A Polynomial Optimization Approach to Principal-Agent Problems[*]

Philipp Renner

Dept. of Business Administration

Universität Zürich

philipp.renner@business.uzh.ch

Karl Schmedders

DBA – Universität Zürich

and Swiss Finance Institute

karl.schmedders@business.uzh.ch

January 10, 2013

Abstract

This paper presents a new method for the analysis of moral hazard principal-agent problems. The new approach avoids the stringent assumptions on 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 rational in the wage, 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.

Keywords: Principal-agent model, moral hazard, polynomial optimization, first-order approach.

JEL codes: C63, D80, D82.

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, it just imposes a lower bound on the agent's expected utility, the incentive-compatibility constraint involves an expected utility maximization problem of the agent. As a consequence, principal-agent problems are a type of bilevel optimization problems,[1] a class of optimization problems that is notoriously difficult. The most popular solution approach to principal-agent problems with a one-dimensional effort set for the agent 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 on the probability distribution of outcomes, which fail to hold in many economic applications.[2] 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 on 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 us to employ the global optimization approach for rational functions of Jibetean and de 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 bilevel 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 optimiza-

---

[1]The major feature of bilevel 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 one. 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.

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

tion problem, which is globally solvable. We can then use the methods implemented in GloptiPoly, see Henrion, Lasserre, and Löfberg (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 on the probability function of outcomes, the monotone likelihood-ratio condition (MLRC) and the convexity of distribution function 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.[3] 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,[4] all of these sufficient sets of conditions are regarded as highly restrictive, see Conlon (2009) and Kadan, Reny, and Swinkels (2011).

Principal-agent models in which the agent's action set is one-dimensional dominate both the literature on the first-order approach as well as the applied and computational literature, see for example, Araujo and Moreira (2001), Judd and Su (2005), Armstrong, Larcker, and Su (2010). However, the analysis of linear multi-task principal-agent models

---

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

[4]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.

in Holmström 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, Reny, and Swinkels (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.

In this paper, we also 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 de Klerk (2006) we encounter a theoretical difficulty. Unlike 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 sequence of nonlinear programs involving the relaxation converge; and, in fact, the limit points yield optimal solutions to the original principal-agent problem.

While our main results are of 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 allows 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 are powerful and have recently received a lot of attention in economics, see for instance Su and Judd (2012) and Dubé, Fox, and Su (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 2 describes the principal-agent model and the classical first-order approach. In Section 3 we introduce our main result for the polynomial optimization approach. Section 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. Section 6 concludes.

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.

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.

$$\max_{\boldsymbol{w} \in W, \boldsymbol{a} \in A} \quad U(\boldsymbol{w}, \boldsymbol{a})$$

$$\text{s.t.} \quad \boldsymbol{a} \in \arg\max_{\boldsymbol{b} \in A} V(\boldsymbol{w}, \boldsymbol{b}) \tag{1}$$

$$V(\boldsymbol{w}, \boldsymbol{a}) \geq \underline{V}$$

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

$$\boldsymbol{a} \in \arg\max_{\boldsymbol{b} \in A} V(\boldsymbol{w}, \boldsymbol{b}) \tag{2}$$

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, Kadan, and Swinkels (2008), Conlon (2009), and others.

2.2 The First-Order Approach

In general it is very difficult to find a global optimal solution to the principal-agent problem (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 (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

$$\frac{\partial}{\partial a} V(\boldsymbol{w}, a) = \sum_{i=1}^{N} \left(\frac{\partial}{\partial a} v(w_i, a) \, \mu(y_i|a) + v(w_i, a) \, \frac{\partial}{\partial a} \mu(y_i|a) \right) = 0. \tag{3}$$

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[5]) 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, Reny, and Swinkels (2011).

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 (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 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 $\boldsymbol{a} \in A$.*

---

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

Assumption 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 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 (1) exists.

Proposition 1. *If Assumptions 1 – 3 hold, then the principal-agent problem (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 2 and 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 (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 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 (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. $\square$

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

The purpose of this section is to state our main result, Theorem 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 4 (Rational Expected Utility Function). *The parameterized probability distribution functions $\mu(y|\bullet) : A \rightarrow [0,1]$ and the agent's Bernoulli utility function $v : J \times A \rightarrow \mathbb{R}$ 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 $c_i, f_i : W \rightarrow \mathbb{R}$ with $\sum_{i=0}^{d} f_i(\boldsymbol{w}) a^i > 0$ for all $(\boldsymbol{w}, a) \in W \times A$.[6] 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 $D = \lceil \frac{d}{2} \rceil$. 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[7] provides us with an equivalent problem to the principal-agent problem (1).

