CentER

Discussion Paper

No. 2005–73

DISCRETE LEAST-NORM APPROXIMATION BY NONNEGATIVE (TRIGONOMETRIC) POLYNOMIALS AND RATIONAL FUNCTIONS

By Alex Y.D. Siem, Etienne de Klerk, Dick den Hertog

June 2005

ISSN 0924-7815

Tilburg University

Discrete least-norm approximation by nonnegative (trigonometric) polynomials and rational functions

A.Y.D. Siem* E. de Klerk‡ D. den Hertog§

14th June 2005

Abstract

Polynomials, trigonometric polynomials, and rational functions are widely used for the discrete approximation of functions or simulation models. Often, it is known beforehand, that the underlying unknown function has certain properties, e.g. nonnegative or increasing on a certain region. However, the approximation may not inherit these properties automatically. We present some methodology (using semidefinite programming and results from real algebraic geometry) for least-norm approximation by polynomials, trigonometric polynomials and rational functions that preserve nonnegativity.

Keywords: (trigonometric) polynomials, rational functions, semidefinite programming, regression, (Chebyshev) approximation.

JEL Classification: C60.

1 Introduction

In the field of approximation theory, polynomials, trigonometric polynomials, and rational functions are widely used; see e.g. Cuyt and Lenin (2002), Cuyt et al. (2004), Fassbender (1997), Forsberg and Nilsson (2005), Jansson et al. (2003), and Yeun et al. (2005). For books on approximation theory, we refer to Powell (1981) and Watson (1980). In the field of computer simulations (both deterministic and stochastic), they are used to approximate the input/output

*Department of Econometrics and Operations Research/ Center for Economic Research (CentER), Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands, Phone:+31 13 4663254, Fax:+31 13 4663280, E-mail: a.y.d.siem@uvt.nl

‡Department of Econometrics and Operations Research/ Center for Economic Research (CentER), Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands, Phone:+31 13 4662031, Fax:+31 13 4663280, E-mail: e.deklerk@uvt.nl

§Department of Econometrics and Operations Research/ Center for Economic Research (CentER), Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands, Phone:+31 13 4662122, Fax:+31 13 4663280, E-mail: d.denhertog@uvt.nl

behavior of a computer simulation. The approximation is also called a metamodel, response surface model, compact model, surrogate model, emulator, or regression model.

We are interested in approximating a function $y : \mathbb{R}^q \mapsto \mathbb{R}$ which is only known up to an error, in a finite set of points $x^1, \ldots, x^n \in \mathbb{R}^q$. We denote the known output values by $y(x^1), \ldots, y(x^n)$. In practice, it is often known beforehand, that the function $y(x)$ has certain properties. Thus it may be known e.g. that the function is nonnegative, increasing or convex. However, it could happen that the approximation does not inherit these properties. It could even be the case that the data does not have the properties, due to errors in the data.

Therefore, in this paper, we construct (trigonometric) polynomial and rational approximations that preserve nonnegativity. For polynomials, we also discuss how to construct increasing polynomial approximations, using the same methodology as for nonnegative approximations. We illustrate the methodology with some examples.

In the field of splines, there is some literature on shape preserving approximations; see e.g. Kuijt (1998) and Kuijt and Van Damme (2001). In Kuijt and Van Damme (2001) a linear approach to shape preserving spline approximation is discussed. Linear constraints are given for shape-preserving univariate B-splines and bivariate tensorproduct B-splines. However, these constraints are only sufficient and in general not necessary. In the field of statistical inference, much work has been done in the estimation of univariate functions restricted by monotonicity; see e.g. Barlow, Bartholomew, Bremner, and Brunk (1972) and Robertson, Wright, and Dykstra (1988). However, these methods cannot be used for least-norm approximation, since they are parameter free.

This paper is organized as follows. In Section 2 we will discuss least-norm approximation by nonnegative and increasing polynomials. Subsequently, in Section 3 we show that we can use the same methodology for least-norm approximation by nonnegative univariate trigonometric polynomials. In Section 4, we discuss least-norm approximation by nonnegative rational functions. In Section 5, we show how to exploit the structure of the problem to speed up the computation of the solution. Finally in Section 6, we summarize our results and discuss possible directions for further research.

2 Approximation by polynomials

We are interested in approximating a function $y : \mathbb{R}^q \mapsto \mathbb{R}$ by a polynomial $p : \mathbb{R}^q \mapsto \mathbb{R}$ of degree d, given input data $x^1, \ldots, x^n \in \mathbb{R}^q$ and corresponding output data $y^1, \ldots, y^n \in \mathbb{R}$ (i.e. $y^i = y(x^i)$). Here, $p(x)$ is defined in terms of a given basis of $m + 1$ monomials:

$$p(x) = \sum_{j=0}^{m} \alpha_j p_j(x),$$

where α_j is the coefficient of the j-th monomial $p_j(x)$.

2.1 General least norm approximation by polynomials

Define $p_\alpha = \left[p(x^1), \dots, p(x^n)\right]^T$ and $y = \left[y(x^1), \dots, y(x^n)\right]^T$. The coefficients α_j are determined by solving the following least-norm optimization problem:

$$\min_\alpha \|p_\alpha - y\|. \tag{1}$$

It is well-known from statistics that the solution for the ℓ_2-norm in (1) is given by

$$\alpha = (D^T D)^{-1} D^T y,$$

where $\alpha = [\alpha_0, \dots, \alpha_m]^T$, and

$$D = \begin{bmatrix} p_0(x^1) & p_1(x^1) & \cdots & p_m(x^1) \\ p_0(x^2) & p_1(x^2) & \cdots & p_m(x^2) \\ \vdots & \vdots & & \vdots \\ p_0(x^n) & p_1(x^n) & \cdots & p_m(x^n) \end{bmatrix}.$$

If we use the ℓ_1-norm or the ℓ_∞-norm, problem (1) can be reformulated as a linear program. Note that by solving (1), we cannot guarantee that $p(x)$ will be nonnegative, even if the data y are nonnegative.

2.2 Approximation by nonnegative polynomials

If we know that the function $y(x)$ is nonnegative on a certain bounded region $\mathcal{U}$, we would like $p(x)$ to be nonnegative on $\mathcal{U}$ as well. We could force this by solving the following mathematical program:

$$\begin{array}{ll} \min\limits_\alpha & \|p_\alpha - y\| \\ \text{s.t.} & p(x) \geq 0 \quad \forall x \in \mathcal{U}. \end{array} \tag{2}$$

Note that using the ℓ_2-norm, (2) is a nonlinear optimization problem with infinitely many constraints, which can be rewritten as

$$\begin{array}{lll} \min\limits_{\alpha,t} & t & \\ \text{s.t.} & \|p_\alpha - y\|_2 \leq t & \\ & p(x) \geq 0 & \forall x \in \mathcal{U}, \end{array}$$

and gives an semi-infinite LP with an additional second order cone constraint. By using the ℓ_1-norm or the ℓ_∞-norm, we obtain a linear program. In case we use the ℓ_1-norm the mathematical program becomes:

$$
\begin{aligned}
\min_{\alpha, t_1, \ldots, t_n} \quad & \sum_{i=1}^{n} t_i \\
\text{s.t.} \quad & \sum_{j=0}^{m} \alpha_j p_j(x) \geq 0 && \forall x \in \mathcal{U} \\
& \sum_{j=0}^{m} \alpha_j p_j(x^i) - t_i \leq y(x^i) && \forall i = 1, \ldots, n \\
& -\sum_{j=0}^{m} \alpha_j p_j(x^i) - t_i \leq -y(x^i) && \forall i = 1, \ldots, n.
\end{aligned}
\tag{3}
$$

In case we use the ℓ_∞-norm the mathematical program becomes:

$$
\begin{aligned}
\mathcal{E} := \min_{\alpha, t} \quad & t \\
\text{s.t.} \quad & \sum_{j=0}^{m} \alpha_j p_j(x) \geq 0 && \forall x \in \mathcal{U} \\
& \sum_{j=0}^{m} \alpha_j p_j(x^i) - t \leq y(x^i) && \forall i = 1, \ldots, n \\
& -\sum_{j=0}^{m} \alpha_j p_j(x^i) - t \leq -y(x^i) && \forall i = 1, \ldots, n.
\end{aligned}
\tag{4}
$$

In the rest of this paper we will only treat the ℓ_∞-norm. This kind of approximation is also called Chebyshev approximation. The methods that we will present in this paper, can also be used in the ℓ_1 and the ℓ_2 case.

