Advanced Modeling and Simulation in Engineering Sciences

RESEARCH ARTICLE

A generalized Fourier transform by means of change of variables within multilinear approximation

Mathilde Chevreuil[1]* and Myriam Slama[2]

*Correspondence:
mathilde.chevreuil@univ-nantes.fr
[1]GeM UMR CNRS 6183,
Université de Nantes, Nantes,
France
Full list of author information is
available at the end of the article

Abstract

The paper deals with approximations of periodic functions that play a significant role in harmonic analysis. The approach revisits the trigonometric polynomials, seen as combinations of functions, and proposes to extend the class of models of the combined functions to a wider class of functions. The key here is to use structured functions, that have low complexity, with suitable functional representation and adapted parametrizations for the approximation. Such representation enables to approximate multivariate functions with few eventually random samples. The new parametrization is determined automatically with a greedy procedure, and a low rank format is used for the approximation associated with each new parametrization. A supervised learning algorithm is used for the approximation of a function of multiple random variables in tree-based tensor format, here the particular Tensor Train format. Adaptive strategies using statistical error estimates are proposed for the selection of the underlying tensor bases and the ranks for the Tensor-Train format. The method is applied for the estimation of the wall pressure for a flow over a cylinder for a range of low to medium Reynolds numbers for which we observe two flow regimes: a laminar flow with periodic vortex shedding and a laminar boundary layer with a turbulent wake (sub-critic regime). The automatic re-parametrization enables here to take into account the specific periodic feature of the pressure.

Keywords: Adapted parametrization, Statistical learning, High-dimensional approximation, Tree tensor networks, Hierarchical tensor format Tensor train format, Adaptive algorithms

Introduction

The approximation of periodic functions plays a significant role in harmonic analysis. In the case of the dynamical response of structures, these responses can notably be highly perturbed by low variability on the model and it then becomes necessary to develop reliable and efficient tools for the prediction of the dynamical random response. We are here interested in constructing an approximation of a multivariate function with periodicity in one or more dimensions based on observations. This is of special interest for instance in uncertainty quantification for vibroacoustic problems where the structure is excited with a harmonic wall pressure field. The wall pressure field is a multivariate function which

depends on the time and on a set of variables, such as the Reynolds number. In practice, the wall pressure is computed for different instances of the variables and consequently on different discrete time grids depending on the instances. When fine discrete models are involved, the evaluations of the model are costly and we may have access only to sparse information in terms of instances of the variables and of observation time interval.

The set of trigonometric polynomials is well adapted for representing periodic functions. Indeed real trigonometric functions of degree m are written in the following form:

$$v(t) = a_0 + \sum_{n=1}^{m} (a_n \cos(nt) + b_n \sin(nt))$$

where the terms are periodic functions with period 2π. This class of functions has nice properties for approximation use, especially if a function f is continuous and given a tolerance ε, there exists a trigonometric polynomial v such that $|f(t) - v(t)| < \varepsilon$ for all t. As mentioned above, in many applications we have access to samples of the polynomial from which we want to determine its coefficients. The trigonometric polynomials are therefore linked to discrete-time signal processing, e.g. the Discrete-Time Fourier Transform (DTFT) converts a sequence of length N on an equally spaced time grid into a trigonometric polynomial of degree $N - 1$. The DTFT is extended to the d-dimensional case in the same manner. Given a sample of a multivariate function, the construction of an approximation of the function in the class of trigonometric functions has been widely addressed and the methods for constructing such representation generally depend on the discretization (see [1–3] and the references herein).

In the present paper an alternative approach is proposed in order to tackle such problems using a sample that is not necessarily structured. It is based on statistical learning methods [4] for multi-dimensional problems with s variables where the multivariate output function $u(x_1, \cdots, x_s)$ of the model, identified with an order-s tensor, is approximated in a parametrized subset of functions

$$\mathcal{M} = \big\{ v = \Psi(\mathbf{a}); \mathbf{a} \in A \big\} \tag{1}$$

where the parameter $\mathbf{a}$ belongs to some set of parameters A and Ψ is a multilinear function with respect to the parameters $\mathbf{a}$. The key idea is to propose an adapted parametrization with m new variables $z_i = g_i(x_1, \ldots, x_s), i = 1, \ldots, m$, for the computation of the response so to obtain structured approximations with low complexity by exploiting the periodicity of the function in some dimensions. In the last decade, active subspace [6] and basis adaptation methods [5] have been proposed to find low dimensional structures using adapted parametrizations with reduced dimension. The first class of methods consist in detecting the directions of strongest variability of a function using gradient evaluations, and then construct an approximation of the function exploiting the defined low-dimension subspace. In [7], active subspaces have been advantageously used for quantifying uncertainty of hypersonic flows around a cylinder. The second class of methods, namely basis adaptation methods, identifies the dominant effects in the form of linear combinations of the input variables and the adapted reduction of the representation is performed through a projection technique. In the current work the change of variables is extended to a wider class of functions g and is done with a method inspired from the projection pursuit method [8] which defines automatically and sequentially the new variables to add. The approximation with this possibly high-dimensional new set of variables is created by exploiting

