MULTIVARIATE ARRIVAL RATE ESTIMATION BY SUM-OF-SQUARES POLYNOMIAL SPLINES AND DECOMPOSITION

DÁVID PAPP AND FARID ALIZADEH

Abstract. A novel method for the arrival rate estimation of multi-dimensional non-homogeneous Poisson processes is presented. The method provides a smooth, piecewise polynomial spline estimator of any prescribed order of differentiability. At the heart of the algorithm is a maximum-likelihood optimization model with sum of squares polynomial constraints, which characterize a proper subset of smooth arrival rate functions. It is investigated when the underlying spline space is dense in the space of continuous arrival rate functions. The method is easily parallelized, exploiting the sparsity of the neighborhood structure of the underlying splines, using an augmented Lagrangian decomposition approach. As a numerical illustration with real-world data, the (spatio-temporal) arrival rate of accidents on the New Jersey Turnpike is estimated.

1. Introduction

The non-homogeneous Poisson process is one of the most frequently used models of stochastic systems in which the rate or intensity of "arrivals" or "events" (often understood in an abstract sense) varies with parameters such as time or location. Consequently, a number of different approaches have already been proposed to the estimation of the arrival rate of such processes, primarily in the univariate case. In studies motivated by a particular application area, where ample prior knowledge is available on the behavior of the system, a parametric model such as [13] or a Bayesian approach such as [10] and [28] may be the most appropriate. General-purpose nonparametric approaches include kernel methods [14] and wavelet-based solutions [12].

Spline estimation approaches have also been successful in univariate arrival rate estimation: [15] describes a method to estimate a linear arrival rate that can be used as a building block for piecewise linear estimation, whereas optimization models involving nonnegative univariate splines of higher degree have been used recently in [1].

Formulating and solving the arrival rate estimation problem as an optimization problem has obvious advantages compared to the computationally less expensive kernel methods and closed form formulae. This approach not only guarantees the optimality of the estimator in a well-defined sense, but it also gives rise to more flexible models, where for instance arbitrary linear constraints, such as the periodicity of the estimator in one or more variables can be added to the model at no additional cost. Furthermore, different objectives, such as maximum likelihood or maximum penalized likelihood, can be treated in a unified manner.

This paper presents a novel optimization-based numerical method for the smooth estimation of arrival rates of non-homogeneous, multi-dimensional Poisson processes from inexact arrivals, using polynomial splines. The approach combines efficient methods of convex optimization with linear and semidefinite (more specifically, sum of squares) constraints and decomposition methods of convex optimization. Its main advantages include the following: (1) it provides a smooth (not piecewise linear) estimator, (2) it exploits the sparsity of the neighborhood structure of the spline, leading to a very efficient method, (3) it is fully parallelizable, which further improves its scalability,

---

Date: May 9, 2011.

The authors are thankful to Andrzej Ruszczyński for his suggestions regarding the decomposition method, and Jonathan Eckstein for a number of valuable comments on an earlier version of this paper.

and (4) it can easily be modified to incorporate additional constraints, including periodicity in one or several variables, and bound constraints on the arrival rate.

While in this paper we concentrate only on arrival rate estimation, many of the ideas presented are applicable to other function approximation and estimation problems, too. For further applications of similar techniques in monotone and concave regression, density estimation, and binary classification, see the thesis of the first author, [19].

The structure of this paper is as follows. In the remainder of this section previous approaches to non-homogeneous Poisson arrival rate estimation are briefly reviewed. The mathematical formulation of the problem is introduced in the next section, where the objective function of our optimization model is also derived, along with an observation of possibly independent interest: the expected number of arrivals of the maximum likelihood estimator (under very general conditions) agrees with the observed number of arrivals. In Section 3 we turn our attention to multivariate nonnegative splines. This functional cone is NP-hard to optimize over, hence, we consider two families of inner approximations, in which maximum likelihood estimators can be found in polynomial time. The first approximation leads to optimization models with semidefinite constraints via "weighted-sum-of-squares" polynomials, the second one uses polynomials with nonnegative coefficients in a nonnegative basis, and results in a linearly constrained optimization model. We prove that (similarly to the cone of nonnegative splines) these spline cones are dense in the cone of nonnegative continuous functions, hence good approximation of the true arrival rate is indeed possible. In Section 4 we adapt a decomposition algorithm of Ruszczyński [24] designed for sparse convex optimization problems, which particularly favors spline estimation problems with their sparse neighborhood structure. Combining the above inner approximation ideas with this decomposition algorithm leads to an efficient, parallel algorithm for the maximum likelihood arrival rate estimation problem. As an illustrative real-world example is provided, we estimate the arrival rate of accidents on the New Jersey Turnpike, as a function of milepost and time, assuming weekly periodicity.

2. Maximum likelihood estimation with inexact arrivals

We consider the following problem. We observe (inexact) arrivals $\mathbf{x}_1, \dots, \mathbf{x}_N \in \Delta \subseteq [0,1]^d$ assumed to have been generated by a non-homogeneous Poisson process with arrival rate $\lambda\colon \Delta \to \mathbb{R}_+$. We assume that λ is deterministic, continuous, and we seek a sufficiently differentiable estimator $\hat{\lambda}$ for it, hence we use piecewise polynomial splines of sufficiently high order of differentiability to approximate the arrival rate. Furthermore, we may require λ to be periodic in any or all of its variables.

In most applications the arrivals are inexact (rounded). With sufficiently many arrivals, this means that a coordinate of multiple arrivals may appear to coincide, invalidating the Poisson assumption. Hence, the effect of rounding cannot be neglected, but a model for rounded arrival times is necessary. Such an approach is outlined next. The possible range of arrivals Δ is divided into small regions Δ_i $(i \in \mathcal{I})$, in which arrivals are rounded to the same point. The data are the number of observed arrivals within each of these regions. Let the number of arrivals recorded in region i be n_i. (Naturally, $\Delta = \bigcup_{i\in\mathcal{I}} \Delta_i$ and $N = \sum_{i\in\mathcal{I}} n_i$.) Then in the non-homogeneous Poisson model the likelihood associated with the arrival rate λ is

$$\begin{aligned} L_{\mathbf{n}}(\lambda) &= \Pr(\text{number of arrivals} = N \,|\, \lambda) \cdot \Pr(\text{distribution of exactly } N \text{ arrivals} = \mathbf{n} \,|\, \lambda) \\ &= \frac{I^N}{N!}e^{-I} \cdot \frac{N!}{\prod_{i\in\mathcal{I}} n_i!} \prod_{i\in\mathcal{I}} \left(\frac{\int_{\Delta_i}\lambda}{I}\right)^{n_i} = \frac{e^{-I}}{\prod_{i\in\mathcal{I}} n_i!} \prod_{i\in\mathcal{I}} \left(\int_{\Delta_i}\lambda\right)^{n_i}, \quad \text{where } I = \int_\Delta \lambda. \end{aligned}$$

Hence, maximizing the likelihood function is equivalent to maximizing

$$f(\lambda) \stackrel{\text{def}}{=} \ln\left(\left(\prod_{i\in\mathcal{I}} n_i!\right) L_{\mathbf{n}}(\lambda)\right) = -\int_\Delta \lambda + \sum_{i\in\mathcal{I}} n_i \ln \int_{\Delta_i} \lambda. \tag{1}$$

This is the objective function of our optimization model.