We will show that we can obtain an upper bound of the solution of optimization problem (4) by using semidefinite programming, and obtain the exact solution in the univariate case. Before we proceed, we first give two theorems. The following theorem gives a characterization of nonnegative polynomials that can be written as sums of squares (SoS) of polynomials.

Theorem 1 (Hilbert (1888))**.** *Any polynomial in q variables with degree d which is nonnegative on $\mathbb{R}^q$ can be decomposed as a sum of squares of polynomials (SoS), for $q = 1$, or $d = 2$, or ($q = 2$ and $d = 4$).*

See Reznick (2000) for a historical discussion and related results. The next theorem gives a useful way to represent SoS polynomials in terms of positive semidefinite matrices.

Theorem 2. *Let $x \in \mathbb{R}^q$ and let $p(x)$, a polynomial of degree $d = 2k$, be SoS. Then there exists a matrix $P \succeq 0$ such that $p(x) = e^T(x)Pe(x)$, where $e(x)$ is a vector consisting of all monomials of degree $d \leq k$.*

Proof. See e.g. Nesterov (2000). □

2.2.1 Univariate nonnegative polynomials

Let us first consider the approximation of a univariate nonnegative function $y(x)$ by a nonnegative polynomial. In this case, Theorem 1 shows that we can write the polynomial as an SoS. Then, using Theorem 2 we can write this nonnegative polynomial as $p(x) = e^T(x)Pe(x)$. For the ℓ_∞-norm, optimization problem (4) can be rewritten as the semidefinite programming problem

$$\begin{array}{ll} \min\limits_{t,P} & t \\ \text{s.t.} & e^T(x^i)Pe(x^i) - t \leq y(x^i) \qquad \forall i = 1, \ldots, n \\ & -e^T(x^i)Pe(x^i) - t \leq -y(x^i) \quad \forall i = 1, \ldots, n \\ & P \succeq 0. \end{array} \tag{5}$$

In practice, however, we are only interested in the polynomial to be nonnegative on a bounded interval; i.e. $\mathcal{U} = [a_0, b_0]$. Without loss of generality we may consider the interval $\mathcal{U} = [-1, 1]$, since we can scale and translate general intervals $[a_0, b_0]$ to $[-1, 1]$.

To construct nonnegative approximation, we use the following theorem.

Theorem 3. *A polynomial $p(x)$ is nonnegative on $[-1, 1]$ if and only if it can be written as*

$$p(x) = f(x) + (1 - x^2)g(x), \tag{6}$$

where $f(x)$ and $g(x)$ are SoS of degree at most $2d$ and $2d - 2$ respectively.

Proof. See e.g. Powers and Reznick (2000). □

Using this, we obtain the following semidefinite programming problem:

$$\begin{array}{lll} \min\limits_{t,P,Q} & t & \\ \text{s.t.} & e_1^T(x^i)Pe_1(x^i) + (1 - (x^i)^2)e_2^T(x^i)Qe_2^T(x^i) - t \leq y(x^i) & \forall i = 1, \ldots, n \\ & -e_1^T(x^i)Pe_1(x^i) - (1 - (x^i)^2)e_2^T(x^i)Qe_2^T(x^i) - t \leq -y(x^i) & \forall i = 1, \ldots, n \\ & P \succeq 0 & \\ & Q \succeq 0, & \end{array} \tag{7}$$

where $e_1(x)$ and $e_2(x)$ are defined in a similar way as $e(x)$; i.e. $e_1(x)$ is a vector consisting of all monomials of degree up to d, and $e_2(x)$ is a vector consisting of all monomials of degree up to $d - 1$. Note that (7) is an exact reformulation of (4) with $\mathcal{U} = [-1, 1]$.

2.2.2 Multivariate nonnegative polynomials

If we are interested in approximating a function on $\mathbb{R}^q$, then we can use Hilbert's theorem in combination with Theorem 2, use semidefinite programming and solve a multivariate version of (5). In this way, we obtain an exact solution of (4), for $q = 1$, $d = 2$ or ($q = 2$ and $d = 4$). In the other cases, by assuming the nonnegative polynomial approximation to be SoS and using Theorem 2, we will merely get an upper bound of the optimal solution of (4).

However, in practice we are primarily interested in nonnegative polynomials on compact regions, instead of $\mathbb{R}^q$. The following theorem describes a property of a polynomial, which is positive on a compact semi-algebraic set.

Theorem 4 (Putinar)**.** *Assume that the semi-algebraic set* $F = \{x \in \mathbb{R}^q | g_\ell(x) \geq 0, \ell = 1, \ldots, \bar{m}\}$*, where* $g_1, g_2, \ldots, g_{\bar{m}}$ *are polynomials, is compact and that there exists a polynomial* $u(x)$ *of the form* $u(x) = u_0(x) + \sum_{\ell=1}^{\bar{m}} u_\ell(x) g_\ell(x)$*, where* $u_0, u_1, \ldots, u_{\bar{m}}$ *are SoS, and for which the set* $\{x \in \mathbb{R}^q | u(x) \geq 0\}$ *is compact. Then, every polynomial* $p(x)$ *positive on* F *has a decomposition*

$$p(x) = p_0(x) + \sum_{\ell=1}^{\bar{m}} p_\ell(x) g_\ell(x),$$

where $p_0, p_1, \ldots, p_{\bar{m}}$ *are SoS.*

Proof. See Putinar (1993). For a more elementary proof, see Schweighofer (2004). □

If $\mathcal{U} = \{x \in \mathbb{R}^q | g_\ell(x) \geq 0, \ell = 1, \ldots, \bar{m}\}$ is compact, and if we know a ball $B(0, R)$ such that $\mathcal{U} \subseteq B(0, R)$, then the condition in Theorem 4 holds. Indeed $\mathcal{U} = \mathcal{U} \cap B(0, R) = \{x \in \mathbb{R}^q : g_\ell(x) \geq 0, \ell = 1, \ldots, \bar{m}, g_{\bar{m}+1}(x) = R^2 - \sum_{i=1}^q x_i^2 \geq 0\}$ and there exists a $u(x) = u_0(x) + \sum_{\ell=1}^{\bar{m}+1} u_\ell(x) g_\ell(x)$, where $u_0, u_1, \ldots, u_{\bar{m}+1}$ are SoS, for which the set $\{x \in \mathbb{R}^q | u(x) \geq 0\}$ is compact. Take $u_0(x) = u_1(x) = \ldots = u_{\bar{m}}(x) = 0$ and $u_{\bar{m}+1}(x) = 1$, to obtain $B(0, R) = \{x \in \mathbb{R}^q | u(x) \geq 0\}$.

