ELSEVIER

Journal of Computational and Applied Mathematics 146 (2002) 459–479

JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS

www.elsevier.com/locate/cam

Fourier series approximation of separable models

U. Amato[a,*], A. Antoniadis[b], I. De Feis[a]

[a] *Istituto per le Applicazioni del Calcolo "Mauro Picone" – Sezione di Napoli CNR, Via Pietro Castellino 111, 80131 Napoli, Italy*

[b] *LMC-IMAG, Université Joseph Fourier, BP 53, 38041 Grenoble Cedex 09, France*

Received 1 December 2001; received in revised form 8 January 2002

Abstract

The approximation of a function affected by noise in several dimensions suffers from the so-called "curse of dimensionality". In this paper a Fourier series method based on regularization is developed both for uniform and random design when a restriction on the complexity of the curve such as additivity is considered in order to circumvent the problem. Optimal convergence theorems are stated and numerical experiments are shown on several test problems available in the literature together with comparisons with alternative methods.

Keywords: Smoothing data; Additive model; Regularization; Generalized cross validation; Fourier series; Uniform data design; Random data design; Nonuniform Fourier transform

1. Introduction

Multivariate data approximation is acknowledged to be a difficult problem for three main reasons: firstly, from a theoretical point of view not all convergence properties of the one-dimensional approximations extend to the higher dimensional setting, especially when issues like positivity or convexity are concerned; secondly the computational burden of the numerical methods heavily increases with the number of dimensions; lastly, there is a sparsity of data inherent to the increased volume in the high-dimensional case that degrades approximations and that could be solved only by enlarging arbitrarily the set of measurements, which can be done only rarely. For these reasons, several assumptions are made on the structure of the set of data or on the model to be approximated. For example efficient algorithms can be developed supposing a regular or line structure of the mesh

* Corresponding author. Tel.: +39-81-560-8101; fax: +39-81-560-8288.
E-mail address: u.amato@tin.it (U. Amato).

0377-0427/02/$ - see front matter
PII: S0377-0427(02)00398-9

[3]. Additive models represent a powerful and consolidated tool for simplifying approximation of multivariate functions from noised samples. They are able to circumvent the "curse of dimensionality" associated with the sparsity of multivariate data, in the sense that they restore the optimal convergence rate of univariate nonparametric regression (according to Stone [28,29] $n^{2r/(2r+1)}$, where r is regularity of the function), whereas for the general p-variate case the optimal convergence rate drops to $n^{2r/(2r+p)}$. The price to pay for the restored optimal convergence rate is the additional assumption of additivity, which restricts the complexity of the curve; several real applications fulfill this assumption. Pioneer works in additive models date to the 1980s, when Breiman and Friedman [6] and Buja et al. [7] proposed a backfitting method; essentially it projects the solution in the space of the additive functions that compose the solution and it solves the corresponding normal equations based on actual data by an iterative method. Backfitting was used extensively in application and simulation studies; however lack of full theoretical properties limited its universal acceptance. Recently, Opsomer [22] and Opsomer and Ruppert [23] proved optimal convergence rate for backfitting, even though under quite strong assumptions on the smoothing matrix and on the design.

In 1990s other methods were proposed for solving additive models. An alternative procedure, marginal integration, was developed in [17,31,19]. It is based on integration of a standard kernel estimator and it involves only averaging instead of iterative solutions. Convergence properties were easier to obtain and actually optimal convergence results at the one-dimensional rate have been established.

Parallel with development of new methods, their software implementation was made available (e.g., Venables and Ripley [32]), which increased their diffusion in applications.

Backfitting comes out to be quite computer intensive due to its iterative nature; on the contrary marginal integration is much more tractable computationally, even though not very efficient in principle; an improvement can be reached by performing one backfitting step from its solution, which however makes the method computationally intensive.

Recently, a wavelet based method (wavelet direct separation) was developed in Amato and Antoniadis [1], able to directly separate the p-dimensional additive problem into p-unidimensional ones, yielding both optimal unidimensional convergence rate and computational efficiency of the algorithm.

The present paper considers Fourier series estimators of additive nonparametric models. Fourier regularization is a consolidated tool in smoothing one-dimensional data; it has optimal convergence properties and computational properties (see, e.g., [13,24,33,14,9,30,2]). Analogously, to the wavelet case, a direct separation of the p-dimensional problem into p-unidimensional problems is performed, giving rise to similar optimal results both in terms of convergence rate and computational efficiency. An approach based on backfitting and Fourier smoothing was already proposed in Bilodeau [5], where however no separation into unidimensional problems was present.

The paper is organized as follows. Section 2 is devoted to the description of the Fourier direct separation method to build the estimator of the regression curve, both for uniform and random design. Theorems of convergence and asymptotic optimality will be proved. In Section 3 an extensive numerical experimentation will be carried out on significant test problems, coming from synthetic data and from real data, both for uniform and random design. Comparisons of performance and computational efficiency with other methods available in literature will be made.

2. Fourier direct separation

A p-order additive smoothing model is defined by

$$Y = m(\mathbf{X}) + U, \tag{1}$$

where $\mathbf{X}=(X^1,\ldots,X^p)^t$ is a vector of p, real-valued coordinates $\{X^j, 1 \leqslant j \leqslant p\}$, Y is a real valued square integrable random variable, U is a zero-mean, square integrable random variable conditionally independent of $\mathbf{X}$, and

$$m(\mathbf{X}) = \sum_{j=1}^{p} m^j(X^j),$$

where $\{m^j(\cdot), 1 \leqslant j \leqslant p\}$ denotes a collection of p unknown functions belonging to Sobolev spaces $H_j = H^{\beta_j}([0,1])$ (β_j positive integer) and periodic with their derivatives up to order β_j.

We will consider first the case of a deterministic regular design on an equally spaced grid of size a power of two for each covariate and then the case of a random design. For the sake of simplicity, the main features of the method will be presented in the bivariate case ($p = 2$), but all results obtained in this paper hold in the more general setting.

2.1. Equispaced design

Let H_1 and H_2 be two real, separable Hilbert spaces of functions defined on $[0,1]$, and let us consider the Hilbert space

$$W = H_1 \otimes H_2,$$

with the tensor product and the scalar product defined as

$$h_1 \otimes h_2 := h_1 \cdot h_2, \quad h_1 \in H_1, \ \ h_2 \in H_2,$$

$$\langle h_1 \otimes h_2, k_1 \otimes k_2 \rangle_W := \langle h_1, k_1 \rangle_{H_1} \langle h_2, k_2 \rangle_{H_2}$$

and the closed linear subspaces of W,

$$W^{\beta_1} = H_1 \otimes K^2, \qquad W^{\beta_2} = K^1 \otimes H_2,$$

where K^1 and K^2 are the linear closed subspaces of H_1 and H_2 spanned, respectively, by the constant functions $\mathbf{1}^1$ and $\mathbf{1}^2$ on $[0,1]$, defined by $\mathbf{1}^j(t) = 1$ for all $t \in [0,1]$. It is easy to see that W^{β_1} and W^{β_2} are isomorphic to H_1 and H_2, respectively.

According to the above notation, each component $m^j(X^j)$ of the additive model can be represented in W by $M^j(\mathbf{X}) \in W^{\beta_j}$ ($j = 1,2$) defined by

$$M^1(\mathbf{X}) = m^1(X^1) \otimes \mathbf{1}^2(X^2), \qquad M^2(\mathbf{X}) = \mathbf{1}^1(X^1) \otimes m^2(X^2).$$

From this it follows that

$$m(\mathbf{X}) = M^1(\mathbf{X}) + M^2(\mathbf{X}). \tag{2}$$

Let us consider the Fourier basis for $L^2([0,1])$

$$\varphi_k(X) := \mathrm{e}^{\mathrm{i}2\pi kX}.$$

It is easy to prove that

$$\left\{\varphi^1_{k_1} \otimes \varphi^2_{k_2}, k_1, k_2 \in Z\right\}$$

is a basis for $L^2([0,1]^2) = L^2([0,1]) \otimes L^2([0,1])$. Therefore $m(\mathbf{X})$ and each $M^j(\mathbf{X})$ can be decomposed as

$$m(\mathbf{X}) = \sum_{k_1,\,k_2} \mu_{k_1,k_2} \varphi^1_{k_1}(X^1)\varphi^2_{k_2}(X^2), \tag{3}$$

$$M^j(\mathbf{X}) = \sum_{k_1,k_2} \mu^j_{k_1,k_2} \varphi^1_{k_1}(X^1)\varphi^2_{k_2}(X^2), \tag{4}$$

where

$$\mu_{k_1,k_2} := \langle m(\mathbf{X}), \varphi^1_{k_1} \otimes \varphi^2_{k_2}\rangle,$$

$$\mu^j_{k_1,k_2} := \langle M^j(\mathbf{X}), \varphi^1_{k_1} \otimes \varphi^2_{k_2}\rangle,$$

$$\mu_{k_1,k_2} = \mu^1_{k_1,k_2} + \mu^2_{k_1,k_2}, \quad k_1, k_2 \in Z.$$

Since the first moment of $\varphi_k(X)$ is zero for $k \neq 0$, we have

$$\mu^1_{k_1,k_2} = 0, \quad \text{if } k_2 \neq 0,$$

$$\mu^2_{k_1,k_2} = 0, \quad \text{if } k_1 \neq 0. \tag{5}$$

In other words the two-dimensional Fourier transform of the additive model is simply given by the coefficients $\mu^1_{k_1,0}$ and μ^2_{0,k_2}, $k_1, k_2 \in Z$, the other ones being null.