specific low-dimensional structures of the function such as sparsity [9] and low rank [10] structures of the function to be approximated that enable the construction of an approximation using few samples as introduced in [11,12]. In the latter reference, the output is approximated in suitable low-rank tensor subsets of the form

$$\mathcal{M} = \left\{ v = \Psi(\mathbf{a}^1, \cdots, \mathbf{a}^L); \mathbf{a}^l \in \mathbb{R}^{n_l} \right\} \tag{2}$$

where Ψ is a multilinear map with parameters $\mathbf{a}^l$, $l = 1, \ldots, L$. This is a special case of (1) with $\mathbf{a} = (\mathbf{a}^1, \ldots, \mathbf{a}^L)$ and $A = \mathbb{R}^{n_1} \times \ldots \times \mathbb{R}^{n_L}$. The dimension of the parametrization $\sum_{l=1}^{L} n_l$ grows linearly with m and thus makes possible the approximation with high dimension m.

The first part of the paper presents an interpretation of the trigonometric functions as a composition of functions $h \circ g$ with specific structured representations for h, and proposes a generalization of the representation for h. The second part is dedicated to the algorithm for constructing the rank structured approximation combined with the change of variable $g(x)$ used to handle periodic functions. Finally the last part illustrates the method on the pressure of a flow around a cylinder.

Periodic functions

Let $u \in \mathcal{H}$ be a multivariate function, with $\mathcal{H}$ a Hilbert space, which depends on a set of independent variables $X = (X_1, \ldots, X_s)$. In the present paper, we consider the specific case where the function u is periodic with respect to one variable denoted τ so that we have $X = (\Xi, \tau)$, with $\Xi = (X_1, \ldots, X_d)$ and $\tau = X_{d+1}$, and $s = d + 1$.

A variable X_ν has values in $\mathcal{X}_\nu$ and an associated measure $\mathrm{d}p_\nu$, $1 \leq \nu \leq s$. The variable X has values in $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_s$ and an associated measure $\mathrm{d}p = \mathrm{d}p_1 \times \cdots \times \mathrm{d}p_s$. The variables X_ν, $\nu = 1, \ldots, d$ can be random variables, $\mathrm{d}p_\nu$ being in that case a probability measure on $\mathcal{X}_\nu$. Let $\mathcal{H}$ be the Hilbert space defined on $\mathcal{X}$, it is a tensor space $\mathcal{H} = \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_s$ with $\mathcal{H}_\nu$ a Hilbert space defined on $\mathcal{X}_\nu$. We consider that $\mathcal{H}_\nu \subset L^2_{p_\nu}(\mathcal{X}_\nu)$ is a finite dimensional subspace of square-integrable functions equipped with the norm $\|u\|_\nu^2 = \int_{\mathcal{X}_\nu} u^2 \mathrm{d}p_\nu$ and $\mathcal{H}$ is a subspace of $L^2_p(\mathcal{X})$ equipped with the canonical norm $\|u\|^2 = \int_{\mathcal{X}} u^2 \mathrm{d}p$. Let $\{\psi_i^\nu\}_{i=1}^{P_\nu}$ be an orthonormal basis of $\mathcal{H}_\nu$ and $\{\psi_{i_1}^1 \otimes \cdots \otimes \psi_{i_s}^s\}_{(i_1,\ldots,i_s) \in [1,\ldots,P_1] \times \cdots \times [1,\ldots,P_d]}$ be an orthonormal basis of $\mathcal{H}$.

A natural representation of the periodic function u can be obtained using the Fourier series. In the following, $x = (\xi, t)$ will denote an observation of X with ξ an observation of $\Xi = (X_1, \ldots, X_d)$ and t an observation of τ, i.e. a point in the periodic dimension.

Trigonometric functions as a composition of functions

Let us consider a T-periodic and continuous real valued function $u : \mathbb{R} \to \mathbb{R}$. It can be represented by its truncated Fourier series which is a sum of harmonic functions:

$$u(t) \approx \sum_{n=0}^{m-1} a_n \cos(\omega_n t) + b_n \sin(\omega_n t) \tag{3}$$

where the circular frequencies are such that $\omega_n = n\omega_f$ (multiplicative constraint on ω_n), with $\omega_f = \frac{2\pi}{T}$, and the coefficients a_n and b_n defined as follows:

$$a_n = \frac{\omega_f}{2\pi} \int_{-\frac{\pi}{\omega_f}}^{\frac{\pi}{\omega_f}} u(t)\cos(\omega_n t)\, dt \quad \text{and} \quad b_n = \frac{\omega_f}{2\pi} \int_{-\frac{\pi}{\omega_f}}^{\frac{\pi}{\omega_f}} u(t)\sin(\omega_n t)\, dt, \tag{4}$$

for $n = 0, \ldots, m-1$.

