A Geometrical Approach to Iterative Isotone Regression

Arnaud GUYADER[1]
Université Rennes 2, INRIA and IRMAR
Campus de Villejean, Rennes, France
arnaud.guyader@uhb.fr

Nicolas JÉGOU
Université Rennes 2
Campus de Villejean, Rennes, France
nicolas.jegou@uhb.fr

Alexander B. NÉMETH
Faculty of Mathematics and Computer Science
Babeş Bolyai University, RO-400084 Cluj-Napoca, Romania
nemab@math.ubbcluj.ro

Sándor Z. NÉMETH
School of Mathematics, The University of Birmingham
Birmingham B15 2TT, United Kingdom
nemeths@for.mat.bham.ac.uk

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.

Index Terms — Nonparametric estimation, Isotonic regression, Additive models, Metric projection onto convex cones.

2010 Mathematics Subject Classification: 52A20, 62G08, 90C33.

1 Introduction

In a statistical setting, consider the nonparametric regression model

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

---

[1]Corresponding author.

where X and Y are both real-valued random variables with X uniform in $[0, 1]$, $\mathbb{E}\left[Y^2\right] < \infty$ and $\mathbb{E}\left[\varepsilon | X\right] = 0$ (see for example [19]). 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 $x_1 = X_{(1)} < \ldots < x_n = X_{(n)}$ 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

$$\min_{u_1 \leq \ldots \leq u_n} \sum_{i=1}^{n} \left(y_i - u_i\right)^2.$$

Since the cone
$$\mathcal{C}^\uparrow = \{u = (u_1, \ldots, u_n) \in \mathbb{R}^n : u_1 \leq \ldots \leq u_n\}$$

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

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

Correspondingly, the antitonic regression of y is the projection of y on the set of vectors with non-increasing coordinates, i.e., $\mathcal{C}^\downarrow = -\mathcal{C}^\uparrow = \{b = (b_1, \ldots, b_n) \in \mathbb{R}^n : b_1 \geq \ldots \geq b_n\}$. 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], [34], [15], and extensions or improvements of the PAVA approach can be found in [16], [28] and [5].

Still in nonparametric statistics, but when $X = (X^1, \ldots, X^d)$ is multidimensional, the additive models were suggested by Friedman and Stuetzle [18] and popularized by Hastie

and Tibshirani [21] 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] and [32]), and when all the smoothers are orthogonal projections, the whole algorithm can be replaced by a global projection operator [20].

There exist other multivariate methods based on repeated fitting of the residuals. Some of them, like L2-boosting [9], boosted kernel regression [14], 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 [13]), 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 iso(y) is defined by (2)):

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

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 [24], 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 [25, 26, 31].

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.

Then, we come back to the statistical framework, the remainder of the paper being organized as follows: the definitions, the analysis and the equivalence of the Iterative Isotone Regression estimator and the Iterative Isotone Bias Reduction estimator are detailed in Section 3. Finally, all the proofs are gathered in the Appendix.

To conclude this introduction, we would like to make a comment on a topic that will **not** be addressed in the present document. Starting from a sample $\mathcal{D}_n = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ of i.i.d. random couples with the same distribution as a generic pair (X, Y), our estimator takes the form $\hat{r}_n^{(k_n)}$, where k_n denotes the number of iterations, possibly depending on the sample size n. In this framework, an important aspect is to specify conditions on the regression model (1) and on the sequence (k_n) so that, for example,

$$\mathbb{E}\left[\left(r(X) - \hat{r}_n^{(k_n)}(X)\right)^2\right] \xrightarrow[n \to \infty]{} 0,$$

where the expectation $\mathbb{E}[.]$ is considered with respect to the sample $\mathcal{D}_n$ and the generic variable X. If this property is satisfied, we say that the estimator $\hat{r}_n^{(k_n)}$ is consistent. This important issue will not be pursued further here and will be addressed elsewhere by the authors.

2 Isotonicity of the projection onto the monotone cones

It turns out that the isotonicity of the projection, as defined in (4) for the specific case of the cone $\mathcal{C}^{\uparrow}$, is in general a very strong requirement which implies the latticiality of the order induced by the convex cone. Thus, the investigation of the isotone projection cones becomes part of the theory of latticially ordered Euclidean and Hilbert spaces. A simple finite method of projection onto isotone projection cones proposed in [30] has become

important in the effective handling of all the problems involving projection onto these cones. Besides nonlinear complementarity, isotone projection cones have applications in other domains of optimization theory. The method proposed in [30] has become important in the effective handling of the problem of map-making from relative distance information, e.g., stellar cartography, see Section 5.13.2.4 in [13] and

www.convexoptimization.com/wikimization/index.php/Projection_on_Polyhedral_Cone

Although we shall not consider projection methods here, we stress that some of the results developed in [30] will be useful in our proofs.

Let us first introduce some notations. If $\mathcal{C}$ is a non-empty, closed convex set in $\mathbb{R}^n$, then for each $x \in \mathbb{R}^n$, there exists a unique nearest point $P_{\mathcal{C}}x$ in $\mathcal{C}$, that is, a point with the property that

$$\|x - P_{\mathcal{C}}x\| = \inf\{\|x - c\| : \ c \in \mathcal{C}\},$$

where $\|.\|$ stands for the Euclidean norm in $\mathbb{R}^n$ [35]. The mapping $P_{\mathcal{C}} : \mathbb{R}^n \to \mathcal{C}$ is called the nearest point mapping of $\mathbb{R}^n$ onto $\mathcal{C}$ or simply the (metric) projection onto $\mathcal{C}$. Let $\mathcal{C}$ be a closed convex cone in $\mathbb{R}^n$, i.e., a closed nonempty set with (i) $\mathcal{C} + \mathcal{C} \subset \mathcal{C}$, and (ii) $t\mathcal{C} \subset \mathcal{C}$, $\forall\, t \in \mathbb{R}_+$. If $\mathcal{C} \cap (-\mathcal{C}) = \{0\}$, then $\mathcal{C}$ is called a closed convex pointed cone. In order to lighten the writings, and since all sets at hand in the following will be closed and convex, we propose to call them respectively "cone" and "pointed cone".

Lemma 1 *Suppose that $\mathcal{C}$ is a cone, denote $L = \mathcal{C} \cap (-\mathcal{C})$ the maximal subspace contained in $\mathcal{C}$, $L^\perp$ its orthogonal complement, and $K = L^\perp \cap \mathcal{C}$. Then, K is a pointed cone in $L^\perp$,*

$$\mathcal{C} = K \oplus L, \tag{5}$$

where $\oplus$ stands for the orthogonal sum, and

$$P_{\mathcal{C}}x = P_K x_{L^\perp} + x_L, \tag{6}$$

where $x = x_{L^\perp} + x_L$ with $x_L \in L$ and $x_{L^\perp} \in L^\perp$.

