldsymbol{x} \in K} p(\boldsymbol{x}). \tag{5.16}$$

Similar to the univariate case, we can rewrite this problem,

$$\begin{aligned} &\sup_{\rho} \ \rho \\ &\text{s.t. } p(\boldsymbol{x}) - \rho > 0 \ \forall \boldsymbol{x} \in K \end{aligned} \tag{5.17}$$

Since Putinar's Positivstellensatz provides a representation for strictly positive polynomials and does not bound the degrees of the sums of squares in the representation, we cannot provide a reformulation of the optimization problem (5.17) in the same simple fashion as we did in the univariate case. Instead we now consider a relaxation of the problem by restricting the degrees of the involved sums of squares. For $d \geq \max\ \{d_p, d_{g_1}, \dots, d_{g_m}\}$ consider the relaxation

$$\begin{aligned} \rho_d = \sup_{\rho, \sigma_0, \sigma_1, \dots, \sigma_m} \quad & \rho \\ \text{s.t.} \quad & p - \rho = \sigma_0 + \sum_{i=1}^{m} \sigma_i g_i \\ & \sigma_0 \in \Sigma_{2d}, \ \sigma_i \in \Sigma_{2(d-d_{g_i})} \end{aligned} \tag{5.18}$$

This problem is again an SDP and thus can be written as

$$\begin{aligned} \rho_d = \sup_{\rho, Q^{(0)}, Q^{(1)}, \dots, Q^{(m)}} \quad & \rho \\ \text{s.t. } & p - \rho = v_d^T Q^{(0)} v_d + \sum_{i=1}^{m} g_i v_{d-d_{g_i}}^T Q^{(i)} v_{d-d_{g_i}} \\ & Q^{(0)}, Q^{(i)} \succcurlyeq 0 \\ & Q^{(0)} \in \mathbb{R}^{\binom{n+d}{d} \times \binom{n+d}{d}}, \ Q^{(i)} \in \mathbb{R}^{\binom{n+d-d_{g_i}}{d-d_{g_i}} \times \binom{n+d-d_{g_i}}{d-d_{g_i}}} \\ & v_d \text{ vector of monomials } \boldsymbol{x}^{\boldsymbol{\alpha}} \text{ up to degree } d, \\ & v_{d-d_{g_i}} \text{ vector of monomials } \boldsymbol{x}^{\boldsymbol{\alpha}} \text{ up to degree } d - d_{g_i} \end{aligned} \tag{5.19}$$

The equality constraint here signifies again equality as polynomials. Thus we just have to compare the coefficients of the polynomials on the left-hand and right-hand side.[12] If the problem is infeasible, then $\rho_d = -\infty$.

[12] To avoid a messy notation we will forgo expressively writing out those equations in the multivariate case.

The optimal value ρ_d then converges from below to the optimal value $p_{\min}$ of $\inf_{\boldsymbol{x}\in K} p(\boldsymbol{x})$. In particular even if we do not obtain an explicit solution we obtain a lower bound on the optimal value $p_{\min}$. In many cases the convergence is finite, that is, for some finite $d \geq \max\ \{d_p, d_{g_1}, \ldots, d_{g_m}\}$ it holds that $\rho_d = p_{\min}$. We have the following theorem:

Proposition 5.7. [**(Lasserre, 2010b, Theorem 5.6)**] *If the assumptions of Putinar's Positivstellensatz hold, then the optimal solution ρ_d of the relaxed problem (5.18) converges (from below) to the optimal value $p_{\min}$ of the original problem (5.16) as $d \to \infty$.*

Rational Objective Function

Jibetean and Klerk, 2006 also prove analogous results for the case of multivariate rational functions. Recall the optimization problem (5.12)

$$p_{\min} = \inf_{\boldsymbol{x}\in K} \frac{p(\boldsymbol{x})}{q(\boldsymbol{x})}.$$

with $p, q \in \mathbb{R}[\boldsymbol{x}]$ and $K = \{\boldsymbol{x} \in \mathbb{R}^n \mid g_1(\boldsymbol{x}) \geq 0, \ldots, g_m(\boldsymbol{x}) \geq 0\}$. For such a set K, the following proposition states that the weak inequality in the definition of $p_{\min}$ in Proposition 5.5 can be replaced by a strict inequality.

Proposition 5.8. [**(Jibetean and Klerk, 2006, Lemma 1)**] *Suppose that K is the closure of some open connected set. Also suppose the assumptions of Proposition 5.5 hold. If p and q have no common factor then*

$$p_{min} = \sup\{\rho \mid p(\boldsymbol{x}) - \rho q(\boldsymbol{x}) > 0,\ \forall \boldsymbol{x} \in K\}.$$

Similar to the polynomial case we define the relaxation for $d \geq \max\ \{d_p, d_{g_1}, \ldots, d_{g_m}\}$,

$$\begin{aligned} \rho_d = & \sup_{\rho,\sigma_0,\sigma_1,\ldots,\sigma_m} \rho \\ & \text{s.t.} \quad p - \rho q = \sigma_0 + \sum_{i=1}^{m} \sigma_i g_i \\ & \qquad\quad \sigma_0 \in \Sigma_{2d},\ \sigma_1 \in \Sigma_{2(d-d_{g_i})} \end{aligned} \tag{5.20}$$

Proposition 5.9. [**(Jibetean and Klerk, 2006, Theorem 9)**] *Under the assumptions of Proposition 5.5 and Putinar's Positivstellensatz, the following statements hold.*

(a) If $p_{min} = -\infty$, then $\rho_d = -\infty$ for all $d = 1, 2, \ldots$.

(b) If $p_{min} > -\infty$, then $\rho_d \leq \rho_{d+1} \leq p_{min}$ for all $d = 1, 2, \ldots,$ and $\lim_{d\to\infty} \rho_d = p_{min}$.

5.5.2 The Multivariate Polynomial Optimization Approach

We now consider the principal-agent problem with a multi-dimensional set of actions, $A \subset \mathbb{R}^L$. We make the following assumption.

Assumption 5.5 (Set of Actions). *The set of actions,* $A = \{\boldsymbol{a} \in \mathbb{R}^L \mid g_1(\boldsymbol{a}) \geq 0, \dots, g_m(\boldsymbol{a}) \geq 0\}$*, is a compact semi-algebraic set with a nonempty interior.*

A multi-dimensional version of Assumption 5.4, the assumption that the agent has a rational expected utility function, imposes

$$-V(\boldsymbol{w}, \boldsymbol{a}) = -\sum_{j=1}^{N} v(w_j, \boldsymbol{a}) p_j(\boldsymbol{a}) = \frac{\sum_{\boldsymbol{\alpha}} c_{\boldsymbol{\alpha}}(\boldsymbol{w}) \boldsymbol{a}^{\boldsymbol{\alpha}}}{\sum_{\boldsymbol{\alpha}} f_{\boldsymbol{\alpha}}(\boldsymbol{w}) \boldsymbol{a}^{\boldsymbol{\alpha}}}.$$

Applying the general relaxation (5.19) to the agent's expected utility optimization problem, we obtain the following relaxation for that problem.

$$\begin{aligned}
\sup_{\rho, Q^{(0)}, Q^{(1)}, \dots, Q^{(m)}} \quad & \rho \\
\text{s.t.} \quad & \sum_{\boldsymbol{\alpha}} c_{\boldsymbol{\alpha}}(\boldsymbol{w}) \boldsymbol{b}^{\boldsymbol{\alpha}} - \rho \sum_{\boldsymbol{\alpha}} f_{\boldsymbol{\alpha}}(\boldsymbol{w}) \boldsymbol{b}^{\boldsymbol{\alpha}} = v_d^T Q^{(0)} v_d + \sum_{i=1}^{m} g_i v_{d-d_{g_i}}^T Q^{(i)} v_{d-d_{g_i}} \\
& Q^{(0)}, Q^{(i)} \succcurlyeq 0 \\
& Q^{(0)} \in \mathbb{R}^{\binom{n+d}{d} \times \binom{n+d}{d}}, \; Q^{(i)} \in \mathbb{R}^{\binom{n+d-d_{g_i}}{d-d_{g_i}} \times \binom{n+d-d_{g_i}}{d-d_{g_i}}} \\
& v_d \text{ vector of monomials } \boldsymbol{b}^{\boldsymbol{\alpha}} \text{ up to degree } d, \\
& v_{d-d_{g_i}} \text{ vector of monomials } \boldsymbol{b}^{\boldsymbol{\alpha}} \text{ up to degree } d - d_{g_i}
\end{aligned} \tag{5.21}$$

The equality in the first constraint signifies an equality of the polynomials on the left-hand and right-hand side in the variables $\boldsymbol{b}$. So, once again we need to equate the coefficients of two polynomials. These equations in turn are polynomials in the matrix elements $Q_{ij}^{(l)}$, $l = 0, 1, \dots, m$, and the variable ρ. Next we use Proposition 5.2 and replace the positive semi-definite matrices $Q^{(i)}$ by $L_{(i)} \left(L_{(i)}\right)^T$, where $L_{(i)}$ are lower triangular matrices (with a nonnegative diagonal). This transformation allows us to drop the explicit constraints on positive semi-definiteness.

For a reformulation of the original principal-agent problem from a bilevel problem to a nonlinear program, we need to characterize the optimal choice of the agent via equations or inequalities. In the case of one-dimensional effort, this reformulation is (5.4d), the generalization of which for multi-dimensional effort would be

$$\sum_{i=0}^{d} c_i(\boldsymbol{w}) \boldsymbol{a}^i - \rho \left(\sum_{i=0}^{d} f_i(\boldsymbol{w}) \boldsymbol{a}^i \right) = 0.$$

Unfortunately, due to the relaxation of the agent's problem we cannot impose this constraint, the resulting nonlinear program would most likely be infeasible. Instead, we use an idea of Couzoudis and Renner, 2013 who allow for solutions of optimization problems to be only approximately optimal; we do not force the left-hand side to be zero but instead only impose a small positive upper bound.

Now we are in the position to state and prove our second theorem, a multivariate extension of Theorem 5.1.