Some optimization methods may benefit from a further simplification which is made possible by our next observation.

Lemma 1. *Let $\mathcal{K}$ be a cone of nonnegative functions over Δ whose restrictions to each Δ_i are integrable. Define $f\colon \mathcal{K} \to \mathbb{R}$ as in (1), and assume that there exists a $\lambda_0 \in \mathcal{K}$ satisfying $\int_\Delta \lambda_0 > 0$. Then every function $\hat{\lambda} \in \arg\min_{\lambda\in\mathcal{K}} f(\lambda)$ satisfies $\int_\Delta \hat{\lambda} = N$. Thus, the expected number of arrivals corresponding to the maximum likelihood estimator from $\mathcal{K}$ equals the observed number of arrivals.*

Proof. Suppose $\hat{\lambda}$ is an optimal solution (implying $\int_\Delta \hat{\lambda} > 0$), and consider feasible solutions $c\hat{\lambda}$ with $c > 0$. We have $f(c\hat{\lambda}) = -c\int_\Delta \hat{\lambda} + N \ln c + \sum_{i\in\mathcal{I}} n_i \ln \int_{\Delta_i} \lambda$, and by assumption $\frac{\mathrm{d}}{\mathrm{d}c} f(c\hat{\lambda})\big|_{c=1} = 0$. The last equation gives $\int_\Delta \hat{\lambda} = N$, in which case $c = 1$ indeed maximizes $f(c\hat{\lambda})$. □

Lemma 1 allows us to remove the first term of the objective function f in (1), provided we add the equation $\int_\Delta \lambda = N$ to our constraints. This, along with the constraint $\lambda \geq 0$ over Δ, renders the set of feasible solutions bounded. In this paper we concentrate on splines, that is, the cone $\mathcal{K}$ in Lemma 1 will be a subset of piecewise polynomial functions nonnegative over Δ.

3. Optimization over multivariate nonnegative splines

Splines are smooth piecewise polynomial functions. They are numerically very easy to handle, and they can approximate continuous functions very well even if the pieces are of small degree. Hence, they are used extensively throughout the estimation and approximation literature. In this paper we consider multivariate splines built from splines with simple knots:

Definition 2. A function $f\colon [a,b] \in R$ is a *univariate polynomial spline of degree m with simple knots* if there exist *knot points* $a_0, \dots, a_\ell$ satisfying $a = a_0 < \cdots < a_\ell = b$ such that f is a polynomial of degree m over each interval $[a_i, a_{i+1}]$, and f has continuous derivatives up to order $m-1$ over (a,b).

For every $d > 2$, a linear space of d-dimensional splines can be constructed by taking the tensor product of d (possibly different) univariate spline spaces. Such splines are called *tensor product splines*. These splines are piecewise d-variate polynomials, the domains of the pieces are axis-parallel rectangular boxes. In this paper all multivariate splines considered are tensor product splines. For an introduction to these basic concepts, as well as their applications in function approximation, see [26].

The log-likelihood function f in (1) is a concave function with easily computable derivatives, whose maximization over reasonably "well-behaved" closed convex sets is straightforward via a number of different convex optimization methods. However, the set of multivariate polynomials nonnegative over a given domain, even though it is convex, is not an easy set to optimize over – in fact, as it is rather well known, even the *recognition* of nonnegative polynomials is difficult.

Proposition 3 ([4])**.** *Deciding whether a k-variate polynomial is nonnegative over $[0,1]^d$ (equivalently, minimizing a d-variate polynomial over the unit cube) is NP-hard, even for multilinear polynomials of degree two.*

Similar statements hold for polynomials nonnegative over other polyhedral sets (including simplexes), as well as for everywhere nonnegative polynomials. This shows that optimization over piecewise nonnegative polynomials is difficult even when the domain of each piece is a rectilinear box, and the degrees of the polynomial pieces are low. To overcome this difficulty, we shall consider inner approximations of the set of nonnegative polynomials: *weighted-sum-of-squares polynomials*, over which optimization of (1) is easy.

3.1. **Weighted-sum-of-squares polynomials.** We say that a polynomial is a *sum-of-squares* (or SOS) polynomial, if it is expressible as a sum of perfect squares. Obviously, d-variate SOS polynomials are nonnegative over the entire $\mathbb{R}^d$. For a fixed semi-algebraic set $\Delta = \{\mathbf{x} \,|\, w_i(\mathbf{x}) \geq 0,\ i = 1, \dots, m\}$, where $w_1, \dots, w_m$ are polynomials, a sufficient (but not necessary!) condition for a polynomial p to be nonnegative over Δ is for it to be expressible as

$$p(\mathbf{x}) = \sum_{I \subseteq \{1,\dots,m\}} \left(\prod_{i \in I} w_i(\mathbf{x}) \right) s_I(\mathbf{x}), \tag{2}$$

or simply as

$$p(\mathbf{x}) = \sum_{i=1}^{m} w_i(\mathbf{x}) s_i(\mathbf{x}), \tag{3}$$

where the polynomials s_i and s_I are SOS polynomials. Generally, polynomials expressible in the form $\sum_{i \in I} w_i s_i$, (I finite), where w_i are fixed polynomials ("weights") and s_i are SOS polynomials, are called *weighted-sum-of-squares polynomials*, or WSOS polynomials for short.

An implication of a theorem of Nesterov [16] is that WSOS polynomials of a given degree admit a good characterization, also suitable for optimization, as long as the set of weights $\{w_i, i \in I\}$ is finite.

Theorem 4 ([16])**.** *Let $W = \{w_i \,|\, i \in I\}$ be a finite set of polynomials nonnegative over $\Delta \subset \mathbb{R}^d$, and consider finite-dimensional linear spaces of polynomials $V_i, i \in I$. Then the set of WSOS polynomials*

$$\Sigma = \left\{ \sum_{i \in I} w_i \sum_j p_{i,j}^2 \,\middle|\, p_{i,j} \in V_i \right\}$$

is a closed convex cone. Moreover, it is representable as the Minkowski sum of $|I|$ linear images of the cone of positive semidefinite matrices of order $\max_i(\dim(V_i))$.

Therefore, the maximization of the log-likelihood function f in (1) over sets defined by linear equations and constraints of the form $A_k(\lambda) \in \Sigma_k$, where each A_k is a linear operator and each Σ_k is a WSOS cone satisfying the conditions of Theorem 4 is a *semidefinite programming problem* [29] with a convex objective function – a tractable optimization problem [22].

