Econometrica Supplementary Material

SUPPLEMENT TO "A POLYNOMIAL OPTIMIZATION APPROACH TO PRINCIPAL–AGENT PROBLEMS": ONLINE APPENDIX

(*Econometrica*, Vol. 83, No. 2, March 2015, 729–769)

BY PHILIPP RENNER AND KARL SCHMEDDERS

Appendix A provides additional mathematical background on multidimensional polynomial optimization. Appendix B gives a brief overview of Chebyshev interpolation. Appendix C provides some additional information on Example 3 in Section 5.2.

APPENDIX A: OPTIMIZATION OF POLYNOMIALS

WE FIRST INTRODUCE THE BASIC CONCEPTS from semidefinite optimization. Then we discuss multivariate polynomial optimization.

A.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 X. Recall that we denote the property of semidefiniteness of a symmetric matrix X by $X \succcurlyeq 0$. A semidefinite optimization problem (in standard form) is defined as follows.

DEFINITION A.1: Let $C, A_j \in \mathbb{R}^{n\times n}$ for all $j = 1, \dots, m$ be symmetric matrices and $b \in \mathbb{R}^m$. We then call the following convex optimization problem a *semidefinite program* (*SDP*):

$$\begin{aligned} (20)\qquad & \sup_X \operatorname{tr}(CX) \\ & \quad \text{s.t.}\quad \operatorname{tr}(A_j X) = b_j, \quad j = 1, \dots, m, \\ & \qquad\qquad X \succcurlyeq 0. \end{aligned}$$

Note that the SDP has a linear objective function and a closed convex feasible region. Thus, semidefinite programs are a special class of convex optimization problems. In fact, semidefinite programs can be solved efficiently both in

© 2015 The Econometric Society DOI: 10.3982/ECTA11351

theory and in practice; see Vandenberghe and Boyd (1996) and Boyd and Vandenberghe (2004).

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

DEFINITION A.2: Let $M = (m_{ij})_{i=1,\dots,n, j=1,\dots,n} \in \mathbb{R}^{n\times n}$ be a matrix and let $I \subset \{1, \dots, n\}$. Then $\det((m_{ij})_{(i,j)\in I\times I})$ is called a *principal minor*. If $I = \{1, \dots, k\}$, then $\det((m_{ij})_{(i,j)\in I\times I})$ is called the kth *leading principal minor*.

PROPOSITION A.1: *Let $Q \in \mathbb{R}^{n\times n}$ be a symmetric matrix with rank m. Then the following statements are equivalent*:

(a) *Q is positive semidefinite*.

(b) *All principal minors of Q are nonnegative*.

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

(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.

A.2. *Optimization of Multivariate Polynomials*

As we previously observed, 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 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 set of nonnegative polynomials and the set of sums of squares 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 polynomials and sums of squares for 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.

A.2.1. *Multivariate Representation and Optimization*

Putinar's Positivstellensatz is the analogue of the univariate sum of squares representation result from Proposition 2 for the multivariate case.

PROPOSITION A.2—Putinar's Positivstellensatz (Lasserre (2010, Theorem 2.14)): *Let* $f, g_1, \dots, g_m \in \mathbb{R}[\mathbf{x}]$ *be polynomials and* $K = \{\mathbf{x} \in \mathbb{R}^n \mid g_1(\mathbf{x}) \geq 0, \dots, g_m(\mathbf{x}) \geq 0\} \subset \mathbb{R}^n$ *a basic semi-algebraic set such that at least one of the following conditions holds*:

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

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

If f *is strictly positive on* K, *then*

$$f = \sigma_0 + \sum_{i=1}^{m} \sigma_i g_i \tag{21}$$

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

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 $\|\mathbf{x}\|_2 \leq B$ for all $\mathbf{x} \in K$, then we can add the redundant ball constraint $B^2 - \sum_i x_i^2 \geq 0$. Note that in contrast to Proposition 2 for univariate polynomials, Putinar's Positivstellensatz does not provide any bounds on the degree of the sums of squares σ_j.

