RUTCOR
RESEARCH
REPORT

Linear and Second Order Cone Programming Approaches to Statistical Estimation Problems

Dávid Papp[a] Farid Alizadeh[b]

RRR 13-2008, September 22, 2008

RUTCOR
Rutgers Center for
Operations Research
Rutgers University
640 Bartholomew Road
Piscataway, New Jersey
08854-8003
Telephone: 732-445-3804
Telefax: 732-445-5472
Email: rrr@rutcor.rutgers.edu
http://rutcor.rutgers.edu/∼rrr

[a]RUTCOR, Rutgers University, 640 Bartholomew Road, Piscataway, NJ 08854-8003; dpapp@rutcor.rutgers.edu
[b]Business School and RUTCOR, Rutgers University, 640 Bartholomew Road, Piscataway, NJ 08854-8003; alizadeh@rutcor.rutgers.edu

Rutcor Research Report
RRR 13-2008, September 22, 2008

Linear and Second Order Cone Programming Approaches to Statistical Estimation Problems

Dávid Papp Farid Alizadeh

Abstract. Estimation problems in statistics can often be formulated as nonlinear optimization problems where the approximating function is also constrained to satisfy certain shape constraints. For example, it may be required that the approximating function be nonnegative, nondecreasing or nonincreasing in some variables, unimodal in a variable, convex, or concave. We show how these shape constraints can be included in models where the approximating function is a polynomial spline of a given degree. Four classes of problems are considered: univariate nonparametric regression with nonnegativity constraints; isotonic, convex, and concave nonparametric regression; density estimation; and univariate arrival rate approximation. It is shown that all of these problems can be modeled by convex optimization models with linear and second-order conic constraints, which can be handled very efficiently both in theory and in practice. It is shown that in certain cases a linearly constrained model is enough, while in other cases nonlinear constraints are unavoidable. Extensive numerical computations show that these models often outperform or match the quality of results obtained using other popular nonparametric statistical methods, such as state-of-the-art kernel methods.

1 Introduction

Polynomials and polynomial splines are important tools in every field that deals with approximation of unknown functions. In particular in statistics and stochastic simulation, prime examples include statistical learning theory, regression of unknown functions and density estimation. In many applications it is required that the approximating function satisfy additional *shape constraints*. For example, it may be required that the approximating function be nonnegative, nondecreasing or nonincreasing in some variables, unimodal in a variable, convex, or concave.

The study of nonnegative approximation and estimation in statistics and simulation has had a long history. Here we simply refer the reader to the dissertation of [18], the survey article of [7], and references in these works for further information. Methodologies used in shape constrained approximation are traditionally based on simple models where often the approximates are determined by closed form formulas. Such approaches are based on various moving averages and weighted versions of it (the kernel methods) which do not involve optimization at all.

It is also possible to formulate optimization based models for shape-constrained problems. The shape constraints by themselves often imply that our search space is a convex cone in a finite dimensional real linear space. The requirements of continuity and smoothness impose additional, usually linear constraints. When we add to our model the requirement of minimizing some loss function or maximizing some likelihood function, then we arrive at an optimization model. Some shape constrained problems are so simple, and their optimization models so special, that they can actually be solved by closed form formulas, see for example the thesis [18]. These types of models involve piecewise constant or linear functions in most cases.

On the other hand, if we wish to use more sophisticated models for regression, estimation, and learning theory based on functional systems such as polynomials, polynomial splines, trigonometric polynomials and splines, wavelets, or more general function spaces, then the requirement of nonnegativity will involve more complicated, though often still computationally tractable, optimization models. A survey of possible approaches is presented in the tutorial [1].

The problems studied in this paper involve shape constrained *polynomial splines*, and hence the resulting optimization models employ constraints that some polynomials with unknown coefficients are nonnegative over a closed interval.

These models have several advantages to approaches with a closed-form formula solution and to kernel methods. They are more flexible: a number of simple constraints, such as periodicity can be trivially added to or removed from the model. An additional advantage over kernel methods is that once the optimization is carried out, the resulting estimate is already in a simple, closed form, independent of the data, that can easily be processed further. Estimates obtained by kernel methods are usually given in a more complicated form, and the evaluation of the estimate at a single point may require computing a sum with as many terms as data points.

Several models have been proposed for problems involving nonnegative splines or polynomials over some interval $[a, b]$, see for example the survey [7], and the references therein. All of these can be fitted in the following framework, to which we will refer in this paper as the *nonnegative basis approach*: we identify a basis of splines or polynomials, $\{b_1, \dots, b_m\}$, which are themselves nonnegative over $[a, b]$, and then confine our search to linear combinations of these basis elements with nonnegative coefficients. Hence, the spline or polynomial we are searching for is represented as $\sum_{i=1}^{m} \lambda_i b_i$. The variables of our optimization model will be the unknown coefficients λ_i, and our shape constraint simplifies to $\lambda_i \geq 0$ for every i. The resulting model has only linear constraints.

Appropriate bases include the basis of B-splines of a fixed order in the case of splines, and the standard polynomial basis, or the Bernstein basis polynomials $B_{n,k} = \binom{n}{k} x^k (1-x)^{n-k}$ (see also Definition 4 below) in the case of polynomials of degree n. Clearly, a similar approach can be used in case of more general functional systems, such as trigonometric polynomials or wavelets, too, wherever we can identify a basis consisting of nonnegative functions.

All models that fit in the above framework are incomplete in some sense. While nonnegative linear combinations of nonnegative functions are nonnegative, the converse, of course, is not true. The set of polynomials of degree n which are nonnegative over the interval $[a, b]$ is a convex but not polyhedral cone for any $n \geq 2$, that is, it cannot be represented by linear equality and inequality constraints only. Hence models in the above framework are bound to miss some nonnegative polynomials, regardless of the choice of basis.

A more sophisticated method is to carry out the optimization over the set of *all* those nonnegative univariate polynomials of degree n (or splines of a given set of knot points, order, and continuity) that are nonnegative over $[a, b]$. It is well-known that this is set is a convex cone, and it can be represented by a combination of linear constraints and a semidefinite constraint on a matrix related to the coefficients of the polynomials, see for example, [16], [8], and [19]. This representation leads to *semidefinite programming (SDP)* or, in some cases, *second order cone programming (SOCP)* models, hence we refer to this approach as the *SDP/SOCP approach*. We review the fundamentals of this theory in Section 2.1.

In Section 2.2 we exhibit a few variants of shape constrained optimization problems when multiple shape constraints are present. In some of these problems the nonnegative basis approach is justified, as (with the right selection of basis) it is equivalent to this more sophisticated approach. We prove that *piecewise (scaled) Bernstein basis polynomial splines* is the right basis for many such problems, and that it is always a better choice than the widely used B-splines.

Nonnegative polynomials have been studied in some areas of applications, for example, in electrical engineering and control theory [15, 21, 23] among others. However, the use of nonnegative polynomials and nonnegative splines in approximation theory and statistical application seems to be recent.

Indeed, a special problem treated in [2] involves the estimation of the arrival rate $\lambda(t)$ of a non-homogeneous Poisson process. In this work, the model required a smooth rate function. As a result, the authors chose cubic splines to estimate $\lambda(t)$ from observed arrival times. The problem was that without any explicit requirement of nonnegativity, the maximum likelihood model was not well behaved, and either could not converge at all, or converged to functions which attained negative values in some intervals. Another recent paper [13] investigates parametric probability density estimation using nonnegative polynomials and semidefinite programming. Here the model requires that the estimated probability density function is nonnegative everywhere. The authors fit products of exponential functions and nonnegative polynomials of small degree (typically up to six or seven), and select the one that appears to fit best.