The truncated Fourier series can be seen as a composition of functions of the form:

$$v(t) = h(g_1(t), \ldots, g_m(t))$$

where h is an additive model $h(z_1, \ldots, z_m) = \sum_{n=1}^{m} h_n(z_n)$ with $h_n(z_n) = a_n \cos(z_n) + b_n \sin(z_n) \in V$, where $V = \text{span}\left\{1, \cos(\cdot), \sin(\cdot)\right\}$, and $g_n(t) = \omega_{n-1}t$ for $n = 1, \ldots, m$.

We propose here to extend the Fourier series to a more generalized framework where the function v is a multivariate function depending both on t and ξ, where the circular frequencies ω_n are chosen adaptively with no multiplicative constraint and where h is chosen in a wider set of functions than additive models.

Generalizing Fourier series for the representation of a multivariate function in tensor format

Let us focus on a function $u(x) = u(\xi, t)$, periodic with respect to t. A natural representation of u is thus by means of a relevant change of variables:

$$u(x) \approx h(g_1(x), \ldots, g_m(x), \xi) \tag{5}$$

where $g_n : \mathbb{R}^{d+1} \to \mathbb{R}$, $n = 1, \ldots, m$ are new variables chosen under the form

$$g_n(\xi, t) = \omega_n(\xi)t \tag{6}$$

and where $h \in V^{\otimes m} \otimes \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_d : \mathbb{R}^{[0,T]^m \times \Xi_1 \times \cdots \times \Xi_d} \to \mathbb{R}$. Representing the periodic function under the form (5) can lead to the definition of a high number m of new variables so that we will consider subsets of low rank tensor formats for the high dimensional function h of dimension $M = m + d$.

A M-dimensional function v in a subset of tensors can be written:

$$v(z) = \Psi(z)(\mathbf{a}_1, \ldots, \mathbf{a}_L) \tag{7}$$

where $\Psi(x)$ is a multilinear map with parameters $(\mathbf{a}_1, \ldots, \mathbf{a}_L)$. We consider here a model class of rank-structured functions associated with a notion of rank. A well known rank is the canonical rank associated to the sum of multiplicative models. The canonical rank of a function h is the minimal integer $\text{rank}_C(v) = r$ such that

$$v(z) = \sum_{k=1}^{r} v_k^1(z_1) \ldots v_k^M(z_M)$$

and we define the subset of canonical tensors:

$$\mathcal{T}_r^C = \{v \in \mathcal{H} : \text{rank}_C(v) \leq r\}. \tag{8}$$

It can be associated to the parametrized representation (7) with $L = M$, where $\mathbf{a}_l \in \mathbb{R}^{r \times n_l}$ with n_l the dimension of the functional basis on which the functions v_k^l are represented, $k = 1, \ldots, r$ and $l = 1, \ldots, M$. We can consider other notions of ranks which provide different models with lower complexity. The α-rank of v, denoted by $\mathrm{rank}_\alpha(v)$, is the minimal integer r_α such that

$$v(z) = \sum_{k=1}^{r_\alpha} v_k^\alpha(z_\alpha) v_k^{\alpha^c}(z_{\alpha^c})$$

with $\alpha \subset \{1, \ldots, M\}$, and z_α and z_{α^c} complementary groups of variables. The T-rank of v, denoted by $\mathrm{rank}_T(v) = \{\mathrm{rank}_\alpha(v) : \alpha \in T\}$, is the tuple $r = \{r_\alpha\}_{\alpha \in T}$ such that

$$v(z) = \sum_{k=1}^{r_\alpha} v_k^\alpha(z_\alpha) v_k^{\alpha^c}(z_{\alpha^c}), \quad \forall \alpha \in T$$

where T is a collection of subsets of $\{1, \ldots, M\}$. We define the subset of rank-structured functions:

$$\mathcal{T}_r^T = \{v \in \mathcal{H} \; : \; \mathrm{rank}_T(v) \leq r\}. \tag{9}$$

The complexity of the associated parametrized representations of tensors is linear with the dimension M and polynomial with the ranks.

Statistical learning method for approximating a function in tensor format with a change of variables

Supervised statistical learning

We consider a model that returns a real-valued variable $Y = u(X)$. An approximation v of the function u, also referred to as metamodel, can be obtained by minimizing the risk

$$\mathcal{R}(v) = \int_\mathcal{X} \ell(u(x), v(x)) \mathrm{d}p(x)$$

over a model class $\mathcal{M}$. The loss function ℓ measures a distance between the observation $u(x)$ and the prediction $v(x)$. In the case of the least squares method, it is chosen as $\ell(y, v(x)) = (y - v(x))^2$.