Theorem 5.2. *Suppose the agent's expected utility maximization problem satisfies Assumption 5.5 and the multi-dimensional version of Assumption 5.4. Let $\boldsymbol{v}_k$ be the vector of monomials in $b_1, \dots, b_L$ up to degree k. Let $d \in \mathbb{N}$ and $\varepsilon > 0$. Including $\rho \in \mathbb{R}$ and lower triangular matrices $L_{(0)} \in \mathbb{R}^{\binom{n+d}{d} \times \binom{n+d}{d}}$ and $L_{(i)} \in \mathbb{R}^{\binom{n+d-d_{g_i}}{d-d_{g_i}} \times \binom{n+d-d_{g_i}}{d-d_{g_i}}}$ for $i = 1, \dots, m$, as additional decision variables, define the following relaxation of the principal-agent problem (5.1):*

$$\max_{\boldsymbol{w}, \boldsymbol{a}, \rho, L_{(0)}, \dots L_{(m)}} U(\boldsymbol{w}, a) \quad \textit{subject to} \tag{5.22}$$

$$\begin{aligned}
\sum_{\boldsymbol{\alpha}} c_{\boldsymbol{\alpha}}(\boldsymbol{w}) \boldsymbol{b}^{\boldsymbol{\alpha}} - \rho \sum_{\boldsymbol{\alpha}} f_{\boldsymbol{\alpha}}(\boldsymbol{w}) \boldsymbol{b}^{\boldsymbol{\alpha}} &= \boldsymbol{v}_d^T L_{(0)} L_{(0)}^T \boldsymbol{v}_d + \sum_{i=1}^{m} g_i \boldsymbol{v}_{d-d_{g_i}}^T L_{(i)} L_{(i)}^T \boldsymbol{v}_{d-d_{g_i}} && \text{(5.22a)} \\
\varepsilon \sum_{\boldsymbol{\alpha}} f_{\boldsymbol{\alpha}}(\boldsymbol{w}) \boldsymbol{a}^{\boldsymbol{\alpha}} &\geq \sum_{\boldsymbol{\alpha}} c_{\boldsymbol{\alpha}}(\boldsymbol{w}) \boldsymbol{a}^{\boldsymbol{\alpha}} - \rho \sum_{\boldsymbol{\alpha}} f_{\boldsymbol{\alpha}}(\boldsymbol{w}) \boldsymbol{a}^{\boldsymbol{\alpha}} && \text{(5.22b)} \\
\sum_{i=0}^{d} c_i(\boldsymbol{w}) \boldsymbol{a}^i &\leq -\underline{V} \left(\sum_{i=0}^{d} f_i(\boldsymbol{w}) \boldsymbol{a}^i \right) && \text{(5.22c)} \\
g_i(\boldsymbol{a}) &\geq 0 \qquad \forall\, i = 1, 2, \dots, m && \text{(5.22d)} \\
\boldsymbol{w} &\in W && \text{(5.22e)}
\end{aligned}$$

This optimization problem has the following properties.

(a) Any feasible point, $\left(\hat{\boldsymbol{w}}, \hat{\boldsymbol{a}}, \hat{\rho}, \hat{L}_{(0)}, \dots, \hat{L}_{(m)}\right)$, satisfies the inequality

$$\max_{\boldsymbol{a} \in A} V(\hat{\boldsymbol{w}}, \boldsymbol{a}) - V(\hat{\boldsymbol{w}}, \hat{\boldsymbol{a}}) \leq \varepsilon. \tag{5.23}$$

(b) Let $(\overline{\boldsymbol{w}}, \overline{\boldsymbol{a}})$ be a solution of the principal-agent problem (5.1)*. Then for any $\varepsilon > 0$ there exists $d(\varepsilon) \in \mathbb{N}$ and $\overline{\rho}, \overline{L}_{(0)}, \dots, \overline{L}_{(m)}$, such that $\left(\overline{\boldsymbol{w}}, \overline{\boldsymbol{a}}, \overline{\rho}, \overline{L}_{(0)}, \dots, \overline{L}_{(m)}\right)$ is feasible for the relaxation* (5.22) *for $d = d(\varepsilon)$.*

(c) Let $(\overline{\boldsymbol{w}}, \overline{\boldsymbol{a}})$ be an optimal solution to (5.1)*. For any ε, let $d(\varepsilon)$ be as in* (b)*. Denote by $u(\varepsilon)$ the optimal value of the relaxation* (5.22) *for given ε and $d = d_\varepsilon$. Then $\lim_{\varepsilon \to 0^+} u(\varepsilon) = U(\overline{\boldsymbol{w}}, \overline{\boldsymbol{a}})$.*

(d) Again, let $(\overline{\boldsymbol{w}}, \overline{\boldsymbol{a}})$ be an optimal solution to (5.1) *and for any ε, let $d(\varepsilon)$ be as in* (b)*. Then, the set of limit points for $\varepsilon \to 0^+$ of any sequence $\left(\boldsymbol{w}^*(\varepsilon), \boldsymbol{a}^*(\varepsilon), \rho^*(\varepsilon), L_{(0)}^*(\varepsilon), \dots, L_{(m)}^*(\varepsilon)\right)$ of optimal solutions to* (5.22)*, projected onto $W \times A$, is contained in the set of optimal solutions to the original principal-agent problem* (5.1).

Before we prove the theorem, we briefly describe the optimization problem (5.22). This problem has the same objective function as the original principal-agent problem (5.1). Constraint (5.22a) uses a sum of squares representation of positive polynomials to ensure that for a contract $\boldsymbol{w}$ chosen by the principal, $-V(\boldsymbol{w},\boldsymbol{a}) \geq \rho$ for all $\boldsymbol{a} \in A$. It is important to emphasize that this equation does not only hold for the optimal choice but in fact for all possible $\boldsymbol{a} \in A$. Therefore, for the purpose of this constraint we need to duplicate the effort vector $\boldsymbol{a}$; in the functional equation (5.22a) we denote effort by $\boldsymbol{b}$. Thus again $\boldsymbol{b}$ is not a variable in the optimization problem. We obtain the equations by comparing the coefficients of the polynomials in $\boldsymbol{b}$. The positive semi-definite matrices in the relaxation of the agent's problem (5.21) are represented via products of lower triangular matrices. Proposition 5.2 shows that any positive semi-definite matrix can be represented in this fashion (even having the property that all diagonal elements are nonnegative). Put differently, constraint (5.22a) ensures that $-\rho$ is an upper bound on the agent's possible expected utility levels. Next, constraint (5.22b) imposes a lower bound on the agent's expected utility level, namely $V(\boldsymbol{w},\boldsymbol{a}) + \varepsilon \geq -\rho$. Therefore, the constraints (5.22a) and (5.22b) force the value of a in any feasible solution to result in a utility for the agent satisfying $-\rho - \varepsilon \leq V(\boldsymbol{w},\boldsymbol{a}) \leq -\rho$. That is, for a given contract $\boldsymbol{w}$ the first two constraints ensure an effort choice by the agent that is within ε of being optimal. The last three constraints are straightforward. Constraint (5.22c) is the transformed participation constraint for the agent's rational expected utility function. Constraint (5.22d) defines the set of the feasible actions and constraint (5.22e) is just the constraint on the compensation scheme from the original principal-agent problem (5.1).

Proof. Under the assumptions of the theorem, the agent's constraints satisfy the conditions of Putinar's Positivstellensatz and so we obtain the sums-of-squares representation for the agent's problem. For fixed d we then restrict the degree of the sum of squares coefficients as is done in the relaxation.

(a) Every feasible point $\left(\hat{\boldsymbol{w}}, \hat{\boldsymbol{a}}, \hat{\rho}, L_{(0)}, \ldots, L_{(m)}\right)$ provides an upper bound $-\hat{\rho}$ on the maximal value of $V(\hat{\boldsymbol{w}},\boldsymbol{a}) = -\frac{\sum_{\boldsymbol{\alpha}} c_{\boldsymbol{\alpha}}(\hat{\boldsymbol{w}})\boldsymbol{a}^{\boldsymbol{\alpha}}}{\sum_{\boldsymbol{\alpha}} f_{\boldsymbol{\alpha}}(\hat{\boldsymbol{w}})\boldsymbol{a}^{\boldsymbol{\alpha}}}$, since (5.22a) implies that

$$\sum_{\boldsymbol{\alpha}} c_{\boldsymbol{\alpha}}(\boldsymbol{w})\boldsymbol{b}^{\boldsymbol{\alpha}} - \hat{\rho}\sum_{\boldsymbol{\alpha}} f_{\boldsymbol{\alpha}}(\boldsymbol{w})\boldsymbol{b}^{\boldsymbol{\alpha}} \geq 0$$

and so, $-\hat{\rho} \geq \max_{\boldsymbol{a}\in A} V(\hat{\boldsymbol{w}},\boldsymbol{a}) \geq V(\hat{\boldsymbol{w}},\hat{\boldsymbol{a}})$. Moreover, constraint (5.22b) implies that

$$\varepsilon \geq -\hat{\rho} - V(\hat{\boldsymbol{w}},\hat{\boldsymbol{a}}) \geq \max_{\boldsymbol{a}\in A} V(\hat{\boldsymbol{w}},\boldsymbol{a}) - V(\hat{\boldsymbol{w}},\hat{\boldsymbol{a}}).$$

Thus, condition (5.23) holds.

(b) Under the assumptions of the theorem, Proposition 5.9 implies that for each fixed $\overline{\boldsymbol{w}}$ and a given $\varepsilon > 0$ there exists a d such that $V(\boldsymbol{w},\boldsymbol{a}) - \rho$ has the representation (5.15) of Putinar's Positivstellensatz with degree d coefficients. For this d, problem (5.22) has a nonempty feasible region.

(c) Recall the agent's optimal value function $\Psi : W \to \mathbb{R}$ from the proof of Proposition 5.1. The projection of the set of feasible points of problem (5.22) to $W \times A$ is a subset of

$$S(\varepsilon) = \{(\boldsymbol{w}, \boldsymbol{a}) \in W \times A \mid \Psi(\boldsymbol{w}) - V(\boldsymbol{w}, \boldsymbol{a}) \leq \varepsilon\}$$

and, by (b), contains $(\overline{\boldsymbol{w}}, \overline{\boldsymbol{a}})$. Let $v(\varepsilon) = \max_{(\boldsymbol{w},\boldsymbol{a}) \in S(\varepsilon)} U(\boldsymbol{w}, \boldsymbol{a})$. Then

$$U(\overline{\boldsymbol{w}}, \overline{\boldsymbol{a}}) \leq u(\varepsilon) \leq v(\varepsilon).$$

Furthermore, since Ψ and V are continuous (Berge's Maximum Theorem), the set $S(\varepsilon)$ is upper hemicontinuous and uniformly compact near 0.[13] By Hogan, 1973, Theorem 5 it follows that v is upper semi-continuous and thus we have

$$U(\overline{\boldsymbol{w}}, \overline{\boldsymbol{a}}) \leq \liminf_{\varepsilon \to 0^+} u(\varepsilon) \leq \limsup_{\varepsilon \to 0^+} u(\varepsilon) \leq \limsup_{\varepsilon \to 0^+} v(\varepsilon) \leq v(0) = U(\overline{\boldsymbol{w}}, \overline{\boldsymbol{a}})$$

Therefore, $\lim_{\varepsilon \to 0^+} u(\varepsilon) = U(\overline{\boldsymbol{w}}, \overline{\boldsymbol{a}})$.

(d) Consider any limit point $(\boldsymbol{w}_0, \boldsymbol{a}_0) \in W \times A$ and any sequence $(\boldsymbol{w}_\varepsilon, \boldsymbol{a}_\varepsilon)$ converging to it for $\varepsilon \to 0$. Condition (c) implies that $U(\boldsymbol{w}_\varepsilon, \boldsymbol{a}_\varepsilon) \to U(\overline{\boldsymbol{w}}, \overline{\boldsymbol{a}})$. By continuity of Ψ and V we also have

$$\lim_{\varepsilon \to 0^+} (\Psi(\boldsymbol{w}_\varepsilon) - V(\boldsymbol{w}_\varepsilon, \boldsymbol{a}_\varepsilon)) = \Psi(\boldsymbol{w}_0) - V(\boldsymbol{w}_0, \boldsymbol{a}_0) = 0.$$

Thus $(\boldsymbol{w}_0, \boldsymbol{a}_0)$ is feasible for (5.1) and attains the optimal value.