In this paper we consider four problems of estimation with nonnegativity constraints, using cubic and quartic splines:

1. Univariate nonparametric regression with nonnegativity constraints,

2. Isotonic, convex and concave nonparametric regression,

3. Density estimation,

4. Univariate arrival rate approximation using cubic splines.

The purpose of this paper is threefold. Firstly, we exhibit the appropriate SDP/SOCP optimization models for each of these problems, as well as the linearly constrained models resulting from the nonnegative basis approach, and we show that when the shape constraints include nonnegativity and concavity, then a linearly constrained model is in fact equivalent to the SDP/SOCP model, while in the most other cases it is not. Secondly, we demonstrate that the SDP/SOCP models are not only theoretically tractable, but are extremely efficient in practice, too. Finally, we show by extensive numerical computations that our shape constrained spline-based models often yield results of comparable quality to other popular nonparametric statistical methods, such as state-of-the-art kernel methods.

Section 2 summarizes the fundamental theory of the representation of nonnegative polynomials over a closed interval, and contains a new theorem on cases when the nonnegative basis approach is equivalent to the SDP approach. In Section 3 we present the statistical applications, recall the main ideas of popular kernel methods, and present our optimization models proposed for these problems. Finally, Section 4 contains the results of our numerical experiments in which all the different methods were compared.

2 Representations of Nonnegative Polynomials and Splines

2.1 Representations involving Semidefinite and Second Order Cone constraints

In this section we summarize some results about the semidefinite representation of nonnegative polynomials, and apply this theory to characterize nonnegative splines. For more details (and proofs in particular) the reader is referred to [19], [1], and [16].

We will use the following notations and conventions throughout the paper: vectors (and matrices) are typeset boldface, their components (rows and columns) are denoted with the corresponding lowercase italic character, and they are indexed starting from 0 rather than 1. That is, for example, the $n+1$ dimensional row vector $\mathbf{p}$ could also be written as $(p_0, \ldots, p_n)$.

The inequality $\mathbf{X} \succcurlyeq 0$ denotes that $\mathbf{X}$ is positive semidefinite.

The main results on the representation of nonnegative polynomials over an interval could be summarized in a single theorem. We split the odd and even degree cases into separate propositions for better readability.

Proposition 1 (Odd degree case, [1, 3, 16, 19, 22])**.** *Let $p = \sum_{i=0}^{n} p_i x^i$ be a polynomial of degree $n = 2k+1$, and $a < b$ be real numbers. Then the following are equivalent.*

1. *$p(x) \geq 0$ for all $x \in [a, b]$.*
2. *$p(x) = (x-a)q^2(x) + (b-x)r^2(x)$ for some polynomials q and r, both having degree at most k.*
3. *$p(x) = (x-a)\sum_{i=1}^{m_1} q_i^2(x) + (b-x)\sum_{j=1}^{m_2} r_j^2(x)$ for some polynomials $q_1, \ldots, q_{m_1}$ and $r_1, \ldots, r_{m_2}$, each having degree at most k.*
4. *There exist symmetric $(k+1) \times (k+1)$ matrices $\mathbf{X} = (x_{ij})_{i,j=0}^{k}$ and $\mathbf{Y} = (y_{ij})_{i,j=0}^{k}$ satisfying*

$$\begin{aligned} \mathbf{X} &\succcurlyeq 0, && \text{(1a)} \\ \mathbf{Y} &\succcurlyeq 0, && \text{(1b)} \\ p_m &= \sum_{i+j=m} (-a x_{ij} + b y_{ij}) + \sum_{i+j=m-1} (x_{ij} - y_{ij}) && \text{(1c)} \end{aligned}$$

for all $m = 0, \ldots, 2k+1$.

Proposition 2 (Even degree case, [1, 3, 16, 19, 22])**.** *Let $p = \sum_{i=0}^{n} p_i x^i$ be a polynomial of degree $n = 2k$, and $a < b$ be real numbers. Then the following are equivalent.*

1. *$p(x) \geq 0$ for all $x \in [a, b]$.*
2. *$p(x) = q^2(x) + (x-a)(b-x)r^2(x)$ for some polynomials q of degree at most k, and r of degree at most $k-1$.*
3. *$p(x) = \sum_{i=1}^{m_1} q_i^2(x) + (x-a)(b-x)\sum_{j=1}^{m_2} r_j^2(x)$ for some polynomials $q_1, \ldots, q_{m_1}$, each of degree at most k, and $r_1, \ldots, r_{m_2}$, each of degree at most $k-1$.*
4. *There exist a symmetric $(k+1) \times (k+1)$ matrix $\mathbf{X} = (x_{ij})_{i,j=0}^{k}$ and a symmetric $k \times k$ matrix $\mathbf{Y} = (y_{ij})_{i,j=0}^{k-1}$ satisfying*

$$\begin{aligned} \mathbf{X} &\succcurlyeq 0, && \text{(2a)} \\ \mathbf{Y} &\succcurlyeq 0, && \text{(2b)} \\ p_m &= \sum_{i+j=m} (x_{ij} - ab y_{ij}) + \sum_{i+j=m-1} (a+b) y_{ij} - \sum_{i+j=m-2} y_{ij} && \text{(2c)} \end{aligned}$$

for all $m = 0, \ldots, 2k$.

Definition 1. *A* univariate polynomial spline $S\colon [a_0, a_m] \to \mathbb{R}$ *of* degree n *with* knot points $a_0 < \cdots < a_m$ *and* continuity $\mathcal{C}^r$ *is a piecewise polynomial function defined by*

$$S(x) = p^{(i)}(x) = \sum_{k=0}^{n} p_k^{(i)}(x - a_i)^k \quad \forall x \in [a_i, a_{i+1}], \tag{3}$$

with continuous derivatives up to order r everywhere in (a_0, a_m), that is, at every internal knot point $a_1, \dots, a_{m-1}$.

We will slightly change this definition, and replace (3) by (5) below to obtain optimization models with better numerical properties.

Cubic splines with continuity $\mathcal{C}^2$ have been extremely popular in approximation theory and in statistics, as well as in engineering.

The above characterization of nonnegative polynomials easily extend to a characterization of nonnegative splines over $[a_0, a_m]$.

Definition 2. *A piecewise polynomial function is a nonnegative spline with continuity $\mathcal{C}^r$ if and only if it consists of nonnegative polynomial pieces $p^{(k)}$, $k = 0, \dots, m-1$, and at each internal knot point the derivatives of the two polynomials meeting at that point agree up to order r:*

$$\frac{\mathrm{d}^j}{\mathrm{d}x^j} p^{(i-1)}(a_i) = \frac{\mathrm{d}^j}{\mathrm{d}x^j} p^{(i)}(a_i) \quad \forall\, i = 1, \dots, m-1; j = 0, \dots, r. \tag{4}$$

Equation (4) translates to linear constraints on the coefficients of the polynomials $p^{(i)}$.

Semidefinite and linear constraints can be handled by some nonlinear optimization and modeling software, such as CVX [14], Sedumi [24], and SDPT3 [25], extremely effectively. However, currently none of these software can handle arbitrary convex, nonlinear objective function, which is necessary for some applications, such as maximum likelihood density estimation. An additional useful property of cubic and quartic splines is that their characterization involve semidefinite constraints of 2×2 matrices. Positive semidefiniteness of 2×2 matrices can be translated to linear and quadratic (second order cone) constraints using the following, well-known facts.

Proposition 3. *The following three statements are equivalent.*

1. *The matrix* $\begin{pmatrix} x_0 & x_1 \\ x_1 & x_2 \end{pmatrix}$ *is positive semidefinite.*
2. $x_0 \geq 0$, $x_2 \geq 0$, *and* $x_0 x_2 - x_1^2 \geq 0$.
3. $x_0 + x_2 \geq 0$ *and* $(x_0 + x_2)^2 \geq (x_0 - x_2)^2 + (2x_1)^2$.

The second system of inequalities can be handled by most state-of-the-art convex optimization software, which also handle arbitrary convex objective function, too. While carrying out the numerical experiments of this paper, we found the solvers KNITRO [20] and LOQO [26] especially effective and useful. The third system is useful for some solvers, such as Sedumi [24], which cannot handle arbitrary convex quadratic constraints, but handle *second order cone constraints* essentially as effectively as linear inequalities.