Let $S = \{(x^k, y^k) : 1 \leq k \leq N\}$ be a sample of n realizations of (X, Y). In practice, the approximation is constructed by minimizing the empirical risk

$$\mathcal{R}_S(v) = \frac{1}{N} \sum_{k=1}^{N} \ell(y^k, v(x^k)), \tag{10}$$

taken as an estimator of the risk. A regularization term R can be used for stability reasons when the training sample is small. An approximation $\tilde{u}$ of u is then solution of

$$\min_{v \in \mathcal{M}} \frac{1}{N} \sum_{k=1}^{N} \left(y^k - v(x^k) \right)^2 + \lambda R(v), \tag{11}$$

with the regularization parameter $\lambda \geq 0$, chosen or computed. The accuracy of the metamodel $\tilde{u}$ is estimated using an estimator of the L^2 error. In practice, the number of numerical experiments is too small to sacrifice part of it for the error estimation. The

error is thus estimated using a k-fold cross validation estimator and more specifically the leave-one-out cross validation estimator [4] which can be easily evaluated by constructing one single metamodel [13]. Cross validation estimators can be used for model selection.

In the following, we present a method to determine $\tilde{u}(x)$ in a sequence of model classes

$$\mathcal{M}(m) = \left\{ v(x) = h(g_1(x), \ldots, g_m(x), \xi); \ h \in \mathcal{M}_h, \ g_i \in \mathcal{M}_g \text{ for } i = 1, \ldots, m \right\} \qquad (12)$$

with $\mathcal{M}_h$ a linear or multilinear model class and $g_i \in \mathcal{M}_g$ a linear model class. We consider here multilinear models and more specifically the tensor subset $\mathcal{T}_r^T$ in order to handle the possibly high dimensional $m + d$ problem. We briefly recall the learning algorithm in a tensor subset [12] and then present the automatic computation of the new variables $z_i = g_i(x)$, $i = 1, \ldots, m$.

Learning with tensor formats

Let $z \in \mathbb{R}^M$, an approximation of u in a tensor subset (2) can be obtained by minimizing the empirical least squares risk:

$$\min_{\mathbf{a}^1, \ldots, \mathbf{a}^L} \frac{1}{N} \sum_{k=1}^{N} \left(y^k - \Psi(z^k)(\mathbf{a}_1, \ldots, \mathbf{a}_L) \right)^2 + \sum_{i=1}^{L} \lambda_i R_i(\mathbf{a}_i) \qquad (13)$$

where $\lambda_i R_i(\mathbf{a}_i)$ are regularization functions. Problem (13) is solved using an alternating minimization algorithm which consists in successively solving an optimization problem on $\mathbf{a}^j$

$$\min_{\mathbf{a}_j} \frac{1}{N} \sum_{k=1}^{N} \left(y^k - \Psi(z^k)(\ldots, \mathbf{a}_j, \ldots) \right)^2 + \lambda_j R_j(\mathbf{a}_j) \qquad (14)$$

for fixed parameters $\mathbf{a}_i$, $i \neq j$. Introducing the linear map $\Psi^j(z)(\mathbf{a}_j) = \Psi(z)(\mathbf{a}_1, \ldots, \mathbf{a}_L)$, problem (14) yields the following learning problem with a linear model

$$\min_{\mathbf{a}_j} \frac{1}{N} \sum_{k=1}^{N} \left(y^k - \Psi^j(z^k)(\mathbf{a}_j) \right)^2 + \lambda_j R_j(\mathbf{a}_j). \qquad (15)$$

If $R_j(\mathbf{a}_j) = \left\| \mathbf{a}_j \right\|_1$, with $\|\mathbf{v}\|_1 = \sum_{i=1}^{\#v} |v_i|$ the ℓ_1-norm, problem (15) is a convex optimization problem known as Lasso [14] or basis pursuit [15]. The ℓ_1-norm is a sparsity inducing regularization function that yields a solution $\mathbf{a}_j$ of (15) that may have coefficients equal to zero. The Lasso is solved using the modified least angle regression algorithm (LARS in [16]).

The algorithm to solve Problem (13) is described in [11] for the canonical tensor format, and in [12] for the tree-based tensor format, which is a special case of rank-structured tensor where T is a dimension partition tree. Adaptive algorithms are proposed to automatically select the tuple yielding a good convergence of the approximation with respect to its complexity. For tree-based tensor formats, at an iteration i, given an approximation v^i of u with T-rank $(r_\alpha^i)_{\alpha \in T}$, the strategy consists in estimating and studying the truncation error $\min_{\text{rank}_\alpha(v) \leq r_\alpha^i} \mathcal{R}(v) - \mathcal{R}(u)$ for different α in T, and choosing to increase the ranks r_α^i associated with the indices α yielding the highest errors. The algorithm and more details in the tree-based tensor format case can be found in [12,17].