By putting $u \leq_{\mathcal{C}} v$ whenever $v - u \in \mathcal{C}$, the cone $\mathcal{C}$ induces a semi-order $\leq_{\mathcal{C}}$ in $\mathbb{R}^n$ which is translation invariant (i.e., $u \leq_{\mathcal{C}} v$ implies $u + z \leq_{\mathcal{C}} v + z$ for any $z \in \mathbb{R}^n$) and scale invariant (i.e., $u \leq_{\mathcal{C}} v$ implies $tu \leq_{\mathcal{C}} tv$ for any $t \in \mathbb{R}_+$).

The projection $P_{\mathcal{C}}$ is said $\mathcal{C}$-*isotone* if $u \leq_{\mathcal{C}} v$ implies $P_{\mathcal{C}}u \leq_{\mathcal{C}} P_{\mathcal{C}}v$. If $P_{\mathcal{C}}$ is $\mathcal{C}$-isotone, then $\mathcal{C}$ is called an *isotone projection cone*. At this point, we would like to emphasize that in [24], the authors investigate isotone projection properties only for pointed cones. Our purpose here is thus to generalize their results to cones which are not necessarily pointed, hence introducing the decomposition $\mathcal{C} = K \oplus L$ in Lemma 1, where K is a pointed cone.

Theorem 1 *Let $\mathcal{C} \subset \mathbb{R}^n$ be a cone,*

$$\mathcal{C} = K \oplus L,$$

with $L = \mathcal{C} \cap (-\mathcal{C})$ and $K = \mathcal{C} \cap L^\perp$. Then, $\mathcal{C}$ is an isotone projection cone if and only if the pointed cone $K \subset L^\perp$ is an isotone projection cone in $L^\perp$.

A simple geometric characterization of the isotone projection pointed cones was given in [24]. It uses the notion of the polar of a cone. If $\mathcal{C} \subset \mathbb{R}^n$ is a cone, then the set

$$\mathcal{C}^- = \{y \in \mathbb{R}^n : \langle x, y \rangle \leq 0, \ \forall \ x \in \mathcal{C}\}$$

is called the *polar* of the cone $\mathcal{C}$. The set $\mathcal{C}^-$ is obviously a cone. We have the following easily verifiable result:

Lemma 2 *Let $\mathcal{C} \subset \mathbb{R}^n$ be a cone. Using the notations introduced in Theorem 1, and denoting the polar of the pointed cone K in the subspace $L^\perp$ by K^-, we have the relation*

$$\mathcal{C}^- = iK^-,$$

where i is the inclusion mapping of $L^\perp$ into $\mathbb{R}^n$.

If the cone $\mathcal{C}$ is *generating* in the sense that $\mathcal{C} - \mathcal{C} = \mathbb{R}^n$, then it is readily seen that the polar $\mathcal{C}^-$ is a pointed cone. Putting together the main result in [24], Theorem 1 and Lemma 2, we have the following conclusion:

Corollary 1 *The generating cone $\mathcal{C}$ is an isotone projection cone if and only if its polar $\mathcal{C}^-$ is a cone generated by linearly independent vectors forming mutually non-acute angles.*

Let us focus our attention on the monotone cone $\mathcal{C}^\downarrow = \{b \in \mathbb{R}^n : b_1 \geq b_2 \geq ... \geq b_n\}$. It is readily seen that $\mathcal{C}^\downarrow$ is a generating cone, but not a pointed cone. Specifically, let

$$L = \mathcal{C}^\downarrow \cap (-\mathcal{C}^\downarrow) = \{x \in \mathbb{R}^n : x_1 = x_2 = ... = x_n\},$$

then L, the maximal subspace contained in $\mathcal{C}^\downarrow$, is of dimension one. We have also that

$$K = L^\perp \cap \mathcal{C}^\downarrow \tag{7}$$

is an $(n-1)$-dimensional pointed cone in the hyperplane $L^\perp$, and

$$\mathcal{C}^\downarrow = \mathcal{C}^\downarrow \cap (L^\perp \oplus L) = K \oplus L.$$

We will prove that the pointed cone K given by (7) is an isotone projection cone in $L^\perp$, hence the following result.

Corollary 2 *The monotone cones $\mathcal{C}^\downarrow$ and $\mathcal{C}^\uparrow$ admit isotone projections.*

This result will be the basic ingredient to show the equality between the algorithms I.I.R. and I.I.B.R. presented in the next section (see Theorem 2).

3 Iterative Isotone Regression

In this section, we first present the algorithm called Iterative Isotonic Regression (I.I.R. in short) which proceeds in alternating isotonic and antitonic regressions in a backfitting fashion. The connection with Von Neumann's algorithm and recent results in this topic will allow us to exhibit the limit of this estimator when the number of iterations goes to infinity.

Let us first briefly recall the statistical framework of the introduction and the idea behind our algorithm. We consider the nonparametric regression model

$$Y = r(X) + \varepsilon, \tag{8}$$

where we assume that the regression function r is right-continuous and of bounded variation. Then, the Jordan decomposition asserts that such a function can be written as the sum of a non-decreasing and a non-increasing function: $r(x) = u(x) + b(x)$. Specifically, if we impose that u and b have singular associated Stieltjes measures and, for example, that

$$\int_0^1 u(x)\,dx = \int_0^1 r(x)\,dx, \tag{9}$$

then the decomposition is unique. In this case, we adopt the following notation

$$\forall x \in [0,1] \qquad r(x) = u^\star(x) + b^\star(x), \tag{10}$$

and we call this latter writing the Jordan minimum variation decomposition of r.

3.1 Framework, notations and convergence

In a statistical setting, we shall use a sample of i.i.d. random couples $(X_1, Y_1), \ldots, (X_n, Y_n)$ following the model (8) and try to estimate the regression function r. Viewing $u^\star$ and $b^\star$ as the components of an additive model involving two monotone terms, the idea of I.I.R. is to apply alternatively isotonic and antitonic regressions. Recall that $x_1 = X_{(1)} < \ldots < x_n = X_{(n)}$ denotes the ordered sample, that $y_1, \ldots, y_n$ are the corresponding observations and that $\mathrm{iso}(y)$ (resp. $\mathrm{anti}(y)$) is the isotonic (resp. antitonic) regression of $y = (y_1, \ldots, y_n)'$.

Besides, if $z = (z_1, \ldots, z_n)'$, the notation $\Delta(z)$ stands for the $(n-1)$ dimensional vector defined as $\Delta(z) = (z_2 - z_1, \ldots, z_n - z_{n-1})'$. Considering two vectors z and $\tilde{z}$, we may write $\Delta(z) \circ \Delta(\tilde{z})$ for their term-by-term (or Hadamard) product, which means

$$\Delta(z) \circ \Delta(\tilde{z}) = ((z_2 - z_1) \times (\tilde{z}_2 - \tilde{z}_1), \ldots, (z_n - z_{n-1}) \times (\tilde{z}_n - \tilde{z}_{n-1}))'.$$

If $\Delta(z) \circ \Delta(\tilde{z}) = 0$, we will say that z and $\tilde{z}$ have singular variations.