Definition 3. *The* $n+1$ *dimensional* second order cone *(or* Lorentz cone*) is defined as*

$$\mathcal{Q}_{n+1} = \{(x_0, \dots, x_n)\colon x_0^2 \geq x_1^2 + \dots + x_n^2, x_0 \geq 0\},$$

and a second order cone constraint *is one of the form* $\mathbf{x} \in \mathcal{Q}_{n+1}$.

The third system of Proposition 3 could also be written as $(x_0 + x_2, x_0 - x_2, 2x_1) \in \mathcal{Q}_3$.

The equalities in (1) and (2) suggest that we may easily run into serious numerical problems if some of the knot points are of different order of magnitude. The constraints obtained from (4) may also contain coefficients that differ by several orders of magnitude. This can be avoided by *scaling*: apply separately an affine transformation (change of variables) on each polynomial $p^{(i)}$ that maps the interval $[a_i, a_{i+1}]$ to $[0, 1]$, and represent each $p^{(i)}$ by the coefficients of the thus transformed polynomial, rather than by the original coefficients. The resulting scaled representation of splines is then the following.

$$S(x) = p^{(i)}(x) = \sum_{k=0}^{n} p_k^{(i)} \left(\frac{x - a_i}{a_{i+1} - a_i} \right)^k \quad \forall x \in [a_i, a_{i+1}] \tag{5}$$

This way the nonnegativity and continuity conditions are entirely independent of the knot points, and the remaining conditions also depend only on the ratios of differences between consecutive knot points $(a_{i+1} - a_i)/(a_i - a_{i-1})$. If the knot points are evenly distributed, that is, if $a_{i+1} - a_i = a_i - a_{i-1}$ for every i, then this representation is entirely independent from the knot points.

For example, here is the complete list of constraints that characterize a nonnegative cubic spline of continuity $\mathcal{C}^2$, with knot points $a_0, \dots, a_m$:

Theorem 4. *The coefficients* $p_k^{(i)}$, $i = 0, \dots, m-1$, $k = 0, \dots, 3$ *in* (5) *represent a nonnegative cubic spline over* $[a_0, a_m]$ *if and only if there exist real numbers* $x_\ell^{(i)}, y_\ell^{(i)}$, $i = 0, \dots, m-1$, $\ell = 0, \dots, 2$ *satisfying the the following system of equations and inequalities for all* $i = 0, \dots, m-1$.

$$\begin{aligned}
p_0^{(i)} &= y_0^{(i)} && \text{(6a)}\\
p_1^{(i)} &= 2y_1^{(i)} + x_0^{(i)} - y_0^{(i)} && \text{(6b)}\\
p_2^{(i)} &= y_2^{(i)} + 2x_1^{(i)} - 2y_1^{(i)} && \text{(6c)}\\
p_3^{(i)} &= x_2^{(i)} - y_2^{(i)} && \text{(6d)}\\
x_0^{(i)}, x_2^{(i)} &\geq 0 && \text{(6e)}\\
x_0^{(i)} x_2^{(i)} - \left(x_1^{(i)}\right)^2 &\geq 0 && \text{(6f)}\\
y_0^{(i)}, y_2^{(i)} &\geq 0 && \text{(6g)}\\
y_0^{(i)} y_2^{(i)} - \left(y_1^{(i)}\right)^2 &\geq 0 && \text{(6h)}\\
p_0^{(i+1)} &= \sum_{j=0}^{3} p_j^{(i)} && \text{(6i)}\\
\frac{1}{a_{i+2} - a_{i+1}} p_1^{(i+1)} &= \sum_{j=1}^{3} \frac{j}{a_{i+1} - a_i} p_j^{(i)} && \text{(6j)}\\
\frac{2}{(a_{i+2} - a_{i+1})^2} p_2^{(i+1)} &= \sum_{j=2}^{3} \frac{j(j-1)}{(a_{i+1} - a_i)^2} p_j^{(i)} && \text{(6k)}
\end{aligned}$$

Proof. A simple application of Propositions 1 and 3, (4), and (5). □

Notice that only (6f) and (6h) are nonlinear in (6), and that the constraints define a convex region, the intersection of the direct product of two second order cones and a convex polyhedron. After eliminating every $p_j^{(i)}$ from the system, the resulting constraints have 6 variables, 5 linear constraints, and 2 quadratic (or second order cone) constraints for each spline piece.

A similar characterization, consisting of linear and second order cone constraints, of nonnegative *quartic* splines of continuity $\mathcal{C}^2$ or $\mathcal{C}^3$ can be derived using Proposition 2; we leave the details to the reader.

2.2 When the Nonnegative Basis Approach is Sufficient

When multiple shape constraints are present, the SDP/SOCP models may be equivalent to a linearly constrained model, and the nonnegative basis approach outlined in Section 1 can be justified, provided that the right basis is selected. In other cases, LP relaxations can be obtained in which the *additional* shape constraints (monotonicity, convexity, etc.) are relaxed, and the SOCP constraints expressing nonnegativity reduce to *equivalent* linear constraints. As we argue in this section, a particularly useful basis is the set of *Bernstein basis polynomials*.

Definition 4. *The kth Bernstein basis polynomial of degree n is defined as*

$$B_{n,k} = \binom{n}{k} x^k (1-x)^{n-k}, \qquad k = 0, \ldots, n.$$

A polynomial of degree n is a Bernstein polynomial *if it can be expressed as a nonnegative linear combination of the Bernstein basis polynomials of degree n:* $B_{n,0}, \ldots, B_{n,n}$. *A spline is called a* piecewise Bernstein polynomial spline *if in its scaled representation* (5) *the polynomials* $p^{(i)}$ *are Bernstein polynomials.*

Clearly every Bernstein polynomial is nonnegative on $[0,1]$, but not vice versa. The following theorem provides a couple of sufficient conditions for a cubic nonnegative polynomial to be Bernstein polynomial.

Theorem 5. *If a cubic polynomial p is nonnegative on* $[0,1]$, *and it satisfies at least one of the conditions below, then it is a Bernstein polynomial:*

1. $p'(0) \geq 0$ *and* $p'(1) \leq 0$,
2. $p'(0) \geq 0$ *and* $p''(0) \geq 0$,
3. $p'(0) \geq 0$ *and* $p''(1) \leq 0$,
4. $p'(1) \leq 0$ *and* $p''(1) \geq 0$,
5. $p''(0) \leq 0$ *and* $p'(1) \leq 0$,
6. $p''(0) \leq 0$ *and* $p''(1) \leq 0$.

Proof. By simple arithmetic. For example, to prove the first claim, suppose that for some coefficients $\lambda_0, \ldots, \lambda_3$, the nonnegative polynomial p can be expressed as

$$\begin{aligned} p(x) &= \sum_{i=0}^{3} \lambda_i B_{3,i}(x) \\ &= (-\lambda_0 + 3\lambda_1 - 3\lambda_2 + \lambda_3)\, x^3 + (3\lambda_0 - 6\lambda_1 + 3\lambda_2)\, x^2 + (3\lambda_1 - 3\lambda_0)\, x + \lambda_0 \end{aligned}$$

Then, by assumptions,

$$\begin{aligned} p(0) &= \lambda_0 \geq 0, \\ p(1) &= \lambda_3 \geq 0, \\ p'(0) &= 3\lambda_1 - 3\lambda_0 \geq 0, \text{ and} \\ -p'(1) &= 3\lambda_2 - 3\lambda_3 \geq 0. \end{aligned}$$

These inequalities trivially imply $\lambda_i \geq 0$ for $i = 0, \ldots, 3$.

The proofs of the other five sufficient conditions are very similar. □

The reader might find it surprising that while the cone of nonnegative cubic polynomials is not polyhedral, and nor are the cones of increasing and decreasing cubic polynomials, we obtain a polyhedral cone when we intersect some of these cones with each other, or with the (polyhedral) cones of concave and convex cubic polynomials.

The following special case is probably worth emphasizing separately for splines.