Learning method with automatic definition of new variables

We now present the method used to automatically search an adapted parametrization of the problem by looking for favored directions in the space of the $d + 1$ input variables. It consists in writing the approximation under the form (5) where $g_n(x)$ can be represented with a parametrized linear map

$$g_n(x) = \mathbf{w}_n^\intercal \boldsymbol{\varphi}(x) = \sum_{j=1}^{p} w_{n,j} \varphi_j(x) \tag{16}$$

with $\mathbf{w}_n = (w_{n,1}, \ldots, w_{n,p})^\intercal \in \mathbb{R}^p$ the vector of parameters of the representation of g_n on an orthonormal functional basis $\{\varphi_j\}_{j=1}^{p}$ of $\mathcal{H}$, and h a M-dimensional function in the model class of rank structured formats $\mathcal{T}_r^C$ or $\mathcal{T}_r^T$ that can be represented with a parametrized multilinear map

$$h(z) = \Psi(z)(\mathbf{a}^1, \ldots, \mathbf{a}^L) \tag{17}$$

with parameters $\mathbf{a}^l$, $l = 1, \ldots, L$. The new set of variables $z = (z_1, \ldots, z_m, \xi)$ is such that $z_n = g_n(x)$, $n = 1, \ldots, m$.

The method is a Projection Pursuit like method [8] that is generalized to a larger class of models for h than just additive models. It is shown in [18] that under reasonable conditions, for $h \in \mathcal{T}_r^T$ and $g \in \mathcal{H}$ we have $h \circ g \in L_p^2(\mathcal{X})$. The approximation $\tilde{u}$ of the form (5) is thus parametrized as follows:

$$\tilde{u}(x) = \Psi\left(\mathbf{w}_1^\intercal \boldsymbol{\varphi}(x), \ldots, \mathbf{w}_m^\intercal \boldsymbol{\varphi}(x), \xi\right)(\mathbf{a}^1, \ldots, \mathbf{a}^L). \tag{18}$$

Let $(z_1, \ldots, z_{m-1})$ be an initial set of variables. A new variable $z_m = g_m(x)$ is introduced using Algorithm 1.

Algorithm 1 Addition of a new variable z_m.

Require: orthonormal set of functions $\{\varphi_j\}_{j=1}^p$, parametrization $\mathbf{w} = (\mathbf{w}_1, \ldots, \mathbf{w}_{m-1})$ of the set of variables $(z_1, \ldots, z_{m-1})$
Ensure: parametrization $\mathbf{w} = (\mathbf{w}_1, \ldots, \mathbf{w}_m)$ of $(z_1, \ldots, z_m)$
 1: set $\mathbf{w}_m$ such that $\mathbf{w}_m^\intercal \boldsymbol{\varphi}(x) = 1 \; \forall x \in \mathbb{R}^d$
 2: set $\mathbf{w} = (\mathbf{w}_1, \ldots, \mathbf{w}_m)$
 3: set $z_m = \mathbf{w}_m^\intercal \boldsymbol{\varphi}(\cdot)$

The parameters $\mathbf{a}_l$, $l = 1, \ldots, L$, and $\mathbf{w}_n$, $n = 1, \ldots, m$, solve the minimization problem (11) over the model class $\mathcal{M}(m)$:

$$\min_{\{\mathbf{a}_l\}_{l=1}^L, \{\mathbf{w}_n\}_{n=1}^m} \frac{1}{N} \sum_{k=1}^{N} \left(y^k - \Psi\left(\mathbf{w}_1^\intercal \boldsymbol{\varphi}(x^k), \ldots, \mathbf{w}_m^\intercal \boldsymbol{\varphi}(x^k), \xi^k\right)(\mathbf{a}^1, \ldots, \mathbf{a}^L)\right)^2 + \sum_{i=1}^{L} \lambda_i R_i(\mathbf{a}_i), \tag{19}$$

with $x^k = (\xi^k, t^k)$. The solution of this problem is found by alternatively solving

$$\min_{\mathbf{a}_1, \ldots, \mathbf{a}_L} \frac{1}{N} \sum_{k=1}^{N} \left(y^k - \Psi\left(\mathbf{w}_1^\intercal \boldsymbol{\varphi}(x^k), \ldots, \mathbf{w}_m^\intercal \boldsymbol{\varphi}(x^k), \xi^k\right)(\mathbf{a}^1, \ldots, \mathbf{a}^L)\right)^2 + \sum_{i=1}^{L} \lambda_i R_i(\mathbf{a}_i) \tag{20}$$

for fixed $(\mathbf{w}_1, \ldots, \mathbf{w}_m)$ using a learning algorithm with rank adaptation [11,12] and