We are now in a position to specify the Iterative Isotonic Regression algorithm.

Algorithm 1 Iterative Isotonic Regression (I.I.R.)

---

(1) Initialization: $\hat{b}^{(0)} = \left(\hat{b}_1^{(0)}, \ldots, \hat{b}_n^{(0)} \right)' = 0$.

(2) Cycle: for $k \geq 1$,

$$
\begin{aligned}
\hat{u}^{(k)} &= \mathrm{iso}\left(y - \hat{b}^{(k-1)} \right) \\
\hat{b}^{(k)} &= \mathrm{anti}\left(y - \hat{u}^{(k)} \right) \\
\hat{y}^{(k)} &= \hat{u}^{(k)} + \hat{b}^{(k)}.
\end{aligned}
$$

(3) Iterate (2) until a stopping condition to be specified.

It can be proved that at each iteration $k \geq 1$,

$$
\begin{cases}
\Delta\left(\hat{u}^{(k)} \right) \circ \Delta\left(\hat{b}^{(k)} \right) = 0 \\
\bar{\hat{u}}^{(k)} = \bar{y} \\
\bar{\hat{b}}^{(k)} = 0,
\end{cases}
\tag{11}
$$

where $\bar{y} := \left(\sum_{i=1}^{n} y_i \right)/n$ stands for the empirical mean of y (see [27], Appendix C). These equations ensure the identifiability of the decomposition $\hat{y}^{(k)} = \hat{u}^{(k)} + \hat{b}^{(k)}$. One might indeed consider them as the translation of conditions (9) in a discrete context. In other words, given $\hat{y}^{(k)}$, both vectors $\hat{u}^{(k)}$ and $\hat{b}^{(k)}$ are uniquely specified.

Figure 1 illustrates the application of the I.I.R. algorithm on the example of the fonction r which is drawn on the top left hand side of the figure. Still on the top left hand side of the figure, we have also plotted $n = 100$ points (x_i, y_i) to obtain our sample: for each point, one has $y_i = r(x_i) + \varepsilon_i$, where the ε_i's are Gaussian centered random variables. Besides these sample points, the three other figures show the estimations $\hat{y}^{(k)}$ with $k = 1, 10, 1000$ iterations.
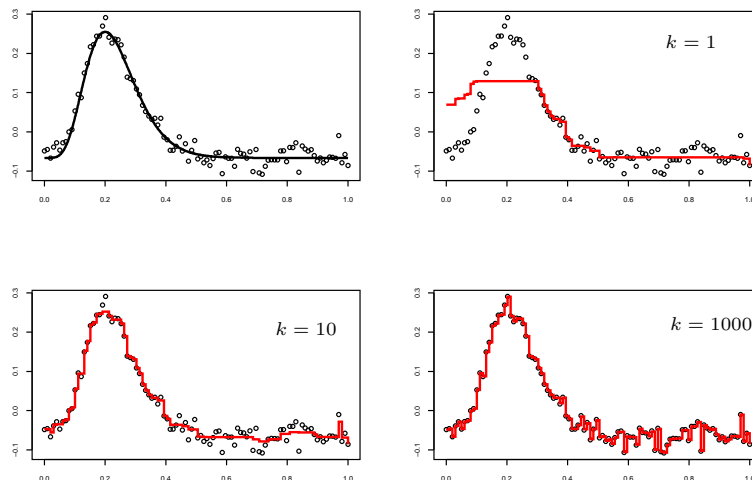

Figure 1: Application of the I.I.R. algorithm for $k = 1, 10, 1000$.

By convention, for each iteration, the method fits a piecewise constant function. One can notice that increasing the number of iterations leads to increase the number of jumps. It also appears on this example that iterating the algorithm leads to the interpolation of the original data. This property is always true, as established by the following result.

Proposition 1 *With the previous notations, one has $\lim_{k \to \infty} \hat{y}^{(k)} = y$.*

A few remarks are in order:

1. The take-home message here is that iterating the algorithm leads to overfitting, which is of course not desired in practice. We will comment on this issue at the end of Section 3.2.

2. Note that Proposition 1 ensures the convergence of the sum $\hat{u}^{(k)} + \hat{b}^{(k)}$, but does not say anything about the convergence of its individual terms $\hat{u}^{(k)}$ and $\hat{b}^{(k)}$. However, Corollary 4.9 in [4] asserts that the sequences $\hat{u}^{(k)}$ and $\hat{b}^{(k)}$ are also convergent. In fact, the specification of these limits will be possible in our situation thanks to the introduction of another equivalent algorithm, called I.I.B.R. This will be the topic of the next Section 3.2.

3. In [29], Mammen and Yu consider isotonic functions in the multivariate additive model (3). They rely on the analysis of Dykstra's algorithm sequences [17] to show that for fixed n, the backfitting procedure converges to the solution of $\arg \min \sum_{j=1}^{n} (y_j - u_j)^2$, where u is the sum of d vectors in $\mathcal{C}^{\uparrow}$. Hence, our case is rather comparable to the one considered by these authors when $d = 2$. Correspondingly, in our setting, one can see the vectors $y - \hat{b}^{(k-1)} - \hat{u}^{(k)}$ and $y - \hat{u}^{(k)} - \hat{b}^{(k)}$ as Dykstra's sequences and consider Mammen and Yu's alternate projections on the polar cones of $\mathcal{C}^{\uparrow}$ and $\mathcal{C}^{\downarrow}$ to prove Proposition 1. However, as we are alternating projections onto opposite cones, considering $\hat{r}^{(k)}$ and $y - \hat{b}^{(k)}$ yields the easier Von Neumann's type interpretation given in our proof.

3.2 The connection with Iterative Isotonic Bias Reduction

In this section, we propose another algorithm inspired by bias reduction techniques in regression [9][10]. In a nutshell, the idea here is to work on the updated residuals $y - \hat{y}^{(k)}$ in order to refine the estimator at each step. Specifically, this algorithm takes the following form.

Algorithm 2 Iterative Isotonic Bias Reduction (I.I.B.R.)

---

(1) Initialization: $\hat{y}^{(0)} = 0$.
(2) Cycle: for $k \geq 1$,

$$
\begin{aligned}
\tilde{u}^{(k)} &= \text{iso}\left(y - \hat{y}^{(k-1)}\right) \\
\tilde{b}^{(k)} &= \text{anti}\left(y - \hat{y}^{(k-1)} - \hat{u}^{(k)}\right) \\
\hat{y}^{(k)} &= \hat{y}^{(k-1)} + \tilde{u}^{(k)} + \tilde{b}^{(k)}.
\end{aligned}
$$

(3) Iterate (2) until a stopping condition to be specified.