The approximability of nonnegative polynomials over semi-algebraic sets by WSOS polynomials is a well studied problem [25, 20, 27, 17]. Unfortunately, to achieve good approximation of p, the SOS polynomials s_I and s_i in (2) and (3) need to have considerably higher degree than the degree of p itself, which is not practical for the large-scale optimization models of our concern. Since we are only interested in the approximation power of the piecewise polynomial spline estimator, we shall choose a different approach, and investigate the approximation power of *piecewise WSOS polynomial splines*, as the subdivision of the domain is refined. In what follows, to keep the discussion simple, we assume $\Delta = [0,1]^d$, and confine our discussion to tensor product splines, defined above as the tensor product of (linear) spaces of univariate splines with free knots.

3.2. **Piecewise weighted-sum-of-squares splines.**

3.2.1. *Some terminology and notation.* Let us assume that Δ is the d-dimensional interval $[0,1]^d$, and consider splines over a rectilinear subdivision of Δ. Such a subdivision can be given by a list of vectors (of possibly different dimensions) $\mathbf{A} = (a_{i,j}), i = 1, \dots d, j = 1, \dots \ell_i$ such that the knot point sequence $(a_{i,1}, \dots, a_{i,\ell_i})$ defines the subdivision of Δ along the ith axis. Then each region of the subdivision is affinely similar to Δ, and we can represent a spline by the coefficients of its polynomial pieces scaled to Δ.

Formally, for each dimension $i = 1, \dots, d$ we fix a basis $\{u_0^{(i)}, \dots, u_{m_i}^{(i)}\}$ of polynomials of degree m_i, with domain $[0, 1]$. The spline s is then given piecewise; for each multi-index $\mathbf{j} = (j_1, \dots, j_d)$, s over the $\mathbf{j}$-th piece is given by

$$s(\mathbf{x}) = q^{(\mathbf{j})}(\mathbf{x}) \quad \forall\, \mathbf{x} = [a_{1,j_1}, a_{1,j_1+1}] \times \cdots \times [a_{d,j_d}, a_{1,j_d+1}].$$

Each polynomial piece $q^{(\mathbf{j})}$ is represented by the coefficients $p_{\mathbf{k}}^{(\mathbf{j})}$, $\mathbf{k} \in \{0, \dots, m_1\} \times \cdots \times \{0, \dots, m_k\}$, of its affinely scaled counterpart $p^{(\mathbf{j})} \colon [0, 1]^d \to \mathbb{R}$ satisfying

$$q^{(\mathbf{j})}(\mathbf{x}) = \sum_{\mathbf{k}} p_{\mathbf{k}}^{(\mathbf{j})} \prod_{i=1}^{d} u_{k_i}^{(i)} \left(\frac{x_i - a_{i,j_i}}{a_{i,j_i+1} - a_{i,j_i}} \right). \tag{4}$$

Note that each $p^{(\mathbf{j})}$ has the same domain, $[0, 1]^d$. It is clear that $s(\mathbf{x}) \geq 0$ for every $\mathbf{x} \in \Delta$ if and only if $p^{(\mathbf{j})}(\mathbf{x}) \geq 0$ for every $\mathbf{j}$ and $\mathbf{x} \in [0, 1]^d$. We refer to this representation of s by the coefficients $p_{\mathbf{k}}^{(\mathbf{j})}$ as the *scaled representation* of s.

3.2.2. *The approximation power of piecewise weighted-sum-of-squares splines.* The nonnegativity of a spline s over Δ reduces to the nonnegativity of each polynomial $p^{(\mathbf{j})}$ over Δ, and our goal now is to identify proper subsets of polynomials nonnegative over Δ that give rise to piecewise polynomial splines with good approximation power. First, we need to introduce some more notation.

Let the subdivision of Δ be defined by a list of vectors $\mathbf{A}$, as above, and the *mesh size* of such a subdivision be defined as $\|\mathbf{A}\| = \max_{i,j}(a_{i,j+1} - a_{i,j})$. We say that a sequence of subdivisions $\mathbf{A}_1, \mathbf{A}_2, \dots$ is *nested* if each subdivision in the sequence refines the previous ones, that is, the knot points (along each axis) of each subdivision in the sequence contain the knot points on the respective axes of all previous subdivisions. A sequence of subdivisions is *asymptotically nested* if each of its elements is included in an infinite nested subsequence. For example, every sequence of uniform subdivisions with an increasing number of knot points along each axis is an asymptotically nested sequence.

Let us denote by Σ a fixed cone of WSOS polynomials with weights nonnegative over Δ. Finally, let $\mathcal{P}(\Sigma, \mathbf{A})$ denote the set of piecewise WSOS polynomial splines over the subdivision $\mathbf{A}$ whose pieces (in their scaled representation) all belong to Σ. We have the following theorem.

Theorem 5. *Assume that $1 \in \operatorname{int} \Sigma$, where 1 denotes the constant one polynomial. Furthermore, let $\mathbf{A}_1, \mathbf{A}_2, \dots$ be an asymptotically nested sequence of subdivisions of $\Delta = [0, 1]^d$ with mesh sizes approaching zero. Then the set $\bigcup_i \mathcal{P}(\Sigma, \mathbf{A}_i)$ is a dense subcone of the cone of nonnegative continuous functions over Δ.*

We do not prove this theorem directly; instead, we shall prove a stronger assertion using polyhedral cones below.

A special case of the approach discussed so far is the following. A sufficient (but obviously not necessary) condition for a polynomial to be nonnegative over Δ is for it to have nonnegative coefficients in a basis $U = \{u_0, \dots, u_m\}$ that consists of polynomials nonnegative over Δ, that is, for it to belong to $\operatorname{cone}(U)$ for a nonnegative basis U. With the notation of Theorem 4, $\operatorname{cone}(U)$ can be considered a WSOS cone Σ with weights in U, whose spaces V_i are the one-dimensional linear spaces consisting only of constant polynomials. Similarly to the piecewise WSOS polynomial splines above, we can define a *piecewise U-spline* as a piecewise polynomial spline whose pieces (in the scaled representation) belong to $\operatorname{cone}(U)$. The set of piecewise U-splines with subdivision $\mathbf{A}$ is denoted by $\mathcal{P}(U, \mathbf{A})$.

Theorem 6. *Consider a basis $U = \{u_0, \dots, u_{\mathbf{m}}\}$ of d-variate polynomials of multi-degree $\mathbf{m} = (m_1, \dots, m_d)$ such that each u_i is nonnegative over $\Delta = [0, 1]^d$, and assume that $1 \in \operatorname{int}\operatorname{cone}(U)$, where 1 denotes the constant one polynomial. Furthermore, let $\mathbf{A}_1, \mathbf{A}_2, \dots$ be an asymptotically* *nested sequence of subdivisions with mesh sizes approaching zero. Then the set* $\bigcup_i \mathcal{P}(U, \mathbf{A}_i)$ *is a dense subcone of the cone of nonnegative functions over* Δ.