Corollary 6. *If a nonnegative cubic polynomial spline is concave, then it is also a piecewise Bernstein polynomial spline.*

Other notable cases include monotone convex splines, and nonnegative unimodal splines. In these cases the above theorem gives an LP relaxation: the nonnegative basis approach using Bernstein basis polynomials is sufficient to represent all these splines.

The list provided in Theorem 5 is, of course, by no means exhaustive, but it already shows that there are many instances, when the nonnegative basis approach can be used. Another example is when the number of knot points is so high that each piece of the spline has a "simple enough" shape, and is either far from the x-axis, or falls into one of the categories mentioned above, turning the spline into a piecewise Bernstein polynomial spline. We found this to happen frequently in the non-homogeneous Poisson process arrival rate estimation problems; see also Section 3.5.

Other bases of cubic polynomials may also be useful, but we do not explore this direction any further. Instead, we show an important situation when the SOCP approach is provably superior to the nonnegative basis approach: nonlinear constraints are unavoidable to describe *nonnegative convex cubic* polynomials. Similar proofs can be constructed for other cases as well, such as the case of nonnegative, monotone, convex polynomials.

Lemma 7. *The cone of cubic polynomials that are nonnegative and convex over* $[0, 1]$ *is not polyhedral.*

Proof. It suffices to show that the cone in consideration has infinitely many extreme rays. Consider the polynomials of the form $p(x) = (x + 2x_0)(x - x_0)^2$ for arbitrary $0 < x_0 < 1$. These are clearly nonnegative and convex over $[0, 1]$. We show that these are all extreme polynomials. Suppose that p can be decomposed as a sum of two other polynomials from the cone: $(x + 2x_0)(x - x_0)^2 = q(x) + r(x)$ for some cubic polynomials q and r, which are also nonnegative and convex over $[0, 1]$. Nonnegativity of q and r imply that x_0 is a root of both q and r, and then it clearly must be a double root of both of them: $q(x) = (ax + b)(x - x_0)^2$ and $r(x) = (cx + d)(x - x_0)^2$ for some constants a, b, c, and d.

Comparing the coefficients of p and $q+r$ yields, by easy algebra, $a+c = 1$ and $b+d = 2x_0$. Furthermore, the convexity of q means that its second derivative is nonnegative at 0, which simplifies to $2ax_0 \leq b$. Similarly, the convexity of r simplifies to $2cx_0 \leq d$. Using these two inequalities and the previous two equations we obtain

$$2x_0 = 2x_0(a + c) = 2ax_0 + 2cx_0 \leq b + d = 2x_0,$$

which can only hold if the inequality in the middle also holds with equality. Consequently, $2ax_0 = b$ and $2cx_0 = d$, which can be written as $a(-2x_0) + b = c(-2x_0) + d = 0$. We got that both q and r have $-2x_0$ as their third root, which is also the third root of p. So the only way to decompose p as a sum of two other elements of our cone is to add two of its nonnegative multiples. This proves that p is an extreme element. □

A nonnegative basis of splines that has been extremely popular are *B-splines*. We argue that B-splines can always be replaced by piecewise Bernstein polynomial splines in optimization models.

Lemma 8. *The cone of cubic B-splines with a given set of knot points is a proper subset of the cone of piecewise Bernstein polynomial splines with the same set of knot points.*

Proof. A B-spline segment between two consecutive knot points can be written as a nonnegative linear combination of the same segment of the B-spline basis functions. We can assume without loss of generality that the knot points in question are at 0 and 1. Then the cubic B-spline segments are the polynomials

$$\left\{x^3, 3x^3 - 6x^2 + 4, -3x^3 + 3x^2 + 3x + 1, (1-x)^3\right\},$$

all of which are linear combinations of the Bernstein basis polynomials with nonnegative coefficients. □

This Lemma has two important implications for this paper. First, it is well-known that cubic B-splines over $[a, b]$ form a dense subset of the space of continuous nonnegative functions over $[a, b]$ (equipped with the uniform norm). The Lemma above implies that the same is true for piecewise cubic Bernstein polynomial splines. (This result can be established directly, too, without any reference to B-splines.) Second, it also justifies why we do not consider using B-splines in any of the models of this paper. Piecewise Bernstein polynomial splines have equally simple and compact representation as B-splines, and they lead to linearly constrained models very similar to those obtained when using B-splines. Yet, they form a strictly larger subset of nonnegative continuous function than B-splines. Nonnegative cubic splines form an even larger subset, but the corresponding optimization problems are also more complex, as they involve second-order conic constraints as well as linear constraints. Hence the comparison of the nonnegative basis approach with Bernstein polynomials and the SOCP approach is more meaningful than the comparison of models based on piecewise Bernstein polynomial splines and B-splines.

3 Models of Shape Constrained Estimation

3.1 Nonparametric regression of a nonnegative function

One of the most fundamental problems in statistics is univariate regression, the problem of estimating a univariate function f based on data (x_i, y_i) $i = 1, \ldots, N$, assumed to come from the model

$$y_i = f(x_i) + \epsilon_i \qquad \forall\, i,$$

where f is the unknown function to be estimated, and ϵ_i are independent, identically distributed random variables with mean zero.

The function f is assumed to belong to a class of functions $\mathcal{F}$, and is estimated by finding a function in $\mathcal{F}$ that fits the data at least as well as any other function in $\mathcal{F}$. The goodness-of-fit of f to the data is measured by some error functional typically of the form $d(f)+s(f)$, where the term $d(f)$ measures the distance of the function values $f(x_i)$ and y_i, while $s(f)$ is a penalty term that penalizes "rough" solutions.

The most common choice for $d(f)$ is the *residual sum of squares* $d(f) = \sum_{i=1}^{N}(f(x_i)-y_i)^2$, which is popular for many reasons: it is simple, easy to handle analytically, often leads to models that can be solved by a closed form formula, and it appears in maximum likelihood models if ϵ_i are normally distributed.

The smoothing term $s(f)$ is often omitted, especially if $\mathcal{F}$ already consists of smooth functions only. If present, it is typically something like $\int |f'|$, $\int |f''|$, or $\int (f'')^2$.

When using cubic splines, all the above choices of $s(f) + d(f)$ lead to optimization models with only linear and second order cone constraints. The nonlinear constraints are unavoidable, as we are optimizing over a non-polyhedral cone.

3.2 Isotonic, Convex and Concave Nonparametric Regression

Isotonic regression is a variant of the regression problem of the previous section, with the additional requirement that the estimated function f is monotone increasing or decreasing. We may also require that f is convex or concave, in which case we speak of convex or concave regression.

Derivatives of polynomials are also polynomials, hence the monotonicity, concavity, and convexity constraints can also be formulated as nonnegativity constraints on polynomials. When using cubic splines, convexity and concavity are linear constraints.

When using cubic splines, Theorem 5 in Section 2.2 provides conditions when the SOCP model is equivalent to a linearly constrained model. Notable special cases include concave nonnegative splines.

3.3 Unconstrained Density Estimation

One can formulate the estimation of probability density functions (*pdf*s) as a shape constrained optimization problem. The problem is to estimate an unknown probability distribution from a finite set of independent samples $X_1, \ldots, X_n$ of that distribution.

A pdf must be nonnegative and integrate to one. We can assume that the pdf to be estimated has finite support, say $[a_0, a_m]$. With this assumption, when using a spline model, the second constraint simplifies to a linear constraint, since the integral of a polynomial of a given interval is a linear function of the coefficients of the polynomial. For example, a cubic spline model can be constructed by adding to (6) the constraint

$$\sum_{i=0}^{m-1}\sum_{j=0}^{3}\frac{a_{i+1}-a_i}{j+1}p_j^{(i)}=1$$

Finally, the objective function needs to be determined. The most common and straightforward approach is *maximum likelihood* estimation. If the unknown pdf is denoted by f, this amounts to maximizing the likelihood function $\prod_{i=1}^n f(X_i)$. This objective function is numerically very badly behaving, and is not necessarily concave. Instead, we shall use its logarithm, $\sum_{i=1}^n \log f(X_i)$ as the objective function, which is concave if f is a polynomial spline of a given knot sequence. It is also important to note that by constraining f to be a polynomial spline, the above maximum likelihood optimization problem is always well-defined. (It has an optimal feasible solution.)