Interestingly, it turns out that algorithms I.I.R. and I.I.B.R. coincide. This remarkable fact, which is the purpose of the next theorem, deeply relies on the property of isotonicity of the projection onto the cones we consider (see Corollary 2). As a by-product, it will allow us to make precise the individual limits of the sequences $\hat{u}^{(k)}$ and $\hat{b}^{(k)}$.

Theorem 2 *With the previous notations, for all $k \geq 1$,*

$$\hat{u}^{(k)} = \sum_{j=1}^{k} \tilde{u}^{(j)} \qquad and \qquad \hat{b}^{(k)} = \sum_{j=1}^{k} \tilde{b}^{(j)}.$$

Proposition 1 ensures that $\hat{y}^{(k)}$ tends to y when k goes to infinity. Thanks to Theorem 2, we will go one step further and show in Corollary 3 that the individual terms $\hat{u}^{(k)}$ and $\hat{b}^{(k)}$ tend to some explicit limits $u_y^{\star}$ and $b_y^{\star}$. These limits are simply the discrete analogous of the Jordan minimum variation decomposition of a function with bounded variation (10) for the vector y. As will be stated below, they are indeed characterized by $y = u_y^{\star} + b_y^{\star}$ and $\Delta(u_y^{\star}) \circ \Delta(b_y^{\star}) = 0$, the empirical means of y and $u_y^{\star}$ being the same.

Before proceeding, let us illustrate this on the example of Figure 2. The left-hand side represents the vector y and the decomposition $y = u_y^{\star} + b_y^{\star}$. One can see that

$$y_{i+1} - y_i > 0 \Rightarrow u_{y,i+1}^{\star} - u_{y,i}^{\star} = y_{i+1} - y_i \text{ and } b_{y,i+1}^{\star} - b_{y,i}^{\star} = 0,$$

and conversely, which, in the discrete case, amounts to say that $u^{\star}$ and $b^{\star}$ have singular associated Stieltjes measures in equation (10). The right-hand side of the figure illustrates that, when the number of iterations k goes to infinity, $\hat{u}^{(k)}$ and $\hat{b}^{(k)}$ converge respectively to $u_y^{\star}$ and $b_y^{\star}$.
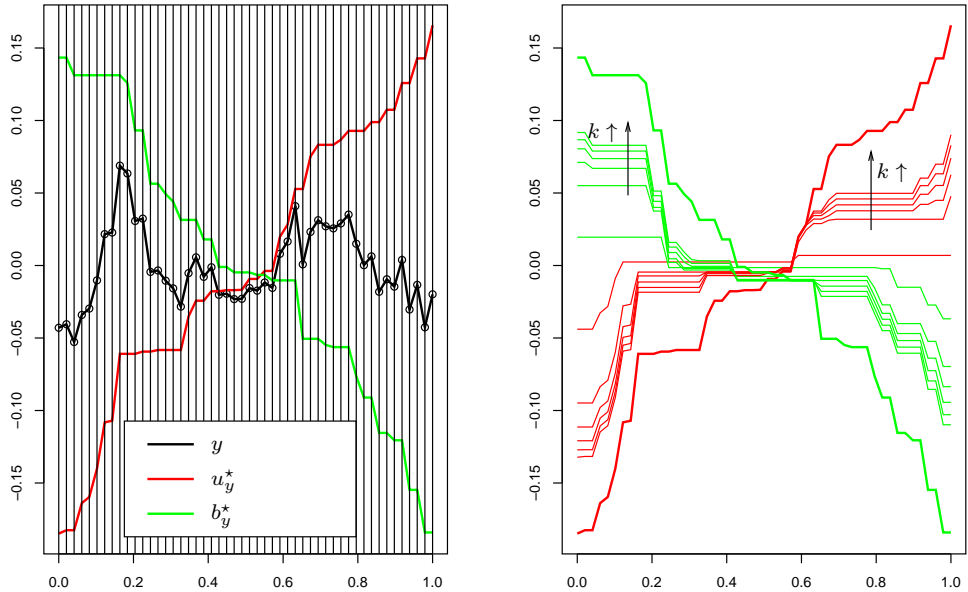

Figure 2: Respective convergences of $\hat{u}^{(k)}$ and $\hat{b}^{(k)}$ to $u_y^{\star}$ and $b_y^{\star}$.

Corollary 3 *The sequences $\left(\hat{u}^{(k)}\right)_{k\geq 1}$ and $\left(\hat{b}^{(k)}\right)_{k\geq 1}$ admit the following limits*

$$\lim_{k\to\infty} \hat{u}^{(k)} = u_y^\star \qquad and \qquad \lim_{k\to\infty} \hat{b}^{(k)} = b_y^\star$$

where $u_y^\star$ and $b_y^\star$ are such that $y = u_y^\star + b_y^\star$, $\bar{u}_y^\star = \bar{y}$, $\bar{b}_y^\star = 0$ and $\Delta(y) = \Delta(u_y^\star) \circ \Delta(b_y^\star)$.

As increasing the number of iterations leads to overfit the data, iterating the procedure until convergence is not desirable. This brings up to the choice of a stopping rule which could be used in practice. Viewing the latter question as a model selection issue suggests stopping criteria based for example on Akaike Information Criterion (AIC, see [1]), modified AIC criterion (AICc, see [23]), Bayesian Information Criterion (BIC, see [33]) and Generalized Cross Validation (GCV, see [12]).

We have applied and compared these stopping rules to I.I.R. through simulated data. For most investigated cases, it appears that AICc shows the best performances among the three other criteria. Then, for this specific stopping criterion, we have compared I.I.R. to non-parametric competitors, namely local polynomial regression (R package locpol) and smoothing spline regression (R package stats, function smooth.spline). For further details on this topic, we refer the interested reader to the R package dedicated to I.I.R. at the following address:

http://www.sites.univ-rennes2.fr/laboratoire-statistique/JEGOU/index.html

The take-home message is that I.I.R. can not compete with spline or local polynomial regression for smooth functions. However, when the functions have discontinuities, our estimator compares favorably with his competitors, in particular when the sample size increases. This suggests that our method could be used to locate discontinuities in a regression framework. Applications arise for example in genomics where the detection of breakpoints from Array Comparative Genomic Hybridization (array CGH) profiles is of crucial importance to identify genes involved in cancer progression [22].