Now, we can obtain an upper bound for the solution of (4) by solving semidefinite programming problem:

$$\begin{array}{lll} \min\limits_{t, P_0, \ldots, P_{\bar{m}+1}} & t & \\ \text{s.t.} & \sum\limits_{\ell=0}^{\bar{m}+1} e_\ell^T(x^i) P_\ell e_\ell(x^i) g_\ell(x^i) - t < y(x^i) & \forall i = 1, \ldots, n \\ & -\sum\limits_{\ell=0}^{\bar{m}+1} e_\ell^T(x^i) P_\ell e_\ell(x^i) g_\ell(x^i) - t < -y(x^i) & \forall i = 1, \ldots, n \\ & P_\ell \succeq 0 & \ell = 0, \ldots, \bar{m}+1, \end{array} \tag{8}$$

where $g_0 \equiv 1$ and $g_{\bar{m}+1}(x) = R^2 - \sum_{i=1}^q x_i^2$. Note that Theorem 4 does not state which degree d the polynomials $p_0, p_1, \ldots, p_{\bar{m}+1}$ have. In practice we have to choose a fixed degree d. Therefore, by solving (8), we get an upper bound of the maximum error $\mathcal{E}$ in (4). Consequently, by restricting ourselves to polynomials of degree d, we cannot guarantee in Theorem 4 that all positive polynomials $p(x)$ of degree d can be written as $p(x) = p_0(x) + \sum_{\ell=1}^{\bar{m}+1} p_\ell(x) g_\ell(x)$, with $p_0, p_1, \ldots, p_{\bar{m}+1}$ polynomials of degree d. Note that in the univariate case, Theorem 3 gives upper bounds for the degrees of $f(x)$ and $g(x)$, so for this case we can solve (4) exactly.

Example 2.1

We consider a two-dimensional example. Given the data in Table 1, we are interested in finding a nonnegative polynomial of degree $d = 3$ on $[0,1]^2$ for which the maximal error at the data points is minimized. First we exclude the nonnegativity constraint, i.e., we solve (1) for the ℓ_∞-norm. This yields a polynomial on $[0,1]^2$ that takes negative values. It turns out that $\mathcal{E} = 0.025$ in (4) and the optimal polynomial is given by $p(x_1, x_2) = 0.9747 - 2.3155x_1 - 7.1503x_2 + 0.8921x_1^2 + 5.1606x_1x_2 + 15.2446x_2^2 + 0.5334x_1^3 - 2.9790x_1^2x_2 - 0.8033x_1x_2^2 - 9.4827x_2^3$ and shown in Figure 1. Now we include the nonnegativity constraint by solving semidefinite optimization

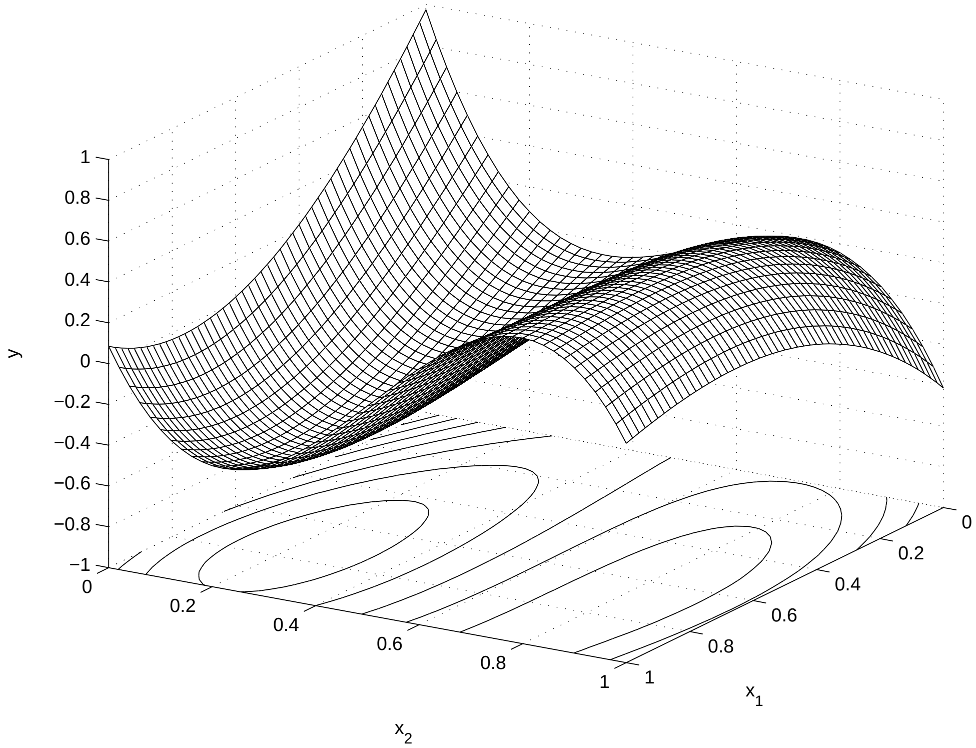

Figure 1: Optimal polynomial of Example 2.1 *without* nonnegativity-constraint.

problem (8), i.e., we take $R = \sqrt{2}$, $g_1(x_1, x_2) = 1 - x_1$, $g_2(x_1, x_2) = 1 - x_2$, $g_3(x_1, x_2) = x_1$, $g_4(x_1, x_2) = x_2$, $e_\ell^T(x_1, x_2) = \begin{bmatrix} 1 & x_1 & x_2 \end{bmatrix}$, for $\ell = 0, \ldots, 4$, and $e_5(x_1, x_2) = 1$. To solve the semidefinite optimization problem, we use SeDuMi; see Sturm (1999). This gives $\mathcal{E} = 0.108$. The corresponding optimal polynomial is $p(x_1, x_2) = 0.8917 - 2.5084x_1 - 3.6072x_2 + 3.2103x_1^2 + 4.2274x_1x_2 + 5.4395x_2^2 - 1.4647x_1^3 - 1.9329x_1^2x_2 - 1.5377x_1x_2^2 - 2.7181x_2^3$, as shown in Figure 2. Note that the polynomial has real roots as expected.

2.3 Approximation by increasing polynomials

We can easily extend the methodology developed in Section 2.2 to increasing polynomials by introducing nonnegativity constraints for the (partial) derivatives.

