Applied Mathematics and Computation 227 (2014) 359–369

Contents lists available at ScienceDirect

Applied Mathematics and Computation

journal homepage: www.elsevier.com/locate/amc

A geometrical approach to Iterative Isotone Regression

Arnaud Guyader [a,*], Nicolas Jégou [b], Alexander B. Németh [c], Sándor Z. Németh [d]

[a] Université Rennes 2, INRIA and IRMAR, Campus de Villejean, Rennes, France
[b] Université Rennes 2, Campus de Villejean, Rennes, France
[c] Faculty of Mathematics and Computer Science, Babeş Bolyai University, RO-400084 Cluj-Napoca, Romania
[d] School of Mathematics, The University of Birmingham, Birmingham B15 2TT, United Kingdom

ARTICLE INFO

Keywords:
Nonparametric estimation
Isotonic regression
Additive models
Metric projection onto convex cones

ABSTRACT

In the present paper, we propose and analyze a novel method for estimating a univariate regression function of bounded variation. The underpinning idea is to combine two classical tools in nonparametric statistics, namely isotonic regression and the estimation of additive models. A geometrical interpretation enables us to link this iterative method with Von Neumann's algorithm. Moreover, making a connection with the general property of isotonicity of projection onto convex cones, we derive another equivalent algorithm and go further in the analysis.

1. Introduction

In a statistical setting, consider the nonparametric regression model

$$Y = r(X) + \varepsilon, \tag{1}$$

where X and Y are both real-valued random variables with X uniform in $[0,1]$, $\mathbb{E}\left[Y^2\right] < \infty$ and $\mathbb{E}[\varepsilon|X] = 0$ (see for example [18]). Assume, in addition, that the regression function r is of bounded variation and, without loss of generality, right-continuous. With this respect, the Jordan decomposition asserts that such a function can be written as the sum of a non-decreasing function u and a non-increasing function b,

$$r(x) = u(x) + b(x).$$

Viewing this latter equation as an additive model involving the increasing part and the decreasing part of r, we propose a new estimator which combines two well-established tools in nonparametric regression: the isotonic regression related to the estimation of monotone functions, and the backfitting algorithm devoted to the estimation of additive models.

The estimation of a monotone regression function dates back to the 50's and the early work by Ayer et al. [2]. Given a sample of independent and identically distributed (i.i.d.) random couples $(X_1, Y_1), \ldots, (X_n, Y_n)$ following the general model (1), denoting $x_1 = X_{(1)} < \cdots < x_n = X_{(n)}$ the ordered sample, and $y_1, \ldots, y_n$ the corresponding observations, the Pool-Adjacent-Violators Algorithm (PAVA) determines a collection of non-decreasing level sets solution to the minimization problem

$$\min_{u_1 \leqslant \cdots \leqslant u_n} \sum_{i=1}^{n} (y_i - u_i)^2.$$

Since the cone

* Corresponding author.
E-mail addresses: arnaud.guyader@uhb.fr (A. Guyader), nicolas.jegou@uhb.fr (N. Jégou), nemab@math.ubbcluj.ro (A.B. Németh), nemeths@for.mat.bham.ac.uk (S.Z. Németh).

0096-3003/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.amc.2013.11.048

$$\mathcal{C}^{\uparrow} = \{u = (u_1, \ldots, u_n) \in \mathbb{R}^n : u_1 \leqslant \cdots \leqslant u_n\}$$

is a closed convex set in $\mathbb{R}^n$, there exists a unique solution to this minimization problem. This solution, called the isotonic regression of y and denoted $\mathrm{iso}(y)$, is the metric projection of $y = (y_1, \ldots, y_n)$ on $\mathcal{C}^{\uparrow}$ with respect to the Euclidean norm, that is

$$\mathrm{iso}(y) = \arg\min_{u \in \mathcal{C}^{\uparrow}} \|y - u\|^2 = \arg\min_{u \in \mathcal{C}^{\uparrow}} \sum_{i=1}^{n} (y_i - u_i)^2. \tag{2}$$

Correspondingly, the antitonic regression of y is the projection of y on the set of vectors with non-increasing coordinates, i.e., $\mathcal{C}^{\downarrow} = -\mathcal{C}^{\uparrow} = \{b = (b_1, \ldots, b_n) \in \mathbb{R}^n : b_1 \geqslant \cdots \geqslant b_n\}$. From now on, $\mathcal{C}^{\uparrow}$ and $\mathcal{C}^{\downarrow}$ will be called monotone cones.