4   Appendix

4.1   Proof of Lemma 1

The relation (5) follows directly from $\mathcal{C} = \mathcal{C} \cap (L^\perp \oplus L)$. It is known (see [35]) that the projection $P_\mathcal{C} x$ of x onto the cone $\mathcal{C}$ is characterized by the couple of relations:

$$\langle x - P_\mathcal{C} x, y \rangle \;\leq\; 0, \; \forall\, y \in \mathcal{C}, \tag{12}$$
$$\langle x - P_\mathcal{C} x, P_\mathcal{C} x \rangle \;=\; 0. \tag{13}$$

Hence, we have to verify the above relations for $P_K x_{L^\perp} + x_L$ instead of $P_\mathcal{C} x$. By the relation (5), $P_K x_{L^\perp} + x_L \in \mathcal{C}$. Then, take an arbitrary $y \in \mathcal{C}$ represented by (5) in the form $y = y_K + y_L$ with $y_K \in K$ and $y_L \in L$. Then, we have

$$\langle x_{L^\perp} + x_L - (P_K x_{L^\perp} + x_L), y_K + y_L \rangle = \langle x_{L^\perp} - P_K x_{L^\perp}, y_K \rangle \leq 0, \;\; \forall\, y = y_K + y_L \in \mathcal{C},$$

because y_L is perpendicular to $x_{L^\perp} - P_K x_{L^\perp} \in L^\perp$, and because of the relation similar to (12) characterizing the projection of $x_{L^\perp}$ onto the cone K in $L^\perp$. Thus, relation (12) holds for $P_K x_{L^\perp} + x_L$ in place of $P_\mathcal{C} x$. We further have

$$\langle x_{L^\perp} + x_L - (P_K x_{L^\perp} + x_L), P_K x_{L^\perp} + x_L \rangle = \langle x_{L^\perp} - P_K x_{L^\perp}, P_K x_{L^\perp} \rangle = 0,$$

because x_L is perpendicular to $x_{L^\perp} - P_K x_{L^\perp}$ and because of the relation similar to (13) applied to $x_{L^\perp} \in L^\perp$ and its projection onto K. The obtained relation is exactly (13) for $P_K x_{L^\perp} + x_L$ instead of $P_\mathcal{C} x$.

4.2 Proof of Theorem 1

Take $u, v \in L^\perp$. Then, $u \leq_K v$ is equivalent to $u \leq_\mathcal{C} v$. If $P_\mathcal{C}$ is $\mathcal{C}$-isotone, then $u \leq_\mathcal{C} v$ implies by Lemma 1 that $P_K v - P_K u = P_\mathcal{C} v - P_\mathcal{C} u \in \mathcal{C}$. Since $P_K u, P_K v \in L^\perp$, it follows that $P_K v - P_K u \in L^\perp \cap \mathcal{C} = K$. The obtained relation shows that P_K is K-isotone, concluding the proof of the necessity of the theorem.

Suppose now that P_K is K-isotone and take $u, v \in \mathbb{R}^n$ with $u \leq_\mathcal{C} v$. If $u = u_{L^\perp} + u_L$ and $v = v_{L^\perp} + v_L$, then using formula (5), we are led to

$$v - u = v_{L^\perp} - u_{L^\perp} + v_L - u_L \in K \oplus L,$$

and hence $v_{L^\perp} - u_{L^\perp} \in K$, that is $u_{L^\perp} \leq_K v_{L^\perp}$, and by the K-isotonicity of P_K it follows that $P_K v_{L^\perp} - P_K u_{L^\perp} \in K$. Hence, using formula (6), we have

$$P_\mathcal{C} v - P_\mathcal{C} u = P_K v_{L^\perp} + v_L - P_K u_{L^\perp} - u_L = P_K v_{L^\perp} - P_K u_{L^\perp} + v_L - u_L \in K \oplus L = \mathcal{C}.$$

That is, $P_\mathcal{C} u \leq_\mathcal{C} P_\mathcal{C} v$, which proves the isotonicity of $P_\mathcal{C}$.

4.3 Proof of Corollary 2

It clearly suffices to prove that the monotone cone $\mathcal{C}^\downarrow = \{b \in \mathbb{R}^n : b_1 \geq b_2 \geq ... \geq b_n\}$ admits an isotone projection. For this, we have to introduce some notations. Let us first take the following basis in $\mathbb{R}^n$:

$$e_1 = (1, 0, ..., 0), e_2 = (1, 1, 0, ..., 0), ..., e_{n-1} = (1, ..., 1, 0), e_n = (1, 1, ..., 1).$$

Then, an arbitrary element $b = (b_1, ..., b_n) \in \mathbb{R}^n$ can be represented in the form

$$b = (b_1 - b_2)e_1 + (b_2 - b_3)e_2 + ... + (b_{n-1} - b_n)e_{n-1} + b_n e_n, \tag{14}$$

the relation $b \in \mathcal{C}^\downarrow$ being equivalent with

$$b_{j-1} - b_j \geq 0, \quad j = 2, ..., n. \tag{15}$$

Let us consider further the following basis in $L^\perp$:

$$e_1' = (n-1, -1, -1, ..., -1), e_2' = (n-2, n-2, -2, ..., -2), ..., e_{n-1}' = (1, ..., 1, -(n-1)).$$

The following notation is standard in the convex geometry and ordered vector space theory: If $M \subset \mathbb{R}^n$ is a non-empty set, then let

$$\text{cone}(M) = \{t_1 n_1 + \dots + t_k n_k : \ n_i \in M, \ t_i \in \mathbb{R}_+, \ i = 1, \dots, k; \ k \in \mathbb{N}\}.$$

The set $\text{cone}(M)$ is the minimal cone containing the set M and it is called *the cone generated by M*. Denoting $L = \mathcal{C}^{\downarrow} \cap (-\mathcal{C}^{\downarrow}) = \{b \in \mathbb{R}^n : \ b_1 = b_2 = \dots = b_n\}$, and $K = L^{\perp} \cap \mathcal{C}^{\downarrow}$, we will see next that

$$K = \text{cone}(\{e'_1, \dots, e'_{n-1}\}). \tag{16}$$

Since $e'_j \in L^{\perp} \cap \mathcal{C}^{\downarrow}$, we have obviously that

$$\text{cone}(\{e'_1, \dots, e'_{n-1}\}) \subset K. \tag{17}$$

Comparing the vectors e_i and e'_j we get

$$\frac{1}{n}(e'_j + j e_n) = e_j, \ \ j = 1, \dots, n-1.$$

By substitution of e_j, $j = 1, \dots, n-1$, the representation (14) of b becomes

$$b = \frac{b_1 - b_2}{n}(e'_1 + e_n) + \frac{b_2 - b_3}{n}(e'_2 + 2e_n) + \dots + \frac{b_{n-1} - b_n}{n}(e'_{n-1} + (n-1)e_n) + b_n e_n. \tag{18}$$

Suppose now that $b \in \mathcal{C}^{\downarrow}$, that is, relations (15) hold. Then, the coefficients of the vectors e'_j in the representation (18) are non-negative. Thus, we have

$$b \in \mathcal{C}^{\downarrow} \Leftrightarrow b = \sum_{j=1}^{n-1} t_j e'_j + t_n e_n, \ t_j \in \mathbb{R}_+, \ j = 1, \dots, n-1, \ t_n \in \mathbb{R}. \tag{19}$$