Proof. First we show that for every polynomial p of degree $\mathbf{m}$, 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 a piece in the subdivision from the knot point sequence $\mathbf{A}_i$:

$$[a_{1,j_1}, a_{1,j_1+1}] \times \cdots \times [a_{d,j_d}, a_{d,j_d+1}].$$

The polynomial p can be represented as a piecewise polynomial spline of degree $\mathbf{m}$ with knot point sequence $\mathbf{A}_i$; its scaled representation is

$$p^{(\mathbf{j})}(x_1, \ldots, x_d) = p((a_{1,j_1+1} - a_{1,j_1})x_1 + a_{1,j_1}, \ldots, (a_{d,j_d+1} - a_{d,j_d})x_d + a_{d,j_d}).$$

Collecting terms in the standard basis, we see that every coefficient in the above expression is of order $\mathcal{O}(\|\mathbf{A}_i\|)$, except for the constant term, which is $p(a_{1,j_1}, \ldots, a_{d,j_d})$. By assumption, this constant term is positive, because p is strictly positive on $[0,1]$. By the assumption on U, $\sum_{k=0}^m \alpha_k u_k \equiv p(a_{1,j_1}, \ldots, a_{d,j_d})$ for some positive $\alpha_0, \ldots, \alpha_m$. Now, if we express $p^{(\mathbf{j})}$ in the basis U: $p^{(\mathbf{j})} = \sum p_k^{(\mathbf{j})} u_k$, we have that $p_k^{(\mathbf{j})} = \alpha_k - \delta_k^{(\mathbf{j})}$ with $|\delta_k^{(\mathbf{j})}| = \mathcal{O}(\|\mathbf{A}\|)$, consequently $p^{(\mathbf{j})} + p(a_{1,j_1}, \ldots, a_{d,j_d}) \max_k(|\delta_k^{(\mathbf{j})}|/\alpha_k)$ has positive coefficients in the basis U. Applying the same argument for every $\mathbf{j}$, we obtain that $p + C_i \in \mathcal{P}(U, \mathbf{A}_i)$ for

$$C_i = \max_{\mathbf{j}} \left(p(a_{1,j_1}, \ldots, a_{d,j_d}) \max_k (|\delta_k^{(\mathbf{j})}|/\alpha_k) \right).$$

Finally, as $|\delta_k^{(\mathbf{j})}| = \mathcal{O}(\|\mathbf{A}\|)$ and p is bounded, $C_i \to 0$ as $\|\mathbf{A}_i\| \to 0$.

The same argument also proves that for every strictly positive spline 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 multi-degree $\mathbf{m}$.

Now our assertion follows from the fact that tensor product splines (of every given order of differentiability) in $\Delta = [0,1]^d$ are dense in the space of continuous functions over Δ; see [26, Theorem 13.21]. □

Example 7 (Bernstein polynomial basis). Let $B = \{b_0, \ldots, b_n\}$ with $b_i(t) = \binom{n}{i} t^i (1-t)^{n-i}$. The polynomial b_i is the ith *Bernstein polynomial* of degree n. Multivariate Bernstein polynomials are defined as products of univariate ones. Observe that $\sum_{i=0}^n b_i(t) = 1$ for every t, hence $1 \in \operatorname{int} \operatorname{cone}(B)$. The same holds for the cone of multivariate Bernstein polynomials. Thus, the multivariate Bernstein polynomial basis satisfies the conditions of Theorem 6.

Note for every WSOS cone Σ satisfying the conditions of Theorem 5 one can find a basis U such that $\operatorname{cone}(U) \subseteq \Sigma$ and $\operatorname{cone}(U)$ satisfies the conditions of Theorem 6.

We also remark that the conditions $1 \in \operatorname{int} \Sigma$ and $1 \in \operatorname{int} \operatorname{cone}(U)$ are sufficient *and necessary* for the desired conclusion. For example, the cone of polynomials with nonnegative coefficients in the standard monomial basis is a WSOS cone with weights nonnegative over $[0,1]^d$. It does not satisfy the condition, as the constant 1 is on the boundary of this polynomial cone. Incidentally, the corresponding cone of splines consists of functions that are monotone nondecreasing in every variable, hence it is not dense in the cone of nonnegative continuous functions over $[0,1]^d$.

3.3. **B-splines and Bernstein polynomials.** We must mention that, unlike the use of piecewise WSOS splines, the above polyhedral approach to the arrival rate estimation problem (and other shape-constrained estimation problems) is not completely new in the statistical and function estimation literature, even though the most common approach is somewhat different: typically a nonnegative basis of spline functions is chosen, such as *B-splines* [3], and the optimization is carried out over the nonnegative linear combination of this basis.

Similar approaches have primarily been used in univariate shape-constrained estimation problems, where the complicating constraint is not the nonnegativity, but the monotonicity or convexity of the estimator. For instance, in [11], convex regression with residual sum of squares loss function is approached by searching over piecewise linear functions with fixed knot points; [5] employs a similar approach for monotone regression, using piecewise constant functions with knot points aligned with the data points. In [21], monotone regression is translated to an optimization problem over the cone generated by *I-splines*, which are integral functions of B-splines, forming a basis of monotone non-decreasing functions. In [9], quadratic B-splines are used for monotone regression, while [7] employs cubic splines in convex/concave regression. It is clear that nonnegativity, monotonicity and convexity in these spline spaces are expressible by finitely many linear inequalities on the spline coefficients: it is sufficient to require that the piecewise linear first (resp., second) derivative of the spline estimator be nonnegative at its knot points. The same does not hold if the degree of the spline is increased by one.

With the exception of Brunk's above result [5], no theoretical justification is provided that the optimal estimator belongs to the polyhedral space considered, or that the optimal estimator can be approximated by a spline chosen from the polyhedral cone of splines considered, for instance in the sense of Theorem 6. To our knowledge, no analogous results to Theorems 5 and 6 are known for multivariate estimation problems.

The perhaps most popular splines used in the estimation, approximation, and engineering literature are B-splines. Since the B-spline basis functions are nonnegative, the polyhedral approach analogous to the previous section is to optimize over splines whose coefficients in the B-spline basis are nonnegative. However, as the following theorem shows, cones generated by B-splines of a given degree and given subdivision form a proper subcone of splines with nonnegative coefficients in their scaled representation (4) with the Bernstein polynomial basis as U. Hence, we do not consider B-splines any further.

Theorem 8. *For every given degree and subdivision of domain, the cone generated by B-splines is a subset of the cone of piecewise multivariate Bernstein polynomial splines of the same degree and subdivision. For degrees greater than one, this containment is strict.*

Proof. Since both multivariate B-splines and piecewise U-splines are defined as tensor product splines, it suffices to prove the assertion in the univariate case, and the multivariate statement immediately follows.

Let n be the degree of the splines, and $\mathbf{a}$ be the given knot point sequence. B-splines are linear combinations of the B-spline basis functions, whose supports are intervals between non-consecutive knot points. Recall the well-known Cox–de Boor recursion formula [3]: if $f_{n,i}$, $i = 0, \dots, m$ denote the B-spline basis functions of degree n with knot points $0, \dots, m-1$, then for every $i = 0, \dots, m$,

$$f_{0,i}(x) = \begin{cases} 1 & i \le x < i+1 \\ 0 & \text{otherwise} \end{cases}, \quad \text{and} \tag{5a}$$

$$f_{n,i}(x) = \frac{x-i}{n} f_{n-1,i}(x) + \frac{(i+n+1)-x}{n} f_{n-1,i+1}(x) \qquad n \ge 1. \tag{5b}$$

To prove our claim, we need a characterization of B-splines of degree n as splines of the form $\mathcal{P}(U_n, \mathbf{a})$ with some basis U_n of degree n polynomials; then we compare U_n to the Bernstein polynomial basis. First, we obtain a recursion on the appropriate U_n from (5), and then we establish a recursive formula on the basis transformation matrix between U_n and the degree n Bernstein polynomial basis. We complete the proof by showing that each entry of that matrix is nonnegative.