Alternatively, one may wish to estimate a cumulative distribution function (*cdf*) with support $[a_0, a_m]$. Then the constraints to be added are: $p(a_0) = 0$, $p(a_m) = 1$ (these are two simple linear constraints), and the nonnegativity of p', another system of constraints that express the nonnegativity of a polynomial. The nonnegativity of the cdf itself is then, of course, superfluous, and the objective function has to be changed, too. We leave the details to the reader. Since the two problems are clearly equivalent, in this paper we only show numerical results for pdf estimation.

3.4 Unimodal Density Estimation

Further constraints can be added to these models for *unimodal* density estimation. If the mode is known, we can place one of the knot points to the mode, and then add constraints that the spline is increasing from the first knot point to the mode, and decreasing from the mode to the last knot point. If the mode is unknown, an approximate solution to the problem can be found by solving several optimization models with different fixed modes, and comparing the optimal solutions that correspond to the different modes. Finding the exact solution this way is still not trivial, as the maximum likelihood function (as a function of the mode) is not unimodal, let alone concave.

Two important remarks are due here. Firstly, owing to the unimodality constraints, the nonnegativity constraints may be simplified: it is sufficient to require the pdf to be nonnegative at the first and last knot point. Secondly, the presence of multiple shape constraints may result in problems where the nonnegative basis approach is equivalent to the SOCP approach.

3.5 Arrival Rate Approximation

The recent paper [2] gives the details of several maximum likelihood models for estimating the arrival rate of a non-homogeneous Poisson process using nonnegative splines. The models are different primarily in how the data is given: by exact arrival times, or by the number of arrivals in given intervals. Furthermore, slightly different models can be obtained if we assume that the arrival rate function is periodic with a fixed period length.

Compared to the regression model with a nonnegativity constraint, the nonnegative spline model for the arrival rate approximation problem is only different in its objective function. The least squares objective function is replaced by a likelihood (or log-likelihood) function, and minimization by maximization. The natural log-likelihood approach directly leads to a model with concave objective function, the details can be found in [2].

The same paper also compares the shape constrained spline approach to some other approaches, hence in Section 4 we only compare this approach to the nonnegative basis approach, using both simulated data, and the same dataset of email arrivals as the one used in [2].

3.6 Selection of Knot Points

In each of the spline models above we have assumed that a fixed sequence of knot points $a_0, \dots, a_m$ is given. Finding the best selection of knot points is a central, but very difficult, problem. Ideally, we would make the knot points variables, and optimize over them as well as the coefficients of the polynomials, but this would result in a hardly tractable, non-convex optimization problem.

A common knot point selection "strategy" is to use evenly distributed knot points. While this is a very crude method, it is also very simple, and it results in simplified optimization models. (For example, from (6) we can eliminate the constants a_i entirely.) Since our primary goal is to show that our shape constrained spline models are competitive with the most commonly used kernel methods in all of the applications mentioned above, we decided on this simple knot point selection method. As shown in Section 4 this was already sufficient for good results.

Another knot point selection method commonly used in regression problems is to place knot points at the data points. A theoretical result supporting this idea is that the optimal solutions to certain optimization models (least squares regression with a penalty term for smoothing) are natural cubic splines [27].

If evenly distributed knot points are chosen, then the number of knot points also have to be decided on. This can be done by any of the common model selection procedures: validation on a test set (if there is one), cross-validation or k-folding (if there is none), or by measuring the trade-off between goodness-of-fit and the number of estimated parameters using either the Akaike or the Bayesian Information Criterion (AIC and BIC, respectively) [4, 5]. Cross-validation and k-folding are much more time consuming than computing a score such as AIC or BIC, hence we decided on using AIC. Following [5], we used the AIC_c score instead of AIC; this is a variant of AIC with an additional second-order term that becomes necessary when the number of estimated parameters is small. When the number of estimated parameters tends to infinity, the difference between the values of AIC_c and AIC tends to zero, so AIC_c is justified regardless of the number of parameters.

The numerical experiments of Section 4 were carried out in the following manner: the number of knot points were increased one by one from 2 to 30 (this proved to be plenty in each of the examples), and the optimal solution to the respective optimization model was determined. The values of the AIC_c scores were determined, and the optimal solution with the best (lowest) score was selected as the final solution.

4 Numerical Examples

In this section we compiled results of numerical experiments in which the SOCP-based cubic (and sometimes quartic) spline models were compared to the piecewise Bernstein polynomial spline model, to kernel methods, and in some cases to parametric methods.

The following acronyms are used to refer to the models in the tables:

1. PM: Parametric model; the precise meaning is explained in each section, as different parametric models are used in different problems.
2. NCS: The best (least squares) fitting nonnegative cubic spline with evenly distributed knot points. The number of knot points is chosen using AIC_c .
3. NICS: The best (least squares) fitting nonnegative, increasing cubic spline with evenly distributed knot points. The number of knot points is chosen using AIC_c .
4. NVCS: The best (least squares) fitting nonnegative, concave cubic spline with evenly distributed knot points. The number of knot points is chosen using AIC_c .
5. NIVCS: The best (least squares) fitting nonnegative, increasing, concave cubic spline with evenly distributed knot points. The number of knot points is chosen using AIC_c .
6. NIVQS: The best (least squares) fitting nonnegative, increasing, concave quartic spline with evenly distributed knot points. The number of knot points is chosen using AIC_c .

The optimization models described in Sections 2 and 3 were implemented using the AMPL modeling language [12], and were solved using the nonlinear solver KNITRO [20]. Most of the optimization problems could be solved very fast, in less than a second, using an ordinary PC.

4.1 Isotonic and Convex/Concave Regression

We compared various shape-constrained spline models to some parametric approaches using a real dataset (Section 4.1.1) and simulated data (Section 4.1.2). The parametric model for the real data is the one used in the first paper on the analysis of the data. The parametric models for the simulated data are simple one- and two-parameter families of functions which contain the true function.

4.1.1 The Rabbit Data

The data is from [10], collected in a study to find the relationship between the weight of the eye lens and the age of rabbits in Australia. The 71-sample data set was retrieved from the OzDASL data library [6].

The original solution was based on parametric regression, using the model $a = A \cdot 10^{-B/(w+C)}$, where the response variable a is the age and the covariate w is the dried eye lens weight of the rabbits; A, B, and C are the unknown nonnegative parameters.

The solution reported in [10] was $A = 288.27$, $B = 60.1912$, $C = 41$; close to the best least squares fit of this model, which is $A = 279.8278$, $B = 55.4063$, $C = 36.0437$. Below we demonstrate that a shape-constrained nonparametric approach can yield a comparably good fit.

Both the interpretation and a plot of the data suggests that the relationship between a and w is likely to be monotone increasing and concave. (The parametric model above satisfies both constraints whenever $B \leq 2C$.) We added both of these constraints, separately and together, to our nonnegative cubic spline models (Bernstein polynomial-based and SOCP). Wherever the two models are equivalent, we used the Bernstein polynomial-based model. We also compared our results to the best fitting increasing and concave quartic spline.

The results are shown in Table 1. L_2 is the residual sum of squares, L_∞ and L_1 are the maximum and the sum of absolute differences, and AIC_c is the value of the Akaike Information Criterion [5] with a second order corrective term needed because of the small number of parameters [4]. PM1 is the solution to the parametric model suggested in [10]. PM2 is the above least-squares solution to the same model. The remaining acronyms in this table are explained in the beginning of Section 4.

Plots of the estimated regression curves are shown on Figure 1.

model	effective number of parameters	L_2	L_∞	L_1	AIC_c
PM1	3	4533.38	21.17	428.554	502.962
PM2	3	4320.65	21.45	423.163	499.550
NCS	6	4379.06	21.72	421.533	507.421
NICS	7	4277.90	22.75	415.037	508.227
NVCS	7	4277.90	22.75	415.037	508.227
NIVCS	7	4277.90	22.75	415.037	508.227
NIVQS	7	4274.97	22.46	414.516	508.178

Table 1: Comparison of the parametric and shape-constrained spline estimators for the regression curves of the Rabbit data. The acronyms in the table headings are explained in the beginning of this section.