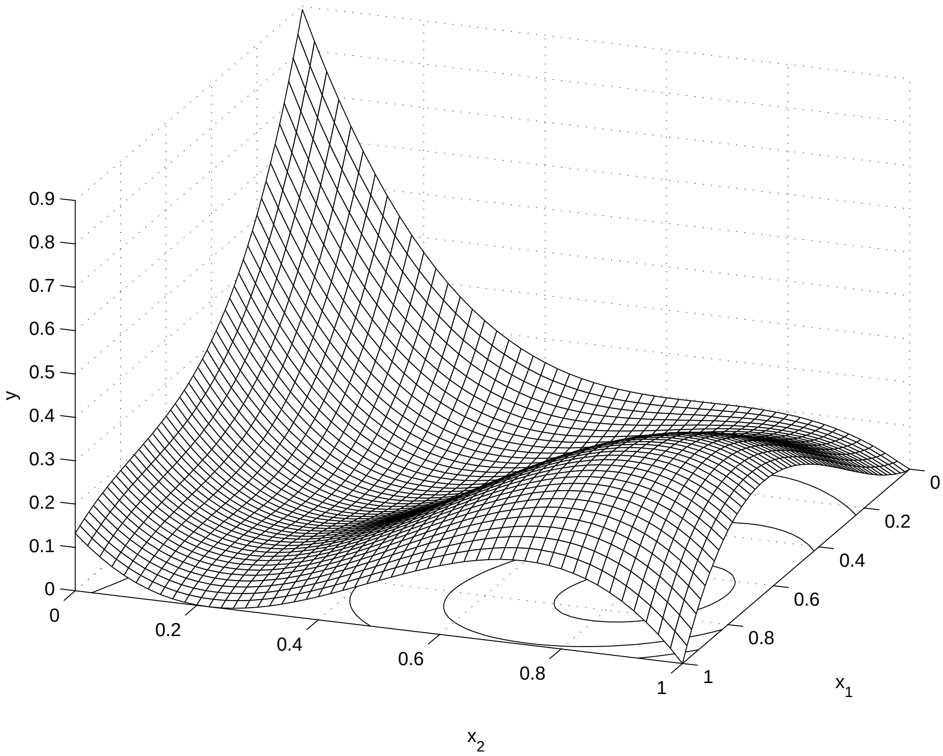

Figure 2: Nonnegative polynomial of Example 2.1 *with* nonnegativity-constraint.

no.	x_1	x_2	y
1	0	0	1
2	0.5	0.5	0
3	1	0	0.11
4	0	0.9	0
5	1	1	0.1
6	0	0.5	0
7	0.4	0	0.2
8	1	0.5	0
9	0.6	1	0.15
10	0.25	1	0
11	0.478	0.654	0.3

Table 1: Dataset of Example 2.1.

no.	Temperature(Kelvin)	Coefficient of Thermal Expansion
1	24.41	0.591
2	54.98	4.703
3	70.53	7.03
4	127.08	12.478
5	271.97	16.549
6	429.66	17.848
7	625.55	19.111

Table 2: Dataset of Example 2.2 (thermal expansion of copper).

Suppose we know that the function $y(x)$ is increasing on a certain region $\mathcal{U}$ and with respect to coordinates x_i with $i \in I \subseteq \{1, \ldots, n\}$. Then, instead of (4), we need to solve the following mathematical program:

$$\begin{array}{lll} \min\limits_{\alpha,t} & t & \\ \text{s.t.} & \frac{\partial p(x)}{\partial x_i} \geq 0 & \forall i \in I,\ \forall x \in \mathcal{U} \\ & \sum\limits_{j=1}^{m} \alpha_j p_j(x^i) - t \leq y(x^i) & \forall i = 1, \ldots, n \\ & -\sum\limits_{j=1}^{m} \alpha_j p_j(x^i) - t \leq -y(x^i) & \forall i = 1, \ldots, n. \end{array} \tag{9}$$

Since a partial derivative of a polynomial is also a polynomial, we can use similar techniques as in Section 2.2 to solve optimization problem (9).

Example 2.2

In this example we consider data of the coefficient of thermal expansion of copper. This data is taken from Croarkin and Tobias (2005). The coefficient of thermal expansion of copper is an increasing function of the temperature of copper. In this example we only use a selection of the data, which is given in Table 2. A scatterplot of this selection of the data is given in Figure 3. First, we apply Chebyshev approximation with a polynomial of degree $d = 5$ without requiring the approximation to be increasing. We get $\mathcal{E} = 0.1486$ in (4), and obtain the polynomial $p(x) = -3.3051 + 0.1545x + 0.2490 \cdot 10^{-4}x^2 - 0.2920 \cdot 10^{-5}x^3 + 0.8014 \cdot 10^{-8}x^4 - 0.6227 \cdot 10^{-11}x^5$. This is the solid line in Figure 3. Note that the approximation is not increasing everywhere. We observe an oscillating behavior that is one of the well-known drawbacks of using polynomials for approximations. A method that reduces oscillating behaviour is ridge regression; see e.g. Montgomery and Peck (1992). Ridge regression, however, cannot guarantee monotonicity. If we use our method, i.e., if we require the approximation to be increasing, we get $\mathcal{E} = 0.2847$. We obtain the polynomial $p(x) = -4.2922 + 0.2054x - 0.7234 \cdot 10^{-3}x^2 + 0.1063 \cdot 10^{-5}x^3 -$

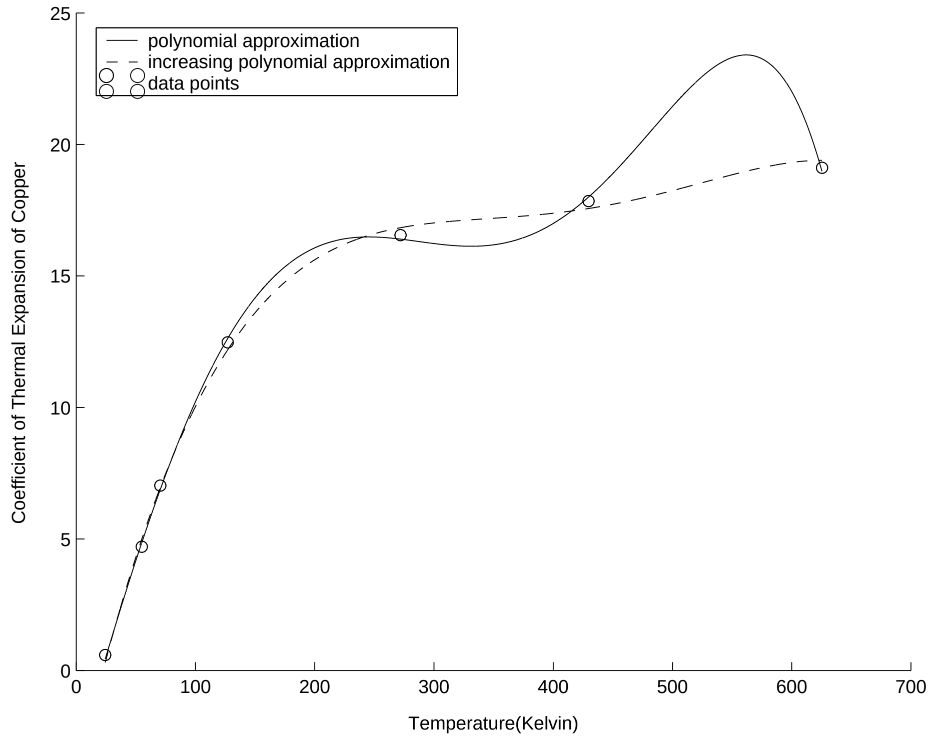

Figure 3: Example of increasing and non-increasing polynomial approximation.

$0.4369 \cdot 10^{-9}x^4 - 0.1578 \cdot 10^{-12}x^5$, which is shown by the dashed line in Figure 3. Indeed, the approximation is increasing in the input variable.

3 Approximation by trigonometric polynomials

We are interested in approximating a function $y : \mathbb{R} \mapsto \mathbb{R}$ by a univariate nonnegative trigonometric polynomial, given input data $x^1, \ldots, x^n \in \mathbb{R}$ and corresponding output data $y^1, \ldots, y^n \in \mathbb{R}$. We can again write a nonnegative trigonometric polynomial as a sum of squares, in a similar way as done with polynomials.

A trigonometric polynomial of degree d has the form

$$p(x) = \alpha_0 + \sum_{k=1}^{d} \left(\alpha_k \sin(kx) + \beta_k \cos(kx)\right), \tag{10}$$

where α_0, α_k and β_k are the coefficients.