Theorem 1. *Let $A = [-1, 1]$ and suppose Assumption 4 holds. Then $(\boldsymbol{w}^*, a^*)$ solves the principal-agent problem (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:*

$$\max_{\boldsymbol{w}, a, \rho, Q^{(0)}, Q^{(1)}} U(\boldsymbol{w}, a) \quad subject\ to \tag{4}$$

---

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

[7] 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 in Section 4.1.3.

$$c_0(\boldsymbol{w}) - \rho f_0(\boldsymbol{w}) = Q_{0,0}^{(0)} + Q_{0,0}^{(1)} \tag{4a}$$

$$c_l(\boldsymbol{w}) - \rho f_l(\boldsymbol{w}) = \sum_{i+j=l} Q_{ij}^{(0)} + \sum_{i+j=l} Q_{ij}^{(1)} - \sum_{i+j=l-2} Q_{ij}^{(1)}, \quad l = 1, \dots, d \tag{4b}$$

$$Q^{(0)}, Q^{(1)} \succcurlyeq 0 \tag{4c}$$

$$\rho \left(\sum_{i=0}^{d} f_i(\boldsymbol{w}) a^i \right) = \sum_{i=0}^{d} c_i(\boldsymbol{w}) a^i \tag{4d}$$

$$\sum_{i=0}^{d} c_i(\boldsymbol{w}) a^i \leq -\underline{V} \left(\sum_{i=0}^{d} f_i(\boldsymbol{w}) a^i \right) \tag{4e}$$

$$-a^2 + 1 \geq 0 \tag{4f}$$

$$\boldsymbol{w} \in W \tag{4g}$$

The new optimization problem (4) has the same objective function as the original principal-agent problem (1). Unlike the original problem, the new problem (4) is not a bilevel optimization problem. Instead the constraint involving the agent's expected utility maximization problem has been replaced by inequalities and equations. Problem (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 (4) will be $-V(\boldsymbol{w}^*, a^*)$, the negative of the agent's maximal expected utility. Constraints (4a)–(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 (4a) and (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 (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 (4c) can be written as a set of polynomial inequalities.) Next, constraint (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 (4a)–(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 (4e) is the transformed participation constraint for the agent's rational expected utility function. Constraint (4f) is a polynomial representation of the feasible action set and constraint (4g) is just the constraint on the compensation scheme from the original principal-agent problem (1).

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

Example 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 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

$$U(\boldsymbol{w}, a) = 2 + 2a - \frac{w_1}{4} + \frac{aw_1}{2} - \frac{a^2w_1}{4} - \frac{w_2}{2} + \frac{a^2w_2}{2} - \frac{w_3}{4} - \frac{aw_3}{2} - \frac{a^2w_3}{4}$$

$$V(\boldsymbol{w}, a) = -\frac{5}{2} + \frac{\sqrt{w_1}}{2} + \sqrt{w_2} + \frac{\sqrt{w_3}}{2} - a - a\sqrt{w_1} + a\sqrt{w_3} - a^2\sqrt{w_2} + \frac{a^2\sqrt{w_1}}{2} - \frac{a^2}{2} + \frac{a^2\sqrt{w_3}}{2}$$

We observe that $V(\boldsymbol{w}, a)$ is a quadratic polynomial in a. The representation of $-V(\boldsymbol{w}, a)$ according to Assumption 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 1, $D = 1$ and so 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 as follows.

$$\max_{w_1,w_2,w_3,a,\rho,n_{00},n_{01},n_{11},m} U(w_1,w_2,w_3,a)$$

$$\text{s.t.} \quad \frac{5}{2} - \frac{\sqrt{w_1}}{2} - \sqrt{w_2} - \frac{\sqrt{w_3}}{2} - \rho = n_{00} + m_{00}$$

$$1 + \sqrt{w_1} - \sqrt{w_3} = 2n_{01}$$

$$\frac{1}{2} - \frac{\sqrt{w_1}}{2} + \sqrt{w_2} - \frac{\sqrt{w_3}}{2} = n_{11} - m_{00}$$

$$\rho = -V(w_1,w_2,w_3,a)$$

$$n_{00} \geq 0, n_{11} \geq 0, n_{00}n_{11} - n_{01}^2 \geq 0, m_{00} \geq 0$$

$$V(w_1,w_2,w_3,a) \geq 0$$

$$-a^2 + 1 \geq 0$$

$$w_1, w_2, w_3 \geq 0$$

We can solve this nonlinear optimization problem with Gloptipoly, see Henrion, Lasserre, and Löfberg (2009), and obtain the globally optimal contract $\boldsymbol{w}^* = (0.3417, 1.511, 3.511)$ and the resulting optimal effort $a^* = 0.6446$. Table 1 reports solutions for different levels of the agent's risk aversion η. For completion the table also reports the corresponding first-best solutions[8] 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.