Assume now to have a discretization of size $n = n_1 n_2$ of m on a regular lattice $(X^1_{i_1}, X^2_{i_2})_{\substack{0 \leqslant i_1 < n_1 \\ 0 \leqslant i_2 < n_2}}$ in $[0,1] \times [0,1]$. Assume further that $n_1 = 2^{S_1}$ and $n_2 = 2^{S_2}$. The discretized values of m are denoted by

$$m_{i_1,i_2} = M^1_{i_1,i_2} + M^2_{i_1,i_2},$$

where

$$m_{i_1,i_2} = m(X^1_{i_1}, X^2_{i_2}), \qquad M^j_{i_1,i_2} = M^j(X^1_{i_1}, X^2_{i_2}), \quad j = 1,2.$$

Let $\mu^{1,n}_{k_1,k_2}$ and $\mu^{2,n}_{k_1,k_2}$ be the discrete Fourier transforms (DFT) of $M^1_{i_1,i_2}$ and $M^2_{i_1,i_2}$, respectively; similar to the continuous Fourier transform, it is also true for the discrete Fourier transform that

$$\mu^{1,n}_{k_1,k_2} = 0, \quad -n_2/2 \leqslant k_2 \neq 0 < n_2/2,$$

$$\mu^{2,n}_{k_1,k_2} = 0, \quad -n_1/2 \leqslant k_1 \neq 0 < n_1/2.$$

This means that the matrix $(\mu^n_{k_1,k_2})_{k_1,k_2}$ of the Fourier coefficients of m_{i_1,i_2} has the following form:

$$
\begin{array}{c}
\\
k_1 = 0 \rightarrow
\end{array}
\begin{array}{c}
\begin{array}{cccccccc}
 & & & k_2 = 0 & & & & \\
 & & & \downarrow & & & &
\end{array}
\\
\begin{pmatrix}
0 & \cdots & 0 & \mu^{1,n}_{-n_1/2,0} & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & \mu^{1,n}_{-1,0} & 0 & \cdots & 0 \\
\mu^{2,n}_{0,-n_2/2} & \cdots & \mu^{2,n}_{0,-1} & \mu^{1,n}_{0,0} + \mu^{2,n}_{0,0} & \mu^{2,n}_{0,1} & \cdots & \mu^{2,n}_{0,n_2/2-1} \\
0 & \cdots & 0 & \mu^{1,n}_{1,0} & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & \mu^{1,n}_{n_1/2-1,0} & 0 & \cdots & 0
\end{pmatrix}
\end{array}
. \tag{6}
$$

Remark 1. The elements of the matrix corresponding to the nonzero coefficients do not overlap (except for the central element, where both $\mu^{1,n}_{0,0}$ and $\mu^{2,n}_{0,0}$ are present). In practice the central row of the matrix, $\mu^{2,n}_{0,k_2}$, is related only to the discrete Fourier transform of the component m^2, while the central column of the same matrix, $\mu^{1,n}_{k_1,0}$, is related only to the discrete Fourier transform of the component m^1. This will have some important consequences in the sequel.

2.1.1. Regularization

When noise affects data, the Fourier transform for the additive model developed previously, leads to estimators with small biases but with high variances. A regularization step is therefore necessary to smooth data and reduce the variability of the approximation of the function m. Fourier regularization has been widely analyzed in the literature (e.g., [13,24,33,14,9,30,2]) for one-dimensional problems, but the theory can be extended to the p-dimensional case. We defer to the above-mentioned papers for a full description of the method. Here we apply the method to models having an additive structure. Basically, this allows us to estimate all components of the unknown function m and to derive convergence results for the approximation in a much easier way by using the properties of the Fourier transform. Let us suppose to observe a noisy sample of a discretized version of m on a uniform grid of size n,

$$y_{i_1,i_2} = m_{i_1,i_2} + \varepsilon_{i_1,i_2},$$

$0 \leqslant i_1 < n_1$, $0 \leqslant i_2 < n_2$, the ε_{i_1,i_2}'s being independent random variables with mean 0 and variance σ^2. Our aim is to estimate a finite dimensional approximation $m_n(\mathbf{X})$ of $m(\mathbf{X})$ (Eq. (3)) by regularization. We shall denote by $M^j_n(\mathbf{X})$ the finite dimensional approximation of $M^j(\mathbf{X})$ $(j = 1, 2)$ (Eq. (4)). Then $m_n(\mathbf{X})$ can be written according to Eq. (2) as $m_n(\mathbf{X}) = M^1_n(\mathbf{X}) + M^2_n(\mathbf{X})$. Let us consider the "raw"

(unbiased) approximation of m

$$Y_n(\mathbf{X}) = \sum_{k_1,\,k_2} Y_{k_1,k_2}\varphi^1_{k_1}(X^1)\varphi^2_{k_2}(X^2), \tag{7}$$

$$Y_{k_1,k_2} = \mathrm{DFT}((y_{i_1,i_2})_{0\leqslant i_1<n_1, 0\leqslant i_2<n_2}),$$

for $-n_1/2 \leqslant k_1 < n_1/2$, $-n_2/2 \leqslant k_2 < n_2/2$. While this approximation is unbiased, its variance is unacceptably large and some smoothing is necessary. To estimate $m_n(\mathbf{X})$ by regularization one considers the following regularization problem:

$$\min_{G_n=G_n^1+G_n^2} \|G_n - Y_n\|^2_{L^2([0,1]^2)} + \lambda_1\|G_n^1 - \bar{G}_n^1\|^2_W + \lambda_2\|G_n^2 - \bar{G}_n^2\|^2_W, \tag{8}$$

where λ_1 and λ_2 are regularization parameters to be chosen properly and $\bar{G}_n^j := \mathrm{ave}(G_n^j)$, $j = 1, 2$. Invoking Eq. (7) and expanding $G_n(\mathbf{X})$ in terms of the Fourier basis, the continuous regularization problem can be expressed in terms of Fourier coefficients as

$$\begin{aligned}
&\min_{\substack{G^1_{k_1,k_2},G^2_{k_1,k_2}\\ -n_1/2\leqslant k_1<n_1/2\\ -n_2/2\leqslant k_2<n_2/2}} \sum_{k_1=-n_1/2}^{n_1/2-1}\sum_{k_2=-n_2/2}^{n_2/2-1} |\omega^1_{k_1,k_2}G^1_{k_1,k_2} + \omega^2_{k_1,k_2}G^2_{k_1,k_2} - Y_{k_1,k_2}|^2\\
&\quad + \lambda_1 \sum_{k_1=-n_1/2}^{n_1/2-1}\sum_{\substack{k_2=-n_2/2\\ (k_1,\,k_2)\neq(0,0)}}^{n_2/2-1} (2k_1\pi)^{2\beta_1}(2k_2\pi)^{2\beta_2}|G^1_{k_1,k_2}|^2\\
&\quad + \lambda_2 \sum_{k_1=-n_1/2}^{n_1/2-1}\sum_{\substack{k_2=-n_2/2\\ (k_1,\,k_2)\neq(0,0)}}^{n_2/2-1} (2k_1\pi)^{2\beta_1}(2k_2\pi)^{2\beta_2}|G^2_{k_1,k_2}|^2,
\end{aligned} \tag{9}$$

where according to Eq. (6) and Remark 1 we can set

$$\omega^1_{k_1,k_2} = \delta_{0,k_2},$$

$$\omega^2_{k_1,k_2} = \delta_{k_1,0}.$$

It is easy to prove that the solution of the regularization problem (9) is explicitly given by

$$G_{0,0} = Y_{0,0},$$

$$G^1_{k_1,0} = \frac{Y_{k_1,0}}{1+(2\pi k_1)^{2\beta_1}\lambda_1}, \quad -n_1/2 \leqslant k_1 \neq 0 < n_1/2,$$

$$G^2_{0,k_2} = \frac{Y_{0,k_2}}{1+(2\pi k_2)^{2\beta_2}\lambda_2}, \quad -n_2/2 \leqslant k_2 \neq 0 < n_2/2, \tag{10}$$

all other coefficients being estimated by 0.

It is worth making the following remarks:

Remark 2. Solution (10) of the regularization problem (9) is obtained by processing separately each component and, in particular, an optimal regularization parameter can be estimated for each direction.

This is a direct consequence of the particular structure of the matrix of the Fourier coefficients, see (6) and Remark 1.

Remark 3. Solution (10) of the regularization problem (8) has exactly the same expression as for the one-dimensional case for each coordinate.