3.1 General least norm approximation by trigonometric polynomials

Approximation by trigonometric polynomials is similar to approximation by ordinary polynomials. We again define $p_{\alpha,\beta} = \left[p(x^1), \ldots, p(x^n)\right]^T$ and $y = \left[y(x^1), \ldots, y(x^n)\right]^T$, and are interested in finding α and β that solve

$$\min_{\alpha,\beta} \quad \|p_{\alpha,\beta} - y\|,$$

where $\|\cdot\|$ is some norm. In Fassbender (1997) efficient numerical methods for least-squares approximation by trigonometric polynomials are developed. For the ℓ_∞-norm we obtain the following linear program:

$$\begin{array}{lll}
\min\limits_{t,\alpha,\beta} & t & \\
\text{s.t.} & \alpha_0 + \sum\limits_{k=1}^{d} (\alpha_k \sin(kx^i) + \beta_k \cos(kx^i)) - t \leq y(x^i) & \forall i = 1, \ldots, n \\
& -\alpha_0 - \sum\limits_{k=1}^{d} (\alpha_k \sin(kx^i) + \beta_k \cos(kx^i)) - t \leq -y(x^i) & \forall i = 1, \ldots, n.
\end{array}$$

We can easily adapt the methods that we will present, to the cases of the ℓ_1-norm and the ℓ_2-norm.

3.2 Approximation by nonnegative trigonometric polynomials

The following theorem states that nonnegative univariate trigonometric polynomials can be expressed in terms of a positive definite matrix.

Theorem 5. *If $p(x)$ is a nonnegative trigonometric polynomial of degree d, then there exists a decomposition $p(x) = e^T(x)Qe(x)$, where $Q \succeq 0$. If $d = 2k+1$ is odd, then*

$$e(x) = \left[\cos\left(\frac{x}{2}\right), \sin\left(\frac{x}{2}\right), \ldots, \cos\left(kx + \frac{x}{2}\right), \sin\left(kx + \frac{x}{2}\right)\right]^T,$$

otherwise $d = 2k$, and

$$e(x) = [1, \cos(x), \sin(x), \ldots, \cos(kx), \sin(kx)]^T.$$

Proof. A sketch of a proof is given in Lofberg and Parrilo (2004). □

We can use this theorem to construct nonnegative trigonometric polynomial approximations by solving the SDP

$$\begin{array}{rll}
\mathcal{E} := \min\limits_{t,Q} & t & \\
\text{s.t.} & e^T(x^i)Qe(x^i) - y(x^i) \leq t & \forall i = 1, \ldots, n \\
& -e^T(x^i)Qe(x^i) + y(x^i) \leq t & \forall i = 1, \ldots, n \\
& Q \succeq 0. &
\end{array} \tag{11}$$

Note that (10) is a periodic function with period 2π. However, the data is in general non-

no.	Time(min)	Concentration(%)
1	5	0.0
2	7	0.0
3	10	0.7
4	15	7.2
5	20	11.5
6	25	15.8
7	30	20.9
8	40	26.6

Table 3: Oil shale dataset (Example 3.1).

periodic. Nevertheless, we can still approximate a non-periodic function on a compact interval by a trigonometric function by scaling and translating the data to $[0, \pi]$.

Example 3.1

We consider data on the pyrolysis of oil shale, taken from Bates and Watts (1988). This data, obtained by Hubbard and Robinson (1950) represents the relative concentration of oil versus time during pyrolysis of oil shale. We used a selection of the data as given in Table 3. This data concerns the relative concentration of oil versus time at a temperature of $673K$. A scatterplot of the data is given in Figure 4. Obviously the concentration of oil is nonnegative. However, if we approximate the concentration as a function of time by a trigonometric polynomial of degree 2, we get $\mathcal{E} = 0.7348$ in (11), and we obtain the trigonometric polynomial

$$\begin{aligned} p(x) &= 12.6303 + 12.1492\sin(-1.7054 + 0.0898x) - 8.0262\cos(-1.7054 + 0.0898x) + \\ &\quad 6.3258\cos^2(-1.7054 + 0.0898x) - \\ &\quad 0.2234\sin(-1.7054 + 0.0898x)\cos(-1.7054 + 0.0898x). \end{aligned}$$

This trigonometric polynomial is plotted in Figure 4 with a solid line. This trigonometric polynomial takes negative values. However, if we use the new methodology to obtain a nonnegative trigonometric polynomial, we obtain the trigonometric polynomial

$$\begin{aligned} p(x) &= 7.0570 - 9.6844\cos(-1.7054 + 0.0898x) + 11.2141\sin(-1.7054 + 0.0898x) + \\ &\quad 13.5710\cos^2(-1.7054 + 0.0898x) + \\ &\quad 1.0457\sin(-1.7054 + 0.0898x)\cos(-1.7054 + 0.0898x) + \\ &\quad 6.3186\sin^2(-1.7054 + 0.0898x), \end{aligned}$$

which is represented by the dashed line in Figure 4. In this case, $\mathcal{E} = 0.8187$.

We cannot extend this methodology to construct increasing trigonometric polynomial ap-

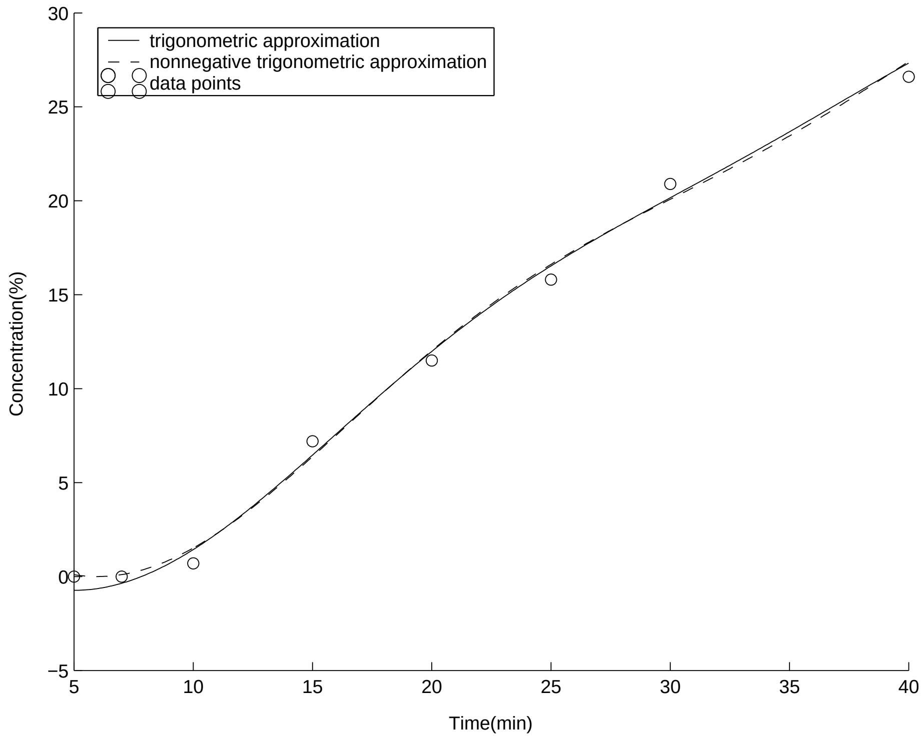

Figure 4: Example of nonnegative and general trigonometric approximation.

proximations in a similar way as done for polynomials, because trigonometric polynomials are periodic functions.