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

$$p_{\min} = \inf_{\mathbf{x} \in K} p(\mathbf{x}). \tag{22}$$

Similarly to the univariate case, we can rewrite this problem:

$$\begin{aligned} &\sup_{\rho} \rho \\ &\quad \text{s.t.} \quad p(\mathbf{x}) - \rho > 0 \quad \forall \mathbf{x} \in K. \end{aligned} \tag{23}$$

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 (23) 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

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

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

$$
\begin{aligned}
(25) \qquad \rho_d = & \sup_{\rho, Q^{(0)}, Q^{(1)}, \ldots, Q^{(m)}} \rho \\
& \text{s.t.} \quad p - \rho = v_d^T Q^{(0)} v_d + \sum_{i=1}^{m} g_i v_{d-d_{g_i}}^T Q^{(i)} v_{d-d_{g_i}}, \\
& Q^{(0)} \succcurlyeq 0, \quad Q^{(i)} \succcurlyeq 0 \quad \forall i = 1, 2, \ldots, m, \\
& Q^{(0)} \in \mathbb{R}^{\binom{n+d}{d} \times \binom{n+d}{d}}, \\
& Q^{(i)} \in \mathbb{R}^{\binom{n+d-d_{g_i}}{d-d_{g_i}} \times \binom{n+d-d_{g_i}}{d-d_{g_i}}} \quad \forall i = 1, 2, \ldots, m, \\
& v_d \text{ vector of monomials } \mathbf{x}^{\boldsymbol{\alpha}} \text{ up to degree } d, \\
& v_{d-d_{g_i}} \text{ vector of monomials } \mathbf{x}^{\boldsymbol{\alpha}} \text{ up to degree } d - d_{g_i}.
\end{aligned}
$$

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

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

PROPOSITION A.3—Lasserre (2010, Theorem 5.6): *If the assumptions of Putinar's Positivstellensatz hold, then the optimal solution ρ_d of the relaxed problem* (24) *converges* (*from below*) *to the optimal value $p_{\min}$ of the original problem* (22) *as $d \to \infty$.*

For the rate of convergence, we refer to Nie and Schweighofer (2007).

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

A.2.2. *Rational Objective Function*

Jibetean and De Klerk (2006) also proved analogous results for the case of multivariate rational functions. Recall the optimization problem (10) from the main body of the paper,

$$p_{\min} = \inf_{\mathbf{x}\in K, q(\mathbf{x})\neq 0} \frac{p(\mathbf{x})}{q(\mathbf{x})}$$

with $p, q \in \mathbb{R}[\mathbf{x}]$ and $K = \{\mathbf{x} \in \mathbb{R}^n \mid g_1(\mathbf{x}) \geq 0, \dots, g_m(\mathbf{x}) \geq 0\}$. For such a set K, the following proposition states that the weak inequality in the definition of $p_{\min}$ in Proposition 4 can be replaced by a strict inequality.

PROPOSITION A.4—Jibetean and De Klerk (2006, Lemma 1): *Suppose that K is the closure of some open connected set. Also suppose the assumptions of Proposition* 4 *hold and q does not change sign on K. If p and q have no common factor, then*

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

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

$$\begin{aligned} \rho_d = & \sup_{\rho, \sigma_0, \sigma_1, \dots, \sigma_m} \rho \\ & \text{s.t.} \quad p - \rho q = \sigma_0 + \sum_{i=1}^{m} \sigma_i g_i, \\ & \qquad \sigma_0 \in \Sigma_{2d}, \quad \sigma_1 \in \Sigma_{2(d-d_{g_i})}. \end{aligned} \tag{26}$$

PROPOSITION A.5—Jibetean and De Klerk (2006, Theorem 9): *Under the assumptions of Proposition* 4 *and Putinar's Positivstellensatz, the following statements hold*:

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

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

APPENDIX B: CHEBYSHEV INTERPOLATION

Many applications of the principal–agent model, such as those examples in Sections 5.1 and 5.2, include utility or probability functions that are not rational. Such functions must first be approximated by polynomial or rational functions before the polynomial optimization approach becomes applicable. The tools of polynomial approximation are manifold. Trefethen (2013) provided an excellent overview of one-dimensional approximation theory and practice. In this paper, we restrict ourselves to Chebyshev approximation due to its popularity in economics. In this appendix, we briefly outline Chebyshev interpolation. For many more details, see Judd (1998) and Trefethen (2013).