This completes the proof of Theorem 5.2. □

Some comments on the technical convergence results of Theorem 5.2 are in order. For the one-dimensional effort case, Theorem 5.1 provides a single well-defined optimization problem that is equivalent to the original principal-agent problem. Ideally, we would like to obtain a similar result for the multi-dimensional effort case. Unfortunately, in general that is impossible. A comparison of the sum of squares representation results for univariate and multivariate polynomials reveals the critical difference between the two cases. Proposition 5.3, the 'Positivstellensatz' for univariate polynomials, provides a sum of squares representation of nonnegative univariate polynomials with an explicit (small) bound on the degree of the involved sums of squares. Proposition 5.6, Putinar's Positivstellensatz, provides a sum of squares representation of positive multivariate polynomials; however, there is no a-priori upper bound on the degree of the involved sums of squares. In fact, from a purely theoretical viewpoint, the necessary degree may be infinite. As a result, any finite-degree representation as in (5.18) may only constitute a relaxation of the original polynomial optimization problem.

Once we have computed a solution we can always verify if it is feasible. To accomplish this we fix $\boldsymbol{w}$ and solve the polynomial optimization problem for the agent to global

[13] Upper hemicontinuity at 0 means that for any sequence $\varepsilon^k \to 0$, $s^k \in S(\varepsilon^k)$ and $s^k \to s$ implies $s \in S(0)$.

optimality. We use GloptiPoly (Henrion et al., 2009), a program written for Matlab, that employs the moment relaxation approach to solving polynomial optimization problems.

In light of the theoretical difficulties for general multivariate polynomials, it is of great interest to characterize polynomial optimization problems that offer a guaranteed convergence of the relaxation for finite d. If both the objective function and the constraints are s.o.s. convex, then the convergence is finite, see Lasserre, 2010b, Theorem 5.15.[14] Also, if the objective function is strictly convex and the constraints are convex, then convergence is finite, see Lasserre, 2010b, Theorem 5.16. The problem of finite convergence continues to be an active research issue in algebraic geometry. For example, Nie, 2012 proved finite convergence under a regularity condition on the set of constraints. His approach requires a reformulation of the problem by adding constraints consisting of minors of a Jacobian derived from the KKT conditions. Unfortunately, it appears to be rather difficult to check the regularity condition in applications.

As a final remark, we point out that Schmüdgen's Positivstellensatz, see Schmüdgen, 1991, yields a representation of multivariate positive polynomials that is different than that of Putinar's Positivstellensatz. This representation is slightly more general but requires higher degree sums of squares. Therefore, it appears to be less attractive for economic applications.

5.5.3 A Multivariate Example

Example 5.2. Let the set of outcomes be $\{0, 3, 6\}$ with probabilities

$$\left\{\frac{1 + a/2 + b}{1 + a + b}, \frac{b}{1 + a + b}, \frac{a/2 - b}{1 + a + b}\right\},$$

satisfying the constraints

$$b \geq 0, \quad a - 2b \geq 0,$$

which assure that the probability functions are nonnegative. The outcome distribution has mean and variance

$$\frac{3(a - b)}{1 + a + b} \quad \text{and} \quad \frac{9\left(2a + a^2 - 3b + ab - 4b^2\right)}{(1 + a + b)^2},$$

respectively. Note that the effort a increases both the expected value and the variance of the outcome. On the contrary, the effort b decreases the expectation and the variance.

The principal's and the agent's Bernoulli utility functions are

$$u(y, w) = -\left(-6 - w + y\right)^2 \quad \text{and} \quad v(a, b, w) = (1 + a + b)\left(-a - \frac{b}{10} + \log\left(1 + w\right)\right),$$

respectively. The expected utility of the agent is

$$\frac{1}{10}\left(-10a - 10a^2 - b - 11ab - b^2 + 10b\log\left(1 + w_2\right) + 5(a - 2b)\log\left(1 + w_3\right)\right) + \left(1 + \frac{a}{2} + b\right)\log\left(1 + w_1\right)$$

[14] A polynomial f is called *s.o.s.* convex, iff $\nabla^2 f = WW^T$ for some matrix W.

and the expected utility of the principal is

$$-\frac{a\left(36+12w_1+w_1^2+w_3^2\right)+2\left((6+w_1)^2+b\left(45+12w_1+w_1^2+6w_2+w_2^2-w_3^2\right)\right)}{2(1+a+b)}.$$

We observe that the largest degree in the variables a and b is two. So, we can choose the relaxation order to be one, that is, all the matrices appearing will be of size 3×3, $L_k=(s_{k,i,j})_{i,j=1,2,3}$, where L_k is a lower triangular matrix with nonnegative diagonal. The sum of squares multipliers now appear as follows

$$\begin{aligned}\sigma_k =& s_{k,1,1}^2+2as_{k,1,1}s_{k,2,1}+a^2(s_{k,2,1}^2+s_{k,2,2}^2)+2bs_{k,1,1}s_{k,3,1}+b^2(s_{k,3,1}^2+s_{k,3,2}^2+s_{k,3,3}^2)+\\ & ab(2s_{k,2,1}s_{k,3,1}+2s_{k,2,2}s_{k,3,2}).\end{aligned}$$

Thus the coefficients in the variables a,b of the following polynomial have to be zero

$$V(a,b,w_1,w_2,w_3)+\rho+\sigma_0+b\sigma_1+(a-2b)\sigma_2+(1-a)\sigma_3.$$

This leads to the following equations

$$\begin{aligned}
0 =& s_{1,3,1}^2+s_{1,3,2}^2+s_{1,3,3}^2-s_{2,3,1}^2-s_{2,3,2}^2-s_{2,3,3}^2\\
0 =& \frac{1}{2}\left(s_{2,2,1}^2+s_{2,2,2}^2\right)-s_{3,2,1}^2-s_{3,2,2}^2\\
0 =& -1+s_{0,2,1}^2+s_{0,2,2}^2+s_{2,1,1}s_{2,2,1}-2s_{3,1,1}s_{3,2,1}+s_{3,2,1}^2+s_{3,2,2}^2\\
0 =& s_{1,2,1}^2+s_{1,2,2}^2-s_{2,2,1}^2-s_{2,2,2}^2+s_{2,2,1}s_{2,3,1}+s_{2,2,2}s_{2,3,2}-2(s_{3,2,1}s_{3,3,1}+s_{3,2,2}s_{3,3,2})\\
0 =& -\frac{11}{10}+2(s_{0,2,1}s_{0,3,1}+s_{0,2,2}s_{0,3,2})+2s_{1,1,1}s_{1,2,1}-2s_{2,1,1}s_{2,2,1}+\\
& s_{2,1,1}s_{2,3,1}-2s_{3,1,1}s_{3,3,1}+2(s_{3,2,1}s_{3,3,1}+s_{3,2,2}s_{3,3,2})
\end{aligned}$$

$$\begin{aligned}
0 =& 2(s_{1,2,1}s_{1,3,1}+s_{1,2,2}s_{1,3,2})-2(s_{2,2,1}s_{2,3,1}+s_{2,2,2}s_{2,3,2})+\frac{1}{2}\left(s_{2,3,1}^2+s_{2,3,2}^2+s_{2,3,3}^2\right)-\\
& s_{3,3,1}^2-s_{3,3,2}^2-s_{3,3,3}^2\\
0 =& -\frac{1}{10}+s_{0,3,1}^2+s_{0,3,2}^2+s_{0,3,3}^2+2s_{1,1,1}s_{1,3,1}-2s_{2,1,1}s_{2,3,1}+s_{3,3,1}^2+s_{3,3,2}^2+s_{3,3,3}^2\\
0 =& \rho+s_{0,1,1}^2+s_{3,1,1}^2+\log\left(1+w_1\right)\\
0 =& -\frac{1}{10}+2s_{0,1,1}s_{0,3,1}+s_{1,1,1}^2-s_{2,1,1}^2+2s_{3,1,1}s_{3,3,1}+\log\left(1+w_1\right)+\\
& \log\left(1+w_2\right)-\log\left(1+w_3\right)\\
0 =& -1+2s_{0,1,1}s_{0,2,1}+\frac{s_{2,1,1}^2}{2}-s_{3,1,1}^2+2s_{3,1,1}s_{3,2,1}+\frac{1}{2}\log\left(1+w_1\right)+\frac{1}{2}\log\left(1+w_3\right).
\end{aligned}$$

We set the reservation utility to $\frac{3}{2}$ and solve this problem with Ipopt. We cannot use Gloptipoly here since the number of variables is too large. We obtain the following solution,

$$a=0.34156,\ b=0.17078,\ w_1=2.7295,\ w_2=4.0491,\ w_3\geq 0.$$

The principal's expected utility is -73.210 and the agent's is $\frac{3}{2}$.

5.6 Conclusion

In this paper we have presented a polynomial optimization approach to moral hazard principal-agent problems. Under the assumption that the agent's expected utility function is a rational function of his effort, we can reformulate the agent's maximization problem as an equivalent system of equations and inequalities. This reformulation allows us to transform the principal-agent problem from a bilevel optimization problem to a nonlinear program. Furthermore, under the assumptions that the principal's expected utility is polynomial and the agent's expected utility is rational in wages (as well as mild assumptions on the effort set and the set of wage choices), we show that the resulting NLP is a polynomial optimization problem. Therefore, techniques from global polynomial optimization enable us to solve the NLP to global optimality. After this analysis of principal-agent problems with a one-dimensional effort choice for the agent, we have also presented a polynomial optimization approach for problems with multi-dimensional effort sets. The solution approach for solving such multi-dimensional problems rests on the same ideas as the approach for the one-dimensional effort model, however, it is technically more difficult. Most importantly, we cannot provide an exact reformulation of the agent's problem but only a relaxation of that problem. Despite this theoretical limitation, the relaxation appears to be often exact in applications.

Our polynomial optimization approach has a number of attractive features. First, we need neither the Mirrlees-Rogerson (or Jewitt) conditions of the classical first-order approach nor the assumption that the agent's utility function is separable. Second, under the additional aforementioned assumptions on the utility functions, the final NLP is a polynomial problem that can be solved to global optimality without concerns about constraint qualifications. Third, unlike the first-order approach, the polynomial approach extends to models with multi-dimensional effort sets.

The technical assumptions underlying the polynomial approach, while limiting, are not detrimental. The most serious limitation of our polynomial optimization approach is that it is not suited for a subsequent traditional theoretical analysis of the principal-agent model. Despite this shortcoming, the polynomial approach can serve as a useful tool to examine the generality of the insights derived from the very restrictive first-order approach. The ability of the approach to find global solutions to principal-agent problems is one of its hallmarks.

6 Computing Generalized Nash Equilibria by Polynomial Programming [1][2]

Abstract

We present a new way of solving generalized Nash equilibrium problems. We assume the feasible set to be compact. Furthermore all functions are assumed to be polynomials. However we do not impose convexity on either the utility functions or the action sets. The key idea is to use Putinar's Positivstellensatz, a representation result for positive polynomials, to replace each agent's problem by a convex optimization problem. The Nash equilibria are then feasible solutions to a system of polynomial equations and inequalities. Our application is a model of the New Zealand electricity spot market with transmission losses based on a real dataset.