4 Approximation by rational functions

Given input data $x^1, \ldots, x^n \in \mathbb{R}^q$ and corresponding output data $y^1, \ldots, y^n \in \mathbb{R}$, we are interested in approximating a function $y : \mathbb{R}^q \mapsto \mathbb{R}$. In this section we consider approximation by rational functions. We first show how to approximate a function $y(x)$ by a rational function, without preserving characteristics. A rational function is a quotient of two polynomials $p(x) = \sum_{j=1}^{m} \alpha_j p_j(x)$ and $q(x) = \sum_{k=1}^{\hat{m}} \beta_k q_k(x)$; i.e. $r(x) = \frac{\sum_{j=0}^{m} \alpha_j p_j(x)}{\sum_{k=0}^{\hat{m}} \beta_k q_k(x)}$. Here m and $\hat{m}$ are the number of monomials of the polynomials $p(x)$ and $q(x)$ respectively.

4.1 General least norm approximation by rational functions

Analogous to p_α, we define $r_{\alpha,\beta} = [r(x^1), \ldots, r(x^n)]^T$. We are interested in solving

$$\min_{\alpha,\beta} \quad \|r_{\alpha,\beta} - y\|,$$

where $\|\cdot\|$ is some norm. In the following, we will discuss the methodology for the ℓ_∞-norm, as done in Powell (1981), Chapter 10, and then extend this with a method to prevent the denominator from being zero. A similar methodology can be used for the ℓ_1-norm and the ℓ_2-norm.

For the ℓ_∞-norm, we obtain the following optimization problem by multiplying each term by the denominator of $r(x)$:

$$\begin{aligned}
\min_{t,\alpha,\beta} \quad & t \\
\text{s.t.} \quad & \sum_{j=0}^{m} \alpha_j p_j(x^i) - y(x^i) \sum_{k=0}^{\hat{m}} \beta_k q_k(x^i) \leq t \sum_{k=0}^{\hat{m}} \beta_k q_k(x^i) \qquad i = 1, \ldots, n \\
& -\sum_{j=0}^{m} \alpha_j p_j(x^i) + y(x^i) \sum_{k=0}^{\hat{m}} \beta_k q_k(x^i) \leq t \sum_{k=0}^{\hat{m}} \beta_k q_k(x^i) \quad i = 1, \ldots, n.
\end{aligned} \tag{12}$$

Note that (12) is a nonlinear optimization problem. However, we can solve this problem efficiently by using binary search. We choose an upper bound for t, say $\bar{t}$, and a lower bound $\underline{t} = 0$, and consider the interval $[\underline{t}, \bar{t}]$. Then we define $\hat{t} = \frac{\bar{t}+\underline{t}}{2}$, and check whether the constraints in (12) are met for this value of t; i.e. we check whether there exist α and β, for which

$$\left\{ \begin{aligned}
& \sum_{j=0}^{m} \alpha_j p_j(x^i) - y(x^i) \sum_{k=0}^{\hat{m}} \beta_k q_k(x^i) \leq \hat{t} \sum_{k=0}^{\hat{m}} \beta_k q_k(x^i) \qquad i = 1, \ldots, n \\
& -\sum_{j=0}^{m} \alpha_j p_j(x^i) + y(x^i) \sum_{k=0}^{\hat{m}} \beta_k q_k(x^i) \leq \hat{t} \sum_{k=0}^{\hat{m}} \beta_k q_k(x^i) \quad i = 1, \ldots, n.
\end{aligned} \right. \tag{13}$$

This is a linear feasibility problem. If the answer is 'yes', then our new interval becomes $[\underline{t}, \frac{\bar{t}+\underline{t}}{2}]$, and otherwise our new interval becomes $[\frac{\bar{t}+\underline{t}}{2}, \bar{t}]$. We repeat this until the interval is sufficiently small.

Instead of just checking the constraints (13), we can also introduce a new variable ε and solve the linear program

$$\begin{aligned}
\min_{\varepsilon,\alpha,\beta} \quad & \varepsilon \\
\text{s.t.} \quad & \sum_{j=0}^{m} \alpha_j p_j(x^i) - y(x^i) \sum_{k=0}^{\hat{m}} \beta_k q_k(x^i) \leq \hat{t} \sum_{k=0}^{\hat{m}} \beta_k q_k(x^i) + \varepsilon \qquad i = 1, \ldots, n \\
& -\sum_{j=0}^{m} \alpha_j p_j(x^i) + y(x^i) \sum_{k=0}^{\hat{m}} \beta_k q_k(x^i) \leq \hat{t} \sum_{k=0}^{\hat{m}} \beta_k q_k(x^i) + \varepsilon \quad i = 1, \ldots, n \\
& \sum_{k=0}^{\hat{m}} \beta_k q_k(\zeta) = 1,
\end{aligned} \tag{14}$$

where $\zeta \in \mathbb{R}^q$ is a constant. Let ε_{opt} be the optimal ε in (14). The last constraint is added to prevent the optimization problem from being unbounded if $\varepsilon_{\text{opt}} < 0$. A common choice is $\zeta = 0$. Now we can distinguish three cases. If $\varepsilon_{\text{opt}} < 0$, then $\hat{t}$ is greater than the least maximum error, and we can tighten the bounds of our interval to $[\underline{t}, \hat{t}]$. In fact, by using the value of

ε_{opt}, we can even tighten the interval to $\left[\underline{t}, \hat{t} - \frac{\varepsilon_{\text{opt}}}{\max_{i=1,\ldots,n}\{\sum_{k=0}^{\hat{m}} \beta_k^{\text{opt}} q_k(x^i)\}}\right]$, where β_k^{opt} are the optimal β_k in optimization problem (14). If $\varepsilon_{\text{opt}} = 0$, then the corresponding $\frac{p(x)}{q(x)}$ is the optimal approximation, and finally if $\varepsilon_{\text{opt}} > 0$, then our upper bound $\hat{t}$ is too small, and we can tighten our interval to $[\hat{t}, \bar{t}]$.

Note that $q(x) = \sum_{k=0}^{\hat{m}} \beta_k q_k(x)$ possibly becomes zero, which is not desirable if we want to avoid poles. We can easily prevent $q(x)$ from becoming zero on a predefined compact set $\mathcal{U} = \{x \in \mathbb{R}^q | g_\ell(x) \geq 0, \forall \ell = 1, \ldots, \bar{m}\}$, where g_ℓ are polynomials, by again using Theorem 2 and Theorem 4. Then, we obtain the following semidefinite optimization problem:

$$
\begin{array}{lll}
\min\limits_{\varepsilon,\alpha,P_0^d,\ldots,P_n^d} & \varepsilon & \\
\text{s.t.} & \displaystyle\sum_{j=0}^{m} \alpha_j p_j(x^i) - \big(y(x^i) - \hat{t}\big)\left(\sum_{\ell=0}^{\bar{m}+1} e_\ell^T(x^i) P_\ell^d e_\ell(x^i) g_\ell(x^i) + \delta\right) \leq \varepsilon & i = 1, \ldots, n \\
& \displaystyle -\sum_{j=0}^{m} \alpha_j p_j(x^i) + \big(y(x^i) - \hat{t}\big)\left(\sum_{\ell=0}^{\bar{m}+1} e_\ell^T(x^i) P_\ell^d e_\ell(x^i) g_\ell(x^i) + \delta\right) \leq \varepsilon & i = 1, \ldots, n \\
& P_\ell^d \succeq 0 & \ell = 0 \ldots, n \\
& \displaystyle\sum_{\ell=0}^{\bar{m}+1} e_\ell^T(\zeta) P_\ell^d e_\ell(\zeta) g_\ell(\zeta) = 1, &
\end{array}
$$

where $\delta > 0$ is a small number, which prevents the denominator $q(x)$ from becoming too small.