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

---

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

4 Derivation of the Polynomial Optimization Approach

In this section we first review the mathematical foundation of Theorem 1 and then prove the theorem. We also discuss the assumptions of the theorem as well as the limitations of the polynomial optimization approach.

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.

4.1.1 Semidefinite Programming

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

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

is called the trace of M. Note that

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

for matrices $C, X \in S^n$ is a linear function on the set S^n of symmetric $n \times n$ matrices. Recall from the beginning of Section 3 the notation $X \succcurlyeq 0$ for positive semidefinite matrices. A semidefinite optimization problem (in standard form) is defined as follows.

Definition 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)*.

$$
\begin{aligned}
&\sup_{X} \operatorname{tr}(CX) \\
&\text{s.t. } \operatorname{tr}(A_j X) = b_j \quad j = 1, \ldots, m \\
&\quad X \succcurlyeq 0
\end{aligned}
\tag{5}
$$

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).

For the characterization of symmetric positive semidefinite matrices we need the following definition.

Definition 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 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.

4.1.2 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 (2010).

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 $\boldsymbol{x} = (x_1,\ldots,x_n)$. 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 $\boldsymbol{x^\alpha}$. The degree of this monomial is $|\boldsymbol{\alpha}| = \sum_{i=1}^n \alpha_i$. A polynomial $p \in \mathbb{R}[\boldsymbol{x}]$, $p = \sum_{\boldsymbol{\alpha}} a_{\boldsymbol{\alpha}} \boldsymbol{x^\alpha}$ is a sum of terms $a_{\boldsymbol{\alpha}} \boldsymbol{x^\alpha}$ 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 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, that is, 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 1. [Lasserre (2010, 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.*

We illustrate this result in $\mathbb{R}$.

4.1.3 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 $(d+1)(d+2)/2$ 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.

4.1.4 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 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 3. [Lasserre (2010, 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 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

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

respectively.

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

4.1.5 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,

$$p_{\min} = \inf_{x \in K} p(x). \tag{6}$$

We can rewrite Problem (6) as follows,

$$\begin{aligned} &\sup_{\rho} \rho \\ &\text{s.t. } p(x) - \rho \geq 0 \; \forall x \in K. \end{aligned} \tag{7}$$

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

$$\rho \leq p_{\min}. \tag{8}$$

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 3 enables us to rewrite the infinitely many constraints of Problem (7). With the polynomial g defined by $g(x) = 1 - x^2$ we obtain

the following optimization problem,

$$\sup_{\rho,\sigma_0,\sigma_1} \rho$$

$$\text{s.t. } p - \rho = \sigma_0 + \sigma_1 g \tag{9}$$

$$\sigma_0, \sigma_0 \in \Sigma[x]$$

Note that the equality constraint here signifies equality as polynomials. Lemma 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 $p \in \mathbb{R}[x]$. According to Proposition 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.

$$\sup_{\rho,Q^{(0)},Q^{(1)}} \rho$$

$$\text{s.t. } p - \rho = v_{d_p}^T Q^{(0)} v_{d_p} + g v_{d_p-1}^T Q^{(1)} v_{d_p-1}$$

$$Q^{(0)}, Q^{(1)} \succcurlyeq 0 \tag{10}$$

$$Q^{(0)} \in \mathbb{R}^{(d_p+1)\times(d_p+1)}, \ Q^{(1)} \in \mathbb{R}^{d_p \times d_p}$$

$$v_{d_p} = (1, x, \ldots, x^{d_p})^T, \ v_{d_p-1} = (1, x, \ldots, x^{d_p-1})^T$$

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 $Q_{ij}^{(0)}$, $i, j = 0, 1, \ldots, d_p$, for the (i,j)-th entry of the matrix $Q^{(0)}$ (similarly for $Q^{(1)}$). Then we can rewrite the first constraint of Problem (10),

$$c_0 - \rho = Q_{0,0}^{(0)} + Q_{0,0}^{(1)},$$

$$c_l = \sum_{i+j=l} Q_{ij}^{(0)} + \sum_{i+j=l} Q_{ij}^{(1)} - \sum_{i+j=l-2} Q_{ij}^{(1)} \quad l = 1, \ldots, d. \tag{11}$$

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 2.

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

Proposition 4. [Lasserre (2010, 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 (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.

4.1.6 Rational Objective Function

Jibetean and de 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,

$$p_{\min} = \inf_{\boldsymbol{x} \in K} \frac{p(\boldsymbol{x})}{q(\boldsymbol{x})}. \tag{12}$$

We can rewrite this problem in polynomial form.

Proposition 5. [Jibetean and de 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 (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 3 and reformulate problem (12),

$$\begin{aligned} \sup_{\rho, \sigma_0, \sigma_1} \quad & \rho \\ \text{s.t.} \quad & p - \rho q = \sigma_0 + g\sigma_1 \\ & \sigma_0 \in \Sigma_{2d}, \ \sigma_1 \in \Sigma_{2(d-1)} \end{aligned} \tag{13}$$

And so we can solve the constrained optimization problem (12) also as an SDP.