The statement in Remark 3 permits to circumvent the curse of dimensionality: all theorems proved in the one-dimensional case easily hold also in the two-dimensional case; in particular, convergence of the estimator can be proved by the following theorem (here stated for the general p-dimensional case).

Theorem 4. *Let $M_n^j(\mathbf{X})$ be the finite-dimensional approximation of $m^j(X^j) : [0,1] \to H^{\beta_j}$, β_j positive integer, and $G_n^j(\mathbf{X})$ the regularized solution of problem* (8) *along the jth coordinate, $j \leqslant p$, with corresponding regularization parameter λ_j. Then*

$$E\|G_n^j - M_n^j\|^2_{L^2([0,1])} = \mathrm{O}(n^{-2\beta_j/(2\beta_j+1)}),$$

for $\lambda_j = \mathrm{O}(n^{-2\beta_j/(2\beta_j+1)})$.

Proof. See Amato and De Feis [2, Theorem 1, p. 264]. □

Note that convergence holds if $n \to \infty$, that is if any of the $n_j \to \infty, j = 1, \ldots, p$.

This proposition serves therefore as an asymptotic justification of our approach. For finite, reasonable sample sizes, it is necessary to define a data-dependent method to appropriately choose the regularization parameters λ_j. The data-based optimal regularization parameter can be estimated through the generalized cross validation (GCV) criterion developed in Craven and Wahba [8]. The estimate λ_j^{GCV} is given by the minimum over λ of the functional $\mathrm{GCV}^j(\lambda)$,

$$\mathrm{GCV}^j(\lambda) = \frac{\sum_{k=-n_j/2}^{n_j/2-1} [(1 - 1/(1 + (2\pi k)^{2\beta_j}\lambda))Y_k^j]^2}{[1/n_j \sum_{k=-n_j/2}^{n_j/2-1}(1 - 1/(1 + (2\pi k)^{2\beta_j}\lambda))]^2}, \tag{11}$$

with Y_k^j denoting the Fourier coefficients of the jth component. For the GCV estimate the following theorem can be proved (see Amato and De Feis [2, Theorem 8, p. 272]):

Theorem 5. *The choice of λ_j^{GCV} provided by the* GCV *criterion is asymptotically optimal in the average, in the sense that if λ_{n_j} is any minimizer of* $\mathrm{GCV}^j(\lambda)$, *then*

$$\lim_{n_j \to \infty} \frac{E\|(G_n^j(\mathbf{X}; \lambda_{n_j}) - M_n^j\|^2_{L^2([0,1])}}{\min_{\lambda \geqslant 0} E\|G_n^j(\mathbf{X}; \lambda) - M_n^j\|^2_{L^2([0,1])}} = 1.$$

Note that convergence for the jth coordinate holds when $n_j \to \infty$.

Summarizing, the procedure for computing the estimator of the two-dimensional additive model in the case of an equispaced design goes through the following algorithm:

Algorithm 1.

Step 1: Calculate the two-dimensional Fourier transform of input data y_{i_1,i_2}, $0 \leqslant i_1 < n_1$, $0 \leqslant i_2 < n_2$ and let Y_{k_1,k_2}, $-n_1/2 \leqslant k_1 < n_1/2$, $-n_2/2 \leqslant k_2 < n_2/2$, be the matrix of Fourier coefficients;

Step 2: Estimate the regularization parameters λ_1 and λ_2 using the GCV criterion (11) applied to the middle row Y_{0,k_2}, $-n_2/2 \leqslant k_2 \neq 0 < n_2/2$, and to the middle column, $Y_{k_1,0}$, $-n_1/2 \leqslant k_1 \neq 0 < n_1/2$, of the Fourier matrix;

Step 3: Compute the estimates of the Fourier coefficients of the components m^1 and m^2 of m ($G^1_{k_1,0}$, $-n_1/2 \leqslant k_1 \neq 0 < n_1/2$ and G^2_{0,k_2}, $-n_2/2 \leqslant k_2 \neq 0 < n_2/2$), by Eq. (10) with the regularization parameters chosen in Step 2;

Step 4: Evaluate the inverse discrete Fourier transform of $G^1_{k_1,0}$, $-n_1/2 \leqslant k_1 \neq 0 < n_1/2$, and G^2_{0,k_2}, $-n_2/2 \leqslant k_2 \neq 0 < n_2/2$; they are the approximations of the functions $m^1(X^1) - \bar{m}^1(X^1)$ and $m^2(X^2) - \bar{m}^2(X^2)$, respectively; together with the estimate $\bar{m}^1(X^1) + \bar{m}^2(X^2) = Y_{0,0}$, they give the final estimate of $m(\mathbf{X})$.

Remark 4. Thanks to the linearity of averaging and of the Fourier transform, the whole procedure is equivalent to compute the one-dimensional Fourier transform for all rows (or for all columns), to average them and then to smooth by regularization the computed Fourier coefficients. Moreover, it also follows that the method is equivalent to project (i.e., to average) directly data y_{i_1,i_2} over each coordinate independently, and to then perform an ordinary Fourier transform on the one-dimensional averaged arrays. This makes the algorithm really fast, since only averages and one-dimensional Fourier transforms are required.

2.2. Random design

We now assume observing an independent nonidentically distributed sample $(\mathbf{X}_t, Y_t)$, $t = 1, \ldots, n$, of size n, from model (1),

$$Y_t = \sum_{i=1}^{p} m_i(X_t^i) + U_t, \quad t = 1, \ldots, n. \tag{12}$$

All subsequent results hold conditionally on the $\mathbf{X}_t$'s, and therefore, by assuming as in Stone [28], that the distribution of $\mathbf{X}$ is absolutely continuous with a density that is bounded away from 0 and ∞ on its support, there is no loss of generality in supposing that the regression design is deterministic but irregular. We will also assume that the density $q(x)$ of the noise variables U_t admits a twice continuously differentiable Fisher score function $q(x)/q'(x)$ that is bounded above by a differentiable real function $\mathscr{L}(x)$ such that $\mathscr{L}(U)$ and $\mathscr{L}'(U)$ admit finite moments up to the order 3. Note that if the noise is Gaussian these conditions are automatically fulfilled. Under these conditions, following arguments similar to those of Low [18], our models under the

deterministic or the random design are asymptotically equivalent to a continuous time white noise model. The equivalence is understood in the sense of Le Cam [15] as that of statistical experiments.

According to the above, it follows that our random design setting does not involve particular modifications with respect to the deterministic design when the continuous time regression problem is considered. However significant complications arise from the discrete point of view in terms of computational efficiency. Fortunately, in the literature several algorithms have been proposed for the fast computation of the Fourier transform even in a nonuniform design setting ([4, 10–12,20,27]). All of them keep the computational cost ($O(n\log n)$) of the famous FFT algorithm. Moreover the algorithms admit an extension to the general p-dimensional case. Hereafter, in analogy with the arguments of Remark 4 for the fixed equispaced setting, we develop a method for additive models that keeps the $O(n\log n)$ computational cost and reduces the whole p-dimensional problem to the solution of p-unidimensional problems. Moreover, the appropriate choice of the relevant FFT extension, allows us to obtain without efforts comparable convergence results for the derived estimators.

Starting point of the method are Eqs. (3)–(5), that still hold in the nonequispaced setting. Therefore again only the central row, $k_1 = 0$, and column, $k_2 = 0$, are needed for restoring the second and the first components, respectively.

Under discretization the classical nonuniform discrete Fourier transforms have to be computed

$$Y_{k_j}^j = \sum_{t=1}^{n} Y_t \varphi_{k_j}^j(X_t^j), \quad k_j = -\frac{n}{2}, \ldots, \frac{n}{2} - 1, \tag{13}$$

by means of one of the known above-mentioned algorithms.

Note that analogously to Remark 4 for the fixed equispaced design, Eq. (13) is equivalent to project data Y_t on each coordinate independently, so that noise removal can proceed independently as well. We considered the same tools as in Section 2.1 for the noise removal phase, namely regularization endowed with the GCV criterion, suitably corrected for the spatial nonequispacity of input data.

The regularization problem will be

$$\min_{\substack{Y_{k_j}^{j,\lambda} \\ -\sqrt{n}/2 \leqslant k_j < \sqrt{n}/2}} \sum_{k_j=-\sqrt{n}/2}^{\sqrt{n}/2-1} |Y_{k_j}^{j,\lambda} - Y_{k_j}^j|^2 + \lambda \sum_{k_j=-\sqrt{n}/2}^{\sqrt{n}/2-1} (2\pi k_j)^{2\beta_j} |Y_{k_j}^{j,\lambda}|^2. \tag{14}$$

Note that the sum is limited to $\sqrt{n}$ elements (in practice only the central $\sqrt{n}$ frequencies are considered significant).

The solution of Eq. (14) is straightforward

$$Y_{k_j}^{j,\lambda} = \frac{Y_{k_j}^j}{1 + \lambda(2\pi k_j)^{2\beta_j}}, \quad j = 1, 2, -\frac{\sqrt{n}}{2} \leqslant k_j < \frac{\sqrt{n}}{2}. \tag{15}$$

The regularization parameter is estimated by the GCV criterion.