The NCS spline is monotone increasing everywhere except in a small segment towards the right-hand side boundary of the domain. The spline is not concave.

The table shows that the NICS, NVCS, and NIVCS models gave exactly the same result. This is because the NICS spline, surprisingly, turned out to be concave, too, but only for the optimal number of knot points. (This is purely accidental.) At this experiment, naturally, we used the SOCP-based method, but since nonnegative concave cubic splines are automatically

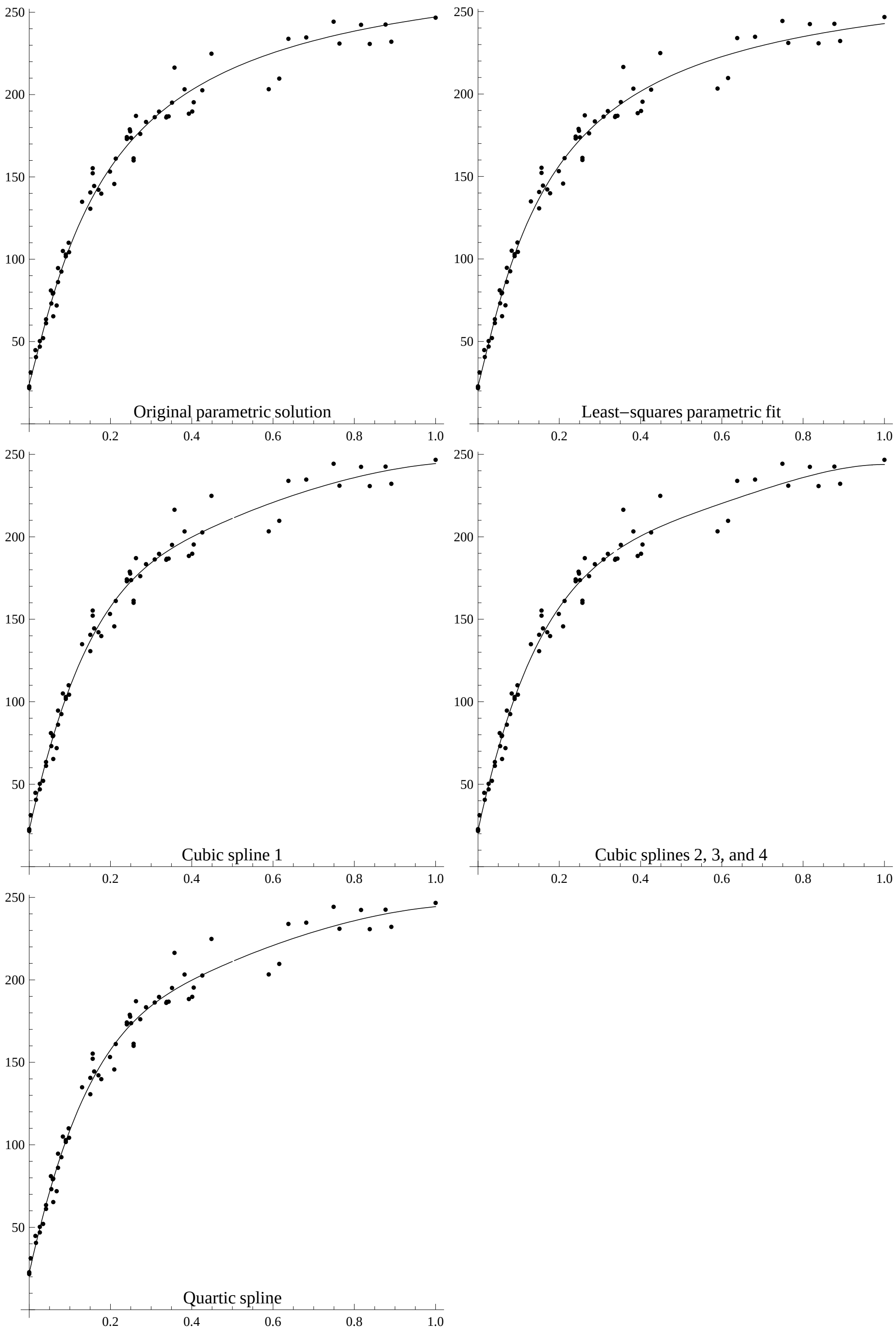

Figure 1: Parametric and shape-constrained spline estimators for the regression curves of the Rabbit data.

piecewise Bernstein polynomial splines, the nonnegative basis approach with Bernstein basis polynomials would have given the same result.

Similarly, the NVCS was automatically monotone increasing for every number of knot points tested. (This is accidental, too.) It follows that the optimal NIVCS spline coincides with the optimal NVCS and NICS splines in this example.

Using quartic splines in this example yields only a very minor improvement compared to cubic splines. (Notice that the same number of effective parameters is achieved, meaning that one fewer knot points was used.) The fitted shape-constrained cubic splines are comparable to the fitted parametric models.

4.1.2 Simulated Data

We simulated noisy data using the model $Y = f(X) + \epsilon$, where $f(x) = \frac{1}{1+e^{-10x}}$, $x \in [0,1]$, and ϵ is normally distributed with mean 0 and standard deviation 0.2. This function was chosen so that the function has a nearly linear, increasing, and a long, nearly horizontal part on the domain – this way it is likely that explicit monotonicity and concavity constraints will be required for a good quality fit.

Random samples of size 50 were drawn uniformly from the interval $[0,1]$. For each sample the least-squares optimal model from the one-parameter family $f_b(x) = \frac{1}{1+e^{-bx}} + \epsilon$ and the two-parameter family $f_{a,b}(x) = \frac{1}{a+e^{-bx}} + \epsilon$ were computed. In this section we refer to these reference models as parametric models PM1 and PM2. Since these least-squares problems are nonconvex optimization problems, we verified (by brute-force) that the locally optimal solutions found are indeed globally optimal. We compared these models to different shape-constrained polynomial spline models by measuring the L_2, L_1, and L_∞ distances of the true function f and the estimators, together with their estimated numbers of parameters and their AIC$_c$. This process was repeated 100 times. The means and standard deviations of the distances are reported in Table 2. In the table headings μ_p and σ_p denote the estimated mean and standard deviation of the L_p norm of the estimators; $p \in \{1, 2, \infty\}$. Similarly, μ_{AIC} and σ_{AIC} are the mean and standard deviation of the AIC$_c$ scores of the estimators. PM1 and PM2 are the least-squares solution to the Parametric models above, the remaining acronyms in this table are explained in the beginning of Section 4.

It is immediate from Table 2 that there is a huge benefit from each added shape constraint. Both the increasing and the concave cubic splines are considerably better estimates than the (otherwise unconstrained) nonnegative splines, and the concave increasing spline is better than both of them, with respect to every distance measure considered. Obviously, the AIC$_c$ scores worsen as we add more constraints to the model. The quartic splines, on the other hand, are not any better than the cubic ones in this example. Notice that the best models outperform even the parametric model PM2 with respect to every measure of goodness-of-fit, except for AIC$_c$, the latter owing to the larger number of parameters. The optimal NIVCS model usually had 5 or 6 effective parameters compared to the 2 fixed parameters of PM2. The single-parameter PM1 model proved to be hard to match.

model	$10^3 \cdot \mu_2$	$10^3 \cdot \sigma_2$	$10^3 \cdot \mu_1$	$10^3 \cdot \sigma_1$	μ_∞	σ_∞	μ_{AIC}	σ_{AIC}
PM1	0.972	1.516	15.685	12.268	0.054	0.050	−20.727	11.443
PM2	7.552	3.051	65.173	15.911	0.265	0.100	−12.934	11.513
NCS	5.635	4.675	55.616	21.907	0.201	0.148	−13.523	11.418
NICS	2.768	2.992	36.890	18.938	0.133	0.099	−11.022	11.486
NVCS	3.277	2.942	43.397	18.716	0.128	0.067	−11.826	11.423
NIVCS	2.243	2.309	34.466	17.964	0.111	0.076	−10.602	11.480
NIVQS	2.613	2.567	36.431	17.667	0.138	0.102	−8.220	11.385

Table 2: Comparison of parametric and spline estimators for the regression curves of a simulated dataset. μ_i and σ_i are the estimated mean and standard deviation of the L_i error based on 100 experiments. The remaining acronyms in the table headings are explained in the beginning on Section 4.