4.2 Proof of Theorem 1

In the review of the mathematical literature, we followed the notation of the survey by Laurent (2009) and used the expressions *sup* and *inf* in the optimization problems. Since the domain K is a compact interval, we can safely write *max* and *min* in the context of our application. Now we are in the position to prove Theorem 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 (4) corresponds to a feasible point for (1) and vice versa.

Let $(\hat{\boldsymbol{w}}, \hat{a}, \hat{\rho}, \hat{Q}^{(0)}, \hat{Q}^{(1)})$ be a feasible point for problem (4). Then by inequality (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 (1).

Now let $(\hat{\boldsymbol{w}}, \hat{a})$ be a feasible point for (1). So $\hat{a} \in \arg\max_{a \in [-1,1]} V(\hat{\boldsymbol{w}}, a)$. By Proposition 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

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

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 (4). $\qquad\square$

The proof establishes that the feasible region of the original principal-agent problem (1) is a projection of the feasible region of the optimization problem (4). The first four constraints of problem (4) capture the agent's expected utility maximization problem. The constraints (4a)–(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 (4) to global optimality.

Corollary 1. *Suppose Assumption 4 holds and that the functions $c_i, f_i : W \to \mathbb{R}$ (in Assumption 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 (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. $\qquad\square$

If the conditions of the corollary are satisfied, we can use the methods employed in GloptiPoly, see Henrion, Lasserre, and Löfberg (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 1 to ensure global uniqueness.

4.3 Discussion of Assumptions and Limitations

Theorem 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 frequently employed model specification 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 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 popular 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 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

$$\frac{u'(y_i - w_i)}{v'(w_i)} = \lambda + \delta \frac{\mu'(y_i|a)}{\mu(y_i|a)} \tag{14}$$

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, Holmström (1979), Jewitt (1988) and Jewitt, Kadan, and Swinkels (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 (4) stated in Theorem 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 (10).

Since we cannot follow the traditional theoretical route, we would instead have to rely on numerical solutions of many instances of problem (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.[9] Our approach following the corollary and using polynomial methods circumvents these problems. In fact, the approach guarantees a globally optimal solution.

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, Larcker, and Su (2010). However, the analysis of linear multi-task principal-agent models in Holmström 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, Reny, and Swinkels (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.

---

[9]For example, Rogerson (1985) makes no reference to a constraint qualification in his derivation of (14). The same is true for Holmström (1979), Jewitt (1988), Conlon (2009), Sinclair-Desgagné (1994), and Jewitt, Kadan, and Swinkels (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 1. We complete our discussion with an illustration of the multi-dimensional approach by a numerical example.

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 (2010) for a review of mathematical results.

5.1.1 Multivariate Representation and Optimization

Putinar's Positivstellensatz is the analogue of the univariate sum of squares representation result from Proposition 3 for the multivariate case.

Proposition 6. [Putinar's Positivstellensatz, Lasserre (2010, 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

$$f = \sigma_0 + \sum_{i=1}^{m} \sigma_i g_i \tag{15}$$

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

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

$$p_{\min} = \inf_{\boldsymbol{x} \in K} p(\boldsymbol{x}). \tag{16}$$

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

$$\begin{aligned} &\sup_{\rho} \rho \\ &\text{s.t. } p(\boldsymbol{x}) - \rho > 0 \; \forall \boldsymbol{x} \in K \end{aligned} \tag{17}$$

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 (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 $d \geq \max \{d_p, d_{g_1}, \ldots, d_{g_m}\}$ consider the relaxation

$$\begin{aligned} \rho_d = &\sup_{\rho, \sigma_0, \sigma_1, \ldots, \sigma_m} \rho \\ &\text{s.t.} \quad p - \rho = \sigma_0 + \sum_{i=1}^{m} \sigma_i g_i \\ &\sigma_0 \in \Sigma_{2d}, \; \sigma_i \in \Sigma_{2(d - d_{g_i})} \end{aligned} \tag{18}$$

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

$$\rho_d = \sup_{\rho, Q^{(0)}, Q^{(1)}, \dots, Q^{(m)}} \rho$$

$$\text{s.t. } p - \rho = v_d^T Q^{(0)} v_d + \sum_{i=1}^{m} g_i v_{d-d_{g_i}}^T Q^{(i)} v_{d-d_{g_i}}$$

$$Q^{(0)}, Q^{(i)} \succcurlyeq 0 \tag{19}$$

$$Q^{(0)} \in \mathbb{R}^{\binom{n+d}{d} \times \binom{n+d}{d}}, \; Q^{(i)} \in \mathbb{R}^{\binom{n+d-d_{g_i}}{d-d_{g_i}} \times \binom{n+d-d_{g_i}}{d-d_{g_i}}}$$

$$v_d \text{ vector of monomials } \boldsymbol{x^{\alpha}} \text{ up to degree } d,$$

$$v_{d-d_{g_i}} \text{ vector of monomials } \boldsymbol{x^{\alpha}} \text{ up to degree } d - d_{g_i}$$

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.[10] If the problem is infeasible, then $\rho_d = -\infty$.

