Appendix to “Shape constrained estimation using nonnegative splines”
published in the Journal of Computational and Graphical Statistics

Dávid Papp and Farid Alizadeh

June 14, 2012

A Second order cone programming (SOCP) and semidefinite programming (SDP)

Semidefinite programming is a generalization of the well-known linear optimization (or linear programming) problem. A *semidefinite program* (or SDP for short) is the abstract problem of finding the optimum of a linear function subject to the constraint that an affine combination of matrices is positive semidefinite. In other words, it is an optimization problem of the form

$$\begin{aligned} & \underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} && \sum_i c_i x_i \\ & \text{subject to} && \mathbf{A}_0 + \sum_{i=1}^n \mathbf{A}_i x_i \succcurlyeq \mathbf{0}, \end{aligned}$$

where the vector $\mathbf{c} \in \mathbb{R}^n$ and the $m \times m$ symmetric matrices $\mathbf{A}_i$, ($i = 0, \dots, n$) are given; x_i denotes the i th component of the vector $\mathbf{x}$; these are the variables. To simplify presentation, we shall now assume that the sought minimum exists (the infimum is finite, and it is attained); this indeed holds for all the SDPs considered in this paper.

Similarly, a *second order cone program* (or SOCP for short) is an optimization problem

$$\begin{aligned} & \underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} && \sum_i c_i x_i \\ & \text{subject to} && \mathbf{a}_0 + \sum_{i=1}^n \mathbf{a}_i x_i \in \mathcal{Q}_{k+1}, \end{aligned}$$

with given vectors $\mathbf{c} \in \mathbb{R}^n$ and $\mathbf{a}_i$, ($i = 0, \dots, n$). Recall that $\mathcal{Q}_{k+1} = \{(z_0, \dots, z_k) : z_0 \geq \|(z_1, \dots, z_k)^\top\|_2\}$ is the $(k+1)$ -dimensional *second order cone*.

The constraints that appear in the above problems are called *semidefinite constraints* and *second order cone constraints* (the former are sometimes also called *linear matrix inequalities*).

It can be shown that SOCPs are special cases of SDPs, and in turn, linear programming is a special case of SOCP. This also implies that SOCP and SDP constraints can be mixed in the same optimization problem, the resulting problem is also an SDP.

The above formulations are considered “standard form” SDPs and SOCPs, but other, seemingly more general, optimization problems that can be converted to the above form are also called SDPs and SOCPs. With this terminology, one of the key observations used in this paper can be summarized as follows: *The constraint that a piecewise polynomial spline with given knot points and degree is nonnegative over a given interval can be expressed as the constraints of an SDP, and if the spline degree is at most three, this SDP simplifies to an SOCP.*

A.1 Remarks on software

Several efficient algorithms for the solution of second order cone programs and semidefinite programs are reviewed in (Alizadeh and Goldfarb, 2003) and (Wolkowicz et al., 2000); these references also provide a general introduction to these areas of optimization. The theoretical complexity of optimization problems involving only linear and second order cone constraints is the same as that of linear optimization (Alizadeh and Goldfarb, 2003). This also implies that the models presented in this paper are *polynomial time solvable*, which is the standard abstraction of the “efficient” solution of “tractable” problems (Cormen et al., 2009, sec. 34.1).

Linear constraints, second order cone constraints, and semidefinite constraints can be handled by some nonlinear optimization and modeling software, such as CVX (Grant and Boyd, 2007) and SeDuMi (Sturm, 2001) very 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.

Second order cone constraints can also be handled by most state-of-the-art convex optimization software, which can solve problems with arbitrary convex objective functions. While carrying out the numerical experiments of this paper, we found the software KNITRO (Nocedal and Waltz, 2003), and CPLEX (CPLEX, 2011) especially effective and useful. The models of this paper can be solved in a fraction of a second without any numerical issues by the software mentioned above. Modeling languages (such as AMPL) also allow the user to input the mathematical optimization models of this paper, such as the representation of cubic splines shown in Theorem 3 in the

Appendix, almost *verbatim* in the solvers. Only minimal effort is required from the user to obtain the estimators. The supplementary material contains AMPL model files for all the shape constrained estimation problems considered in the paper.

B An SOCP characterization of nonnegative cubic splines

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

$$p_0^{(i)} = y_0^{(i)} \tag{6a}$$

$$p_1^{(i)} = 2y_1^{(i)} + x_0^{(i)} - y_0^{(i)} \tag{6b}$$

$$p_2^{(i)} = y_2^{(i)} + 2x_1^{(i)} - 2y_1^{(i)} \tag{6c}$$

$$p_3^{(i)} = x_2^{(i)} - y_2^{(i)} \tag{6d}$$

$$(x_0^{(i)} + x_2^{(i)}, x_0^{(i)} - x_2^{(i)}, 2x_1^{(i)})^\top \in \mathcal{Q}_3 \tag{6e}$$

$$(y_0^{(i)} + y_2^{(i)}, y_0^{(i)} - y_2^{(i)}, 2y_1^{(i)})^\top \in \mathcal{Q}_3 \tag{6f}$$

$$p_0^{(i+1)} = \sum_{j=0}^3 p_j^{(i)} \tag{6g}$$

$$\frac{1}{a_{i+2} - a_{i+1}} p_1^{(i+1)} = \sum_{j=1}^3 \frac{j}{a_{i+1} - a_i} p_j^{(i)} \tag{6h}$$

$$\frac{2}{(a_{i+2} - a_{i+1})^2} p_2^{(i+1)} = \sum_{j=2}^3 \frac{j(j-1)}{(a_{i+1} - a_i)^2} p_j^{(i)} \tag{6i}$$