In particular, if $b \in K$, then, by (7), we have $b \in L^{\perp}$. Hence, by multiplying (19) scalarly by e_n and by using $\langle b, e_n \rangle = 0$ (which follows from $b \in L^{\perp}$ and $e_n \in L$) and $\langle e'_j, e_n \rangle = 0$ (which follows from $e'_j \in L^{\perp}$ and $e_n \in L$), we get $t_n = 0$. This reasoning shows the inclusion $K \subset \text{cone}(\{e'_1, \dots, e'_{n-1}\})$, which together with (17) proves (16). We consider now the vectors

$$u_1 = (-1, 1, 0, \dots, 0), u_2 = (0, -1, 1, 0, \dots, 0), \dots, u_{n-1} = (0, \dots, 0, -1, 1).$$

Then, $u_i \in L^{\perp}$, $i = 1, \dots, n-1$, and we have

$$\langle u_i, e'_j \rangle = 0 \ \text{if} \ i \neq j, \ \ \langle u_i, e'_i \rangle < 0, \ i, j = 1, \dots, n-1. \tag{20}$$

According to the reasonings in [30], the relations (20) show that $\text{cone}(\{u_1, \dots, u_{n-1}\})$ is the polar of K in the subspace $L^{\perp}$. Furthermore, we have $\langle u_i, u_j \rangle \leq 0$ if $i \neq j$. By the main result in [24], this shows that K is an isotone projection pointed cone in $L^{\perp}$.

4.4 Proof of Proposition 1

In the following, y is held fixed and $y+\mathcal{C}^\uparrow$ stands for the translated cone. All the notations are recalled on Figure 3. To understand the geometrical forces driving Proposition 1, Figure 3 also provides a very simple interpretation of the algorithm, as it illustrates that the sequences of vectors $\hat{u}^{(k)}$ and $y - \hat{b}^{(k)}$ might be seen as alternate projections on the cones $\mathcal{C}^\uparrow$ and $y+\mathcal{C}^\uparrow$. In what follows, we justify this illuminating geometric interpretation in a rigorous way, and we explain its key role in the proof of the convergence as k goes to infinity.

First, by definition, we have $\hat{u}^{(1)} = P_{\mathcal{C}^\uparrow}(y)$, and a moment's thought shows that

$$P_{y+\mathcal{C}^\uparrow}(\hat{u}^{(1)}) = y + P_{\mathcal{C}^\uparrow}(\hat{u}^{(1)} - y) = y - P_{\mathcal{C}^\downarrow}(y - \hat{u}^{(1)}).$$

Then, the very definition of $\hat{b}^{(1)} = P_{\mathcal{C}^\downarrow}(y - \hat{u}^{(1)})$ leads to $y - \hat{b}^{(1)} = P_{y+\mathcal{C}^\uparrow}(\hat{u}^{(1)})$. In the same manner, since $\hat{b}^{(2)} = P_{\mathcal{C}^\downarrow}(y - \hat{u}^{(2)})$, we get

$$y - \hat{b}^{(2)} = y - P_{\mathcal{C}^\downarrow}(y - \hat{u}^{(2)}) = y + P_{\mathcal{C}^\uparrow}(y - \hat{u}^{(2)}) = P_{y+\mathcal{C}^\uparrow}(\hat{u}^{(2)}).$$

More generally, denoting $\hat{b}^{(0)} = 0$, this yields for all $k \geq 1$,

$$\hat{u}^{(k)} = P_{\mathcal{C}^\uparrow}(y - \hat{b}^{(k-1)}) \qquad \text{and} \qquad y - \hat{b}^{(k)} = P_{y+\mathcal{C}^\uparrow}(\hat{u}^{(k)}).$$
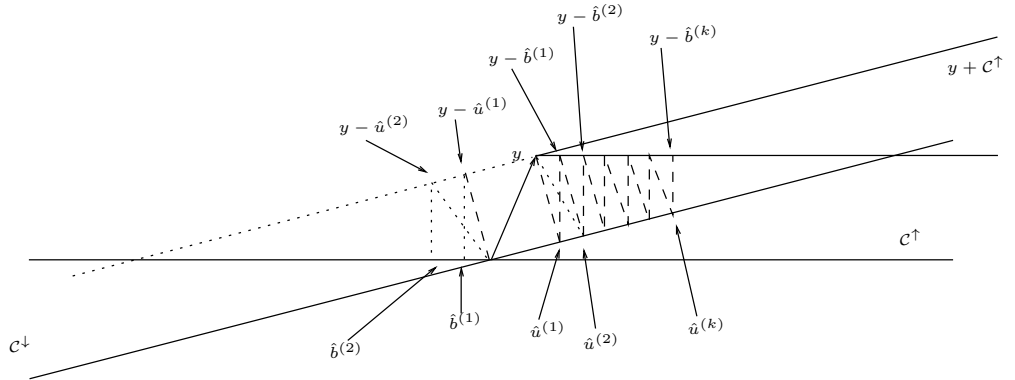

Figure 3: Interpretation of the I.I.R. method as a Von Neumann's algorithm.

It remains to invoke Theorem 4.8 in [4] to conclude that

$$(y - \hat{b}^{(k)}) - \hat{u}^{(k)} = y - \hat{y}^{(k)} \xrightarrow[k\to\infty]{} 0,$$

which ends the proof of Proposition 1.

4.5 Proof of Theorem 2

First we prove that for all $k \geq 1$,

$$\hat{u}^{(k+1)} - \hat{u}^{(k)} \in \mathcal{C}^{\uparrow} \qquad \text{and} \qquad \hat{b}^{(k+1)} - \hat{b}^{(k)} \in \mathcal{C}^{\downarrow}. \tag{21}$$

For $k = 1$, we have $\hat{u}^{(2)} - \hat{u}^{(1)} = \text{iso}(y - \hat{b}^{(1)}) - \text{iso}(y)$. Since $-\hat{b}^{(1)} \in \mathcal{C}^{\uparrow}$ and as $\mathcal{C}^{\uparrow}$ is an isotone projection cone by Corollary 2, we deduce that $\hat{u}^{(2)} - \hat{u}^{(1)}$ belongs to $\mathcal{C}^{\uparrow}$. In the same manner,

$$\hat{b}^{(2)} - \hat{b}^{(1)} = \text{anti}(y - \hat{u}^{(2)}) - \text{anti}(y - \hat{u}^{(1)}),$$

which is equivalent to

$$\hat{b}^{(2)} - \hat{b}^{(1)} = \text{anti}\left(y - \hat{u}^{(1)} - (\hat{u}^{(2)} - \hat{u}^{(1)})\right) - \text{anti}(y - \hat{u}^{(1)}).$$

Now, as was just noticed, $-(\hat{u}^{(2)} - \hat{u}^{(1)})$ belongs to $\mathcal{C}^{\downarrow}$, which is also an isotone projection cone, so that $\hat{b}^{(2)} - \hat{b}^{(1)}$ is indeed in $\mathcal{C}^{\downarrow}$. We can iterate this reasoning. For example, at the next step:

$$\hat{u}^{(3)} - \hat{u}^{(2)} = \text{iso}(y - \hat{b}^{(2)}) - \text{iso}(y - \hat{b}^{(1)}) = \text{iso}\left(y - \hat{b}^{(1)} - (\hat{b}^{(2)} - \hat{b}^{(1)})\right) - \text{iso}(y - \hat{b}^{(1)}),$$

and we may go on with the same arguments as before.