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 $d \geq \max \{d_p, d_{g_1}, \dots, d_{g_m}\}$ it holds that $\rho_d = p_{\min}$. We have the following theorem:

Proposition 7. [Lasserre (2010, Theorem 5.6)] *If the assumptions of Putinar's Positivstellensatz hold, then the optimal solution ρ_d of the relaxed problem (18) converges (from below) to the optimal value $p_{\min}$ of the original problem (16) as $d \to \infty$.*

5.1.2 Rational Objective Function

Jibetean and de Klerk (2006) also prove analogous results for the case of multivariate rational functions. Recall the optimization problem (12)

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

with $p, q \in \mathbb{R}[\boldsymbol{x}]$ and $K = \{\boldsymbol{x} \in \mathbb{R}^n \mid g_1(\boldsymbol{x}) \geq 0, \dots, g_m(\boldsymbol{x}) \geq 0\}$. For such a set K, the following proposition states that the weak inequality in the definition of $p_{\min}$ in Proposition 5 can be replaced by a strict inequality.

Proposition 8. [Jibetean and de Klerk (2006, Lemma 1)] *Suppose that K is the closure of some open connected set. Also suppose the assumptions of Proposition 5 hold. If p and q have no common factor then*

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

---

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

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

$$\rho_d = \sup_{\rho, \sigma_0, \sigma_1, \ldots, \sigma_m} \rho$$

$$\text{s.t.} \quad p - \rho q = \sigma_0 + \sum_{i=1}^{m} \sigma_i g_i \tag{20}$$

$$\sigma_0 \in \Sigma_{2d}, \ \sigma_1 \in \Sigma_{2(d - d_{g_i})}$$

Proposition 9. **[Jibetean and de Klerk (2006, Theorem 9)]** *Under the assumptions of Proposition 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 $\lim_{d \to \infty} \rho_d = p_{min}$.

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

A multi-dimensional version of Assumption 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 (19) to the agent's expected utility optimization problem, we obtain the following relaxation for that problem.

$$\sup_{\rho, Q^{(0)}, Q^{(1)}, \ldots, Q^{(m)}} \rho$$

$$\text{s.t.} \quad \sum_{\boldsymbol{\alpha}} c_{\boldsymbol{\alpha}}(\boldsymbol{w}) \boldsymbol{b}^{\boldsymbol{\alpha}} - \rho \sum_{\boldsymbol{\alpha}} f_{\boldsymbol{\alpha}}(\boldsymbol{w}) \boldsymbol{b}^{\boldsymbol{\alpha}} = v_d^T Q^{(0)} v_d + \sum_{i=1}^{m} g_i v_{d-d_{g_i}}^T Q^{(i)} v_{d-d_{g_i}}$$

$$Q^{(0)}, Q^{(i)} \succcurlyeq 0 \tag{21}$$

$$Q^{(0)} \in \mathbb{R}^{\binom{n+d}{d} \times \binom{n+d}{d}}, \ Q^{(i)} \in \mathbb{R}^{\binom{n+d-d_{g_i}}{d-d_{g_i}} \times \binom{n+d-d_{g_i}}{d-d_{g_i}}}$$

$$v_d \text{ vector of monomials } \boldsymbol{b}^{\boldsymbol{\alpha}} \text{ up to degree } d,$$

$$v_{d-d_{g_i}} \text{ vector of monomials } \boldsymbol{b}^{\boldsymbol{\alpha}} \text{ up to degree } d - d_{g_i}$$

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 $Q_{ij}^{(l)}$, $l = 0, 1, \ldots, m$, and the variable ρ. Next we use Proposition 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 (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 (forthcoming) 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 1. As before, we can safely replace the expressions *sup* and *inf* by *max* and *min* in the context of our application.