6.1 Introduction

There has been a lot of interest in the computation of normalized Nash equilibria since Rosen, 1965 introduced them. In essence the approach is to reformulate the problem either as a variational inequality or to use penalty functions or Nikaido-Isoda-type functions. The one thing all computational papers have in common is the assumption of player convexity on the utility functions of the players (Facchinei and Kanzow, 2007).

There are some attempts in the literature to extend the solution approach to quasi-convex problems. However error bounds are only provided under some strong monotonicity assumptions (Aussel et al., 2011; Aussel and Dutta, 2011).

Unlike the usual approaches to generalized Nash equilibrium problems (GNEPs), we do not need any convexity assumptions on our functions and sets. We are also not restricted to normalized equilibria. However, since we employ tools from real algebraic geometry we require every constraint and objective function to be polynomial. Note that the KKT conditions do not provide sufficient conditions in the case of non-convex functions. Thus, instead of the usual approach, we replace each agent's problem with the convex relaxation obtained by Putinar's Positivstellensatz. Each of these problems is then a parametrized, semi-definite optimization problem. The corresponding optimality conditions however are a system of polynomial equations and inequalities. We show that some equilibria

[1] (Couzoudis and Renner, 2013)

[2] We thank Didier Aussel, Hans-Jakob Lüthi, Cordian Riener and two anonymous referees.

are feasible points to this system and that any other feasible point is almost optimal. In cases in which we have a representation for a non-negative polynomial, only equilibria are feasible points. We find those points with the solver Ipopt.

We could also do this, with just slight modification, for rational functions. The relevant theorems to apply this approach for rational functions can be found in Jibetean and Klerk, 2006. Our attention is on a non-cooperative, single stage game in normal form, a onetime situation without reoccurrence, and a finite number of players who move simultaneously.

Let $N \in \mathbb{N}$ be the finite number of players in the examined N-person game. Every player $\nu \in I$ with $I := \{1, \dots, N\}$ chooses his strategy x^ν from the strategy set $X_\nu(x^{-\nu}) \subseteq \mathbb{R}^{n_\nu}$, where n_ν is a positive integer. For the sake of simplicity the strategy set of all other players except ν is given by $x^{-\nu} := (x^{\nu'})_{\nu'=1,\nu'\neq\nu}^{N} \in \mathbb{R}^{n_{-\nu}}$. The complete strategy vector of all players is specified with $x := (x^\nu)_{\nu=1}^N \in \mathbb{R}^n$ and $n := \sum_{\nu=1}^N n_\nu$. Hence $n_{-\nu} = n - n_\nu$ and therefore the tuple of strategies for the whole game has dimension n:

$$x := (x^\nu, x^{-\nu})^T = (x^1, \dots, x^{\nu-1}, x^\nu, x^{\nu+1}, \dots, x^N)^T \in \mathbb{R}^n.$$

The scope of action $X_\nu(x^{-\nu})$ for every player ν in this game is influenced by the strategies of the opponents $x^{-\nu}$. For $\nu = 1, \dots, N$ let $X_\nu : \mathbb{R}^{n_{-\nu}} \rightrightarrows \mathbb{R}^{n_\nu}$ be a point-to-set-mapping and for every fixed $x^{-\nu}$ a subset of $\mathbb{R}^{n_\nu}$.The allowed strategy set of the player ν has then the following form:

$$X_\nu(x^{-\nu}) := \{x^\nu \mid (x^\nu, x^{-\nu}) \in X\}.$$

Thereby $X \subseteq \mathbb{R}^n$ is assumed to be nonempty and compact which implies the compactness of every set $X_\nu(x^{-\nu})$. For this work X has the following structure

$$X := \{x \in \mathbb{R}^n \mid g_\nu(x) \geq 0, h_\nu(x^\nu) \geq 0 \ \forall \nu = 1, \dots, N\}$$

where the functions $g_\nu : \mathbb{R}^n \to \mathbb{R}^{l_\nu}$ are constraints influenced by other players, $h_\nu : \mathbb{R}^{n_\nu} \to \mathbb{R}^{m_\nu}$ are constraints specific to each player ν and l_ν, m_ν are positive integers. Combining the two set declarations X and X_ν gives us

$$X_\nu(x^{-\nu}) := \{x^\nu \mid g_\nu(x^\nu, x^{-\nu}) \geq 0,\ h_\nu(x^\nu) \geq 0\} \qquad \forall \nu = 1, \dots, N$$

and at the same time the possibility to define the feasible set for any point $x \in \mathbb{R}^n$:

$$\Omega(x) := X_1(x^{-1}) \times \cdots \times X_N(x^{-N}).$$

The last and yet missing basic element is the payoff. The assessment of the player's strategy set X_ν and therefore the choice of action x^ν of player ν depends on the corresponding utility or payoff function $\theta_\nu : \mathbb{R}^n \to \mathbb{R}$. A finite N-person game is defined by the triple $(I, (X_\nu)_{\nu \in I}, (\theta_\nu)_{\nu \in I})$.

Furthermore a theoretical construct is needed to decide which player choices are rational and optimal. In this case the payoff function θ_ν is assumed to be a cost or loss function. Every player ν tries to minimize his loss given the exogenous decision of the competition:

$$R_v(x^{-\nu}): \qquad \min_{x^\nu \in \mathbb{R}^{n_\nu}} \theta_\nu(x^\nu, x^{-\nu}) \\ s.t. \quad x^\nu \in X_\nu(x^{-\nu}).$$

The solution set mapping R_v is also the best-response mapping of player ν.

Definition 6.1. A strategy $x^\star \in \Omega(x^\star)$ is a Generalized Nash Equilibrium (GNE), if and only if $x^{\star,\nu}$ satisfies the following inequality:

$$\theta_\nu(x^{\star,\nu}, x^{\star,-\nu}) \ \leq \ \theta_\nu(x^\nu, x^{\star,-\nu}) \qquad \forall x^\nu \in X_\nu(x^{\star,-\nu}),\ \forall \nu \in \{1,\ldots,N\}.$$

The Nash Equilibrium is therefore, for any player ν, the optimal decision given the expected choice $x^{\star,-\nu}$ of the fellow players.

6.2 The Model

"The [Electricity] Authority [of New Zealand] is responsible for ensuring the effective day-to-day operation of the electricity system and markets through the operation of core system and market services in accordance with the [Electricity Industry Participation] Code [2010]."[3] The corresponding tasks are assigned to these market operation service providers: Registry Manager, Reconciliation Manager, Pricing Manager, Clearing Manager, Information System Manager, System Operator and Inter-island Financial Transmission Rights Manager.[4] Our focus here lies on the Independent System Operator (ISO) and on Part 13 of the Electricity Industry Participation Code 2010 which sets out the trading arrangements.

New Zealand consists of two main islands, the energy demanding north and the energy producing south with a High Voltage Direct Current (HVDC) Link between them. Our highly abstract network is therefore composed of only two nodes and two directed edges. The arc t_1 is directed from the north island to the south island and t_2 is the inverted arc of t_1. The power flow capacity of both arcs is $\bar{t}$.

"Each [Energy Producer (EP) or] generator ... must submit an offer to the system operator for each trading period in the schedule period, under which the generator is prepared to sell electricity to the clearing manager, and ensure that the system operator receives an offer at least 71 trading periods before the beginning of the trading period to which the offer applies."[5] An offer submitted by a generator may have a maximum of 5 price bands for each trading period and may not exceed, for each trading period, the generator's reasonable estimate of the quantity of electricity capable of being supplied at that node. The price offered in each band must increase progressively from band to band as the aggregate quantity increases. An exception are intermittent generators with

[3] Electricity Authority NZ, 2012
[4] Electricity Authority NZ, 2012
[5] Electricity Authority NZ, 2010b, p. 13.6

a maximum of 1 price band for each trading period and co-generators with a maximum of 2 price bands for each trading period.[6] "For each price band, an ... offer must specify a quantity expressed in megawatt [MW] to not more than 3 decimal places. The minimum quantity that may be bid or offered in a price band for a trading period is 0.000 MW."[7] "Prices in ... offers must be expressed in dollars and whole cents per megawatthour [MWh] ... There is no upper limit on the prices that may be specified and the lower limit is \$0.00 per MWh ...."[8]

It is clear that the inverse supply function of the power producers would be piecewise linear and an extension to the here presented model is straightforward. In the literature the approach for the New Zealand electricity spot market is to make an quadratic approximation of the inverse supply function. This is shown in Aussel et al., 2012 which in turn is based on Hobbs et al., 2000 and yields in Aussel's example depending on the control variables quadratic or even non-convex cubic polynomials as revenue functions.

For this proof of concept we restrict the offers from energy producers to one price band with a lower level of zero and a static demand δ_N for the north island and δ_S for the south island. Hence, in this scenario there are only two kinds of players, the Independent System Operator and the Energy Producers. The price per Gigawatt-hour [GWh] of each EP is denoted by a^ν and q^ν is the maximum amount of power offered in GWh. The ISO can choose for any energy offer q^ν the effective fraction c^ν. Then the revenue function for each EP is given by $a^\nu c^\nu q^\nu$. All terms are with respect to one time period which is one day. This is in accordance to the pricing manager who sets the final prices on a daily basis.

The ISO wants to achieve the social optimum in minimizing the expenditure for the needed energy and the loss through transporting energy.

$$\begin{aligned}
\min_{c^\nu,\ t_1,\ t_2} \quad & \sum_{\nu=1}^{N} a^\nu c^\nu q^\nu \\
\text{s.t.} \quad & \sum_{\nu=1}^{i} c^\nu q^\nu - t_1 - \lambda(t_1)^2 + t_2 - \lambda(t_2)^2 - \delta_N \geq 0 \\
& \sum_{\nu=i+1}^{N} c^\nu q^\nu - t_2 - \lambda(t_2)^2 + t_1 - \lambda(t_1)^2 - \delta_S \geq 0 \\
& 0 \leq c^\nu \leq 1 \quad \forall \nu = 1, \ldots, N \\
& \bar{t} \geq t_1 \geq 0,\ \bar{t} \geq t_2 \geq 0.
\end{aligned} \tag{6.1}$$

The objective function is linear and the constraint set is convex, so the ISO problem is clearly convex. Both network constraints are pretty self explanatory except for the quadratic terms $\lambda\left(t_{\{1,2\}}\right)^2$. Transmitting power is not lossless hence we form a term for

[6] Electricity Authority NZ, 2010b, pp. 13.9, 13.12
[7] Electricity Authority NZ, 2010b, p. 13.16
[8] Electricity Authority NZ, 2010b, p. 13.15

the arising heat loss using Joule's and Ohm's laws (Tipler et al., 2000):

$$\left.\begin{aligned} P &= UI \\ P_{loss} &= I^2 R \end{aligned}\right\} P_{loss} = R \cdot \left(\frac{P}{U}\right)^2 = R' \cdot l \cdot \left(\frac{P}{U}\right)^2 = \frac{R' \cdot l}{U^2} \cdot P^2.$$

The unit of P is watt and not equal to watt hour unit of $t_{\{1,2\}}$ thus a conversion is necessary:

$$\frac{R' \cdot l}{U^2} \cdot \left(\frac{P_h}{24}\right)^2 \cdot 24 = \frac{R' \cdot l}{U^2 \cdot 24} \cdot (P_h)^2 = \lambda \cdot (P_h)^2 .$$

The meanings of the different terms are:

P:	power	R:	electrical resistivity at 20°C
U:	voltage	R':	electrical resistivity/km at 20°C
I:	current	l:	length of transmission line.