4.2 Density Estimation

The following tests were conducted to compare our methods to a wider range of kernel methods. The benchmark distributions, as well as the experimental design is based on [11, Chapter 8]. The probability density functions (pdf) of the benchmark distributions are:

$$\begin{aligned}
f_1(x) &= \frac{9}{10}\phi_{1/2}(x-5) + \frac{1}{10}\phi_{1/2}(x-7),\\
f_2(x) &= \phi_1(x-5),\\
f_3(x) &= \frac{1}{5}U([3,8]),\\
f_4(x) &= \frac{1}{5}\psi_{1.4,2.6}\left(\frac{1}{5}(x-0.3)\right)\\
f_5(x) &= \frac{1}{4}\phi_{9/5}(x-6) + \frac{4}{5}\phi_{1/10}(x-2)\\
f_6(x) &= \frac{1}{2}\phi_{1/2}(x-3.5) + \frac{1}{2}\phi_{1/2}(x-6.5),
\end{aligned}$$

where $\phi_\sigma(x)$ is the pdf of the normal distribution with mean zero and standard deviation σ, $U([a,b])$ is the pdf of the uniform distribution on $[a,b]$, and $\psi_{\alpha,\beta}$ is the pdf of the Beta density with parameters α and β.

Following [9], the authors of [11] argue that the most meaningful comparison between estimates of probability densities is the L_1 distance of their pdfs, and they report only this measure of goodness-of-fit in their benchmark. Hence in this section we do the same when comparing our methods and theirs.

For each benchmark density, random samples of size 100 were generated. Then the optimal cubic spline densities were determined using the Bernstein polynomial based and the SOCP based methods outlined in Sections 2.1 and 2.2, and compared (by its L_1 distance to the known pdf) to the kernel estimates of [11], which employ the Epanechnikov kernel and the normal density kernel, and thirteen bandwidth selection methods, including the "optimal method", which simply determines the bandwidth that minimizes the L_1 error. This is clearly not a rational method, as it requires the knowledge of the estimated pdf, but it serves as a good benchmark, as no other method bandwidth selection can possibly beat it. The process were repeated 100 times, the means and standard deviations of the L_1 distances are reported in Table 4.2.

To avoid repeating all the results of the book, we refer the reader to the tabulated numerical results of [11, pp. 326–327]. In our table we only show the L_1 errors of the "optimal method". Columns OP-E and OP-N show the results of the optimal method using the Epanechnikov and the normal density kernel, respectively. The columns Bernstein and SOCP contain the results of the least-squares best fitting piecewise Bernstein polynomial and nonnegative splines, respectively. The rows μ_i and σ_i show the mean and the standard deviation of the L_1 errors from the experiments with the density f_i.

	OP-E	OP-N	Bernstein	SOCP
μ_1	0.158	0.160	0.150	0.159
σ_1	0.053	0.053	0.049	0.045
μ_2	0.127	0.129	0.064	0.075
σ_2	0.054	0.054	0.069	0.060
μ_3	0.217	0.218	0.171	0.179
σ_3	0.042	0.042	0.069	0.075
μ_4	0.162	0.164	0.148	0.153
σ_4	0.041	0.041	0.065	0.064
μ_5	0.335	0.335	0.649	0.643
σ_5	0.055	0.055	0.022	0.051
μ_6	0.187	0.189	0.227	0.226
σ_6	0.053	0.053	0.031	0.026

Table 3: Estimated means and standard deviations of the L_1 errors of kernel and spline density estimates. See the text for the interpretation of the table headings.

It is safe to conclude that in all but one of these examples both the Bernstein polynomial-based and the SOCP methods give results comparable to those of the kernel methods. Except for the density f_5 (a bad mixture of normal distributions) both methods nearly match, break even with, or outperform even the optimal kernel method. The results are easy to interpret: the less smooth the estimated pdf is, the higher the disadvantage of cubic splines to the kernel methods. The difference between Bernstein polynomial-based and the SOCP methods is too small to be considered significant, their performance is essentially identical in all the examples.

4.3 Arrival Rate Estimation

The paper [2] gives detailed numerical evaluation of the SDP/SOCP approach and they show that it compares very favorably to some wavelet-based methods. We only compare the SDP/SOCP approach and the nonnegative basis approach using Bernstein polynomials.

In Section 4.3.1 we compare the two methods on the same email arrival data that was used introduced in [2]. In Section 4.3.2 we use simulated arrival data.

4.3.1 Email Arrival Data

In this section we present results obtained for a dataset of 10150 e-mail arrivals recorded in a period of 446 days. As described in [2], the arrival times were inexact, as they were rounded to the nearest second. It was assumed that the arrival rate is periodic with a period of one week. The details of the appropriate maximum likelihood model using nonnegative cubic splines is also given in [2], and using cross-validation the authors of [2] found that the optimal number of evenly distributed knot points to be used is 48. This is a large number in the sense that it gives about 7 pieces for each day. With so many pieces it can be expected that Theorem 5 will apply to many of the spline pieces.

We determined the optimal nonnegative spline as well as the optimal piecewise Bernstein polynomial spline with $7k, (k = 3, \ldots, 14)$ knot points, and with the quoted 48 knot points. We found that with 28 knot points there was some small but visible difference between the two corresponding splines. Not surprisingly, the differences occur during night times (when the arrival rate is close to zero) and in the subsequent mornings. In all the other experiments, with a different number of knot points, the two splines were identical.

4.3.2 Simulated Data

Simulated experiments were carried out in the following manner. Arrival times were simulated by our implementation of the well-known thinning method [17], using a simple arrival rate function,

$$4000(\cos(6\pi t + 0.2) + 2.16 \sin(4\pi t + 3.8) + 2.6),$$

which demonstrates several properties that can potentially make the approximation difficult: it gets near zero, it has a high and narrow peak, and it also has a long flat part. It was also chosen to be periodic with period 1, as most practical applications assume periodic arrival rate function.

Assuming that the interval $[0, 1]$ represents a week, arrival times were rounded to the nearest minute. Then the optimal cubic spline arrival rates were determined using the Bernstein polynomial based and the SOCP based methods outlined in Sections 2.1 and 2.2, with AIC_c used in knot point selection. This process was repeated a hundred times. Figure 2 shows the 95% empirical confidence bands plotted against each other and the true arrival rate function. The red curve is $\lambda(t)$, the blue band corresponds to the SOCP approach, and the green band corresponds to the nonnegative basis approach. It is apparent that the difference between the results of the two approaches is quite small. Figure 3 shows the width of the confidence bands plotted against each other. It is clear from this plot that the quality of the SOCP-based model is pointwise better than that of the nonnegative basis approach.