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, and to simplify the calculations we consider scaled Bernstein polynomials with leading coefficients ± 1, that is, for the purposes of this proof, the ith Bernstein polynomial of degree n is $b_{n,i}(x) = x^i(1-x)^{n-i}$.

Note that the basis functions $f_{n,0}, \dots, f_{n,m}$ are shifted copies of each other: $f_{n,i}(x) = f_{n,0}(x-i)$. Hence the B-spline segment between knot points 0 and 1 is a nonnegative linear combination of the functions $f_{n,0}(x+i)$, $i = 0, \dots, m$, or more precisely, of the functions

$$g_{n,i}(x) \stackrel{\text{def}}{=} f_{n,0}(x+i) \quad \forall\, i = 0, \dots, n,$$

as $f_{n,0}(x+i)$ is identically zero for $i > n$. Hence, $U_n = \{g_{n,0}, \dots, g_{n,n}\}$. Rewriting the recursion (5b) in terms of $g_{n,i}$ we obtain

$$\begin{aligned} g_{n,i}(x) &= f_{n,0}(x+i) \\ &= \frac{x+i}{n} f_{n-1,0}(x+i) + \frac{(n+1)-(x+i)}{n} f_{n-1,1}(x+i) \\ &= \frac{x+i}{n} f_{n-1,0}(x+i) + \frac{(n+1)-(x+i)}{n} f_{n-1,0}(x+(i-1)) \\ &= \frac{x+i}{n} g_{n-1,i}(x) + \frac{(n+1)-(x+i)}{n} g_{n-1,i-1}(x), \quad \forall\, i = 0, \dots, n, \end{aligned} \tag{6}$$

with $g_{0,0} = 1$.

Using this recursion we can also find a recursive formula for the matrix of the basis transformation from the polynomials $g_{n,i}$ to the Bernstein polynomials $b_{n,i}$. Let $A_{n,k,i,j}$ be the coefficient of $b_{n,j}$ in the unique representation of $g_{k,i}$ in the basis $\{b_{n,0}, \dots, b_{n,n}\}$, and $B_{n,k,i,j}$ be the coefficient of $b_{n,j}$ in the unique representation of $x \to x g_{k,i}(x)$ in the same basis. Then we have

$$A_{n+1,n,i,j} = A_{n,n,i,j} + A_{n,n,i,j-1} \tag{7}$$

from the identity $b_{n+1,j-1} + b_{n+1,j} = b_{n,j}$, and

$$B_{n+1,k,i,j} = A_{n,k,i,j-1} \tag{8}$$

from the identity $x b_{n,j-1}(x) = b_{n+1,j}(x)$. (The coefficients $A_{n,k,i,j}$ are defined to be zero whenever the indices are "out of bounds", in particular when $j < 0$.) Using (7) and (8), the recursion (6) on $g_{n,i}$ translates to

$$\begin{aligned} nA_{n,n,i,j} &= B_{n,n-1,i,j} + iA_{n,n-1,i,j} + (n+1-i)A_{n,n-1,i-1,j} - B_{n,n-1,i-1,j} \\ &= A_{n-1,n-1,i,j-1} + iA_{n,n-1,i,j} + (n+1-i)A_{n,n-1,i-1,j} - A_{n-1,n-1,i-1,j-1} \\ &= A_{n-1,n-1,i,j-1} + iA_{n-1,n-1,i,j} + iA_{n-1,n-1,i,j-1} + (n+1-i)A_{n-1,n-1,i-1,j} + \\ &\quad + (n+1-i)A_{n-1,n-1,i-1,j-1} + A_{n-1,n-1,i-1,j-1} \\ &= iA_{n-1,n-1,i,j} + (n+1-i)A_{n-1,n-1,i-1,j} + \\ &\quad + (i+1)A_{n-1,n-1,i,j-1} + (n-i)A_{n-1,n-1,i-1,j-1} \end{aligned}$$

for every $n \geq 1$ and $0 \leq i, j \leq n$. Every coefficient on the right-hand side is nonnegative. Since $A_{0,0,0,0} = 1 \geq 0$, this implies that $A_{n,n,i,j} \geq 0$ for every $0 \leq i, j \leq n$ by induction.

To see that the containment is strict when $n \geq 2$, it is sufficient to exhibit a Bernstein polynomial that is not a nonnegative linear combination of U_n. Since $b_{n,k}(0) = b_{n,k}(1) = 0$ for every $k = 1, \dots, n-1$, but $g_{n,i}(0) + g_{n,i}(1) > 0$ for every $i = 0, \dots, n$, it follows that every $b_{n,k}$ is such a polynomial for $k = 1, \dots, n-1$. □

4. A decomposition method for multivariate spline estimation

The size of the optimization models involving multivariate splines prohibits the solution of models of high dimension or small mesh size. On the other hand, these problems have a very regular and sparse structure that makes them potentially amenable to decomposition methods. In this section we outline an augmented Lagrangian decomposition method with particularly good convergence properties for spline estimation problems. We illustrate it, in the next section, by estimating the (two-dimensional) weekly arrival rate of car accidents on the New Jersey Turnpike.

There is a vast literature on decomposition methods for both linear and nonlinear convex optimization problems, an area initiated by Dantzig and Wolfe [6] and Benders [2]. Several existing methods can be adapted to our problem. The method we propose is a considerably simplified version of the augmented Lagrangian-based method of [24] specifically designed for sparse problems. Since some of the details of our specific estimation problem might obscure the main ideas of the algorithm, we shall discuss the method in a slightly more abstract form than necessary for our purposes.

4.1. **Augmented Lagrangian decomposition for sparse problems.** Let $L \geq 2$, and let $\mathcal{X}_i$ ($i = 1, \dots, L$) be a nonempty compact subset of $\mathbb{R}^{n_i}$. Finally, let $f_i \colon \mathcal{X}_i \to \mathbb{R}$ be convex. With these given, we consider the convex optimization problem

$$\text{minimize} \quad f(\mathbf{x}) \stackrel{\text{def}}{=} \sum_{i=1}^{L} f_i(\mathbf{x}_i) \tag{9a}$$

$$\text{subject to} \quad \sum_{i=1}^{L} \mathbf{A}_i \mathbf{x}_i = \mathbf{b} \tag{9b}$$

$$\mathbf{x}_i \in \mathcal{X}_i \quad i = 1, \dots, L. \tag{9c}$$

As a specific example, we can model arrival rate estimation problems in this framework: the integrals in the objective function (1) can be computed piecewise, and the nonnegativity constraints are also defined piecewise. The nonnegativity constraints are then replaced by constraints that the scaled representation of each piece belongs to a WSOS cone. In the rest of the paper we will refer to these constraints as "the WSOS constraints" for short. The WSOS constraints are translated to semidefinite constraints via Theorem 4 or to linear constraints if the polyhedral approach and Theorem 6 are used.