Table 1
Summary of average and standard error of (16) for all experiments with the `addfit` predictor

Test	$n = 64$		$n = 256$		$n = 1024$		
	Error	Ratio	Error	Ratio	Error	Ratio	
Regular design							
1	0.17 ± 0.04	1	0.14 ± 0.02	1	0.13 ± 0.01	1	m^1
	0.10 ± 0.04	1	0.06 ± 0.02	1	0.05 ± 0.01	1	m^2
2	0.17 ± 0.04	1	0.14 ± 0.02	1	0.13 ± 0.01	1	m^1
	0.10 ± 0.04	1	0.06 ± 0.02	1	0.05 ± 0.01	1	m^2
312	0.04 ± 0.01	1	0.020 ± 0.005	1	0.009 ± 0.003	1	m^1
	0.05 ± 0.02	1	0.03 ± 0.01	1	0.021 ± 0.005	1	m^2
334	0.04 ± 0.01	1	0.042 ± 0.005	1	0.042 ± 0.002	1	m^1
	0.4 ± 0.1	1	0.22 ± 0.06	1	0.15 ± 0.04	1	m^2
Random design							
1	0.16 ± 0.05	1.1 ± 0.1	0.14 ± 0.02	1.02 ± 0.03	0.13 ± 0.01	1.00 ± 0.01	m^1
	0.11 ± 0.04	1.1 ± 0.2	0.06 ± 0.02	1.1 ± 0.1	0.05 ± 0.01	1.01 ± 0.08	m^2
2	0.4 ± 0.2	1.1 ± 0.2	0.22 ± 0.07	1.01 ± 0.07	0.16 ± 0.03	1.00 ± 0.03	m^1
	0.3 ± 0.1	1.1 ± 0.2	0.17 ± 0.07	1.01 ± 0.09	0.09 ± 0.03	1.00 ± 0.05	m^2
312	0.04 ± 0.01	1.1 ± 0.2	0.020 ± 0.006	1.02 ± 0.09	0.010 ± 0.003	1.00 ± 0.06	m^1
	0.05 ± 0.02	1.1 ± 0.2	0.030 ± 0.009	1.02 ± 0.06	0.020 ± 0.005	1.00 ± 0.02	m^2
334	0.04 ± 0.01	1.1 ± 0.1	0.042 ± 0.005	1.01 ± 0.03	0.041 ± 0.003	1.004 ± 0.008	m^1
	0.4 ± 0.1	1.1 ± 0.3	0.24 ± 0.07	1.1 ± 0.2	0.16 ± 0.04	1.0 ± 0.1	m^2
412	0.05 ± 0.03	2 ± 2	0.03 ± 0.01	1.5 ± 0.9	0.016 ± 0.009	1.4 ± 0.6	m^1
	0.3 ± 0.1	1.02 ± 0.04	0.24 ± 0.05	1.01 ± 0.02	0.24 ± 0.03	1.002 ± 0.002	m^2
434	0.21 ± 0.05	1.02 ± 0.05	0.24 ± 0.05	1.01 ± 0.02	0.25 ± 0.04	1.001 ± 0.007	m^1
	0.7 ± 0.3	1.1 ± 0.3	0.7 ± 0.2	1.1 ± 0.2	0.7 ± 0.1	1.0 ± 0.1	m^2

Therefore the algorithm for solving the additive model in the nonuniform design goes through the following steps:

Algorithm 2.
Input: Nonuniform two-dimensional sample of size n: $(X_t^1, X_t^2, y_t)_{1 \leqslant t \leqslant n}$
for $j = 1, 2$ **do**
Step 1: Project the sample on coordinate X^j;
Step 2: Compute the one-dimensional nonuniform Fourier transform (Eq. (13)), $Y_{k_j}^j$;
Step 3: Choose the regularization parameter by GCV criterion;
Step 4: Regularize the computed transform by Eq. (15) in correspondence of the regularization parameter estimated in Step 3;
Step 5: Inverse Fourier transform the smoothed coefficients, which yields the approximation of the jth component.

Table 2
Summary of average and standard error of (16) for all experiments with the `funfits` predictor

Test	$n = 64$		$n = 256$		$n = 1024$		
	Error	Ratio	Error	Ratio	Error	Ratio	
Regular design							
1	0.13 ± 0.04	1.03 ± 0.09	0.07 ± 0.02	1.0 ± 0.1	0.04 ± 0.01	1.04 ± 0.08	m^1
	0.11 ± 0.04	1.0 ± 0.1	0.06 ± 0.02	1.0 ± 0.1	0.03 ± 0.01	1.05 ± 0.08	m^2
2	0.3 ± 0.1	1.5 ± 0.2	0.14 ± 0.03	1.20 ± 0.05	0.07 ± 0.01	1.18 ± 0.04	m^1
	0.21 ± 0.05	1.24 ± 0.07	0.11 ± 0.02	1.16 ± 0.04	0.06 ± 0.01	1.17 ± 0.04	m^2
312	0.02 ± 0.01	0.9 ± 0.2	0.008 ± 0.005	0.9 ± 0.2	0.004 ± 0.003	0.9 ± 0.2	m^1
	0.05 ± 0.02	1.0 ± 0.1	0.031 ± 0.008	1.0 ± 0.1	0.022 ± 0.004	1.05 ± 0.07	m^2
334	0.04 ± 0.01	1.03 ± 0.06	0.022 ± 0.005	1.04 ± 0.06	0.013 ± 0.002	1.06 ± 0.05	m^1
	0.4 ± 0.1	1.03 ± 0.08	0.21 ± 0.06	1.0 ± 0.1	0.12 ± 0.03	1.04 ± 0.08	m^2
Random design							
1	0.13 ± 0.05	1.1 ± 0.2	0.07 ± 0.02	1.1 ± 0.1	0.04 ± 0.01	1.04 ± 0.09	m^1
	0.11 ± 0.04	1.1 ± 0.2	0.06 ± 0.02	1.1 ± 0.1	0.035 ± 0.009	1.0 ± 0.1	m^2
2	0.4 ± 0.2	1.1 ± 0.3	0.20 ± 0.07	1.0 ± 0.2	0.10 ± 0.04	1.0 ± 0.1	m^1
	0.3 ± 0.1	1.1 ± 0.3	0.17 ± 0.07	1.0 ± 0.2	0.09 ± 0.03	1.0 ± 0.1	m^2
312	0.02 ± 0.01	1.1 ± 0.7	0.008 ± 0.006	1.0 ± 0.5	0.004 ± 0.003	1.0 ± 0.6	m^1
	0.05 ± 0.02	1.1 ± 0.2	0.031 ± 0.009	1.0 ± 0.1	0.018 ± 0.004	1.04 ± 0.08	m^2
334	0.03 ± 0.01	1.1 ± 0.2	0.020 ± 0.005	1.1 ± 0.1	0.013 ± 0.002	1.05 ± 0.06	m^1
	0.4 ± 0.2	1.2 ± 0.3	0.23 ± 0.06	1.1 ± 0.1	0.12 ± 0.03	1.1 ± 0.1	m^2
412	0.03 ± 0.02	1.1 ± 0.8	0.01 ± 0.01	1.0 ± 0.7	0.007 ± 0.005	1.0 ± 0.4	m^1
	0.11 ± 0.05	1.2 ± 0.3	0.07 ± 0.03	1.2 ± 0.2	0.05 ± 0.02	1.2 ± 0.1	m^2
434	0.08 ± 0.04	1.2 ± 0.3	0.05 ± 0.02	1.2 ± 0.2	0.03 ± 0.02	1.2 ± 0.1	m^1
	0.6 ± 0.3	1.2 ± 0.4	0.4 ± 0.2	1.1 ± 0.2	0.24 ± 0.08	1.0 ± 0.1	m^2

As before, the mean value is estimated through the coefficient $Y_{0,0}$. The computational cost of Algorithm 2 when the Steidl method is considered for evaluating the nonuniform Fourier transform (Step 2) is $O(\alpha n \log \alpha n)$, with integer $\alpha > 1$ being the oversampling factor (see [27] for details). There are two reasons to consider the Steidl's algorithm: first of all estimates of the approximation error are given, and second these estimates improve those obtained by Dutt and Rokhlin [10].