For the cost function of the EPs we seek a technology-driven approach. If something is burned up, the produced heat can be measured in Joule [J]. The heat conversion rate η_ν now answers the question how much of that heat can be converted into electric energy. This principle can of course be extended into non heat based technologies. The player-specific cost function for the EPs has now the following form:

$$C_\nu(q^\nu) = \overbrace{(\eta_\nu \cdot q^\nu)}^{GJ/GWh \times GWh} \cdot \underbrace{(\phi_\nu + \tau_\nu)}_{\text{fuel price: } \$/GJ} + \overbrace{o_\nu}^{\$/GWh} q^\nu .$$

The parameter η_ν is technology dependent. The fuel price consists of the commodity price ϕ_ν and the delivery costs τ_ν. The coefficient for other variable costs such as operation and maintenance is denoted as o_ν. Furthermore each EP ν has a limited power generating capacity κ_ν given in GWh. The optimization problem for the profit maximizing EPs is then as follows:

$$\begin{aligned} \max_{a^\nu,\ q^\nu}\ & a^\nu c^\nu q^\nu - (\eta_\nu c^\nu q^\nu)(\phi_\nu + \tau_\nu) - o_\nu c^\nu q^\nu \\ \text{s.t. }\ & \kappa_\nu \geq q^\nu \geq 0, \\ & a^\nu - \eta_\nu(\phi_\nu + \tau_\nu) - o_\nu \geq 0, \\ & \sum_{i \neq \nu}^{N} a^i c^i q^i - a^\nu c^\nu \sum_{i=1}^{N} c^i q^i \geq 0, \\ & \sum_{i=1}^{N} q^i - \lambda\left((t_1)^2 + (t_2)^2\right) - \delta_N - \delta_S \geq 0. \end{aligned} \tag{6.2}$$

A static upper bound on the price is an invitation to a price agreement. All producers would settle at this upper bound. On the other hand competitive behavior is to underbid your opponents, as long as your price is above your marginal costs, and given you are chosen by the ISO. So the two constraints on α^ν do exactly that. The last constraint is introduced to avoid the infeasibility of the ISO's problem.

Obviously the constraint set is compact. The Eigenvalues of the Hessian of the objective function are $\{-c^\nu, c^\nu\}$ with $c^\nu \in [0,1]$. Thus the Hessian is indefinite for non trivial c^ν and therefore this simple model does not have a distinct curvature. With this game we can obviously take full advantage of our approach.

6.3 Sum of Squares Optimization

In outlining the theory behind our solution approach we follow Laurent, 2009 and Lasserre, 2010b. We will use the following notations:

- $\mathbb{R}[x_1, \dots, x_n]$ the ring of polynomials in n variables over the real numbers.
- $\boldsymbol{x} = (x_1, \dots, x_n)$ as shorthand notation.
- $f \in \mathbb{R}[\boldsymbol{x}]$, $f(\boldsymbol{x}) = \sum_\alpha a_\alpha \boldsymbol{x}^\alpha$ where $\alpha \in \mathbb{N}^n$ and $\boldsymbol{x}^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$. Then $\deg(f) = \max_{\{\alpha | a_\alpha \neq 0\}} |\alpha| = \max_{\{\alpha | a_\alpha \neq 0\}} \sum_i \alpha_i$.
- For any polynomial $g \in \mathbb{R}[\boldsymbol{x}]$ denote $d_g = \left\lceil \frac{\deg(g)}{2} \right\rceil$.

6.3.1 Basic Definitions and Theorems

First we will introduce the class of convex optimization problems that is of interest to us here. To do this we need a few basic notions from linear algebra.

Definition 6.2. A symmetric matrix $Q \in \mathbb{R}^{n \times n}$ is called *positive semidefinite*, if and only if $w^T Q w \geq 0$ for all $w \in \mathbb{R}^n$. It is denoted $Q \succcurlyeq 0$.

In the next proposition we recall a condition for a matrix to be positive semidefinite.

Proposition 6.1. *Let $Q \in \mathbb{R}^{n \times n}$ be a symmetric matrix with rank m. Then the following statements are equivalent*

(a) Q is positive semidefinite.

(b) There exists a lower triangular matrix $L \in \mathbb{R}^{n \times n}$ with nonnegative diagonal such that $Q = LL^T$.

Note here that the equivalent statement for positive semidefiniteness can be expressed by polynomial equations and inequalities.

Now we introduce some basic notions from real algebraic geometry.

Definition 6.3. Let $g_1, \dots, g_k \in \mathbb{R}[\boldsymbol{x}]$. We call the set

$$K = \{\boldsymbol{x} \mid g_1(\boldsymbol{x}) \geq 0, \dots, g_k(\boldsymbol{x}) \geq 0\}$$

a *basic semi-algebraic set.*

Definition 6.4. (a) A polynomial $\sigma \in \mathbb{R}[\boldsymbol{x}]$ of degree $2d$ is called a *sum of squares*, if and only if there exists polynomials $p_1, \dots, p_m \in \mathbb{R}[\boldsymbol{x}]$ such that $\sigma = \sum_i p_i^2$.

(b) Let $\Sigma[\boldsymbol{x}] \subset \mathbb{R}[\boldsymbol{x}]$ denote the set of sum of squares.

(c) Let $\Sigma_{2d}[\boldsymbol{x}] \subset \mathbb{R}[\boldsymbol{x}]$ denote the set of sum of squares up to degree $2d$.

Since a sum of squares is always a nonnegative function it is easy to see that a polynomial can only be a sum of squares if it has even degree. The question whether a positive polynomial is a sum of squares however is much more involved. There is the following representation result.

Theorem 6.1 (Putinar's Positivstellensatz)**.** *Lasserre, 2010b, Th.2.14*
Let $f, g_1 \dots, g_m \in \mathbb{R}[\boldsymbol{x}]$ be polynomials and

$$K = \{\boldsymbol{x} \in \mathbb{R}^n \mid g_1(\boldsymbol{x}) \geq 0, \dots, g_m(\boldsymbol{x}) \geq 0\} \subset \mathbb{R}^n$$

a semi-algebraic set such that one of the following holds,

- *g_i are affine and K is a bounded polyhedron.*
- *For some j the set $\{\boldsymbol{x} \in \mathbb{R}^n \mid g_j(\boldsymbol{x}) \geq 0\}$ is compact.*

If f is strictly positive on K then

$$f = \sigma_0 + \sum_{i=1}^{m} \sigma_i g_i \tag{6.3}$$

for some $\sigma_0, \dots, \sigma_m \in \Sigma[\boldsymbol{x}]$.

The conditions for this theorem to hold are not as restrictive as it might seem at first glance. If we know an N such that $\|\boldsymbol{x}\|_2 \leq N$ for all $\boldsymbol{x} \in K$, we then can add the redundant ball constraint $\sum_i x_i^2 \leq N^2$. The problem with the theorem is that we do not know the degrees of the coefficients σ_i. There exist bounds on their degree Lasserre, 2010b, Th.2.16, however for practical purposes these are of no use to us.

In an example below we will later see that we can use this kind of representation even without the compactness assumption. Additionally in many models one has an intuitive idea about the form of the solution and can pick an N accordingly.

6.3.2 Optimization of Polynomials over Semialgebraic Sets

Let $p, g_1, \dots, g_m \in \mathbb{R}[\boldsymbol{x}]$ and $K = \{\boldsymbol{x} \mid g_1(\boldsymbol{x}) \geq 0, \dots, g_m(\boldsymbol{x}) \geq 0\}$. We want to solve the following optimization problem

$$\inf_{\boldsymbol{x} \in K} p(\boldsymbol{x}). \tag{6.4}$$

Assuming the supremum over the empty set is $-\infty$, this is equivalent to the following semi infinite optimization problem

$$\begin{aligned} &\sup_{\rho} \rho \\ &\text{s.t. } p(\boldsymbol{x}) - \rho > 0 \ \forall \boldsymbol{x} \in K \end{aligned} \tag{6.5}$$

In general this is a difficult problem. However there are several representation results for positive polynomials over basic semi-algebraic sets. If K satisfies the conditions of theorem 6.1 then we can reformulate the problem into

$$\begin{aligned} \sup_{\rho,\sigma_0,\sigma_1,\ldots,\sigma_m} \quad & \rho \\ \text{s.t. } p - \rho &= \sigma_0 + \sum_i \sigma_i g_i \\ \sigma_0, \sigma_i &\in \Sigma[\boldsymbol{x}]. \end{aligned} \tag{6.6}$$

Note that the equality here means equality as polynomial functions. This is equivalent to the coefficients on the left and on the right hand side being equal.

Since Putinar's Positivestellensatz does not give a degree bound on the coefficients we have to look at a relaxation of the previous problem. Fix $d \in \mathbb{N}$

$$\begin{aligned} \rho_d = \sup_{\rho,\sigma_0,\sigma_1,\ldots,\sigma_m} \quad & \rho \\ \text{s.t. } p - \rho &= \sigma_0 + \sum_i \sigma_i g_i \\ \sigma_0 \in \Sigma_{2d}[\boldsymbol{x}], & \ \sigma_i \in \Sigma_{2(d-d_{g_i})}[\boldsymbol{x}]. \end{aligned} \tag{6.7}$$

Let w_k be the vector of monomials up to degree k in the variables $\boldsymbol{x}$ then (6.7) can be formulated as the following semidefinite program (SDP)[9]

$$\begin{aligned} \sup_{\rho,M_0,M_1,\ldots,M_m} \quad & \rho \\ \text{s.t. } p - \rho &= w_d^T M_0 w_d + \sum_i w_{d-d_{g_i}}^T M_i w_{d-d_{g_i}} g_i \\ M_0 \in \mathbb{R}^{\binom{n+d}{d}}, & \ M_i \in \mathbb{R}^{\binom{n+d-d_{g_i}}{d-d_{g_i}} \times \binom{n+d-d_{g_i}}{d-d_{g_i}}} \\ M_0 \succcurlyeq 0, & \ M_i \succcurlyeq 0. \end{aligned} \tag{6.8}$$

Again note that the equality in (6.8) is an equality as functions. Thus we have only to compare coefficients. The equality constraints we obtain this way are linear. In particular $\boldsymbol{x}$ is not a variable in the above optimization problem.

The relaxation (6.8) gives in general a lower bound on the objective of (6.4). We have the following theorem relating the solutions of (6.8) to (6.6).

Theorem 6.2. *Lasserre, 2010b, Th.5.6 Let the assumptions of Putinar's Positivstellensatz hold. Then the optimal solution of the relaxed problem ρ_d converges for $d \to \infty$ to the optimal solution.*

[9]See e.g. Boyd and Vandenberghe, 2004.

6.4 Reformulating the GNEP

We now return to the GNEP problem from section 1. We have to deal with the parametrized optimization problem of each agent. In Lasserre, 2010a the dual problem is considered.