Thus, we can set L to be the number of pieces, (9c) are the WSOS constraints, and (9b) includes the continuity of the estimator and its derivatives, as well as periodicity constraints, as needed for the problem. See [19] for models of various shape-constrained estimation problems in the same framework.

Alternatively, if the estimator is a polynomial spline over a rectilinear grid, as in Theorems 5 and 6, we can set $L = 2$, since all constraints connect only pieces of two disjoint classes, following a chessboard-like pattern. The same is true for some other regular subdivisions as well, including a regular simplicial subdivision. This property can be exploited by some methods. We shall focus on a direct consequence of this observation: that in our arrival rate estimation model *all of the coupling constraints in* (9b) *involve variables corresponding to only two different* $\mathcal{X}_i$.

The condition that the sets $\mathcal{X}_i$ be bounded is a rather technical condition, as one can always find reasonable bounds on the spline coefficients based on the data. In our arrival rate estimation model this condition is always satisfied, since the optimal estimator is a piecewise nonnegative polynomial function whose integral is given by Lemma 1.

The method proposed in [24] associates multipliers $\boldsymbol{\pi}$ to the linear coupling constraints, and considers a separable approximation Λ_{apx} of the augmented Lagrangian of (9),

$$\Lambda(\mathbf{x}, \boldsymbol{\pi}) = f(\mathbf{x}) + \langle \boldsymbol{\pi}, \mathbf{b} - \mathbf{A}\mathbf{x} \rangle + \frac{\rho}{2} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|^2,$$

in which the bilinear terms in the quadratic penalties are linearized around a point $\tilde{\mathbf{x}} \in \mathbb{R}^{\sum_{i=1}^{L} n_i}$:

$$\Lambda_{\text{apx}}(\mathbf{x}, \tilde{\mathbf{x}}, \boldsymbol{\pi}) \stackrel{\text{def}}{=} \sum_{i=1}^{L} \Lambda_i(\mathbf{x}_i, \tilde{\mathbf{x}}_i, \boldsymbol{\pi}) \stackrel{\text{def}}{=} \sum_{i=1}^{L} f_i(\mathbf{x}_i) - \langle \mathbf{A}_i^{\mathrm{T}} \boldsymbol{\pi}, \mathbf{x}_i \rangle + \frac{\rho}{2} \Big\| \mathbf{b} - \mathbf{A}_i \mathbf{x}_i - \sum_{j \neq i} \mathbf{A}_j \tilde{\mathbf{x}}_j \Big\|^2. \tag{10}$$

The approach then is to fix a set of multipliers, and find an approximate minimizer of the augmented Lagrangian $\Lambda(\cdot, \boldsymbol{\pi})$ by iteratively minimizing $\Lambda_{\text{apx}}(\cdot, \tilde{\mathbf{x}}, \boldsymbol{\pi})$ and updating $\tilde{\mathbf{x}}$ so as to approximate the optimal solution better. The components Λ_i of the approximate Lagrangian can be optimized separately, and even in parallel (in a Jacobi, rather than a Gauss–Seidel, fashion), which is a very desirable property when infrastructure for massively parallel computations is available. Once an approximately optimal solution to the augmented Lagrangian is found, the multipliers are updated as in the classic multiplier method [23]. The formal definition of the method is given in Algorithm 1, which requires two parameters: the augmented Lagrangian coefficient ρ, and a step size parameter τ.

Algorithm 1: Partially linearized augmented Lagrangian decomposition

parameters: $\rho > 0, \tau > 0$
1 initialize $\boldsymbol{\pi}$ `/* arbitrary initial value */`
2 repeat
3 $\boldsymbol{\pi} \leftarrow \boldsymbol{\pi} + \rho(\mathbf{b} - \mathbf{A}\mathbf{x})$
4   foreach $i = 1, \ldots, L$ do solve $\min_{\mathbf{x}_i \in \mathcal{X}_i} \Lambda_i(\mathbf{x}_i, \tilde{\mathbf{x}}, \boldsymbol{\pi})$ `/* parallel */`
5   if $\mathbf{A}_i\mathbf{x}_i \neq \mathbf{A}_i\tilde{\mathbf{x}}_i$ *for any* $i = 1, \ldots, L$ then
6 $\tilde{\mathbf{x}} \leftarrow \tilde{\mathbf{x}} + \tau(\mathbf{x} - \tilde{\mathbf{x}})$
7   go to step 4
8   end if
9 until $\mathbf{A}\mathbf{x} = \mathbf{b}$
10 return $\mathbf{x}$

It can be shown that for every $\rho > 0$ and small enough $\tau > 0$ Algorithm 1 is convergent; see below. For the inner loop this is a special case of Theorem 1 in [24], and for the outer loop this follows from the convergence of the method of multipliers [23]. Perhaps the most attractive feature of Algorithm 1 is that its rate of convergence depends very highly on the largest number M of variable blocks $\mathbf{x}_i$ linked by a coupling equality constraint, favoring problems with a sparse neighborhood structure. Let us measure the progress of the algorithm by the difference

$$\Delta^{(k)} \stackrel{\text{def}}{=} \Lambda(\tilde{\mathbf{x}}^{(k)}, \boldsymbol{\pi}^{(k)}) - \min_{\mathbf{x} \in \mathcal{X}} \Lambda(\mathbf{x}, \boldsymbol{\pi}^{(k)}),$$

where the superscript k refers to the number of times the outer loop of Algorithm 1 has been executed. The following theorem establishes the rate of convergence under a technical assumption.

Theorem 9. *Assume that there is a $\gamma > 0$ such that for every $\boldsymbol{\pi}$ and every $\mathbf{x} \in \mathcal{X}$,*

$$\Lambda(\mathbf{x}, \boldsymbol{\pi}) - \min_{\mathbf{x} \in \mathcal{X}} \Lambda(\mathbf{x}, \boldsymbol{\pi}) \geq \gamma \operatorname{dist}(\mathbf{x}, \arg\min_{\mathbf{x} \in \mathcal{X}} \Lambda(\mathbf{x}, \boldsymbol{\pi}))^2,$$

and let $\alpha = \max_i \|\mathbf{A}_i\|_2$. Then for every $\rho > 0$ and $0 < \tau < 1/(M-1)$,

$$\Delta^{(k+1)} \leq \left(1 - \frac{\tau(1 - \tau(M-1))}{1 + 2\rho\alpha^2(M-1)^2\gamma^{-1}}\right) \Delta^{(k)}. \tag{11}$$

Proof. This is a special case of Theorem 2 in [24]. □

We remark again that in our arrival rate estimation model each coupling constraint connects only two $\mathbf{x}_i$'s, that is, we have $M = 2$. Thus Algorithm 1 has particularly rapid convergence when applied to our problem.