$$\min_{\mathbf{w}_1 \in \mathbb{R}^p, \ldots, \mathbf{w}_m \in \mathbb{R}^p} \frac{1}{N} \sum_{k=1}^{N} \left(y^k - \Psi \left(\mathbf{w}_1^\mathsf{T} \boldsymbol{\varphi}(x^k), \ldots, \mathbf{w}_m^\mathsf{T} \boldsymbol{\varphi}(x^k), \xi^k \right) (\mathbf{a}^1, \ldots, \mathbf{a}^L) \right)^2 \tag{21}$$

for fixed $(\mathbf{a}^1, \ldots, \mathbf{a}^L)$. The optimization problem (21) is a nonlinear least squares problem that is solved with a Gauss-Newton algorithm. The overall algorithm is presented in Algorithm 2.

Algorithm 2 Learning algorithm for an approximation with automatic definition of the new variables.

Require: sample $S = \{(x^k, y^k) : 1 \leq k \leq N\}$, collection of M parametrizations $\Psi_i(\cdot)(\mathbf{a}^1, \ldots, \mathbf{a}^{L_i})$ of functions of $i + d$ variables, initial values for the parameters $(\mathbf{a}^1, \ldots, \mathbf{a}^{L_i})$, basis for the representation of the g_i, $i = 1, \ldots, m_{\max}$
Ensure: approximation $v(x) = \Psi(\mathbf{w}_1^\mathsf{T} \boldsymbol{\varphi}(x), \ldots, \mathbf{w}_m^\mathsf{T} \boldsymbol{\varphi}(x), \xi)(\mathbf{a}^1, \ldots, \mathbf{a}^L)$ with selected optimal $m \in \{1, \ldots, m_{\max}\}$
  1: **for** $i = 1, \ldots, m_{\max}$ **do**
  2:    add one variable $g_i(x) = \mathbf{w}_i^\mathsf{T} \boldsymbol{\varphi}(x)$ using Algorithm 1
  3:    set $z = (\mathbf{w}_1^\mathsf{T} \boldsymbol{\varphi}(x), \ldots, \mathbf{w}_i^\mathsf{T} \boldsymbol{\varphi}(x), \xi)$
  4:    **while** not converged **do**
  5:       compute the approximation $\Psi_i(z)(\mathbf{a}^1, \ldots, \mathbf{a}^{L_m})$ solution of (20)
  6:       compute $\mathbf{w}_1, \ldots, \mathbf{w}_i$, solutions of (21)
  7:       set $z = (\mathbf{w}_1^\mathsf{T} \boldsymbol{\varphi}(x), \ldots, \mathbf{w}_i^\mathsf{T} \boldsymbol{\varphi}(x), \xi)$
  8:    **end while**
  9: **end for**
 10: select m which minimizes $\mathcal{R}_V(v_m)$ with V a validation set independent of S
 11: set $\Psi \leftarrow \Psi_m$, $L \leftarrow L_m$

In step 5 of algorithm 2, the parametrization of the model is to be selected in a collection of parametrizations. In this paper, we consider the rank structured function $\mathcal{T}_r^T$ where T is a dimension partition tree over $\{1, \ldots, m\}$, which corresponds to the model class of functions in tree-based tensor format [19], a particular class of tensor networks [20]. The new node associated with the new variable $z_i = g_i(x)$ is added at the top of the tree. The representation of a function v in $\mathcal{T}_r^T(\mathcal{H})$ requires the storage of a number of parameters that depends both on the collection T and on the associated T-rank r, as well as on the dimensions of the functional spaces in each dimension. To reduce the number of coefficients that need to be computed during the learning process of a tensor v, one can then represent $v \in \mathcal{T}_r^T(\mathcal{H})$ for different collections T and associated T-ranks r, choosing the ones yielding the smallest storage complexity. Furthermore, this adaptation can prove useful when dealing with changes of variables, as introduced in the previous subsection, because it can remove the difficulty of how to add a variable: no matter what the initial ordering of the variables is, an adaptation procedure may be able to find an optimal one yielding a smaller storage complexity. When T is a dimension partition tree, a stochastic algorithm is presented in [12] for trying to reduce the storage complexity of a tree-based tensor format at a given accuracy. This adaptation is not considered in the paper.

Change of variables for periodic functions

In this section we present a generalization of the Fourier series where one does not need a structured sample (e.g. a grid) in the variable t. Indeed when the function to approximate is known to be periodic with respect to t, the periodicity of the approximation can be forced. It is done on the one hand by introducing a functional basis φ such that $g_n = \mathbf{w}_n^\mathsf{T} \boldsymbol{\varphi}$

in (16) can be identified to (6) by choosing

$$\varphi_j(x) = \varphi_j^{(\xi)}(\xi)t \tag{22}$$

where $\{\varphi_j^{(\xi)}\}$ is a d-dimensional tensorized orthogonal basis of $\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_d$. The circular frequencies in (6) are then expressed as:

$$\omega_n(\xi) = \sum_{j=1}^{p} w_{n,j}\varphi_j^{(\xi)}(\xi) = \mathbf{w}_n^{\mathsf{T}}\boldsymbol{\varphi}^{(\xi)}(\xi). \tag{23}$$

On the other hand, we choose bases of trigonometric functions $\{\psi_i^n\}_{i=1}^{P_n}$ for the representation of h in the dimensions associated with the new variables z_n, $n = 1, \ldots, m$.