Fix ν and $d \in \mathbb{N}$ sufficiently large. We look at player ν's optimization problem

$$\begin{aligned}&\min_{x^\nu} \theta_\nu(x^\nu, x^{-\nu})\\ &\text{s.t. } h_\nu(x^\nu) \geq 0,\ g_\nu(x^\nu, x^{-\nu}) \geq 0\end{aligned} \tag{6.9}$$

We now regard $x^{-\nu}$ as a parameter and formulate a relaxation as in (6.7). We obtain the following parametrized optimization problem

$$\begin{aligned}\sup_{\rho^\nu, M_i^\nu, N_j^\nu} &\ \rho^\nu\\ \text{s.t. } &\theta_\nu(*, x^{-\nu}) - \rho^\nu = w_d^T M_0^\nu w_d + \sum_i w_{d-d_{g_{\nu,i}}}^T M_i^\nu w_{d-d_{g_{\nu,i}}} g_{\nu,i}(*, x^{-\nu}) +\\ &\sum_j w_{d-d_{h_{\nu,j}}}^T N_j^\nu w_{d-d_{h_{\nu,j}}} h_{\nu,j}(*)\\ &M_0^\nu \in \mathbb{R}^{\binom{n+d}{d}},\ M_i^\nu \in \mathbb{R}^{\binom{n+d-d_{g_{\nu,i}}}{d-d_{g_{\nu,i}}} \times \binom{n+d-d_{g_{\nu,i}}}{d-d_{g_{\nu,i}}}},\ M_i^\nu \succcurlyeq 0,\\ &N_j^\nu \in \mathbb{R}^{\binom{n+d-d_{h_{\nu,i}}}{d-d_{h_{\nu,i}}} \times \binom{n+d-d_{h_{\nu,i}}}{d-d_{h_{\nu,i}}}},\ N_i^\nu \succcurlyeq 0.\end{aligned} \tag{6.10}$$

Note that, since $x^{-\nu}$ is regarded as a parameter, $d_{h_{\nu,j}}$ only refers to the degree of $h_{\nu,j}$ in the x^ν. Furthermore in an abuse of notation let w_k denote the monomials in the variables x^ν up to degree k. The $*$ signifies an equality as functions in the entries it replaces. Here this means that we compare the coefficients of the variables x^ν which themselves depend on the parameters $x^{-\nu}$. In particular those equations are polynomials in the variables $M_i^\nu, N_j^\nu, \rho^\nu$ and $x^{-\nu}$.

Now we use Proposition 6.1 and perform a change of coordinates replacing M_i^ν with $L_i^\nu (L_i^\nu)^T$ and N_j^ν with $T_j^\nu (T_j^\nu)^T$, where L_i^ν and T_j^ν are lower triangular matrices with nonnegative diagonal. This results in a system of polynomial equations and inequalities in the variables $\rho^\nu, L_i^\nu, T_j^\nu$ and $x^{-\nu}$.

For each ν we now have a problem of the form (6.10) but without the positive semidefinite constraint. To find an equilibrium we need an optimality condition or a relaxation thereof to replace the optimization. We propose the following:

Theorem 6.3. *Let the assumptions of Putinar's Positivstellensatz hold and $\varepsilon > 0, d \in \mathbb{N}$ sufficiently large. Furthermore let $(x^{-\nu}, \rho^\nu, L_i^\nu, T_j^\nu)$ for all ν, j, i be values in $\mathbb{R}$ satisfying the following polynomial system of equations and inequalities*

$$\begin{aligned}\theta_\nu(*, x^{-\nu}) - \rho_d^\nu = &w_d^T L_0^\nu (L_0^\nu)^T w_d +\\ &\sum_i w_{d-d_{g_{\nu,i}}}^T L_i^\nu (L_i^\nu)^T w_{d-d_{g_{\nu,i}}} g_{\nu,i}(*, x^{-\nu}) +\\ &\sum_j w_{d-d_{h_{\nu,j}}}^T T_j^\nu (T_j^\nu)^T w_{d-d_{h_{\nu,j}}} h_{\nu,j}(*)\end{aligned} \tag{6.11}$$

$$
\begin{aligned}
\varepsilon \geq & \theta_\nu(x^\nu, x^{-\nu}) - \rho^\nu \\
g_{\nu,i}(x^\nu, x^{-\nu}) \geq & 0 \\
h_{\nu,j}(x^\nu) \geq & 0 \\
& (T_j^\nu)_{l,l} \geq 0, \quad (L_i^\nu)_{k,k} \geq 0 \\
& L_0^\nu \in \mathbb{R}^{\binom{n+d}{d}}, \; L_i^\nu \in \mathbb{R}^{\binom{n+d-d_{g_{\nu,i}}}{d-d_{g_{\nu,i}}} \times \binom{n+d-d_{g_{\nu,i}}}{d-d_{g_{\nu,i}}}}, \\
& T_j^\nu \in \mathbb{R}^{\binom{n+d-d_{h_{\nu,i}}}{d-d_{h_{\nu,i}}} \times \binom{n+d-d_{h_{\nu,i}}}{d-d_{h_{\nu,i}}}}, \\
& L_0^\nu, L_i^\nu, T_j^\nu \textit{ lower triangular.}
\end{aligned}
\tag{6.12}
$$

Then for all ν the point $x = (x^1, \ldots, x^N)$ satisfies the following inequality

$$
|\theta_\nu(x^\nu, x^{-\nu}) - \min_{y \in X_\nu(x^{-\nu})} \theta_\nu(y, x^{-\nu})| \leq \varepsilon \tag{6.13}
$$

Additionally let $\overline{x}$ be any equilibrium then there exists $k \in \mathbb{N}$ and $\rho^\nu, L_i^\nu, T_j^\nu$ for all ν such that $\overline{x}$ is a solutions to the above system of equations and inequalities.

Proof. For any feasible point $\hat{x}^\nu, \hat{\rho}_d^\nu, \hat{L}_i^\nu, \hat{T}_j^\nu$ we have that $\hat{\rho}_d^\nu$ is a lower bound on the optimum of $\theta_\nu(x^\nu, \hat{x}^{-\nu})$. So

$$
\hat{\rho}_d^\nu \leq \min_{x^\nu \in X^\nu(\hat{x}^{-\nu})} \theta_\nu(x^\nu, \hat{x}^{-\nu}) \leq \theta_\nu(\hat{x}^\nu, \hat{x}^{-\nu}).
$$

Since $\varepsilon \geq \theta_\nu(\hat{x}^\nu, \hat{x}^{-\nu}) - \hat{\rho}_d^\nu$, constraint (6.12), we obtain the following

$$
0 \leq \min_{x^\nu \in X^\nu(\hat{x}^{-\nu})} \theta_\nu(x^\nu, \hat{x}^{-\nu}) - \hat{\rho}_d^\nu \leq \theta_\nu(\hat{x}^\nu, x^{-\nu}) - \hat{\rho}_d^\nu \leq \varepsilon.
$$

Thus we know that $\theta_\nu(\hat{x}^\nu, \hat{x}^{-\nu})$ satisfies the inequality (6.13).

If $\overline{x}$ is an equilibrium, then due to Theorem 6.2 we know that for any $\varepsilon > 0$ and any ν there exists a k_ν such that $\theta_\nu(y, \overline{x}^{-\nu}) - \rho_{k_\nu}^\nu$ has a Putinar representation with $|\min_{y \in X_\nu(\overline{x}^{-\nu})} \theta_\nu(y, \overline{x}^{-\nu}) - \rho_{k_\nu}^\nu| < \varepsilon$ and relaxation order k_ν. Now we just have to set k to the maximum of the k_ν and then $\overline{x}$ is feasible. $\square$

Once we computed a feasible point we can check whether it is a true equilibrium. To accomplish this we solve the polynomial optimization problem for each player to global optimality. We use Gloptipoly (Henrion et al., 2009) a program written for Matlab that employs the moment relaxation approach to solving polynomial optimization problems. This is the dual approach to the here presented sum of squares method.

Next we want to illustrate how to reformulate a GNEP into a system of equations and inequalities using our approach.

Example 6.1. We are looking at a simplified version of our model. To avoid confusion with exponents, we will write the players' number in the index. Let the number of players be two with objective function

$$
\begin{aligned}
\max_{a_\nu, q_\nu} \; & a_\nu q_\nu - k_\nu q_\nu \\
\text{s.t. } & g(a_1, q_1, a_2, q_2) \geq 0,
\end{aligned}
\tag{6.14}
$$

where $(k_1, k_2) = (1, \frac{1}{2})$ and $g(a_1, q_1, a_2, q_2) = 1 - a_1^2 - q_1^2 - a_2^2 - q_2^2$. We obtain a relaxation of order 1, i.e. the degree of σ_0 and $g_i\sigma_i$ does not exceed 2. Therefore the multiplier of our constraint is just a nonnegative real number denoted my m_ν. We look at the following equation.

$$-a_1q_1 + k_1q_1 - \rho_1 = w_1^T LL^T w_1 + m_1 g(a_1, q_1, a_2, q_2) \tag{6.15}$$
$$-a_2q_2 + k_2q_2 - \rho_2 = w_2^T MM^T w_2 + m_2 g(a_1, q_1, a_2, q_2), \tag{6.16}$$

where $L = \begin{pmatrix} L_{1,1} & 0 & 0 \\ L_{2,1} & L_{2,2} & 0 \\ L_{3,1} & L_{3,2} & L_{3,3} \end{pmatrix}$, $M = \begin{pmatrix} M_{1,1} & 0 & 0 \\ M_{2,1} & M_{2,2} & 0 \\ M_{3,1} & M_{3,2} & M_{3,3} \end{pmatrix}$ and $w_\nu = \begin{pmatrix} 1 \\ a_\nu \\ q_\nu \end{pmatrix}$.

We write out equation (6.15).

$$\begin{aligned} -q_1 + a_1q_1 + \rho_1 + L_{1,1}^2 + 2a_1L_{1,1}L_{2,1} + a_1^2L_{2,1}^2 + a_1^2L_{2,2}^2 + \\ 2q_1L_{1,1}L_{3,1} + 2a_1q_1L_{2,1}L_{3,1} + q_1^2L_{3,1}^2 + 2a_1q_1L_{2,2}L_{3,2} + \\ q_1^2L_{3,2}^2 + q_1^2L_{3,3}^2 + m_1 - a_2^2m_1 - a_1^2m_1 - q_2^2m_1 - q_1^2m_1 = 0 \end{aligned}$$

Comparing coefficients in a_1, q_1 and adding the other constraints gives for player one the following equations and inequalities.

$$\begin{aligned} \rho_1 + L_{1,1}^2 + 2m_1 - a_2^2m_1 - q_2^2m_1 &= 0 \\ -1 + 2L_{1,1}L_{3,1} &= 0 \\ 2L_{1,1}L_{2,1} &= 0 \\ L_{3,1}^2 + L_{3,3}^2 - m_1 &= 0 \\ L_{2,1}^2 + L_{2,2}^2 - m_1 &= 0 \\ 1 + 2L_{2,1}L_{3,1} + 2L_{2,2}L_{3,2} &= 0 \\ m_1 &\geq 0 \\ g(a_1, q_1, a_2, q_2) &\geq 0 \\ a_1q_1 - q_1 + \rho_1 &= 0 \end{aligned}$$

For $\varepsilon = 0$ and relaxation order 1, we obtain the following equilibrium

$$a_1 = 0.5,\ q_1 = 0.86602,\ a_2 = 0.00002,\ q_2 = 0.00323.$$

6.5 Computational Results

We now return to the model of the New Zealand electricity spot market. First we present a real data set, second solve the model and lastly verify the solutions. All terms are with respect to a single time period which is a day.