We stress that sparseness of data inherent to the multidimensional setting and the nonuniform design affect the solution of the proposed method by some bias. In order to reduce this problem, two different actions can be considered. The first one is to correct Eq. (13) as

$$Y_{k_j}^j = \sum_{t=1}^{n} a^j(t) Y_t \varphi_{k_j}^j(X_t^j), \quad k_j = -\frac{n}{2}, \ldots, \frac{n}{2} - 1,$$

Table 3
Summary of average and standard error of (16) for all experiments with the wavelet direct separation predictor

Test	$n = 64$		$n = 256$		$n = 1024$		
	Error	Ratio	Error	Ratio	Error	Ratio	
Regular design							
1	0.15 ± 0.04	1	0.11 ± 0.02	1	0.08 ± 0.01	1	m^1
	0.14 ± 0.04	1	0.10 ± 0.02	1	0.072 ± 0.009	1	m^2
2	0.15 ± 0.04	1	0.11 ± 0.02	1	0.08 ± 0.01	1	m^1
	0.14 ± 0.04	1	0.10 ± 0.02	1	0.072 ± 0.009	1	m^2
312	0.05 ± 0.01	1	0.035 ± 0.006	1	0.025 ± 0.003	1	m^1
	0.06 ± 0.02	1	0.045 ± 0.008	1	0.032 ± 0.004	1	m^2
334	0.04 ± 0.01	1	0.026 ± 0.004	1	0.018 ± 0.002	1	m^1
	0.5 ± 0.1	1	0.34 ± 0.06	1	0.25 ± 0.04	1	m^2
Random design							
1	0.29 ± 0.09	2.2 ± 0.9	0.20 ± 0.04	2.5 ± 0.7	0.15 ± 0.02	2.4 ± 0.4	m^1
	0.28 ± 0.09	1.7 ± 0.7	0.18 ± 0.04	2.1 ± 0.5	0.13 ± 0.02	2.1 ± 0.4	m^2
2	0.4 ± 0.1	1.4 ± 0.6	0.30 ± 0.07	1.3 ± 0.3	0.22 ± 0.03	1.3 ± 0.2	m^1
	0.4 ± 0.1	1.2 ± 0.3	0.26 ± 0.07	1.2 ± 0.2	0.19 ± 0.04	1.2 ± 0.2	m^2
312	0.19 ± 0.06	$3 \pm 1.$	0.13 ± 0.03	5 ± 1	0.10 ± 0.01	5 ± 1	m^1
	0.3 ± 0.1	2 ± 1	0.23 ± 0.05	4 ± 1	0.16 ± 0.02	4.6 ± 0.9	m^2
334	0.12 ± 0.04	1.01 ± 0.07	0.05 ± 0.01	1.02 ± 0.07	0.033 ± 0.006	1.04 ± 0.07	m^1
	3 ± 1	9 ± 4	2.4 ± 0.5	10 ± 3	1.7 ± 0.2	9 ± 2	m^2
412	0.20 ± 0.07	2 ± 1	0.13 ± 0.06	2 ± 1	0.11 ± 0.05	3 ± 1	m^1
	0.4 ± 0.2	2 ± 1	0.25 ± 0.09	2 ± 1	0.20 ± 0.06	3 ± 1	m^2
434	0.17 ± 0.07	1.1 ± 0.2	0.10 ± 0.05	1.1 ± 0.2	0.08 ± 0.08	1.1 ± 0.1	m^1
	2 ± 2	4 ± 4	2 ± 1	4 ± 3	1.5 ± 1	4 ± 3	m^2

with

$$a^j(t) = \begin{cases} (X_{t+1}^j - X_t^j)/2, & t = 1, \\ (X_{t+1}^j - X_{t-1}^j)/2, & 2 < t < n, \\ (X_t^j - X_{t-1}^j)/2, & t = n, \end{cases}$$

in order to get convergence of the (nonuniform) discrete Fourier transform to the Fourier coefficients. The second action is to reiterate the algorithm in a backfitting fashion, just as made in Linton [16] for marginal integration. Computational speed of Algorithm 2 suggests Fourier direct separation to have a better efficiency.

From the results of Elbel and Steidl [12] and Steidl [27] on the approximation properties of the nonuniform Fourier transform and its inverse, the effect of using a nonuniform Fourier transform in deriving the error of approximation leads, after some tedious by straightforward calculations, to an upperbound of the error by an expression proportional to the error for the equispaced deterministic design, the proportionality constant being independent of the sample size. It follows then, by a reasoning similar to that of Amato and De Feis [2] for the deterministic equispaced case,

Table 4
Summary of average and standard error of (16) for all experiments with the Fourier direct separation predictor

Test	$n = 64$		$n = 256$		$n = 1024$		
	Error	Ratio	Error	Ratio	Error	Ratio	
Regular design							
1	0.14 ± 0.04	1	0.09 ± 0.02	1	0.05 ± 0.01	1	m^1
	0.13 ± 0.04	1	0.09 ± 0.02	1	0.05 ± 0.01	1	m^2
2	0.14 ± 0.04	1	0.09 ± 0.02	1	0.05 ± 0.01	1	m^1
	0.13 ± 0.04	1	0.09 ± 0.02	1	0.05 ± 0.01	1	m^2
312	0.04 ± 0.01	1	0.030 ± 0.007	1	0.017 ± 0.004	1	m^1
	0.06 ± 0.02	1	0.043 ± 0.008	1	0.028 ± 0.004	1	m^2
334	0.04 ± 0.01	1	0.026 ± 0.004	1	0.017 ± 0.002	1	m^1
	0.4 ± 0.1	1	0.23 ± 0.08	1	0.12 ± 0.05	1	m^2
Random design							
1	0.22 ± 0.08	1.4 ± 0.5	0.11 ± 0.02	1.1 ± 0.2	0.06 ± 0.01	1.02 ± 0.08	m^1
	0.30 ± 0.09	1.3 ± 0.3	0.13 ± 0.03	1.1 ± 0.2	0.07 ± 0.01	1.06 ± 0.09	m^2
2	0.4 ± 0.1	1.1 ± 0.3	0.24 ± 0.08	1.0 ± 0.1	0.15 ± 0.04	1.01 ± 0.09	m^1
	0.4 ± 0.2	1.1 ± 0.2	0.25 ± 0.07	1.0 ± 0.1	0.14 ± 0.04	1.01 ± 0.08	m^2
312	0.14 ± 0.04	1.2 ± 0.3	0.05 ± 0.01	1.2 ± 0.2	0.024 ± 0.004	1.1 ± 0.1	m^1
	0.24 ± 0.07	1.3 ± 0.4	0.10 ± 0.02	1.2 ± 0.2	0.043 ± 0.007	1.1 ± 0.1	m^2
334	0.19 ± 0.06	1.0 ± 0.1	0.07 ± 0.02	1.03 ± 0.04	0.031 ± 0.006	1.01 ± 0.02	m^1
	1.1 ± 0.3	2 ± 1	0.5 ± 0.2	1.9 ± 0.9	0.17 ± 0.06	1.3 ± 0.4	m^2
412	0.3 ± 0.1	1.1 ± 0.3	0.19 ± 0.08	1.0 ± 0.2	0.16 ± 0.07	1.1 ± 0.2	m^1
	0.5 ± 0.2	1.4 ± 0.4	0.4 ± 0.1	1.3 ± 0.2	0.3 ± 0.1	1.2 ± 0.1	m^2
434	0.4 ± 0.2	1.0 ± 0.1	0.3 ± 0.2	1.01 ± 0.04	0.3 ± 0.2	1.00 ± 0.05	m^1
	0.8 ± 0.3	1.5 ± 0.9	0.6 ± 0.7	1 ± 2	0.4 ± 0.3	1 ± 1	m^2

that, under the general assumptions in this section, the approximation shares the same asymptotic properties.

3. Numerical experiments

A simulation study was conducted to assess the performance of the approximation proposed in this paper. We have applied the Fourier direct separation method to several test problems considered previously in the literature and to a real problem. In the simulations we make comparisons with other known methods, namely the wavelet direct separation [1], the spline smoothing (`funfits` code [21]), and the local polynomial regression (`addfit` code [23]). The algorithm to generate the Fourier direct separation approximation was implemented in MATLAB. The MATLAB script is available upon request for the interested readers.

For the simulation study we have used several two-dimensional models (purely additive and with an interaction term) and several types of noise. They are the following:

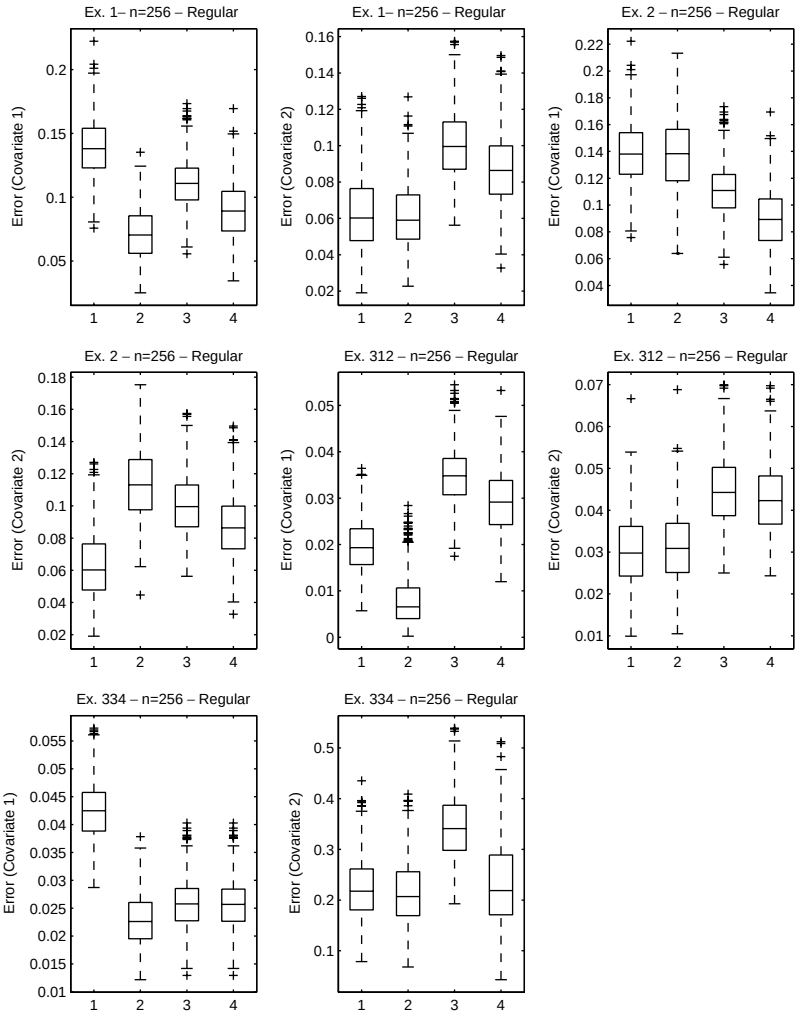

Fig. 1. Boxplot of error (16) for regular design, n=256, Examples 7, 8, 912, 934 and the predictors (1: kernel; 2: splines; 3: wavelet direct separation; 4: Fourier direct separation).

Problem 7 (Sperlich et al. [26]).

$$m = m^1 + m^2 + \varepsilon,$$
$$m^1 = 1.5\sin(-1.5X^1), \quad -2 \leqslant X^1 \leqslant 2,$$
$$m^2 = -(X^2)^2 + E(X^2)^2, \quad -2 \leqslant X^2 \leqslant 2.$$

X^1 and X^2 come from a regular design or from a uniform distribution on $[-2, 2]$. *Noise* ε *is independent Gaussian with zero mean and standard deviation* 0.5.

Problem 8. $m = m^1 + m^2 + X^1X^2 + \varepsilon$, *where* X^1, X^2, m^1, m^2 *and* ε *are the same as Problem* 7.

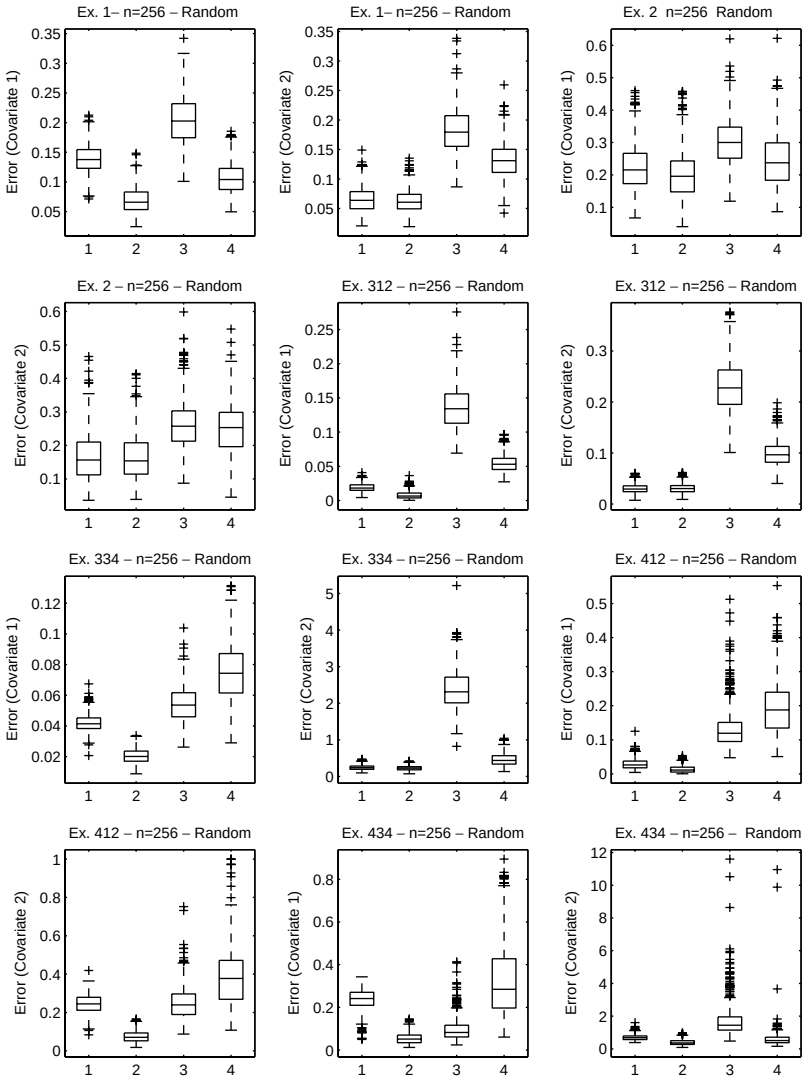

Fig. 2. Boxplot of error (16) for random design, $n = 256$, Examples 7, 8, 912, 934, 1012, 1034 and the predictors (1: kernel; 2: splines; 3: wavelet direct separation; 4: Fourier direct separation).

Problem 9 (Sperlich et al. [25]).

$m = m^1 + m^2 + \varepsilon$,

where m^1 *and* m^2 *are two different functions from the set*

$$f_1(X) = 2X,$$

$$f_2(X) = X^2 - EX^2,$$

$$f_3(X) = \mathrm{e}^X - E\mathrm{e}^X,$$

$$f_4(X) = 0.5\sin(-1.5X).$$

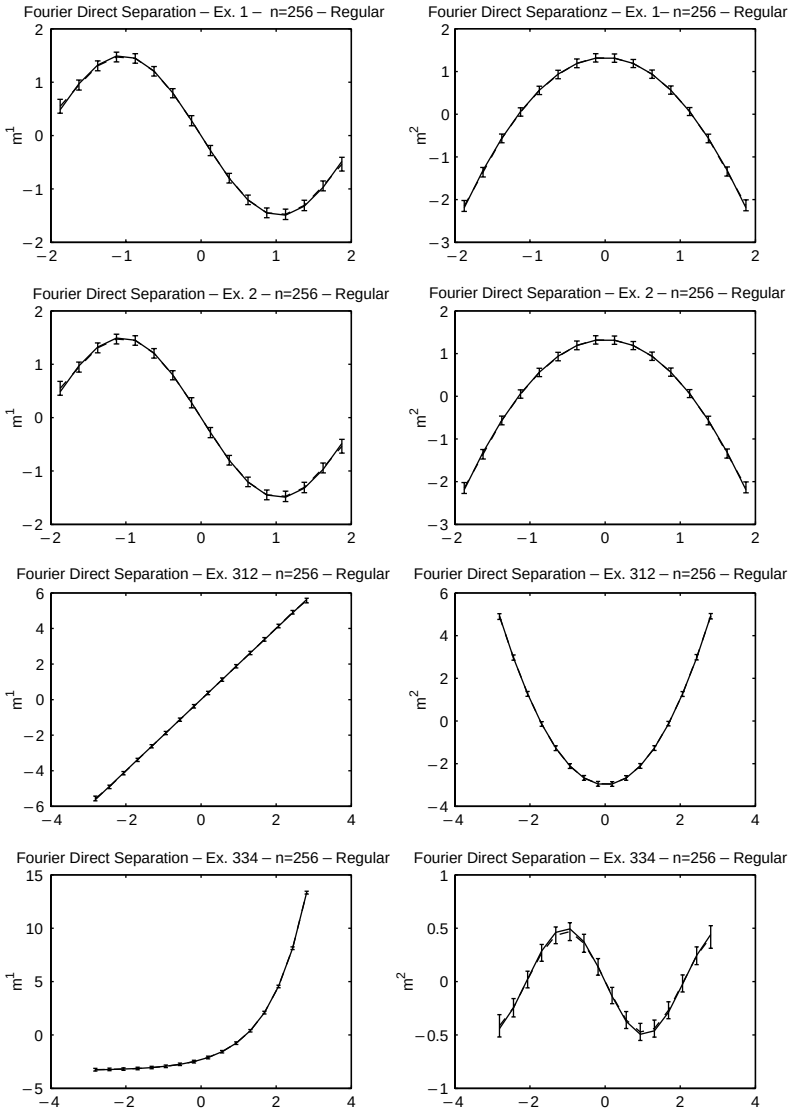

Fig. 3. Average of predicted functions (dashed line) and MSE for Examples 7, 8, 912, 934 $n = 256$, regular design and the Fourier direct separation predictor. True functions are shown by a solid line.

X can come from a regular design or from a uniform random distribution, both on $[-3,3]$. *Then the test problem is individuated by the code* $3pq$, *where* p *and* $q(1 \leqslant p,q \leqslant 4)$ *are the functions chosen among the four listed. Noise* ε *is independent Gaussian with zero mean and standard deviation* 0.5.