Let T_{max} be the maximal width of the observation interval in the dimension t. Supposing this interval is large enough to include the largest periods of the periodic functions that can be learned, the guarantee for the approximation not to have larger periods is to constrain the circular frequencies in (6) such that:

$$\omega_n \geq \frac{2\pi}{\mathsf{T}_{max}}, \ n = 1, \ldots, m. \tag{24}$$

This constraint is imposed for all values taken by ξ in S. Using expression (23), it is recast under the form

$$-A\mathbf{w} \leq -B \tag{25}$$

where $\mathbf{w} = (\mathbf{w}_1, \ldots, \mathbf{w}_m) \in \mathbb{R}^{p \times m}$, $A \in \mathbb{R}^{N \times p}$ is the array of evaluations of $\varphi_j^{(\xi)}(\xi)$ for the values $(x_1^k, \ldots, x_d^k)_{k=1}^N$ of ξ in the training set S:

$$A = \begin{bmatrix} \varphi_1^{(\xi)}(x_1^1, \ldots, x_d^1) & \ldots & \varphi_p^{(\xi)}(x_1^1, \ldots, x_d^1) \\ \vdots & \ddots & \vdots \\ \varphi_1^{(\xi)}(x_1^N, \ldots, x_d^N) & \ldots & \varphi_p^{(\xi)}(x_1^N, \ldots, x_d^N) \end{bmatrix}, \tag{26}$$
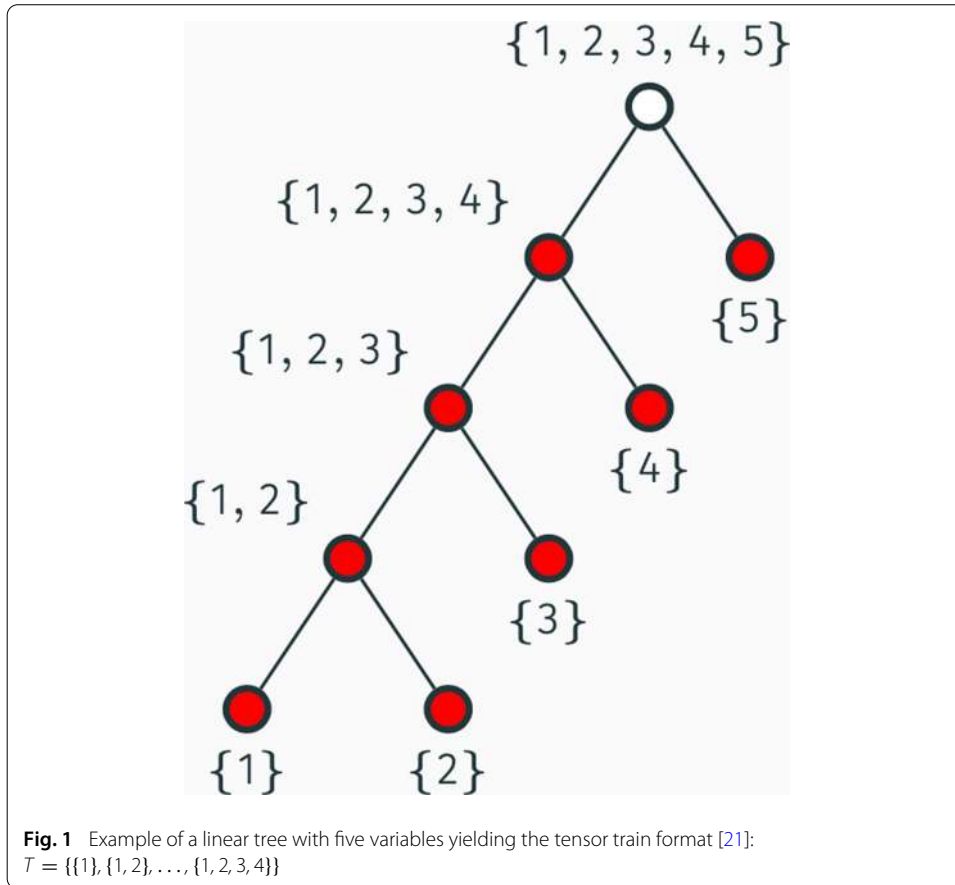
and $B \in \mathbb{R}^{N \times m}$ is a full array with values $-2\pi/\mathsf{T}_{max}$. The optimization problem (21) for the computation of the parameters $\mathbf{w}$ is replaced with the constrained optimization problem