The specific data for the HVDC Link in New Zealand are as follows: The power flow capacity of both arcs is $\bar{t} = 16.8GWh$. Given the value of the electrical resistivity/km

	Technology	κ_ν $[GWh]$	η_ν $[GJ/GWh]$	ϕ_ν $[\$/GJ]$	τ_ν $[\$/GJ]$	o_ν $[\$/GWh]$	δ $[GWh]$
North		93.09					65.67
	Coal	16.84	10500	4.006	0	9600	
	Diesel	0.15	11000	25.000	0	9600	
	Gas	33.74	7686.626	6.500	1.0625	5076	
	Geothermal	16.72	12000	1.000	0	6174	
	Hydro	20.18	3600	1.000	0	0	
	Wind	4.52	3600	1.000	0	12038	
	Wood Waste	0.93	12000	2.000	0	11800	
South		47.02					38.69
	Hydro	46.54	3600	1.000	0	0	
	Wind	0.47	3600	1.000	0	16000	
Total		140.12					104.46

Table 6.1: Coefficients for the New Zealand electricity spot market model based on real data.

at 20°C of 0.0139 Ω/km, the length of transmission line of 607km and the operating voltage of $350'000V$ the result is $\lambda = 0.0689$ for $t_{\{1,2\}}$ in GWh. Listed in table 6.1 are all the coefficients for our electricity spot market model of New Zealand which are based on the Electricity Authority NZ, 2011, 2010a. The demands and the capacities for the climate-based technologies corresponds to the average of the daily data of 2011.

We set $\varepsilon = 0$ and $d = 2$. We solve the model with the Ipopt solver in GAMS with a residual of 10^{-9}. The running time is 17.45 seconds on an Intel E3-1290 with 8GB of RAM. The results for bid, production and profit can be found in table 6.2. We verified the results using GloptiPoly and SeDuMi.

	Technology	Price offer $[\$/GWh]$	Prod. offer $[GWh]$	Cleared $[GWh]$	Profit $[M\ \$]$	Transport to $[GWh]$
North						2.42
	Coal	63206.1	16.8	9.33	0.11	
	Diesel	28233.1	0	0	0	
	Gas	63206.1	28.99	14.64	0	
	Geothermal	35127.3	16.72	16.72	0.28	
	Hydro	34159	20.18	20.18	0.62	
	Wind	63206.1	4.52	2.79	0.13	
	Wood Waste	30714.1	0	0	0	
South						0
	Hydro	31645.5	46.54	41.51	1.16	
	Wind	32260.5	0	0	0	

Table 6.2: Solution for the New Zealand electricity spot market model.

With a homogeneous good one would usually expect a Bertrand competition. In contrast to this our simple model has a constant demand and so the consumer's demand is perfectly inelastic with respect to the price. Additionally we are capacity constrained and so price equal marginal costs is out of question. Basically we are looking at a continuum of non obvious Cournot Equilibria where always some of the EPs do not offer their marginal costs.

The consumption-weighted average price of the here presented equilibrium is 40712.3 $\$/GWh$ and is of the same magnitude as the reference node Stratford ($SFD2201$) with a consumption-weighted average price of 43130.1 $\$/GWh$.

Bibliography

Allgower, E. and K. Georg. *Introduction to Numerical Continuation Methods.* SIAM, 2003.

Alvi, E. "First-Order approach to Principal-Agent problems: a generalization". *The GENEVA Papers on Risk and Insurance* 22.1 (1997), pp. 59–65.

Araujo, A. and H. Moreira. "A general lagrangian approach for non-concave moral hazard problems". *Journal of Mathematical Economics* 35.1 (2001), pp. 17–39.

Armstrong, C. S., D. F. Larcker, and C.-L. Su. "Endogenous selection and moral hazard in compensation contracts". *Operations Research* 58.4-Part-2 (2010), pp. 1090–1106.

Aussel, D. and J. Dutta. "On gap functions for multivalued stampacchia variational inequalities". *Journal of Optimization Theory and Applications* 149.3 (2011), pp. 513–527.

Aussel, D., R. Correa, and M. Marechal. "Gap functions for quasivariational inequalities and generalized nash equilibrium problems". *Journal of Optimization Theory and Applications* 151.3 (2011), pp. 474–488.

Aussel, D., R. Correa, and P. Marechal. "Spot electricity market with transmission losses". *Journal of Industrial and Management Optimization* (2012).

Bajari, P., H. Hong, J. Krainer, and D. Nekipelov. "Computing equilibria in static games of incomplete information using the All-Solution homotopy". Working paper. 2010.

Bates, D., J. Hauenstein, A. Sommese, and C. Wampler. *Bertini: software for numerical algebraic geometry.* Available at http://www.nd.edu/∼sommese/bertini. 2005.

Berge, C. *Topological Spaces.* Oliver & Boyd: Edinburgh and London, 1963.

Berry, S. and A. Pakes. "The pure characteristics demand model". *International Economic Review* 48.4 (2007), pp. 1193–1225.

Besanko, D. and U. Doraszelski. "Capacity dynamics and endogenous asymmetries in firm size". *RAND Journal of Economics* 35.1 (2004), pp. 23–49.

Besanko, D., U. Doraszelski, Y. Kryukov, and M. Satterthwaite. "Learning-by-doing, organizational forgetting, and industry dynamics". *Econometrica* 78.2 (2010), pp. 453–508.

Borkovsky, R. N., U. Doraszelski, and Y. Kryukov. "A user's guide to solving dynamic stochastic games using the homotopy method". *Operations Research* 58.4-Part-2 (2010), pp. 1116–1132.

Borkovsky, R. N., U. Doraszelski, and Y. Kryukov. "A dynamic quality ladder model with entry and exit: exploring the equilibrium correspondence using the homotopy method". *Quantitative Marketing and Economics* 10.2 (2012), pp. 197–229.

Boyd, S. and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004.

Conlon, J. R. "Two new conditions supporting the first-order approach to multisignal principal–agent problems". *Econometrica* 77.1 (2009), pp. 249–278.

Couzoudis, E. and P. Renner. “Computing gerneralized nash equilibria by polynomial programming”. *Mathematical Methods of Operations Research* (2013).

Cox, D. A., J. Little, and D. O’Shea. *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra, 3/e (Undergraduate Texts in Mathematics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2007.

Datta, R. “Finding all nash equilibria of a finite game using polynomial algebra”. *Economic Theory* 42.1 (2010), pp. 55–96.

Doraszelski, U. and K. L. Judd. “Avoiding the curse of dimensionality in dynamic stochastic games”. *Quantitative Economics* 3.1 (2012), pp. 53–93.

Doraszelski, U. and A. Pakes. “A framework for applied dynamic analysis in IO”. Ed. by M. Armstrong and R. Porter. Vol. 3. Handbook of Industrial Organization. Elsevier, 2007. Chap. 30, pp. 1887–1966.

Doraszelski, U. and M. Satterthwaite. “Computable markov-perfect industry dynamics”. *The RAND Journal of Economics* 41.2 (2010), pp. 215–243.

Drexler, F. “Eine Methode zur Berechnung sämtlicher Lösungen von Polynomgleichungssystemen”. *Numer. Math.* 29 (1977), pp. 45–58.

Drexler, F. “A homtopy method for the calculation of all zeros of zerodimensional polynomial ideals”. *Developments in Statistics* 1 (1978), pp. 69–93.

Dubé, J.-P., J. T. Fox, and C.-L. Su. “Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation”. *Econometrica* 80.5 (2012), pp. 2231–2267.

Eaves, B. and K. Schmedders. “General equilibrium models and homotopy methods”. *Journal of Economic Dynamics and Control* 23.9-10 (1999), pp. 1249–1279.

Electricity Authority NZ. GEM: Generation Expansion Model. Accessed: 12.10.2011. 2010.

Electricity Authority NZ. *Electricity Industry Participation Code 2010 - Part 13*. Version: 21.09.2012. 2010.

Electricity Authority NZ. CDS: Centralised Dataset. Accessed: 12.10.2011. 2011.

Electricity Authority NZ. *Market operation service providers*. http://www.ea.govt.nz/industry/mo-service-providers/, Accessed: 01.10.2012. May 2012.

Ericson, R. and A. Pakes. “Markov-perfect industry dynamics: a framework for empirical work”. *Review of Economic Studies* 62.1 (1995), pp. 53–82.

Facchinei, F. and C. Kanzow. “Generalized nash equilibrium problems”. *4OR: A Quarterly Journal of Operations Research* 5.3 (2007), pp. 173–210.

Fair, J. C. and B. Taylor. “Solution and maximum likelihood estimation of dynamic nonlinear rational expectations models”. *Econometrica* 51 (1983), pp. 1169–1185.

Filar, J. A. and K. Vrieze. *Competitive Markov decision processes*. Springer, 1997, p. 411.

Flaherty, M. “Industry structure and cost-reducing investment”. *Econometrica* 48.5 (1980), pp. 1187–1209.

Fudenberg, D. and J. Tirole. “Learning-by-doing and market performance”. *Bell Journal of Economics* 14.2 (1983), pp. 522–530.

Fudenberg, D. and J. Tirole. "Capital as a commitment: strategic investment to deter mobility". *Journal of Economic Theory* 31.2 (1983), pp. 227–250.

García, C. and T. Li. "On the number of solutions to polynomial systems of equations". *SIAM Journal of Numerical Analysis* 17.4 (1980), pp. 540–546.

Garcia, C. and W. Zangwill. *Global continuation methods for finding all solutions to polynomial systems of equations in N variables.* Tech. rep. 7755. University of Chicago, Center for Math. Studies in Business and Economics, 1977.

Greuel, G.-M. and G. Pfister. *A Singular Introduction to Commutative Algebra.* Springer-Verlag, 2002.

Grossman, S. J. and O. D. Hart. "An analysis of the principal-agent problem". *Econometrica* 51.1 (1983), pp. 7–45.

Hao, W., J. D. Hauenstein, B. Hu, Y. Liu, A. J. Sommese, and Y.-T. Zhang. "Multiple stable steady states of a reaction-diffusion model on zebrafish dorsal-ventral patterning". *Discrete and Continuous Dynamical Systems-Series S* 4 (2011), pp. 1413–1428.

Henrion, D., J. Lasserre, and J. Löfberg. "Gloptipoly 3: moments, optimization and semidefinite programming". *Optimization Methods and Software* 24 (4-5 2009), pp. 761–779.

Herings, P. and R. Peeters. "A globally convergent algorithm to compute all nash equilibria for n -Person games". *Annals of Operations Research* 137.1 (2005), pp. 349–368.

Hilbert, D. "Über die Darstellung definiter Formen als Summe von Formenquadraten". *Mathematische Annalen* 32.3 (1888), pp. 342–350.