The first condition of the theorem is the *quadratic growth condition* of [24], and is satisfied by the log-likelihood objective. Theorem 9 also suggests a way to set the parameter τ: minimizing the coefficient on the right-hand side of (11) we obtain that $\tau = (2(M-1))^{-1}$ is the recommended choice, regardless of ρ. The theory does not provide guidance in the selection of ρ.

5. Numerical illustration – NJ Turnpike accidents

We consider the two-dimensional point process of car accidents on the New Jersey Turnpike (NJTP). The two dimensions are time (with an assumed weekly periodicity) and location along the road. It is not entirely clear whether this is indeed a Poisson process, as accidents may change the traffic pattern, which in turn affects the distribution of the accidents. Furthermore, coincidences in the location (which could occur, for example, because of a construction) have likelihood zero in every Poisson model. However, as the accidents are relatively rare (serious accidents that change the traffic pattern for a long period of time are even more so), and major highways are assumed to have no easy-to-hit objects, a Poisson model may be a reasonable approximation. The fact that accidents may occur more frequently close to exits does not contradict the non-homogeneous Poisson model.

The data. We obtained car accident data from the New Jersey Department of Transportation [18]. The raw data contained information on every car accident in 2009 recorded at the accident locations by police officers. The time of the accident is rounded to the nearest minute, but it is not clear whether the recorded time is the approximate time of the accident, the time the police were notified of the accident, or the time the officers attended to the accident. Hence, we consider this as noisy data, despite its apparent precision. The location is given by the Standard Route Identifier of the road segment and an approximate milepost reading (variably rounded, apparently to the nearest 0.05 mile or to the nearest mile).

We removed all entries from the data that corresponded to accidents in roads other than the NJTP segment marked I-95. This is an approximately 78-mile-long segment stretching between two state borders (with Pennsylvania and New York, respectively) with no forks or joins. Date and time were present for each accident, however, the location (milepost) was not always specified. To simplify the analysis, we removed entries with missing milepost information. While we could take these accidents into account directly in the maximum-likelihood approach, such incomplete entries were few, and it is reasonable to assume that accidents whose milepost is missing follow the same time and milepost distribution as the entries with complete information. Hence, we simplify our model, and estimate the arrival rate based only on the entries with complete information, and then we divide the obtained arrival rate with the proportion of entries with complete information to account for the discarded accidents. We also removed all accidents that happened on ramps while entering or leaving the highway, as they are confounding in multiple ways. This left us with 4138 accidents.

Numerical results. In our example $\Delta = [0, T] \times [0, X]$; $T = 1$ week, $X = 77.96$ miles. Considering the format of the data, the regions $\mathcal{I}$ in the objective function can be rectangles no smaller than 1 minute by 0.1 miles, but even considerably larger rectangles are reasonable, given the rounding errors.

Figure 1 shows a biquadratic spline estimator with 28×13 pieces (so each piece corresponds to 6 hours and roughly 6 miles), the regions $\mathcal{I}$ were 1 minute by 1 mile rectangles. The estimator was obtained using the polyhedral model, with the cubic Bernstein polynomial basis, and an AMPL [8] implementation of Algorithm 1, in which the subproblems were solved by the solver KNITRO [30].

A few important features to note: along the time dimension seven daily peaks are clearly distinguishable. Two "hot spots" are also identified: the are with a high number of accidents each day is the vicinity of the Newark airport exit, while the second one, with a large daily variation in the number of accidents, is in the area where the truck lanes and the cars-only lanes merge.

6. Conclusion

We have presented an efficient approach for the spline estimation of non-homogeneous, multi-dimensional Poisson processes from inexact arrivals. The key idea behind the approach was to consider piecewise polynomial splines whose pieces are from a polynomial cone possessing two

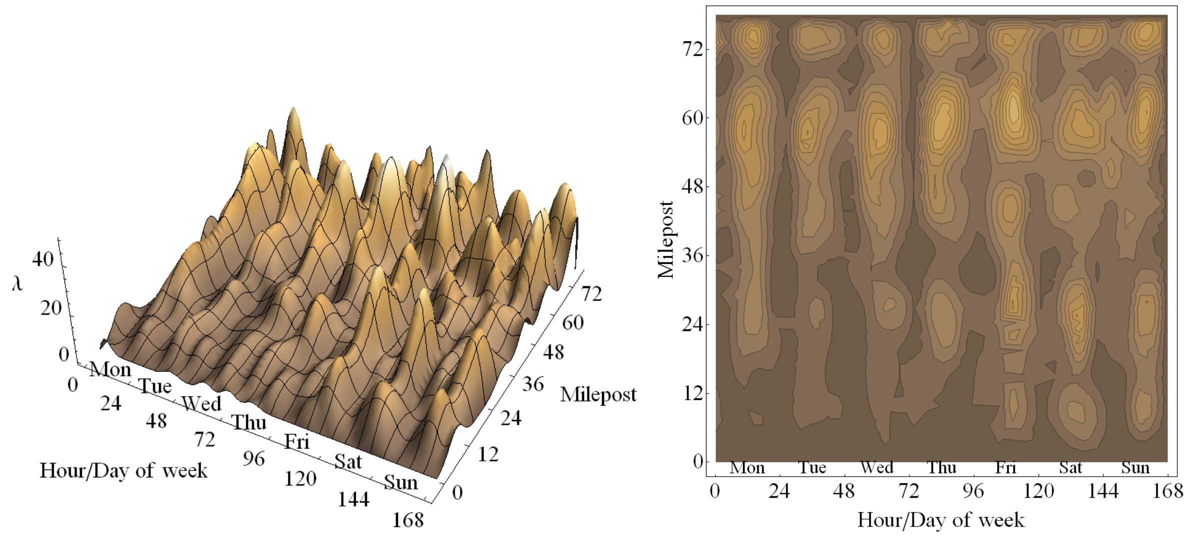

FIGURE 1. A piecewise biquadratic sum-of-squares spline estimator of the NJTP accident rate obtained using Algorithm 1. Left: three-dimensional plot. Right: contour plot. The horizontal axis shows the time (Monday–Sunday), the vertical axis is location.

properties: the cone admits a good characterization suitable for optimization, and the splines built from them can uniformly approximate every continuous function. Two variants of this approach were considered: one that uses weighted-sum-of-squares polynomials and leads to optimization models with semidefinite constraints, and one that uses polynomials with nonnegative coefficients and results in linearly constrained optimization models.

The approach was then combined with a decomposition method whose worst-case running time explicitly depends on the sparsity of the neighborhood structure defined by the constraints coupling subproblems. This neighborhood structure is the sparsest possible for the spline estimation problems of our interest.

We have presented the WSOS (spectrahedral) approach along with its polyhedral special case, and established their theoretical soundness with the density theorems (Theorem 5 and 6). We have shown that they improve a popular method that uses B-splines, and also illustrated their practical feasibility by estimating the arrival rate of accidents on the New Jersey Turnpike, using a basic single-machine configuration. In a parallel computing setting, where the subproblems corresponding to the pieces are solved in parallel, the general WSOS approach is more desirable, since it gives better approximations to nonnegative splines, and the additional time required to solve the semidefinite programming subproblems, as opposed to polyhedral subproblems, is negligible, owing to the relatively small size of these subproblems. The practical difference between the polyhedral and the WSOS approaches will be subject of further computational study.