A major attraction of isotonic regression procedures is their simplicity. Since they are nonparametric and data driven (i.e., they do not require the tuning of any smoothing parameter), these estimators have raised considerable interest since more than fifty years. A comprehensive account on the subject can be found in [3], statistical properties have been studied in [6,7,33,14], and extensions or improvements of the PAVA approach can be found in [15,27,5].

Still in nonparametric statistics, but when $X = (X^1, \ldots, X^d)$ is multidimensional, the additive models were suggested by Friedman and Stuetzle [17] and popularized by Hastie and Tibshirani [20] as a way to get around the so-called "curse of dimensionality". In brief, this means that, in multivariate smoothing, nonparametric estimators have to consider large neighborhoods of a particular point of the space to catch observations, and hence large biases can result. The additive model assumes that the regression function can be written as the sum of smooth terms in the covariates:

$$r(X) = \sum_{j=1}^{d} r_j(X^j). \tag{3}$$

Since each explanatory variable is represented separately in (3), the additive model provides a logical extension of the standard linear regression and once an additive model is fitted, one can easily interpret the role of each variable in predicting the response.

Buja et al. [8] proposed the backfitting algorithm as a practical method for fitting additive models. It consists in iterated fitting of the *partial residuals* from earlier steps until convergence is reached. If the current estimates are $\hat{r}_1, \ldots, \hat{r}_d$, then $\hat{r}_j$ is updated by smoothing $y - \sum_{k \neq j} \hat{r}_k$ against X^j. While backfitting has attracted much attention and is frequently applied, it has been somewhat difficult to analyze theoretically. Nonetheless, when using linear smoothers in each direction, the convergence of the algorithm can be related to the spectrum of the individual smoothing matrices (see, e.g., [8,31]), and when all the smoothers are orthogonal projections, the whole algorithm can be replaced by a global projection operator [19].

There exist other multivariate methods based on repeated fitting of the residuals. Some of them, like L2-boosting [9], boosted kernel regression [13], iterative bias reduction [10], do not assume any particular structure for the regression function. The common principle of these approaches is to start out with a biased smoother or a weak learner, and then to estimate and correct the bias in an iterative manner. Hence, instead of smoothing the *partial residuals*, one smoothes the *global residuals* $y - \sum_{j=1}^{d} \hat{r}_j$, and then correct the previous smoother. Just as for the backfitting, the convergence of these algorithms as well as the statistical properties of these estimators have mainly been studied in the case of linear smoothers.

In our situation, however, it is noteworthy that projections on convex cones are not linear operators. But considering our iterative estimator as the application of Von Neumann's algorithm (see for example [12]), we will show that iterating the procedure tends to reproduce the data. Moreover, we manage to go further in the analysis by proving that the individual terms of the sum converge as well to identified limits. This is in fact possible thanks to a result which is rather unexpected from the statistical point of view: iterating isotonic regression in a backfitting fashion or in a boosting fashion yields the same estimators at each step.

Interestingly, this result stems from a property of the projections onto monotone cones which, in our case, reads as follows (recall that $\mathrm{iso}(y)$ is defined by (2)):

$$\forall (y, \tilde{y}) \in \mathbb{R}^n \times \mathbb{R}^n, \quad y - \tilde{y} \in \mathcal{C}^{\uparrow} \;\Rightarrow\; \mathrm{iso}(y) - \mathrm{iso}(\tilde{y}) \in \mathcal{C}^{\uparrow} \tag{4}$$

From a more general perspective, one can see this equation as a particular case of the property of isotonicity of the projection onto convex cones. Here isotonicity is considered with respect to the order induced by the cone. The idea to relate the ordering induced by a convex cone and the metric projection onto the convex cone goes back to the paper by Isac and Németh [23], where a convex cone in the Euclidean space which admits an isotone projection onto it (called by the authors *isotone projection cone*) was characterized. Thereafter, this notion was considered in the complementarity theory to provide new existence results and iterative methods [24,25,30].

Yet, the notion of the cone in the above cited papers is used in the sense of "closed convex pointed cone". Confronted with the question if the monotone cones $\mathcal{C}^{\downarrow}$ and $\mathcal{C}^{\uparrow}$, which are not pointed, admit or not isotonic metric projections, we shall develop in Section 2 a general theory in order to apply it to this special case. This seems to be the simplest way to tackle this problem. Therefore, Theorem 1 below is interesting by itself. By using this approach, Corollary 2 states that the monotone cones $\mathcal{C}^{\downarrow}$ and $\mathcal{C}^{\uparrow}$ admit isotone metric projections.