Hobbs, B., C. Metzler, and J.-S. Pang. "Strategic gaming analysis for electric power systems: an mpec approach". *Power Systems, IEEE Transactions on* 15.2 (2000), pp. 638–645.

Hogan, W. W. "Point-to-set maps in mathematical programming". *SIAM Review* 15.3 (1973), pp. 591–603.

Hölmstrom, B. "Moral hazard and observability". *The Bell Journal of Economics* 10.1 (1979), pp. 74–91.

Holmstrom, B. and P. Milgrom. "Multitask principal-agent analyses: incentive contracts, asset ownership, and job design". *Journal of Law, Economics, & Organization* 7 (1991), pp. 24–52.

Jewitt, I. "Justifying the first-order approach to principal-agent problems". *Econometrica* 56.5 (1988), pp. 1177–1190.

Jewitt, I., O. Kadan, and J. M. Swinkels. "Moral hazard with bounded payments". *Journal of Economic Theory* 143.1 (2008), pp. 59–82.

Jibetean, D. and E. de Klerk. "Global optimization of rational functions: a semidefinite programming approach". *Mathematical Programming* 106 (1 2006), pp. 93–109.

Judd, K. and C.-L. Su. "Computation of moral-hazard problems" (2005). working paper.

Judd, K., P. Renner, and K. Schmedders. "Finding all pure-strategy equilibria in static and dynamic games with continuous strategies". *Quantitative Economics* 3 (2 2012).

Kadan, O. and J. M. Swinkels. *On the moral hazard problem without the first-order approach.* working paper. 2012.

Kadan, O., P. Reny, and J. M. Swinkels. "Existence of optimal mechanisms in principal-agent problems" (2011). MFI Working Paper No. 2011–002.

Kalaba, R. and L. Tesfatsion. "Solving nonlinear equations by adaptive homotopy continuation". *Applied Mathematics and Computation* 41.2 (1991), pp. 99–115.

Kloek, T. and H. K. van Dijk. "Bayesian estimates of equation system parameters: an application of integration by monte carlo". *Econometrica* 46.1 (1978), pp. 1–19.

Kubler, F. and K. Schmedders. "Competitive equilibria in semi-algebraic economies". *Journal of Economic Theory* 145.1 (2010), pp. 301–330.

Laraki, R. and J. Lasserre. "Semidefinite programming for min–max problems and games". *Mathematical Programming* 131.1 (2012), pp. 305–332.

Lasserre, J. B. "An explicit exact sdp relaxation for nonlinear 0-1 programs". *Integer Programming and Combinatorial Optimization*. Ed. by K. Aardal and B. Gerards. Vol. 2081. Lecture Notes in Computer Science. Springer Berlin, 2001, pp. 293–303.

Lasserre, J. "Global optimization with polynomials and the problem of moments". *SIAM Journal on Optimization* 11.3 (2001), p. 796.

Lasserre, J. *Moments, Positive Polynomials and Their Applications*. Vol. 1. Imperial College Press Optimization Series. Imperial College Press, 2010.

Lasserre, J. "A "Joint+Marginal" approach to parametric polynomial optimization". *SIAM Journal on Optimization* 20 (2010), p. 1995.

Laurent, M. "Sums of squares, moment matrices and optimization over polynomials". Vol. 149. IMA Volumes in Mathematics and its Applications. Springer, 2009, pp. 157–270.

LiCalzi, M. and S. Spaeter. "Distributions for the first-order approach to principal-agent problems". *Economic Theory* 21.1 (2003), pp. 167–173.

Loera, J. A. D., R. Hemmecke, M. Köppe, and R. Weismantel. "Integer polynomial optimization in fixed dimension". *Mathematics of Operations Research* 31.1 (2006), pp. 147–153.

McKelvey, R. and A. McLennan. "Computation of equilibria in finite games". *Handbook of Computational Economics*. Ed. by H. M. Amman, D. A. Kendrick, and J. Rust. Vol. 1. Handbook of Computational Economics. Elsevier, 1996. Chap. 2, pp. 87–142.

McKelvey, R., A. McLennan, and T. Turocy. *Gambit: software tools for game theory, version 0.2007.01.30*. http://www.gambit-project.org. 2007.

Mirrlees, J. A. *The implications of moral hazard for optimal insurance*. mimeo. 1979.

Mirrlees, J. A. "The theory of moral hazard and unobservable behaviour: part i". *The Review of Economic Studies* 66.1 (1999), pp. 3–21.

Morgan, A. and A. Sommese. "A homotopy for solving general polynomial systems that respects m-homogeneous structures". *Applied Mathematics and Computation* 24.2 (1987), pp. 101–113.

Nie, J. "An exact jacobian SDP relaxation for polynomial optimization". *Mathematical Programming* (2012), pp. 1–31.

Pakes, A. and P. McGuire. "Computing markov-perfect nash equilibria: numerical implications of a dynamic differentiated product model". *RAND Journal of Economics* 25.4 (1994), pp. 555–589.

Pakes, A. and P. McGuire. "Stochastic algorithms, symmetric Markov perfect equilibrium, and the 'curse' of dimensionality". *Econometrica* 69.5 (2001), pp. 1261–1281.

Parrilo, P. A. "Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization". PhD thesis. California Institute of Technology, 2000.

Prescott, E. S. "A primer on moral-hazard models". *Economic Quarterly* 85 (1999), pp. 47–78.

Prescott, E. S. "Computing solutions to moral-hazard programs using the dantzig–wolfe decomposition algorithm". *Journal of Economic Dynamics and Control* 28.4 (2004), pp. 777 –800.

Renner, P. and K. Schmedders. "A polynomial optimization approach to principal-agent problems". 2013.

Rogerson, W. P. "The first-order approach to principal-agent problems". *Econometrica* 53.6 (1985), pp. 1357–1367.

Rosen, J. B. "Existence and uniqueness of equilibrium points for concave N-Person games". *Econometrica* 33.3 (1965), pp. 520–534.

Rostalski, P., I. A. Fotiou, D. J. Bates, A. G. Beccuti, and M. Morari. "Numerical algebraic geometry for optimal control applications". *SIAM Journal on Optimization* 21.2 (2011), pp. 417–437.

Schmüdgen, K. "The k-moment problem for compact semi-algebraic sets". *Mathematische Annalen* 289 (1 1991), pp. 203–206.

Shapley, L. S. "Stochastic games". *Proceedings of the National Academy of Sciences of the United States of America* 39.10 (1953). PMID: 16589380 PMCID: 1063912, pp. 1095–1100.

Sinclair-Desgagné, B. "The first-order approach to multi-signal principal-agent problems". *Econometrica* 62.2 (1994), pp. 459–465.

Sommese, A. and C. Wampler. *The Numerical Solution to Systems of Polynomials.* World Scientific, 2005.

Spence, A. "The learning curve and competition". *Bell Journal of Economics* 12.1 (1981), pp. 49–70.

Spence, M. "Cost reduction, competition, and industry performance". *Econometrica* 52.1 (1984), pp. 101–21.

Sturmfels, B. *Solving systems of polynomial equations.* AMS BOOKstore, 2002, p. 164.

Su, C.-L. and K. L. Judd. "Constrained optimization approaches to estimation of structural models". *Econometrica* 80.5 (2012), pp. 2213–2230.

Tipler, P., D. Gerlich, and G. Jerke. *Physik.* Vol. 3. korr. Nachdr. Spektrum Lehrbuch. Spektrum Akad. Verl., 2000.

Turocy, T. L. "Towards a black-box solver for finite games: computing all equilibria with Gambit and PHCpack". *Software for Algebraic Geometry.* 2008, pp. 133–147.

Vandenberghe, L. and S. Boyd. "Semidefinite programming". *SIAM Review* 38.1 (1996), pp. 49–95.

Verschelde, J. "Algorithm 795: PHCpack: a general-purpose solver for polynomial systems by homotopy continuation". *ACM Trans. Math. Softw.* 25.2 (1999), pp. 251–276.

von Stengel, B. "Computation of nash equilibria in finite games: introduction to the symposium". *Economic Theory* 42.1 (2010), pp. 1–7.

Watson, L. T., M. Sosonkina, R. C. Melville, A. P. Morgan, and H. F. Walker. "Algorithm 777: HOMPACK90: a suite of Fortran 90 codes for globally convergent homotopy algorithms". *ACM Trans. Math. Softw.* 23.4 (1997), pp. 514–549.

Wise, S., A. Sommese, and L. Watson. "Algorithm 801: polsys_plp: a partitioned linear product homotopy code for solving polynomial systems of equations". *ACM Trans. Math. Softw.* 26.1 (2000), pp. 176–200.

Philipp Renner

Curriculum Vitae

May 2013

Address: Chair of Quantitative Business Administration, Universität Zürich, Mousssonstrasse 15, CH-8044 Zürich.
Phone: +41 44 634 3782
Email: philipp.renner@business.uzh.ch
WWW: http://www.business.uzh.ch/professorships/qba/members/renner.html

Education and Qualifications

2004	Abitur	Gymnasium am Kaiserdom, Speyer
08/2008	Diploma	in Mathematics, Technische Universität Kaiserslautern, Advisor: Anne Frübis-Krüger
2008-today		Doctoral student, Universität Zürich, Advisor: Karl Schmedders
10/2010		Research stay at CRM Barcelona

Refereed research papers

1. Judd, K., P. Renner, and K. Schmedders (2012). Finding All Pure-Strategy Equilibria in Static and Dynamic Games with Continuous Strategies. *Quantitative Economics* **3** (2).
2. Couzoudis, E. and P. Renner (2013). Computing Gerneralized Nash Equilibria by Polynomial Programming. *Mathematical Methods of Operations Research*.

Submitted papers

1. Renner, P. and K. Schmedders (2013). "A Polynomial Optimization Approach to Principal-Agent Problems".

Master thesis

1. Renner, P. (2008). "Zur Implementierung von Jungs Desingularisierungsalgorithmus für Flächen". MA thesis. Technische Universität Kaiserslautern.

Presentations

09/2010	"Finding All Pure-Strategy Equilibria in Static and Dynamic Games with Continuous Strategies", OR days 2010, Fribourg
11/2010	"Finding All Pure-Strategy Equilibria in Static and Dynamic Games with Continuous Strategies", CRM research seminar, Barcelona
09/2011	"Solving Principal Agent Problems by Polynomial Programming", Euro 2011, Zurich
10/2011	"Solving Principal Agent Problems by Polynomial Programming", AG 2012, Raleigh
06/2012	"Finding Generalized Nash Equilibria by Polynomial Programming", Oberseminar Reelle Algebraische Geometrie, Konstanz
06/2012	"Computing Generalized Nash Equilibria by Polynomial Programming", OR days 2012, Neuchâtel
08/2012	"Computing Generalized Nash Equilibria by Polynomial Programming", 21st International symposium on mathematical programming, Berlin

Teaching

2006-2007	Tutorial to lecture "Einführung in die Mathematik" (Introduction to mathematics), Technische Universität Kaiserslautern
2009-today	Tutorial to lecture "Mathematik I für Ökonomen" (Mathematics for economists I), Universität Zürich
2009-today	Tutorial to lecture "Mathematik II für Ökonomen" (Mathematics for economists II), Universität Zürich