Next, we prove the desired result by induction on k. At the first step ($k = 1$), both algorithms clearly coincide. Now, let us assume that it is still true at step $k \geq 1$, which means

$$\hat{y}^{(k)} = \hat{u}^{(k)} + \hat{b}^{(k)} \qquad \text{with } \hat{u}^{(k)} = \sum_{j=1}^{k} \tilde{u}^{(j)} \text{ and } \hat{b}^{(k)} = \sum_{j=1}^{k} \tilde{b}^{(j)}.$$

Our objective is to prove that $\hat{u}^{(k+1)} = \sum_{j=1}^{k+1} \tilde{u}^{(j)}$. For this, let us show that

$$\|y - \hat{b}^{(k)} - \hat{u}^{(k+1)}\| = \|y - \hat{y}^{(k)} - \tilde{u}^{(k+1)}\|. \tag{22}$$

Due to the fact that $\hat{u}^{(k+1)}$ is the best non-decreasing approximation of $y - \hat{b}^{(k)}$, and since $\hat{u}^{(k)} + \tilde{u}^{(k+1)}$ itself is a non-decreasing sequence, we deduce that

$$\|y - \hat{b}^{(k)} - \hat{u}^{(k+1)}\| \leq \|y - \hat{b}^{(k)} - (\hat{u}^{(k)} + \tilde{u}^{(k+1)})\| = \|y - \hat{y}^{(k)} - \tilde{u}^{(k+1)}\|.$$

Now, as $\tilde{u}^{(k+1)}$ is the best non-decreasing approximation of $y - \hat{y}^{(k)}$, it is in particular a better approximation than $\hat{u}^{(k+1)} - \hat{u}^{(k)}$, this latter being itself non-decreasing as was just established above (see (21)). Thus, we are led to

$$\|y - \hat{y}^{(k)} - \tilde{u}^{(k+1)}\| \leq \|y - \hat{y}^{(k)} - (\hat{u}^{(k+1)} - \hat{u}^{(k)})\| = \|y - \hat{b}^{(k)} - \hat{u}^{(k+1)}\|.$$

Putting all the pieces together, (22) is proved, and we get

$$\|y - \hat{y}^{(k)} - \tilde{u}^{(k+1)}\| = \|y - \hat{b}^{(k)} - \hat{u}^{(k+1)}\| = \|y - \hat{y}^{(k)} - (\hat{u}^{(k+1)} - \hat{u}^{(k)})\|.$$

This indicates that $\tilde{u}^{(k+1)}$ and $\hat{u}^{(k+1)} - \hat{u}^{(k)}$ both realize the minimal distance to $y - \hat{y}^{(k)}$. As a consequence, $\tilde{u}^{(k+1)} = \hat{u}^{(k+1)} - \hat{u}^{(k)}$, and finally $\hat{u}^{(k+1)} = \sum_{j=1}^{k+1} \tilde{u}^{(j)}$. The same arguments may be repeated to establish that $\hat{b}^{(k+1)} = \sum_{j=1}^{k+1} \tilde{b}^{(j)}$. Details are omitted.

4.6 Proof of Corollary 3

In order to lighten the notations a bit, let us assume that the empirical mean of y is equal to zero. Consequently, all the intermediates of both algorithms have also zero mean, as well as $u_y^\star$ and $b_y^\star$. This allows us to work in the hyperplane $\mathcal{E}$, that means the subspace of $\mathbb{R}^n$ consisting in all zero mean vectors. In this hyperplane, we introduce the norm V to quantify the variations of a vector:

$$\begin{array}{rcl} \mathrm{V}: \ \mathcal{E} & \to & \mathbb{R}^+ \\ y & \mapsto & \mathrm{V}(y) = \sum_{i=1}^{n-1} |y_{i+1} - y_i| \end{array}$$

Then, it is routine to check that for any vector $z \in \mathcal{E}$ and any decomposition $z = u_z + b_z$ with u_z and b_z both in $\mathcal{E}$, we have the following equivalence

$$\Delta(u_z) \circ \Delta(b_z) = 0 \ \Leftrightarrow \ \mathrm{V}(z) = \mathrm{V}(u_z) + \mathrm{V}(b_z). \tag{23}$$

Now, from (11), we know that for all $k \geq 1$, $\hat{u}^{(k)}$ and $\hat{b}^{(k)}$ have singular variations (one can also see that on Figure 2), which implies that for all $k \geq 1$,

$$\mathrm{V}(\hat{y}^{(k)}) = \mathrm{V}(\hat{u}^{(k)}) + \mathrm{V}(\hat{b}^{(k)}).$$

Accordingly, we just have to justify that the limits $u^\infty := \lim_{k \to \infty} \hat{u}^{(k)}$ and $b^\infty := \lim_{k \to \infty} \hat{b}^{(k)}$ satisfy $\mathrm{V}(u^\infty) + \mathrm{V}(b^\infty) = \mathrm{V}(y)$ to deduce that $\Delta(u^\infty) \circ \Delta(b^\infty) = 0$, hence $u^\infty = u_y^\star$ and $b^\infty = b_y^\star$. The existence of the limits implies that for $i = 1, \ldots, n$,

$$u_i^\infty := \lim_{k \to \infty} \hat{u}_i^{(k)} \qquad \text{and} \qquad b_i^\infty := \lim_{k \to \infty} \hat{b}_i^{(k)}$$

are well-defined. Then, the continuity of the norm V leads to

$$\lim_{k \to \infty} \mathrm{V}(\hat{u}^{(k)}) = \mathrm{V}(u^\infty) \qquad \text{and} \qquad \lim_{k \to \infty} \mathrm{V}(\hat{b}^{(k)}) = \mathrm{V}(b^\infty),$$

and the same relation holds for $\hat{y}^{(k)}$ and y. Thus,

$$\mathrm{V}(y) = \mathrm{V}\left(\lim_{k \to \infty} \{\hat{u}^{(k)} + \hat{b}^{(k)}\}\right) = \lim_{k \to \infty} \mathrm{V}\left(\hat{u}^{(k)} + \hat{b}^{(k)}\right) = \lim_{k \to \infty} \left\{\mathrm{V}(\hat{u}^{(k)}) + \mathrm{V}(\hat{b}^{(k)})\right\},$$

which yields

$$\mathrm{V}(y) = \lim_{k \to \infty} \mathrm{V}(\hat{u}^{(k)}) + \lim_{k \to \infty} \mathrm{V}(\hat{b}^{(k)}) = \mathrm{V}\left(\lim_{k \to \infty} \hat{u}^{(k)}\right) + \mathrm{V}\left(\lim_{k \to \infty} \hat{b}^{(k)}\right) = \mathrm{V}(u^\infty) + \mathrm{V}(b^\infty).$$

From equivalence (23), we deduce that $u^\infty = u_y^\star$ and $b_y^\star = b^\infty$, where $y = u_y^\star + b_y^\star$ is the Jordan minimum variation decomposition of y.