5 Conclusion

Two optimization-based approaches, both using low-degree splines, are presented, and compared to each other and to popular kernel methods in various applications of shape con-

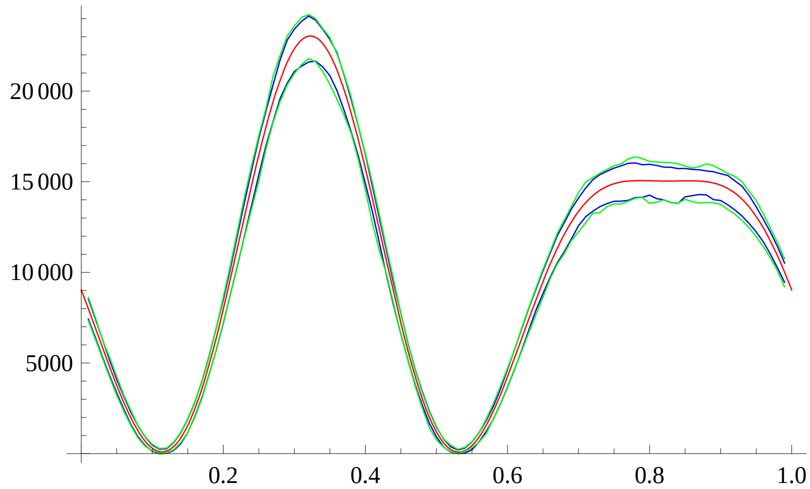

Figure 2: Comparison of the 95% empirical confidence bands of the results from SOCP and nonnegative basis approaches. The blue curve corresponds to the SOCP approach, and the green one to the piecewise Bernstein polynomial splines.

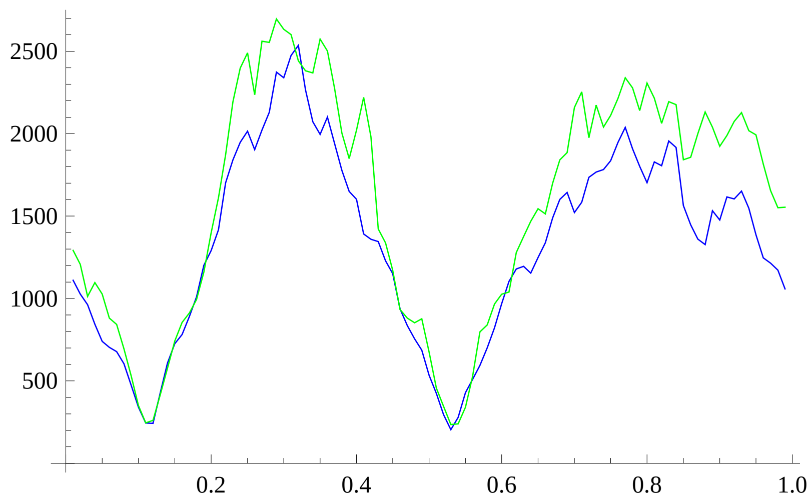

Figure 3: The bandwidths of the 95% empirical confidence bands of the results from SOCP and nonnegative basis approaches. The blue curve corresponds to the SOCP approach, and the green one to the piecewise Bernstein polynomial splines.

strained estimation problems.

Theoretically, the SDP/SOCP approach, which requires solving optimization problems with linear and second order conic constraints, is clearly superior to the basic nonnegative basis approach, and it allows us to optimize precisely over the set that we want. The nonnegative basis approach results in simpler, linearly constrained models, but in many cases it allows us to optimize over a strictly smaller or larger set than what we need.

The optimization models obtained from both approaches were easily solvable in a fraction of a second using readily available nonlinear solvers.

In many instances the two approaches gave very similar results, but we have also found several examples where the SOCP approach is superior in practice, too. On the other hand, we exhibited a few nontrivial cases when the two approaches are provably equivalent.

We found that these approaches can match, and frequently outperform state-of-the-art kernel methods both in regression and density estimation problems.

References

[1] Farid Alizadeh, *Semidefinite and second order cone programming and their application to shape constrained regression and density estimation*, Proceedings of the INFORMS Annual Meeting (Pittsburgh), INFORMS, 2006.

[2] Farid Alizadeh, Jonathan Eckstein, Nilay Noyan, and Gábor Rudolf, *Arrival rate approximation by nonnegative cubic splines*, Operations Research **56** (2008), 140–156.

[3] Louis Brickman and Leon Steinberg, *On nonnegative polynomials*, The American Mathematical Monthly **69** (1962), no. 3, 218–221.

[4] Kenneth P. Burnham and David Anderson, *Model selection and multi-model inference*, 2 ed., Springer, 2003.

[5] ______, *Multimodel inference – understanding aic and bic in model selection*, Sociological Methods & Research **33** (2004), no. 2, 261–304.

[6] OzDASL Australasian Data and Story Library, `http://www.statsci.org/data/index.html`.

[7] Michel Delecroix and Christine Thomas-Agnan, *Spline and Kernel Regression under Shape Restrictions*, Smoothing and Regression, Approaches, Computation, and Applications, Wiley Interscience, 2000.

[8] H. Dette and W. J. Studden, *The Theory of Canonical Moments with Applications in Statistics, Probability, and Analysis*, Wiley Interscience Publishers, 1997.

[9] L. Devroye and L. Györfi, *Nonparametric density estimation – the L_1 view*, John Wiley & Sons, 1985.

[10] M. L. Dudzinski and R. Mykytowycz, *The eye lens as an indicator of age in the wild rabbit in australia*, CSIRO Wildlife Research **6** (1961), 156–159.

[11] P.P.B. Eggermont and V.N. LaRiccia, *Maximum penalized likelihood estimation, volume i: Density estimation*, Springer-Verlag, New York, 2001.

[12] R. Fourer, D. M. Gay, and B. W. Kernighan, *AMPL: A modeling language for mathematical programming*, 2 ed., Boyd and Fraser, Danvers, MA, 2002.

[13] Tadayoshi Fushiki, Shingo Horiuchi, and Takashi Tsuchiya, *A maximum likelihood approach to density estimation with semidefinite programming*, Neural Computation **18** (2006), 2777–2812.

[14] Michael Grant and Stephen Boyd, *CVX: Matlab software for disciplined convex programming (web page and software)*, Stanford university, December 2007, Available from `http://stanford.edu/~boyd/cvx`.

[15] Didier Henrion and Andrea Garulli (eds.), *Positive polynomials in control*, LNCS, vol. 312, Springer, 2005.

[16] S. Karlin and W. J. Studden, *Tchebycheff Systems, with Applications in Analysis and Statistics*, Wiley Interscience Publishers, 1966.

[17] Peter A.W. Lewis and Gerald S. Shedler, *Simulation methods for Poisson processes in nonstationary systems*, Proceedings of the 10th conference on Winter simulation (Miami Beach, FL) (Harold Joseph Highland, ed.), IEEE Press, Piscataway, NJ, USA, December 1978, pp. 155–163.

[18] M. Meyer, *Shape Restricted Inference with Applications to Nonparametric Regression,Smooth Nonparametric Function Estimation, and Density Estimation*, Ph.D. thesis, University of Michigan, Ann Arbor, 1996.

[19] Yurii Nesterov, *Squared functional systems and optimization problems*, High Performance Optimization (H. Frenk, K. Roos, T. Terlaky, and S. Zhang, eds.), Appl. Optim., Kluwer Acad. Publ., Dordrecht, 2000, pp. 405–440.

[20] J. Nocedal and R. A. Waltz, *KNITRO user's manual*, Tech. Report OTC 2003/05, Optimization Technology Center, Northwestern University, Evanston, IL, 2003.

[21] Pablo A. Parrilo and Sanjay Lall, *Semidefinite programming relaxations and algebraic optimization in control*, European Journal of Control **9** (2003), no. 2-3.

[22] George Pólya and Gabor Szegő, *Problems and theorems in analysis*, vol. 2.

[23] Stephen Prajna, Pablo A. Parrilo, and Anders Rantzer, *Nonlinear control synthesis by convex optimization*, IEEE Transactions on Automatic Control **49** (2004), no. 2, 310–314.

[24] Jos F. Sturm, *Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones (updated for version 1.05)*, 2001, Available from `http://sedumi.mcmaster.ca`.

[25] Reha H. Tütüncü, Kim-Chuan Toh, and Michel J. Todd, *Solving semidefinite-quadratic-linear programs using SDPT3*, Mathematical Programming Ser. B **95** (2003), no. 2, 189–217.

[26] R. J. Vanderbei, *LOQO user's manual*, Tech. Report ORFE-99, Operations Research and Financial Engineering, Princeton University, October 2000.

[27] G. Wahba, *Spline models for observational data*, SIAM, 1990.