Several questions and future research directions emerge. While in this paper we concentrate only on arrival rate estimation, many of the ideas presented are also applicable for other estimation problems as well. Many of these ideas appear in the thesis of the first author, but research in this area is far from complete. Another observation worthy of further investigation is the fact that most splines have a bipartite structure: their pieces can be partitioned into two disjoint classes such that only pieces from different classes are adjacent. This makes the arrival estimation model considered in this paper amenable to a number of other iterative decomposition methods, including alternating direction methods. Comparing these approaches and the one presented, especially on a parallel architecture, would be interesting.

References

1. Farid Alizadeh, Jonathan Eckstein, Nilay Noyan, and Gábor Rudolf, *Arrival rate approximation by nonnegative cubic splines*, Operations Research **56** (2008), no. 1, 140–156, doi:10.1287/opre.1070.0443.
2. Jacques F. Benders, *Partitioning procedures for solving mixed-variables programming problems*, Numerische Mathematik **4** (1962), 238–252, doi:10.1007/BF01386316.
3. Carl De Boor, *A practical guide to splines*, revised ed., Applied Mathematical Sciences, vol. 27, Springer, New York, NY, 2001.
4. Endre Boros and Peter L. Hammer, *Pseudo-Boolean optimization*, Discrete Applied Mathematics **123** (2002), no. 1–3, 155–225, doi:10.1016/S0166-218X(01)00341-9.
5. H. D. Brunk, *On the estimation of parameters restricted by inequalities*, Annals of Mathematical Statistics **29** (1958), no. 2, 437–454, doi:10.1214/aoms/1177706621.
6. George B. Dantzig and Philip Wolfe, *Decomposition principle for linear programs*, Operations Research **8** (1960), no. 1, 101–111, doi:10.1287/opre.8.1.101.
7. P. Dierckx, *Algorithm 42. an algorithm for for cubic spline fitting with convexity constraints*, Computing **24** (1980), no. 4, 349–371, doi:10.1007/BF02237820.
8. Robert Fourer, David M. Gay, and Brian W. Kernighan, *AMPL: A modeling language for mathematical programming*, second ed., Brooks/Cole–Thomson Learning, Pacific Grove, CA, 2002.
9. Xuming He and Peide Shi, *Monotone B-spline smoothing*, Journal of the American Statistical Association **93** (1998), no. 442, 643–650, doi:10.2307/2670115.
10. Juha Heikkinen and Elja Arjas, *Non-parametric Bayesian estimation of a spatial Poisson intensity*, Scandinavian Journal of Statistics **25** (1998), 435–450, doi:10.1111/1467-9469.00114.
11. Clifford Hildreth, *Point estimates of ordinates of concave functions*, Journal of the American Statistical Association **49** (1954), no. 267, 598–619, doi:10.2307/2281132, corrigenda doi:10.2307/2281222.
12. Michael E. Kuhl and Prashant S. Bhairgond, *Nonparametric estimation of nonhomogeneous Poisson processes using wavelets*, Proceedings of the 2000 Winter Simulation Conference (Orlando, FL, USA) (J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, eds.), December 2000, pp. 562–571.
13. Michael E. Kuhl and James R. Wilson, *Least squares estimation of nonhomogeneous Poisson processes*, Journal of Statistical Computation and Simulation **67** (2000), 699–712, doi:10.1080/00949650008812036.
14. M. R. Leadbetter and Diane Wold, *On estimation of point process intensities*, Contributions to statistics. Essays in honour of Norman Lloyd Johnson (Pranab Kumar Sen, ed.), North-Holland, Amsterdam, 1983, pp. 299–312.
15. William Massey, Geraldine Parker, and Ward Whitt, *Estimating the parameters of a nonhomogeneous Poisson process with linear rate*, Telecommunication Systems **5** (1996), 361–388, doi:10.1007/BF02112523.
16. Yurii Nesterov, *Squared functional systems and optimization problems*, High Performance Optimization (Hans Frenk, Kees Roos, Tamás Terlaky, and Shuzhong Zhang, eds.), Applied Optimization, vol. 33, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2000, pp. 405–440.
17. Jiawang Nie and Markus Schweighofer, *On the complexity of Putinar's Positivstellensatz*, Journal of Complexity **23** (2007), no. 1, 135–150, doi:10.1016/j.jco.2006.07.002.
18. State of New Jersey Department of Transportation, *Crash records*, `http://www.state.nj.us/transportation/refdata/accident/`, Accessed on October 29, 2010.
19. Dávid Papp, *Optimization models for shape-constrained function estimation problems involving nonnegative polynomials and their restrictions*, Ph.D. thesis, Rutgers University, New Brunswick, NJ, USA, May 2011.
20. Mihai Putinar, *Positive polynomials on compact semi-algebraic sets*, Indiana University Mathematics Journal **42** (1993), 969–984.
21. J. O. Ramsay, *Monotone regression splines in action*, Statistical Science **3** (1988), no. 4, 425–441, doi:10.1214/ss/1177012761.
22. James Renegar, *A mathematical view of interior-point methods in convex optimization*, MPS-SIAM Series on Optimization, no. 03, MPS-SIAM, Phiadelphia, PA, 2001.
23. Ralph Tyrrell Rockafellar, *Augmented Lagrangians and applications of the proximal point algorithm in convex programming*, Mathematics of Operations Research **1** (1976), 97–116, doi:10.1287/moor.1.2.97.
24. Andrzej Ruszczyński, *On convergence of an augmented Lagrangian decomposition method for sparse convex optimization*, Mathematics of Operations Research **20** (1995), 634–656, doi:10.1287/moor.20.3.634.
25. Konrad Schmüdgen, *The K-moment problem for compact semi-algebraic sets*, Mathematische Annalen **289** (1991), 203–206, doi:10.1007/BF01446568.
26. Larry L. Schumaker, *Spline functions: Basic theory*, Wiley-Interscience, 1981.
27. Markus Schweighofer, *On the complexity of Schmüdgen's Positivstellensatz*, Journal of Complexity **20** (2004), no. 4, 529–543, doi:10.1016/j.jco.2004.01.005.

28. Jonathan Weinberg, Lawrence D. Brown, and Jonathan R. Stroud, *Bayesian forecasting of an inhomogeneous Poisson process with applications to call center data*, Journal of the American Statistical Association **102** (2007), 1185–1198, doi:10.1198/016214506000001455.
29. Henry Wolkowicz, Romesh Saigal, and Lieven Vandenberghe (eds.), *Handbook of semidefinite programming: Theory, algorithms, and applications*, International series in operations research & management science, vol. 27, Kluwer, Norwell, MA, 2000.
30. Ziena, *KNITRO documentation*, `http://www.ziena.com/documentation.htm`.