Proof. Eqs. (6a)–(6f) come from Proposition 1, using Proposition 3 to translate the 2×2 semidefinite constraints to second order cone constraints. Eqs. (6g)–(6i) express the continuity of the derivatives up to order two. $\square$

C Proof of Theorem 1

We start the proof by showing that for every polynomial p of degree n , strictly positive over $[0, 1]$, there exist nonnegative constants C_i such that $p + C_i \in \mathcal{P}(U, \mathbf{a}_i)$ for every i , and $\lim C_i = 0$.

Fix i , and consider two adjacent knot points a_k and a_{k+1} from the knot point sequence $\mathbf{a}_i$. The polynomial p can be represented as a piecewise polynomial spline of degree n with knot point sequence $\mathbf{a}_i$; its scaled representation (4) has $p^{(k)}(x) = p((a_{k+1} - a_k)x + a_k)$. Collecting terms in the standard basis, we have

$$p^{(k)}(x) = p((a_{k+1} - a_k)x + a_k) = p(a_k) + \sum_{i=1}^n q_i^{(k)} x^i$$

with some $q_i^{(k)} = \mathcal{O}((a_{k+1} - a_k)^i)$, $i = 1, \dots, n$. By assumption, $p(a_k) > 0$, because p is strictly positive on $[0, 1]$. All other coefficients $q_i^{(k)}$ are of order $\mathcal{O}(a_{k+1} - a_k)$. By the assumption on U , $\sum_{j=0}^n \alpha_j u_j \equiv p(a_k)$ for some positive $\alpha_0, \dots, \alpha_n$. Now, if we express $p^{(k)}$ in the basis U : $p^{(k)} = \sum p_j^{(k)} u_j$, we have that $p_j^{(k)} = \alpha_j - \delta_j^{(k)}$ with $|\delta_j^{(k)}| = \mathcal{O}(a_{k+1} - a_k)$, consequently $p^{(k)} + p(a_k) \max_j (|\delta_j^{(k)}|/\alpha_j)$ has positive coefficients in the basis U . Applying the same argument for every k , we obtain that $p + C_i \in \mathcal{P}(U, \mathbf{a}_i)$ for

$$C_i = \max_{k: a_k \in \mathbf{a}_i} (p(a_k) \max_j (|\delta_j^{(k)}|/\alpha_j)).$$

Finally, as $|\delta_j^{(k)}| = \mathcal{O}(a_{k+1} - a_k)$ and p is bounded, $C_i \rightarrow 0$ as $\|\mathbf{a}_i\| \rightarrow 0$.

The same argument can be used to prove that for every spline, positive over $[0, 1]$, with knot point sequence $\mathbf{a}$, and for every sequence $\{\mathbf{a}_i\}$ consisting of subdivisions of $\mathbf{a}$ satisfying $\lim \|\mathbf{a}_i\| = 0$, there exist nonnegative constants C_i such that $s + C_i \in \mathcal{P}(U, \mathbf{a}_i)$ for every i , and $\lim C_i = 0$.

Consequently, $\bigcup_i \mathcal{P}(U, \mathbf{a}_i)$ is a dense subset of nonnegative splines of degree n .

Finally, let us consider an arbitrary nonnegative function $f \in C([a, b])$. By the approximation theory of splines (see for example (Schumaker, 1981, Theorem 6.27)), for every n there exists a constant $M_n > 0$, depending only on n , but not on f , such that for every knot point sequence $\mathbf{a}$ there exists a (not sign-constrained) spline $s \in \mathcal{S}(n, \mathbf{a})$ satisfying

$$\|f - s\|_\infty \leq M_n \omega_n(f, \|\mathbf{a}\|), \quad (7)$$

where ω_n is the n th modulus of smoothness of f in $L^\infty([a, b])$, satisfying $\lim_{t \searrow 0} \omega_n(f, t) = 0$ for every $n \geq 1$ provided that f is continuous on $[a, b]$. Let ε denote the right-hand side of (7). Because f is nonnegative, the spline $s' = s + \varepsilon$ is nonnegative, and it satisfies

$$\|f - s'\|_\infty \leq 2M_n \omega_n(f, \|\mathbf{a}\|).$$

Hence, there are nonnegative splines $s_i \in \mathcal{P} \cap \mathcal{S}(n, \mathbf{a}_i)$ satisfying $\lim \|f - s_i\|_\infty = 0$.

We already saw that if $\lim \|\mathbf{a}_i\| = 0$, and $\{\mathbf{a}_i\}$ is asymptotically nested, then every nonnegative spline with knot point sequence $\mathbf{a}_i$ can be approximated with arbitrarily small positive error by some spline in $\mathcal{S}(U, \mathbf{a}_j)$ with a sufficiently high j . By the above argument the same holds for f .