Theorem 2. *Suppose the agent's expected utility maximization problem satisfies Assumption 5 and the multi-dimensional version of Assumption 4. Let $\boldsymbol{v}_k$ be the vector of monomials in $b_1, \ldots, b_L$ 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 $i = 1, \ldots, m$, as additional decision variables, define the following relaxation of the principal-agent problem (1):*

$$\max_{\boldsymbol{w}, \boldsymbol{a}, \rho, L_{(0)}, \ldots L_{(m)}} U(\boldsymbol{w}, a) \quad subject\ to \tag{22}$$

$$\sum_{\boldsymbol{\alpha}} c_{\boldsymbol{\alpha}}(\boldsymbol{w}) \boldsymbol{b}^{\boldsymbol{\alpha}} - \rho \sum_{\boldsymbol{\alpha}} f_{\boldsymbol{\alpha}}(\boldsymbol{w}) \boldsymbol{b}^{\boldsymbol{\alpha}} = \boldsymbol{v}_d^T L_{(0)} L_{(0)}^T \boldsymbol{v}_d + \sum_{i=1}^{m} g_i \boldsymbol{v}_{d-d_{g_i}}^T L_{(i)} L_{(i)}^T \boldsymbol{v}_{d-d_{g_i}} \tag{22a}$$

$$\varepsilon \sum_{\boldsymbol{\alpha}} f_{\boldsymbol{\alpha}}(\boldsymbol{w}) \boldsymbol{a}^{\boldsymbol{\alpha}} \geq \sum_{\boldsymbol{\alpha}} c_{\boldsymbol{\alpha}}(\boldsymbol{w}) \boldsymbol{a}^{\boldsymbol{\alpha}} - \rho \sum_{\boldsymbol{\alpha}} f_{\boldsymbol{\alpha}}(\boldsymbol{w}) \boldsymbol{a}^{\boldsymbol{\alpha}} \tag{22b}$$

$$\sum_{i=0}^{d} c_i(\boldsymbol{w}) \boldsymbol{a}^i \leq -\underline{V} \left(\sum_{i=0}^{d} f_i(\boldsymbol{w}) \boldsymbol{a}^i \right) \tag{22c}$$

$$g_i(\boldsymbol{a}) \geq 0 \quad \forall\ i = 1, 2, \ldots, m \tag{22d}$$

$$\boldsymbol{w} \in W \tag{22e}$$

This optimization problem has the following properties.

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

$$\max_{\boldsymbol{a} \in A} V(\hat{\boldsymbol{w}}, \boldsymbol{a}) - V(\hat{\boldsymbol{w}}, \hat{\boldsymbol{a}}) \leq \varepsilon. \tag{23}$$

(b) Let $(\overline{\boldsymbol{w}}, \overline{\boldsymbol{a}})$ be a solution of the principal-agent problem (1). Then for any $\varepsilon > 0$ there exists $d(\varepsilon) \in \mathbb{N}$ and $\overline{\rho}, \overline{L}_{(0)}, \ldots, \overline{L}_{(m)}$, such that $\left(\overline{\boldsymbol{w}}, \overline{\boldsymbol{a}}, \overline{\rho}, \overline{L}_{(0)}, \ldots, \overline{L}_{(m)}\right)$ is feasible for the relaxation (22) for $d = d(\varepsilon)$.

(c) Let $(\overline{\boldsymbol{w}}, \overline{\boldsymbol{a}})$ be an optimal solution to (1). For any ε, let $d(\varepsilon)$ be as in (b). Denote by $u(\varepsilon)$ the optimal value of the relaxation (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 (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 of optimal solutions to (22), $\left(\boldsymbol{w}^*(\varepsilon), \boldsymbol{a}^*(\varepsilon), \rho^*(\varepsilon), L_{(0)}^*(\varepsilon), \ldots, L_{(m)}^*(\varepsilon)\right)$, projected onto $W \times A$, is contained in the set of optimal solutions to the original principal-agent problem (1).

Before we prove the theorem, we briefly describe the optimization problem (22). This problem has the same objective function as the original principal-agent problem (1). Constraint (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 (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 (21) are represented via products of lower triangular matrices. Proposition 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 (22a) ensures that $-\rho$ is an upper bound on the agent's possible expected utility levels. Next, constraint (22b) imposes a lower bound on the agent's expected utility level, namely $V(\boldsymbol{w}, \boldsymbol{a}) + \varepsilon \geq -\rho$. Therefore, the constraints (22a) and (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 (22c) is the transformed participation constraint for the agent's rational expected utility function. Constraint (22d) defines the set of the feasible actions and constraint (22e) is just the constraint on the compensation scheme from the original principal-agent problem (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 $V(\hat{\boldsymbol{w}}, \boldsymbol{a}) = -\frac{\sum_{\alpha} c_{\alpha}(\hat{\boldsymbol{w}})\boldsymbol{a}^{\alpha}}{\sum_{\alpha} f_{\alpha}(\hat{\boldsymbol{w}})\boldsymbol{a}^{\alpha}}$, since (22a) implies that

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

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 (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 (23) holds.

(b) Under the assumptions of the theorem, Proposition 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 (15) of Putinar's Positivstellensatz with degree d coefficients. For this d, problem (22) has a nonempty feasible region.