Consider the problem of approximating a sufficiently smooth function f on a compact interval. Without loss of generality, we consider the interval $[-1, 1]$. Suppose that $x_0, x_1, \dots, x_n$ are $n+1$ distinct points in $[-1, 1]$. There exists a unique polynomial $p_n(x)$ of degree at most n with the property

$$f(x_i) = p_n(x_i) \quad \text{for} \quad i = 0, 1, \dots, n,$$

for the function $f : [-1, 1] \to \mathbb{R}$. That is, the polynomial p_n interpolates the $n+1$ points $(x_0, f(x_0)), (x_1, f(x_1)), \dots, (x_n, f(x_n))$. Next, we define Chebyshev nodes and Chebyshev polynomials.

DEFINITION B.1: Let $n \in \mathbb{N}$ and define the points

$$x_j = \cos\left(\frac{(2j+1)\pi}{2n+2}\right) \quad \text{for} \quad j = 0, \dots, n,$$

in the interval $[-1, 1]$. These points are called the *Chebyshev nodes*.

Define the sequence of polynomials $(T_j)_{j\geq 0}$ by

$$T_0 = 1,$$
$$T_1 = x,$$
$$T_{k+1} = 2xT_k - T_{k-1} \quad \text{for} \quad k \geq 1.$$

We call T_k the *kth Chebyshev polynomial*.

Let $n \in \mathbb{N}$ and $x_0, \dots, x_n$ be the Chebyshev nodes. We call the function

$$P_n(x) = \sum_{i=0}^{n} c_i T_i(x) \tag{27}$$

the *Chebyshev interpolant* of f of degree n, where the coefficients $c_0, \dots, c_n$ are given by

$$c_0 = \frac{1}{n+1} \sum_{j=0}^{n} f(x_j),$$

$$c_k = \frac{2}{n+1} \sum_{j=0}^{n} f(x_j) \cos\left(\frac{k(2j+1)\pi}{2n+2}\right), \quad k = 1, \dots, n.$$

Note that the Chebyshev nodes $x_0, \dots, x_n$ correspond to the $n+1$ zeros of the Chebyshev polynomial T_{n+1}. The interpolant satisfies $p_n(x_j) = f(x_j)$ for all $j = 0, \dots, n$. In particular, the interpolation is exact at each node x_j.

The following proposition describes the convergence of the Chebyshev interpolant for ν-times differentiable functions.

PROPOSITION B.1—Judd (1998, Theorem 6.7.3): *Let* $\nu > 0$ *be an integer and* f *be a* ν*-times differentiable function on* $[-1, 1]$. *Then*

$$\|f - p_n\|_\infty \leq \left(\frac{2}{\pi}\ln(n+2) + 1\right)\frac{(n+1-\nu)!}{(n+1)!}\left(\frac{\pi}{2}\right)^\nu \|f^{(\nu)}\|_\infty.$$

Note that the rate of convergence here is $O(n^{-\nu})$. For analytic functions, the convergence behavior is even better. The Chebyshev interpolation converges at $O(C^{-n})$, where $C > 1$ depends on some properties of the function extended to the complex plane. For details, see Trefethen (2013, Theorem 8.2).