4.2 Approximation by nonnegative rational functions

To construct nonnegative rational approximations, we need a characterization of nonnegative rational functions. The following theorem gives a characterization of nonnegative rational functions on open connected sets or the (partial) closure of such a set. Note that two polynomials $p(x)$ and $q(x)$ are called relatively prime, if they have no common factors.

Theorem 6. *Let $p(x)$ and $q(x)$ be relatively prime polynomials on $\mathbb{R}^q$ and let $U \subseteq \mathbb{R}^q$ be an open connected set or the (partial) closure of such a set. Then the following two statements are equivalent:*

1. *$p(x)/q(x) \geq 0 \ \forall \ x \in U$ such that $q(x) \neq 0$;*

2. *$p(x)$ and $q(x)$ are both nonnegative, or both nonpositive, on U;*

Proof. See Jibetean and De Klerk (2003). □

Therefore, to enforce a rational approximation to be nonnegative on a set U that meets the conditions of Theorem 6, without loss of generality, we may assume that both the numerator $p(x)$ and the denominator $q(x)$ are nonnegative. Note that requiring $q(x)$ to be positive also prevents the rational function from having poles.

Using this characterization, the optimization problem becomes as follows:

$$
\begin{array}{lll}
\min\limits_{\varepsilon,\alpha,\beta} & \varepsilon & \\
\text{s.t.} & \sum\limits_{j=0}^{m} \alpha_j p_j(x^i) - y(x^i) \sum\limits_{k=0}^{\hat{m}} \beta_k q_k(x^i) \leq \hat{t} \sum\limits_{k=0}^{\hat{m}} \beta_k q_k(x^i) + \varepsilon & i = 1, \ldots, n \\
& -\sum\limits_{j=0}^{m} \alpha_j p_j(x^i) + y(x^i) \sum\limits_{k=0}^{\hat{m}} \beta_k q_k(x^i) \leq \hat{t} \sum\limits_{k=0}^{\hat{m}} \beta_k q_k(x^i) + \varepsilon & i = 1, \ldots, n \\
& \sum\limits_{j=0}^{m} \alpha_j p_j(x) \geq 0 & \forall x \in \mathcal{U} \\
& \sum\limits_{k=0}^{\hat{m}} \beta_k q_k(x) \geq \delta & \forall x \in \mathcal{U} \\
& \sum\limits_{k=0}^{\hat{m}} \beta_k q_k(\zeta) = 1, &
\end{array}
\tag{15}
$$

where $\delta > 0$ is a small number and $\zeta \in \mathbb{R}^q$ is a constant.

Now, we use Theorem 2 in combination with Theorem 4, to model optimization problem (15) as a semidefinite programming problem:

$$
\begin{array}{ll}
\min\limits_{\varepsilon,P_\ell^n,P_\ell^d} & \varepsilon \\
\text{s.t.} & \sum\limits_{\ell=0}^{\bar{m}+1} e_\ell^T(x^i) P_\ell^n e_\ell(x^i) g_\ell(x^i) - (y(x^i) + \hat{t}) \left(\sum\limits_{\ell=0}^{\bar{m}+1} e_\ell^T(x^i) P_\ell^d e_\ell(x^i) g_\ell(x^i) + \delta \right) \leq \varepsilon \\
& \hfill i = 1, \ldots, n \\
& -\sum\limits_{\ell=0}^{\bar{m}+1} e_\ell^T(x^i) P_\ell^n e_\ell(x^i) g_\ell(x^i) + (y(x^i) - \hat{t}) \left(\sum\limits_{\ell=0}^{\bar{m}+1} e_\ell^T(x^i) P_\ell^d e_\ell(x^i) g_\ell(x^i) + \delta \right) \leq \varepsilon \\
& \hfill i = 1, \ldots, n \\
& P_\ell^n \succeq 0 \hfill \ell = 0, \ldots, \bar{m}+1 \\
& P_\ell^d \succeq 0 \hfill \ell = 0, \ldots, \bar{m}+1 \\
& \sum\limits_{\ell=0}^{\bar{m}+1} e_\ell^T(\zeta) P_\ell^d e_\ell(\zeta) g_\ell(\zeta) = 1.
\end{array}
\tag{16}
$$

In the multivariate case (16) is just an approximation of (15), since we do not know the degree of the monomials of $e_\ell(x)$. However, in the univariate case (16) is an exact reformulation of (15), since in the univariate case Theorem 3 specifies the degree d of the polynomials $f(x)$ and $g(x)$, we know the degree of the monomials of $e_\ell(x)$ in (16).

Example 4.1

In this example we use the same data on the pyrolysis of oil shale as used in Example 3.1. Note again that the concentration of oil should be nonnegative. However, if we approximate the concentration as a function of time by a rational function by quadratic numerator and quadratic denominator, we get $\mathcal{E} = 0.5962$, and obtain the rational function

$$r(x) = -78400 \frac{0.1162622153 \cdot 10^{19} - 0.3521415490 \cdot 10^{18} x + 0.2544537885 \cdot 10^{17} x^2}{-0.3691739529 \cdot 10^{21} - 0.767285004 \cdot 10^{21} x - 0.319303558 \cdot 10^{20} x^2},$$

which is plotted in Figure 5. Obviously, the rational function is not nonnegative. However, if

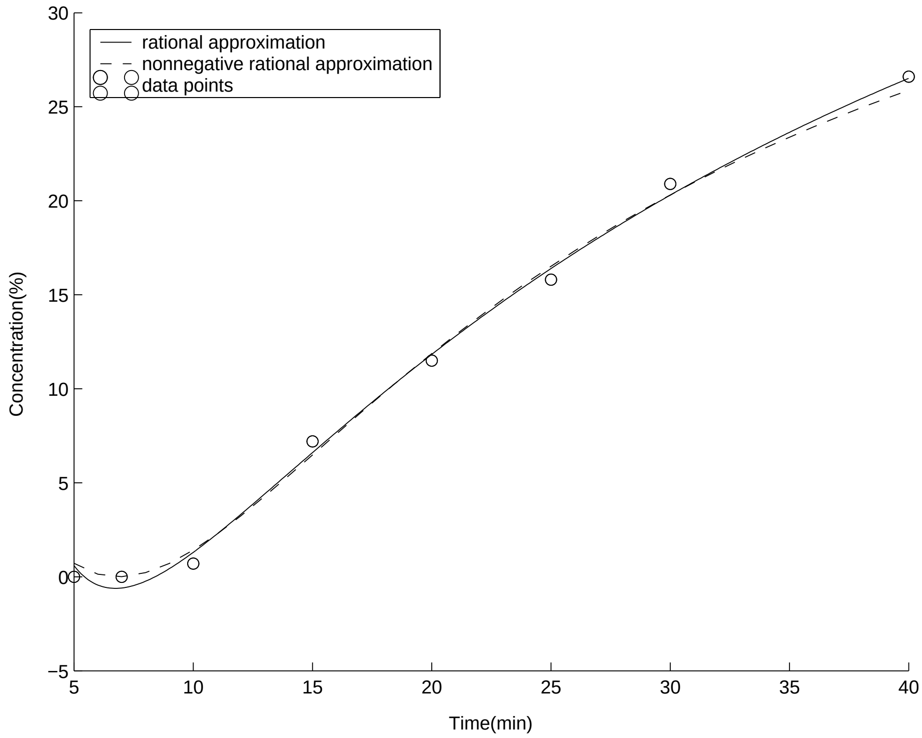

Figure 5: Example of nonnegative and general rational approximation.

we force the rational function to be nonnegative, we obtain the function

$$r(x) = 0.4883 \cdot 10^{-3} \frac{-0.1015375413 \cdot 10^{27} + 0.2962932361 \cdot 10^{26} x - 0.2161508979 \cdot 10^{25} x^2}{-.3859790305 \cdot 10^{22} - 0.126652437 \cdot 10^{21} x - .224221696 \cdot 10^{20} x^2},$$

which is represented by the dashed line in Figure 5. Now, $\mathcal{E} = 0.7178$. The increase in $\mathcal{E}$ is only due to forcing the nonnegativity, since this is a univariate example.