References

[1] H. Akaike (1973). Information theory and an extension of the maximum likelihood principle. *In: B. N. Petrov, B. F. Csaki (eds), Second international symposium on information theory*, 267-281.

[2] M. Ayer, H.D. Brunk, G.M. Ewing, W.T. Reid, and E. Silverman (1955). An empirical distribution function for sampling with incomplete information. *The Annals of Mathematical Statistics*, 641-647.

[3] R.E. Barlow, D.J. Bartholomew, J.M. Bremner, and H.D. Brunk (1972). *Statistical inference under order restrictions: Theory and application of isotonic regression.* John Wiley & Sons.

[4] H.H. Bauschke and J.M. Borwein (1994). Dykstra's alternating projection algorithm for two sets. *Journal of Approximation Theory*, **79**(3), 418-443.

[5] M.J. Best and N. Chakravarti (1990). Active set algorithms for isotonic regression; An unifying framework. *Mathematical Programming*, **47**(1), 425-439.

[6] H.D. Brunk (1955). Maximum likelihood estimates of monotone parameters. *The Annals of Mathematical Statistics*, 607-616.

[7] H.D. Brunk (1970). Estimation of isotonic regression. *Nonparametric Techniques in Statistical Inference*, 177-195.

[8] A. Buja, T.J. Hastie, and R. J. Tibshirani (1989). Linear smoothers and additive models. *The Annals of Statistics*, 453-510.

[9] P. Bühlmann and B. Yu (2003). Boosting with the L2 Loss: Regression and Classification. *Journal of the American Statistical Association*, **462**(98), 324-339.

[10] P.-A. Cornillon, N. Hengartner, and E. Matzner-Løber (2011). Recursive Bias Estimation for Multivariate Regression Smoothers. *arXiv* 1105.3430.

[11] P.-A. Cornillon, N. Hengartner, N. Jégou, and E. Matzner-Løber (2012). Iterative bias reduction: a comparative study. *Statistics and Computing*.

[12] P. Craven and G. Wahba (1978). Smoothing noisy data with spline functions. *Numerische Mathematik*, **31**(4), 377-403.

[13] J. Dattorro (2005). *Convex Optimization and Euclidean Distance Geometry. $\mathcal{M}\varepsilon\beta oo$*, v2012.01.28.

[14] M. Di Marzio and C.C. Taylor (2008). On boosting kernel regression. *Journal of Statistical Planning and Inference*, **138**(8), 2483-2498.

[15] C. Durot (2007). On the Lp-error of monotonicity constrained estimators. *The Annals of Statistics*, **35**(3), 1080-1104.

[16] R.L. Dykstra (1981). An isotonic regression algorithm. *Journal of Statistical Planning and Inference*, **5**(4), 355-363.

[17] R.L. Dykstra (1983). An algorithm for restricted least squares regression. *Journal of the American Statistical Association*, 837-842.

[18] J.H. Friedman and W. Stuetzle (1981). Projection pursuit regression. *Journal of the American Statistical Association*, 817-823.

[19] L. Györfi, M. Kohler, A. Krzyżak, and H. Walk (2002). *A Distribution-Free Theory of Nonparametric Regression*, Springer-Verlag, New York.

[20] W. Härdle and P. Hall (1993). On the backfitting algorithm for additive regression models. *Statistica Neerlandica*, **47**(1), 43-57.

[21] T.J. Hastie and R.J. Tibshirani (1990). *Generalized additive models.* Chapman & Hall/CRC.

[22] P. Hupé, N. Stransky, J.-P. Thiery, F. Radvanyi and E. Barillot (2004). Analysis of array CGH data: from signal ration to gain and loss to DNA regions. *Bioinformatics*, **20**(18), 3413-3422.

[23] C.M. Hurvich, J.S. Simonoff and C.L. Tsai (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **60**(2), 271-293.

[24] G. Isac and A.B. Németh (1986). Monotonicity of metric projections onto positive cones of ordered Euclidean spaces. *Archiv der Mathematik*, **46**(6), 568-576.

[25] G. Isac and A.B. Németh (1990). Isotone projection cones in Hilbert spaces and the complementarity problem. *Bollettino della Unione Matematica Italiana B.*, **7**(4), 773–802.

[26] G. Isac and A.B. Németh (1990). Projection methods, isotone projection cones, and the complementarity problem. *Journal of Mathematical Analysis and Applications*, **153**(1), 258–275.

[27] N. Jégou (2012). Régression Isotonique Itérée. *PhD Thesis*, Université Rennes 2.

[28] C.I.C. Lee (1983). The min-max algorithm and isotonic regression. *The Annals of Statistics*, **11**(2), 467-477.

[29] E. Mammen and K. Yu (2007). Additive isotone regression. *Asymptotics: Particles, Processes and Inverse Problems, IMS Lecture Notes-Monograph Series*,**55**, 179-195.

[30] A.B. Németh and S.Z. Németh (2010). How to project onto an isotone projection cone. *Linear Algebra and its Applications*, **433**(1), 41-51.

[31] S.Z. Németh (2009) Iterative methods for nonlinear complementarity problems on isotone projection cones. *Journal of Mathematical Analysis and Applications*, **350**(1), 340-347.

[32] J.D. Opsomer and D. Ruppert (1997). Fitting a bivariate additive model by local polynomial regression. *The Annals of Statistics*, **25**(1), 186-211.

[33] G. Schwarz (1978). Estimating the dimension of a model. *The Annals of Statistics*, **6**(2), 461-464.

[34] F.T. Wright (1981). The asymptotic behavior of monotone regression estimates. *The Annals of Statistics*, **9**(2), 443-448.

[35] E. Zarantonello (1971). Projections on Convex Sets in Hilbert Space and Spectral Theory, I: Projections on Convex Sets, II: Spectral Theory. *Contributions to Nonlinear Functional Analysis (Proc. Sympos. Univ. Wisconsin, Madison)*, Academic Press, New York, pp. 237-424.