APPENDIX C: FUNCTIONS FOR APPLICATION ON EXECUTIVE COMPENSATION IN SECTION 5.2

The following expression is the function, which is used to approximate the agent's utility function. It is a Chebyshev polynomial in the variable a:

$$\begin{aligned}
&0.005\sqrt{0.14\beta_O + 1.27\beta_S + w}\,a^7 - 0.002\sqrt{0.78\beta_S + w}\,a^7 \\
&\quad - 0.005\sqrt{0.94\beta_O + 2.08\beta_S + w}\,a^7 \\
&\quad + 0.002\sqrt{2.32\beta_O + 3.46\beta_S + w}\,a^7 \\
&\quad - 0.00\sqrt{0.48\beta_S + w}\,a^6 + 0.007\sqrt{0.78\beta_S + w}\,a^6 \\
&\quad - 0.003\sqrt{0.14\beta_O + 1.27\beta_S + w}\,a^6 \\
&\quad - 0.005\sqrt{0.94\beta_O + 2.08\beta_S + w}\,a^6 \\
&\quad + 0.008\sqrt{2.32\beta_O + 3.46\beta_S + w}\,a^6 \\
&\quad - 0.003\sqrt{4.86\beta_O + 5.99\beta_S + w}\,a^6 \\
&\quad - 0.002\sqrt{0.28\beta_S + w}\,a^5 + 0.006\sqrt{0.48\beta_S + w}\,a^5 \\
&\quad + 0.006\sqrt{0.78\beta_S + w}\,a^5 - 0.03\sqrt{0.14\beta_O + 1.27\beta_S + w}\,a^5 \\
&\quad + 0.03\sqrt{0.94\beta_O + 2.08\beta_S + w}\,a^5 - 0.007\sqrt{2.32\beta_O + 3.46\beta_S + w}\,a^5 \\
&\quad + 0.001\sqrt{10.22\beta_O + 11.35\beta_S + w}\,a^5 + 0.006\sqrt{0.28\beta_S + w}\,a^4 \\
&\quad + 0.005\sqrt{0.48\beta_S + w}\,a^4 - 0.05\sqrt{0.78\beta_S + w}\,a^4 \\
&\quad + 0.04\sqrt{0.14\beta_O + 1.27\beta_S + w}\,a^4 + 0.05\sqrt{0.94\beta_O + 2.08\beta_S + w}\,a^4 \\
&\quad - 0.06\sqrt{2.32\beta_O + 3.46\beta_S + w}\,a^4 - 0.18a^3
\end{aligned}$$

$$
\begin{aligned}
&+0.01\sqrt{4.86\beta_O+5.99\beta_S+w}a^4-0.04a^4-0.01\sqrt{0.28\beta_S+w}a^3\\
&-0.04\sqrt{0.48\beta_S+w}a^3+0.05\sqrt{0.78\beta_S+w}a^3\\
&+0.15\sqrt{0.14\beta_O+1.27\beta_S+w}a^3-0.05\sqrt{2.32\beta_O+3.46\beta_S+w}a^3\\
&+0.05\sqrt{4.86\beta_O+5.99\beta_S+w}a^3-0.15\sqrt{0.94\beta_O+2.08\beta_S+w}a^3\\
&+0.005\sqrt{10.22\beta_O+11.35\beta_S+w}a^3\\
&+0.16\sqrt{2.32\beta_O+3.46\beta_S+w}a^2+0.01\sqrt{0.28\beta_S+w}a^2\\
&+0.09\sqrt{0.48\beta_S+w}a^2+0.15\sqrt{0.78\beta_S+w}a^2\\
&-0.24\sqrt{0.14\beta_O+1.27\beta_S+w}a^2\\
&-0.27\sqrt{0.94\beta_O+2.08\beta_S+w}a^2+0.09\sqrt{4.86\beta_O+5.99\beta_S+w}a^2\\
&+0.004\sqrt{10.22\beta_O+11.35\beta_S+w}a^2-0.28a^2\\
&-0.006\sqrt{0.28\beta_S+w}a-0.09\sqrt{0.48\beta_S+w}a-0.38\sqrt{0.78\beta_S+w}a\\
&+0.003\sqrt{10.22\beta_O+11.35\beta_S+w}a^4\\
&-0.34\sqrt{0.14\beta_O+1.27\beta_S+w}a+0.33\sqrt{0.94\beta_O+2.08\beta_S+w}a\\
&+0.41\sqrt{2.32\beta_O+3.46\beta_S+w}a+0.08\sqrt{4.86\beta_O+5.99\beta_S+w}a\\
&+0.002\sqrt{10.22\beta_O+11.35\beta_S+w}a-0.18a+0.001\sqrt{0.28\beta_S+w}\\
&+0.03\sqrt{0.48\beta_S+w}+0.67\sqrt{0.14\beta_O+1.27\beta_S+w}\\
&+0.71\sqrt{0.94\beta_O+2.08\beta_S+w}+0.27\sqrt{2.32\beta_O+3.46\beta_S+w}\\
&+0.03\sqrt{4.86\beta_O+5.99\beta_S+w}-0.04+0.25\sqrt{0.78\beta_S+w}\\
&-0.005\sqrt{4.86\beta_O+5.99\beta_S+w}a^5.
\end{aligned}
$$