(c) Recall the agent's optimal value function $\Psi : W \to \mathbb{R}$ from the proof of Proposition 1. The projection of the set of feasible points of problem (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.[11] 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, $\lim_{\varepsilon \to 0^+} u(\varepsilon) = U(\overline{\boldsymbol{w}}, \overline{\boldsymbol{a}})$.

---

[11] 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 $s \in S(0)$.

(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

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

Thus $(\boldsymbol{w}_0, \boldsymbol{a}_0)$ is feasible for (1) and attains the optimal value.

This completes the proof of Theorem 2. $\qquad\square$

Some comments on the technical convergence results of Theorem 2 are in order. For the one-dimensional effort case, Theorem 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 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 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 (18) may only constitute a relaxation of the original polynomial optimization problem.

Once we have computed a solution we can always verify that it is feasible. To accomplish this we fix $\boldsymbol{w}$ and solve the polynomial optimization problem for the agent to global optimality using GloptiPoly (Henrion, Lasserre, and Löfberg 2009).

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 (2010, Theorem 5.15).[12] Also, if the objective function is strictly convex and the constraints are convex, then convergence is finite, see Lasserre (2010, 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.

---

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

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.

5.3 A Multivariate Example

Example 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 $b \geq 0$ and $a - 2b \geq 0$, 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

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

respectively. The expected utility of the agent is

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

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

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

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

$$0 = s_{1,3,1}^2 + s_{1,3,2}^2 + s_{1,3,3}^2 - s_{2,3,1}^2 - s_{2,3,2}^2 - s_{2,3,3}^2$$

$$0 = \frac{1}{2}\left(s_{2,2,1}^2 + s_{2,2,2}^2\right) - s_{3,2,1}^2 - s_{3,2,2}^2$$

$$0 = -1 + s_{0,2,1}^2 + s_{0,2,2}^2 + s_{2,1,1}s_{2,2,1} - 2s_{3,1,1}s_{3,2,1} + s_{3,2,1}^2 + s_{3,2,2}^2$$

$$0 = s_{1,2,1}^2 + s_{1,2,2}^2 - s_{2,2,1}^2 - s_{2,2,2}^2 + s_{2,2,1}s_{2,3,1} + s_{2,2,2}s_{2,3,2} - 2(s_{3,2,1}s_{3,3,1} + s_{3,2,2}s_{3,3,2})$$

$$0 = -\frac{11}{10} + 2(s_{0,2,1}s_{0,3,1} + s_{0,2,2}s_{0,3,2}) + 2s_{1,1,1}s_{1,2,1} - 2s_{2,1,1}s_{2,2,1} +$$

$$s_{2,1,1}s_{2,3,1} - 2s_{3,1,1}s_{3,3,1} + 2(s_{3,2,1}s_{3,3,1} + s_{3,2,2}s_{3,3,2})$$

$$0 = 2(s_{1,2,1}s_{1,3,1} + s_{1,2,2}s_{1,3,2}) - 2(s_{2,2,1}s_{2,3,1} + s_{2,2,2}s_{2,3,2}) + \frac{1}{2}\left(s_{2,3,1}^2 + s_{2,3,2}^2 + s_{2,3,3}^2\right) -$$

$$s_{3,3,1}^2 - s_{3,3,2}^2 - s_{3,3,3}^2$$

$$0 = -\frac{1}{10} + s_{0,3,1}^2 + s_{0,3,2}^2 + s_{0,3,3}^2 + 2s_{1,1,1}s_{1,3,1} - 2s_{2,1,1}s_{2,3,1} + s_{3,3,1}^2 + s_{3,3,2}^2 + s_{3,3,3}^2$$

$$0 = \rho + s_{0,1,1}^2 + s_{3,1,1}^2 + \log\left(1 + w_1\right)$$

$$0 = -\frac{1}{10} + 2s_{0,1,1}s_{0,3,1} + s_{1,1,1}^2 - s_{2,1,1}^2 + 2s_{3,1,1}s_{3,3,1} + \log\left(1 + w_1\right) +$$

$$\log\left(1 + w_2\right) - \log\left(1 + w_3\right)$$

$$0 = -1 + 2s_{0,1,1}s_{0,2,1} + \frac{s_{2,1,1}^2}{2} - s_{3,1,1}^2 + 2s_{3,1,1}s_{3,2,1} + \frac{1}{2}\log\left(1 + w_1\right) + \frac{1}{2}\log\left(1 + w_3\right).$$

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

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.

References

ALVI, E. (1997): "First-Order Approach to Principal-Agent Problems: A Generalization," *The GENEVA Papers on Risk and Insurance*, 22, 59–65.

ARAUJO, A. AND H. MOREIRA (2001): "A General Lagrangian Approach for Nonconcave Moral Hazard Problems," *Journal of Mathematical Economics*, 35, 17–39.