Problem 10. *Same as Problem* 9, *but for* X^1 *and* X^2 *having uncorrelated Gaussian distributions with mean* 0 *and variance* 1 (*test code* $4pq$).

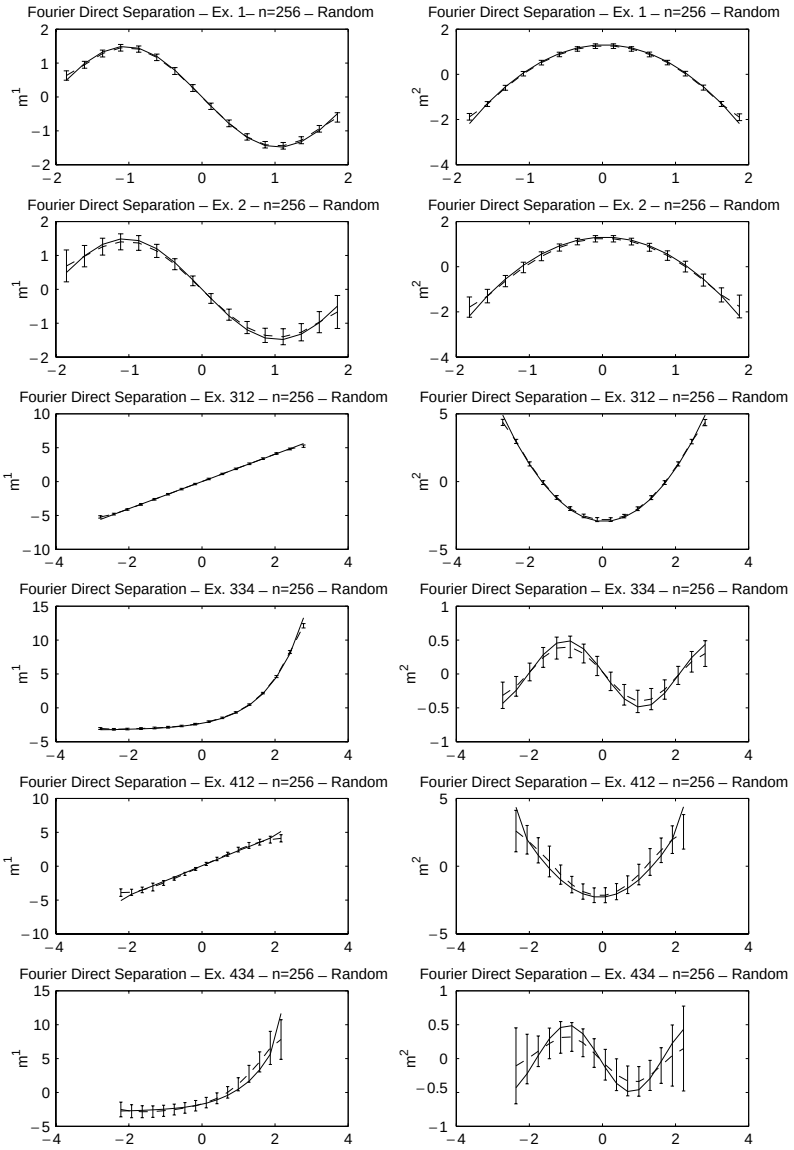

Fig. 4. Average of predicted functions (dashed line) and mean square error for Examples 7, 8, 912, 934, 1012, 1034, $n = 256$, random design and the Fourier direct separation predictor. True functions are shown by a solid line.

We have considered also a real data example. The `minitri` data (available in the `funfits` package) reports the results from a mini triathlon sponsored by Bud Lite, held in Cary, NC, June 1990. Times are in minutes for the male 30–34 group. Components are: swim ($\frac{1}{2}$ mile), bike (15 miles), run (4 miles). We have applied a two-dimensional additive model to this dataset in order to explain the run record times as a function of the swim and bike record times.

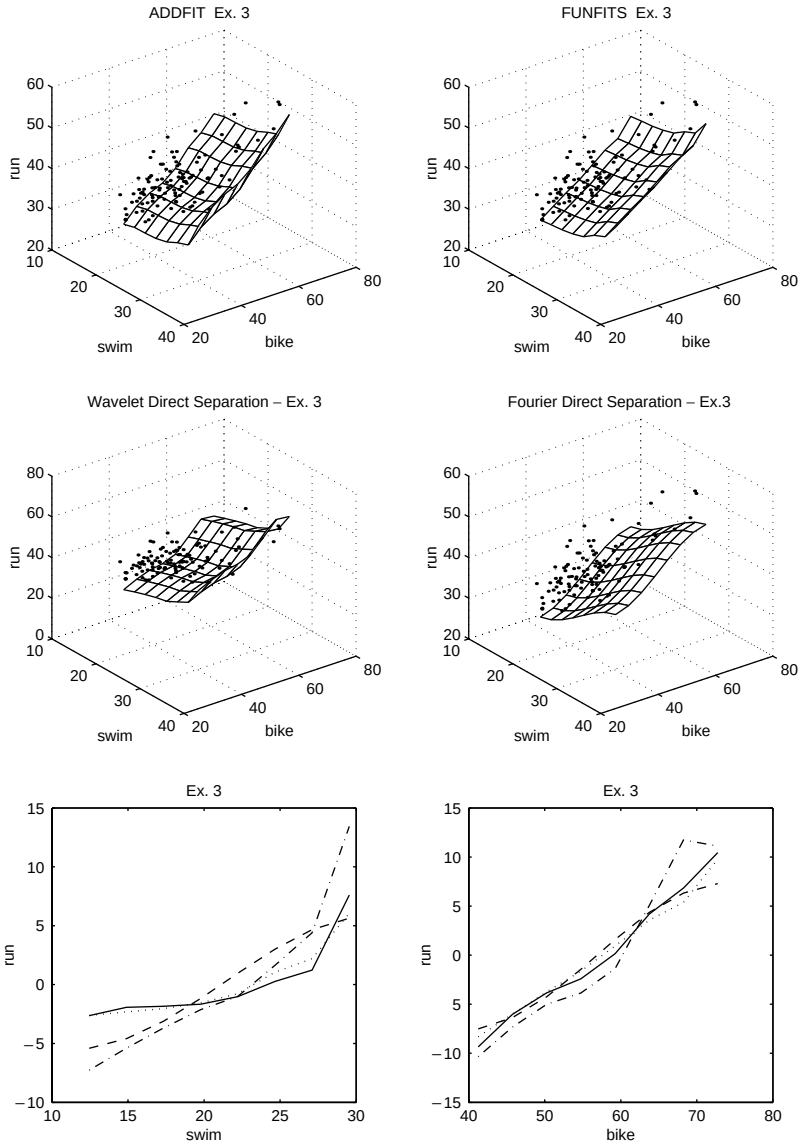

Fig. 5. Surfaces predicted by the methods on the `minitri` data. The lower panel represents plots of the predicted components by kernel (dashed line), spline (dotted line), wavelet (dotted-dashed line), and Fourier (continuous line) approximations.

The Fourier direct separation approximation was built considering the cosine transform, in order to deal with nonperiodic functions.

For each test problem, several sample sizes were considered and for each sample size/test function combination, 500 Monte Carlo replications were performed. To fairly compare the approximation methods, the codes were run on the same set of simulated data.

To measure the quality of the fit of each approximation, $\hat{m}^j$, separately we have computed the mean square error (MSE), s_k^j for each component separately,

$$s_k^j = \frac{1}{n}\sum_{t=1}^{n}(\hat{m}^j(X_t^j) - m^j(X_t^j))^2, \quad j = 1,2,\ k = 1,\ldots,500,$$

together with their averages over Monte Carlo replications (MASE).

Recalling however that the $\hat{m}^j$ and m^j are only unique up to some additive constant and to make sure that the comparisons between methods only reflect substantial differences in shape between the true and the estimated function we have used as a quality of fit criterion

$$s_k^j = \frac{1}{n}\sum_{t=1}^{n}[(\hat{m}^j(X_t^j) - \mathrm{ave}(\hat{m}^j)) - (m^j(X_t^j) - \mathrm{ave}(m^j))]^2. \tag{16}$$

The 500 values of s_k^j for each smoother are displayed graphically by means of boxplots and summarized by their average with a standard error in suitable tables.

Moreover, in order to show the behavior of the approximations with respect to the test functions, we have computed bias and MSE of each predictor as a function of each component. Their average over the samples is shown in suitable error plots.

Finally, in order to rank our smoothers we derived an efficiency measure. In particular the MASE comparisons do not usually provide a simple ranking between estimators. Our approach is to measure the efficiency of the procedures against a similar procedure estimating $m^1(X^1)$ based on a knowledge of $m^2(\cdot)$ (and conversely). In our case the efficiency is to compare the ratio of the s_k^j by the MSE of the estimated m^j when one smooths the partial errors $Y_t - m^2(X_t^2)$ $(Y_t - m^1(X_t^1))$ in the direction of interest X^1 (X^2).