The utility function for the principal, with ten quadrature nodes, is the following:

$$
\begin{aligned}
&0.56e^{-a^2/2-0.51a-0.50}\big(1.27(1-\beta_S)-0.14\beta_O\big)\\
&+0.56e^{-a^2/2+0.46a-0.46}\big(2.082(1-\beta_S)-0.94\beta_O\big)\\
&+0.56e^{-a^2/2+1.48a-1.40}\big(3.46(1-\beta_S)-2.32\beta_O\big)\\
&+0.56e^{-a^2/2+2.58a-3.53}\big(5.99(1-\beta_S)-4.86\beta_O\big)\\
&+0.56e^{-a^2/2+3.85a-7.42}\big(11.35(1-\beta_S)-10.22\beta_O\big)
\end{aligned}
$$

$$+0.44e^{-a^2/2-1.48a-1.47}(1-\beta_S)+0.27e^{-a^2/2-2.46a-3.39}(1-\beta_S)$$
$$+0.16e^{-a^2/2-3.48a-6.36}(1-\beta_S)+0.09e^{-a^2/2-4.58a-10.69}(1-\beta_S)$$
$$+0.04e^{-a^2/2-5.85a-17.14}(1-\beta_S)-w.$$

REFERENCES

BOYD, S., AND L. VANDENBERGHE (2004): *Convex Optimization*. Cambridge: Cambridge University Press. [2]

HILBERT, D. (1888): "Über die Darstellung definiter Formen als Summe von Formenquadraten," *Mathematische Annalen*, 32, 342–350. [2]

JIBETEAN, D., AND E. DE KLERK (2006): "Global Optimization of Rational Functions: A Semidefinite Programming Approach," *Mathematical Programming*, 106, 93–109. [5]

JUDD, K. (1998): *Numerical Methods in Economics*. Cambridge, MA: MIT Press. [6,7]

LASSERRE, J. (2010): *Moments, Positive Polynomials and Their Applications*. Imperial College Press Optimization Series, Vol. 1. London: Imperial College Press. [3,4]

LAURENT, M. (2009): "Sums of Squares, Moment Matrices and Optimization Over Polynomials," in *Emerging Applications of Algebraic Geometry*. IMA Volumes in Mathematics and Its Applications, Vol. 149. New York: Springer, 157–270. [3]

NIE, J., AND M. SCHWEIGHOFER (2007): "On the Complexity of Putinar's Positivstellensatz," *Journal of Complexity*, 23, 135–150. [4]

TREFETHEN, L. N. (2013): *Approximation Theory and Approximation Practice*. Philadelphia, PA: SIAM. [5-7]

VANDENBERGHE, L., AND S. BOYD (1996): "Semidefinite Programming," *SIAM Review*, 38, 49–95. [2]

Hoover Institution, Stanford University, 434 Galvez Mall, Stanford, CA 94305, U.S.A.; phrenner@gmail.com

and

University of Zurich and Swiss Finance Institute, Moussonstrasse 15, 8044 Zurich, Switzerland; karl.schmedders@business.uzh.ch.

Manuscript received January, 2013; final revision received September, 2014.