ARMSTRONG, C. S., D. F. LARCKER, AND C.-L. SU (2010): "Endogenous Selection and Moral Hazard in Compensation Contracts," *Operations Research*, 58, 1090–1106.

BERGE, C. (1963): *Topological Spaces*, Oliver & Boyd: Edinburgh and London.

BOYD, S. AND L. VANDENBERGHE (2004): *Convex Optimization*, Cambridge University Press.

CONLON, J. R. (2009): "Two New Conditions Supporting the First-Order Approach to Multisignal Principal–Agent Problems," *Econometrica*, 77, 249–278.

COUZOUDIS, E. AND P. RENNER (forthcoming): "Computing Gerneralized Nash Equilibria by Polynomial Programming," *Mathematical Methods of Operations Research.*

DUBÉ, J.-P., J. T. FOX, AND C.-L. SU (2012): "Improving the Numerical Performance of Static and Dynamic Aggregate Discrete Choice Random Coefficients Demand Estimation," *Econometrica*, 80, 2231–2267.

GROSSMAN, S. J. AND O. D. HART (1983): "An Analysis of the Principal-Agent Problem," *Econometrica*, 51, 7–45.

HENRION, D., J. LASSERRE, AND J. LÖFBERG (2009): "GloptiPoly 3: Moments, Optimization and Semidefinite Programming," *Optimization Methods and Software*, 24, 761–779.

HILBERT, D. (1888): "Über die Darstellung definiter Formen als Summe von Formenquadraten," *Mathematische Annalen*, 32, 342–350.

HOGAN, W. W. (1973): "Point-to-Set Maps in Mathematical Programming," *SIAM Review*, 15, pp. 591–603.

HOLMSTRÖM, B. (1979): "Moral Hazard and Observability," *The Bell Journal of Economics*, 10, 74–91.

HOLMSTRÖM, B. AND P. MILGROM (1991): "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, & Organization*, 7, 24–52.

JEWITT, I. (1988): "Justifying the First-Order Approach to Principal-Agent Problems," *Econometrica*, 56, 1177–1190.

JEWITT, I., O. KADAN, AND J. M. SWINKELS (2008): "Moral Hazard with Bounded Payments," *Journal of Economic Theory*, 143, 59–82.

JIBETEAN, D. AND E. DE KLERK (2006): "Global Optimization of Rational Functions: A Semidefinite Programming Approach," *Mathematical Programming*, 106, 93–109.

JUDD, K. AND C.-L. SU (2005): "Computation of Moral-Hazard Problems," Working paper.

KADAN, O., P. RENY, AND J. M. SWINKELS (2011): "Existence of Optimal Mechanisms in Principal-Agent Problems," MFI Working Paper No. 2011–002.

KADAN, O. AND J. M. SWINKELS (2012): "On the Moral Hazard Problem without the First-Order Approach," Working paper.

LASSERRE, J. (2010): *Moments, Positive Polynomials and Their Applications*, vol. 1 of *Imperial College Press Optimization Series*, Imperial College Press.

LAURENT, M. (2009): *Sums of Squares, Moment Matrices and Optimization Over Polynomials*, Springer, vol. 149 of *IMA Volumes in Mathematics and its Applications*, 157–270.

LICALZI, M. AND S. SPAETER (2003): "Distributions for the First-order Approach to Principal-agent Problems," *Economic Theory*, 21, 167–173.

MIRRLEES, J. A. (1979): "The Implications of Moral Hazard for Optimal Insurance," Mimeo.

——— (1999): "The Theory of Moral Hazard and Unobservable Behaviour: Part I," *The Review of Economic Studies*, 66, 3–21.

NIE, J. (2012): "An Exact Jacobian SDP Relaxation for Polynomial Optimization," *Mathematical Programming*, 1–31.

PRESCOTT, E. S. (1999): "A Primer on Moral-hazard Models," *Economic Quarterly*, 47–78.

——— (2004): "Computing Solutions to Moral-hazard Programs Using the Dantzig-Wolfe Decomposition Algorithm," *Journal of Economic Dynamics and Control*, 28, 777 – 800.

ROGERSON, W. P. (1985): "The First-Order Approach to Principal-Agent Problems," *Econometrica*, 53, pp. 1357–1367.

SCHMÜDGEN, K. (1991): "The K-moment Problem for Compact Semi-algebraic Sets," *Mathematische Annalen*, 289, 203–206.

SINCLAIR-DESGAGNÉ, B. (1994): "The First-Order Approach to Multi-Signal Principal-Agent Problems," *Econometrica*, 62, pp. 459–465.

SU, C.-L. AND K. L. JUDD (2012): "Constrained Optimization Approaches to Estimation of Structural Models," *Econometrica*, 80, 2213–2230.

VANDENBERGHE, L. AND S. BOYD (1996): "Semidefinite Programming," *SIAM Review*, 38, pp. 49–95.