We cannot easily extend the methodology for least-norm approximation by increasing rational functions, because the coefficients of polynomials in numerator and denominator of the derivative of a rational function $\frac{p(x)}{q(x)}$ are not linear in the coefficients of $p(x)$ and $q(x)$ anymore.

5 Exploiting structure during computation

Semidefinite programming (SDP) solvers usually require the problem to be cast in the form:

$$\min_{X\succeq 0, x\geq 0} \{\mathrm{trace}(CX) + c^T x \mid \mathrm{trace}(A_i X) + a_i^T x = b_i \quad (i = 1, \ldots, m)\},$$

where $C, A_1, \ldots, A_m$ are data matrices and $b, c, a_1, \ldots, a_m$ are data vectors.

The approximation problems we have considered may all be formulated as (SDP) problems in this form, and with the special property that the matrices A_i are rank one matrices. For example, in problem (7), we have $A_i = e(x^i)e(x^i)^T$ — a rank one matrix.

This structure can be exploited by interior point algorithms to speed up the computation. In particular, the solver DSDP (see Benson et al. (2000)) has been designed to do this.

Thus it is possible to solve problem (7) within minutes for up to a thousand data points and with a approximating polynomial of degree up to a thousand. For the other univariate approximation problems we have considered, we can solve instances of similar sizes in the order of a few minutes.

For the multivariate approximation problems, e.g. (8), the size of the monomial vector $e_\ell(x^i)$ is given by $\binom{q+d_\ell-1}{d_\ell}$, where $2d_\ell$ is the degree of the function p_ℓ (see Section 2.2.2) and q is the dimension (number of variables).

If q and the d_ℓ values are such that $\binom{q+d_\ell-1}{d_\ell}$ is at most a hundred, and the number of data points at most $n = 100$, then efficient computation is still possible.

6 Conclusions and further research

We have presented a least-norm approximation method to approximate functions by nonnegative and increasing polynomials, nonnegative trigonometric polynomials, and nonnegative rational functions. This methodology uses semidefinite programming and results from the field of real algebraic geometry. We have given several artificial and real-life examples, which demonstrate that our methodology indeed results in nonnegative or increasing approximations. We also studied how to exploit the structure of the problem to make the problem computationally easier. As a result of this we can deal with relatively large problems.

For further research we are interested in studying least-norm approximation by polynomials to approximate convex functions. In the univariate case, we can easily use the same methodology as presented in this paper, since a polynomial is convex if and only if its second derivative is nonnegative. In the multivariate quadratic case, the problem of approximating a function by a convex quadratic polynomial is already studied by Den Hertog, De Klerk, and Roos (2002).

References

Barlow, R., R. Bartholomew, J. Bremner, and H. Brunk (1972). *Statistical inference under order restrictions.* Chichester: Wiley.

Bates, D. and D. Watts (1988). *Nonlinear regression analysis and its applications.* New York: Wiley.

Benson, S., Y. Ye, and X. Zhang (2000). Solving large-scale sparse semidefinite programs for combinatorial optimization. *SIAM Journal on Optimization*, **10**, 443–461.

Croarkin, C. and P. Tobias (Eds.) (2005). *NIST (National Institute of Standards and Technology)/SEMATECH e-Handbook of Statistical Methods.* http://www.itl.nist.gov/div898/handbook/: World Wide Web.

Cuyt, A. and R. Lenin (2002). Computing packet loss probabilities in multiplexer models using adaptive rational interpolation with optimal pole placement. Technical report, University of Antwerp.

Cuyt, A., R. Lenin, S. Becuwe, and B. Verdonk (2004). Adaptive multivariate rational data fitting with applications in electromagnetics. Technical report, University of Antwerp.

Fassbender, H. (1997). On numerical methods for discrete least-squares approximation by trigonometric polynomials. *Mathematics of Computation*, **66**(218), 719–741.

Forsberg, J. and L. Nilsson (2005). On polynomial response surfaces and Kriging for use in structural optimization of crashworthiness. *Structural and Multidisciplinary Optimization*, **29**, 232–243.

Hertog, D. den, E. de Klerk, and K. Roos (2002). On convex quadratic approximation. *Statistica Neerlandica*, **563**, 376–385.

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

Hubbard, A. and W. Robinson (1950). A thermal decomposition study of Colorado oil shale. Rept. Invest. 4744, U.S. Bureau of Mines.

Jansson, T., L. Nilsson, and M. Redhe (2003). Using surrogate models and response surfaces in structural optimization – with application to crashworthiness design and sheet metal forming. *Structural and Multidisciplinary Optimization*, **25**, 129–140.

Jibetean, D. and E. de Klerk (2003). Global optimization of rational functions: a semidefinite programming approach. *Mathematical Programming A*. To appear.

Kuijt, F. (1998). *Convexity preserving interpolation – Stationary nonlinear subdivision and splines.* Ph. D. thesis, University of Twente, Enschede, The Netherlands.

Kuijt, F. and R. van Damme (2001). A linear approach to shape preserving approximation. *Advances in Computational Mathematics*, **14**, 25–48.

Lofberg, J. and P. Parrilo (2004). From coefficients to samples: a new approach to SOS optimization. Working paper.

Montgomery, D. and E. Peck (1992). *Introduction to linear regression analysis.* New-York: Wiley.

Nesterov, Y. (2000). Squared functional systems and optimization problems. In: H. Frenk, K. Roos, and T. Terlaky (Eds.), *High Performance Optimization*, Volume 13, Chapter 17, pp. 405–440. Dordrecht, The Netherlands: Kluwer.

Powell, M. (1981). *Approximation theory and methods.* Cambridge: Cambridge University Press.

Powers, V. and B. Reznick (2000). Polynomials that are positive on an interval. *Transactions of the American Mathematical Society*, **352**(10), 4677–4692.

Putinar, M. (1993). Positive polynomials on compact semi-algebraic sets. *Indiana University Mathematics Journal*, **42**, 969–0984.

Reznick, B. (2000). Some concrete aspects of Hilbert's 17th problem. *Contemporary Mathematics*, **253**, 251–272.

Robertson, T., F. Wright, and R. Dykstra (1988). *Order restricted statistical inference.* Chichester: John Wiley & Sons.

Schweighofer, M. (2004). Optimization of polynomials on compact semialgebraic sets. *SIAM Journal on Optimization*. To appear.

Sturm, J. (1999). Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optimization Methods and Software*, **11–12**, 625–653.

Watson, G. (1980). *Approximation theory and numerical methods.* Chichester: Wiley.

Yeun, Y., Y. Yang, W. Ruy, and B. Kim (2005). Polynomial genetic programming for response surface modeling Part 1: a methodology. *Structural and Multidisciplinary Optimization*, **29**, 19–34.