Experiments were carried out under different settings: size ($n = 64, 256, 1024$) and design (fixed and random). In the case of random design Steidl's algorithm was considered for the computation of the nonuniform Fourier transform, with parameter $\alpha = 2$.

Tables 1–4 summarize all experiments in terms of average and standard errors of (16) for the four predictors. Ratio with respect to the values obtained in the unidimensional fit as explained above is also shown. In the case of Problem 3 results are not available since true values are not known. In addition, for a fair comparison with respect to both components and across functions, the MASE is normalized with respect to the L_2 norm of the true function, so that values shown in the tables are to be read as the square root of the actual relative MASE. Fourier direct separation is endowed with just one backfitting step in order to reduce the bias of the estimator for the random design. This is enough to give top performance of the method as confirmed by efficiency measure close to 1.

A visual representation of these results is shown in the boxplots of Fig. 1 for regular designs and Fig. 2 for random designs, both for $n = 256$. Figs. 3 and 4 show plots of bias and MSE as a function of covariate for $n = 256$ and for regular and random design, respectively, in the case of the Fourier direct separation predictor. We defer to Amato and Antoniadis [1] for a comparison with analogous plots of the `addfit`, `funfits` and wavelet direct separation predictors. Fig. 5 shows the three-dimensional scatterplots and surfaces as predicted by all the methods on the `minitri` data.

Analysis of tables and figures shows that the `funfits` predictor generally performs better than its competitors; we mention that `addfit` method often fails to find a solution, due to convergence problems in the choice of bandwidth; in these cases empirical methods have been used. Moreover

Table 5
CPU execution times (sec) of `addfit` (`K`), `funfits` (`S`), wavelet direct separation (`W`) and Fourier direct separation (`F`) for different sizes of the sample and regular and random design

Test	$n = 64$				$n = 256$				$n = 1024$			
	K	S	W	F	K	S	W	F	K	S	W	F
Regular design												
1	0.08	0.01	0.01	0.01	0.36	0.02	0.01	0.01	0.96	0.1	0.02	0.01
2	0.06	0.01	0.01	0.01	0.3	0.02	0.01	0.01	0.96	0.11	0.02	0.01
312	0.06	0.01	0.01	0.01	0.27	0.01	0.01	0.01	0.9	0.05	0.02	0.01
334	0.06	0.01	0.01	0.01	0.31	0.01	0.01	0.01	0.97	0.07	0.02	0.01
Random design												
1	0.1	0.02	0.03	0.02	0.34	0.06	0.04	0.03	1.34	0.25	0.06	0.11
2	0.1	0.01	0.03	0.02	0.36	0.06	0.04	0.03	1.35	0.24	0.06	0.11
312	0.1	0.01	0.03	0.02	0.38	0.06	0.04	0.03	1.36	0.21	0.06	0.11
334	0.1	0.01	0.03	0.02	0.36	0.06	0.04	0.03	1.39	0.26	0.06	0.11
412	0.09	0.01	0.03	0.02	0.37	0.06	0.04	0.03	1.36	0.23	0.06	0.11
434	0.09	0.01	0.03	0.02	0.34	0.06	0.04	0.03	1.37	0.26	0.06	0.11

in the case of regular design Fourier direct separation method retrieves functions with less bias than `funfits`, even though variance is larger.

Finally, Table 5 shows CPU execution times for the four predictors under different settings of the experiments (MATLAB code was run on a 800 MHz Pentium III PC). In both regular and random design, Fourier and wavelet direct separation show the best performances, with Fourier direct separation outperforming significantly the corresponding wavelet method. In conclusion we can assert that the Fourier direct separation method provides the best compromise between accuracy and computational cost.

Acknowledgements

Work supported by Agenzia Spaziale Italiana (ASI)

References

[1] U. Amato, A. Antoniadis, Adaptive wavelet series estimation in separable nonparametric regression models, Statist. Comput. 11 373–394.
[2] U. Amato, I. De Feis, Convergence for the regularized inversion of Fourier series, J. Comput. Appl. Math. 87 (1997) 261–284.
[3] I.J. Anderson, J.C. Mason, Efficient matrix methods for the true least-squares approximation of structured multivariate data, in: K. Chui, L. Schumaker (Eds.), Approximation Theory VIII, Vol. 1, World Scientific Publishing, Singapore, 1995, pp. 7–14.
[4] G. Beylkin, On the fast Fourier transforms of functions with singularities, Appl. Comput. Harmonics Anal. 2 (1995) 363–381.

[5] M. Bilodeau, Fourier smoother and additive models, Canad. J. Statist. 20 (1992) 257–269.
[6] L. Breiman, J.H. Friedman, Estimating optimal transformations for multiple regression and correlation (with discussion) J. Amer Statist. Assoc. 80 (1985) 580–619.
[7] A. Buja, T.J. Hastie, R.J. Tibshirani, Linear smoothers and additive models, Ann. Statist. 17 (1989) 453–510.
[8] P. Craven, G. Wahba, Smoothing noisy data with spline functions, Numer. Math. 59 (1979) 377–403.
[9] D. De Falco, M. Frontini, L. Gotusso, Recent results on the regularization of Fourier polynomials, Appl. Math. Comput. 66 (1994) 1–8.
[10] A. Dutt, V. Rokhlin, Fast fourier transforms for nonequispaced data, SIAM J. Sci. Comput. 14 (1993) 1368–1393.
[11] A. Dutt, V. Rokhlin, Fast fourier transforms for nonequispaced data II, Appl. Comput. Harmonics Anal. 2 (1995) 85–100.
[12] B. Elbel, G. Steidl, Fast fourier transforms for nonequispaced grids, in: C.K. Chui, L. Schumaker (Eds.), Approximation Theory IX, Vol. 2, Vanderbilt University Press, Nashville, TN, 1998, pp. 39–46.
[13] M. Frontini, L. Gotusso, On smoothing of approximate Fourier representations, Istituto per Applicazioni del Calcolo, Series III, Vol. 208, Rome, Italy, 1981.
[14] M. Frontini, L. Gotusso, A regularization method for discrete Fourier polynomials, J. Comput. Appl. Math. 49 (1993) 71–78.
[15] L. Le Cam, Asymptotic Methods in Statistical Decision Theory, Springer, New York, 1986.
[16] O. Linton, Efficient estimation of additive nonparametric regression models, Biometrika 84 (1986) 469–473.
[17] O. Linton, J.P. Nielsen, A kernel method of estimating structured nonparametric regression based on marginal integration, Biometrika 82 (1995) 93–100.
[18] M.G. Low, Renormalization and white noise approximation for nonparametric functional estimation, Ann. Statist. 20 (1992) 545–554.
[19] W.K. Newey, Kernel estimation of partial means, Econom. Theory 10 (1994) 233–253.
[20] N. Nguyen, Q.H. Liu, The regular Fourier matrices and nonuniform fast Fourier transforms, SIAM J. Sci. Comput. 21 (1999) 283–293.
[21] D. Nychka, B. Bailey, S. Ellner, P. Haaland, P. O'Connel, FUNFITS: data analysis and statistical tools for estimating functions, Technical Report, Raleigh, North Carolina State University, 1993.
[22] J.D. Opsomer, Optimal bandwidth selection for fitting an additive model by local polynomial regression, Ph.D. Dissertation, Cornell University, 1995.
[23] J.D. Opsomer, D. Ruppert, Fitting a bivariate additive model by local polynomial regression, Ann. Statist. 25 (1997) 186–211.
[24] J. Rice, M. Rosenblatt, Integrated mean squared error of a smoothing spline, J. Approx. Theory 33 (1981) 353–369.
[25] S. Sperlich, O.B. Linton, W.Härdle, A simulation comparison between integration and backfitting methods of estimating separable nonparametric regression models, Technical Report, Humboldt, University of Berlin, 1997.
[26] S. Sperlich, D. Tjøstheim, L. Yang, Nonparametric estimation and testing of interaction additive models, Technical Report, Humboldt, University of Berlin, 1998.
[27] G. Steidl, A note on fast Fourier transforms for nonequispaced grids, Adv. Comput. Math. 9 (1998) 337–352.
[28] C.J. Stone, Additive regression and other nonparametric models, Ann. Statist. 13 (1985) 689–705.
[29] C.J. Stone, The dimensionality reduction principle for generalized additive models, Ann. Statist. 14 (1986) 590–606.
[30] M. Tasche, N. Weyrich, Smoothing inversion of Fourier series using generalized cross-validation, Results Math. 29 (1996) 183–195.
[31] D. Tjøstheim, B. Auestad, Nonparametric identification of nonlinear time series: projections, J. Amer. Statist. Assoc. 89 (1994) 1398–1409.
[32] W.N. Venables, B. Ripley, Modern Applied Statistics with S-Plus, 3rd Edition, Springer, New York, 1999.
[33] G. Wahba, Spline Models for Observational Data, SIAM, Philadelphia PA, 1990.