$$\min_{\substack{\mathbf{w}_1 \in \mathbb{R}^p, \ldots, \mathbf{w}_m \in \mathbb{R}^p \\ -A\mathbf{w} \leq -B}} \frac{1}{N} \sum_{k=1}^{N} \left(y^k - \Psi\left(\mathbf{w}_1\varphi(x^k), \ldots, \mathbf{w}_m\varphi(x^k), \xi^k\right)(\mathbf{a}^1, \ldots, \mathbf{a}^L) \right)^2 \tag{27}$$

which is solved with a NonLinear Programming (NLP) method.

Application

The method is applied to the prediction of the wall pressure p for a flow over a cylinder for two ranges (low and medium) of Reynolds numbers for which we observe two flow regimes: a laminar flow with periodic vortex shedding and a laminar boundary layer with a turbulent wake (sub-critic regime). The automatic re-parametrization enables here to take into account the specific periodic feature of the pressure p with the time. The variables of the problem are $X = (Re, \Theta, \tau)$ with τ the time, Re the Reynolds number and Θ the angular steps, that take values in $\mathcal{X} = \mathcal{X}_{Re} \times \mathcal{X}_{\Theta} \times \mathcal{X}_{\tau}$ and we have $d = 2$. The pressure is evaluated on a tensor grid with

Fig. 1 Example of a linear tree with five variables yielding the tensor train format [21]:
$T = \{\{1\}, \{1, 2\}, \dots, \{1, 2, 3, 4\}\}$

- at low Reynolds numbers: 300 time steps in $\mathcal{X}_\tau = [27\,s, 29.4\,s]$, 50 simulations of Re chosen uniformly on $\mathcal{X}_{Re} = [70, 200]$ and 128 angular steps in $\mathcal{X}_\Theta = [0, 2\pi[$,
- at higher Reynolds numbers: 1500 time steps, 11 simulations of Re chosen on $\mathcal{X}_{Re} = [7000, 13000]$ and 320 angular steps in $\mathcal{X}_\Theta = [0, 2\pi[$.

A representation (5) with new variables $g_n(\xi, t)$, where $\xi = (Re, \theta)$, is computed with Algorithm 2 where the optimization problem (21) of step 6 is replaced with the constrained optimization problem (27). The new variables g_n are represented on a basis of functions as in (22) where $\{\varphi_j^{(\xi)}\}_{j=1}^P$ is a polynomial basis with maximal partial degree 2 ($P = 9$). We choose for the model class of h the tensor train (TT) format associated to the linear tree $T = \{\{1\}, \{1, 2\}, \dots, \{1, \dots, M - 1\}\}$ (see Fig. 1) where $M = m + d$, the associated parametrized subset is

$$\left\{ v = \Psi(\mathbf{a}^1, \dots, \mathbf{a}^M); \mathbf{a}^l \in \mathbb{R}^{r_{l-1} \times n_l \times r_l}, l = 1, \dots, M \right\}$$

with $r = (r_0, r_1, \dots, r_M)$ the TT-rank where $r_0 = r_M = 1$, and n_l is the dimension of the functional basis $\{\psi_l\}_{l=1}^{n_l}$ in dimension l for the representation of h. Here we use trigonometric bases $\{1, \cos(x), \sin(x)\}$ in the dimensions of $z_i = g_i(t, \theta, Re)$ and polynomial bases in the dimensions of θ and Re. The sequential quadratic programming (SQL) method is used to solve (27).

The accuracy of the metamodel $\tilde{u}$ is estimated using the unbiased test error estimator based on a test set S_{test} of N_{test} realizations of (X, Y) independent of the training set S:

$$e_{\text{test}}^2 = \frac{\mathcal{R}_{S_{test}}(\tilde{u})}{\mathcal{R}_{S_{test}}(0)} \tag{28}$$

where $\mathcal{R}_{S_{test}}(v)$ is the empirical risk defined in (10). Considering the least squares estimator, we have $e_{\text{test}}^2 = \frac{\frac{1}{N_{test}}\sum_{k=1}^{N_{\text{test}}}(y^k-\tilde{u}(x^k))^2}{\frac{1}{N_{test}}\sum_{k=1}^{N_{\text{test}}}(y^k)^2}$.

Low Reynolds numbers

We first consider the approximation of the wall pressure for low Reynolds numbers for which the data is given for 50 values of the Reynolds number in $\mathcal{X}_{Re} = [70, 200]$. The approximation is constructed with the following setting:

- for g_i: polynomial bases with a maximal degree of 1 for t and 2 for θ and Re,
- for h: polynomial bases of degree 14 for θ and 3 for Re,
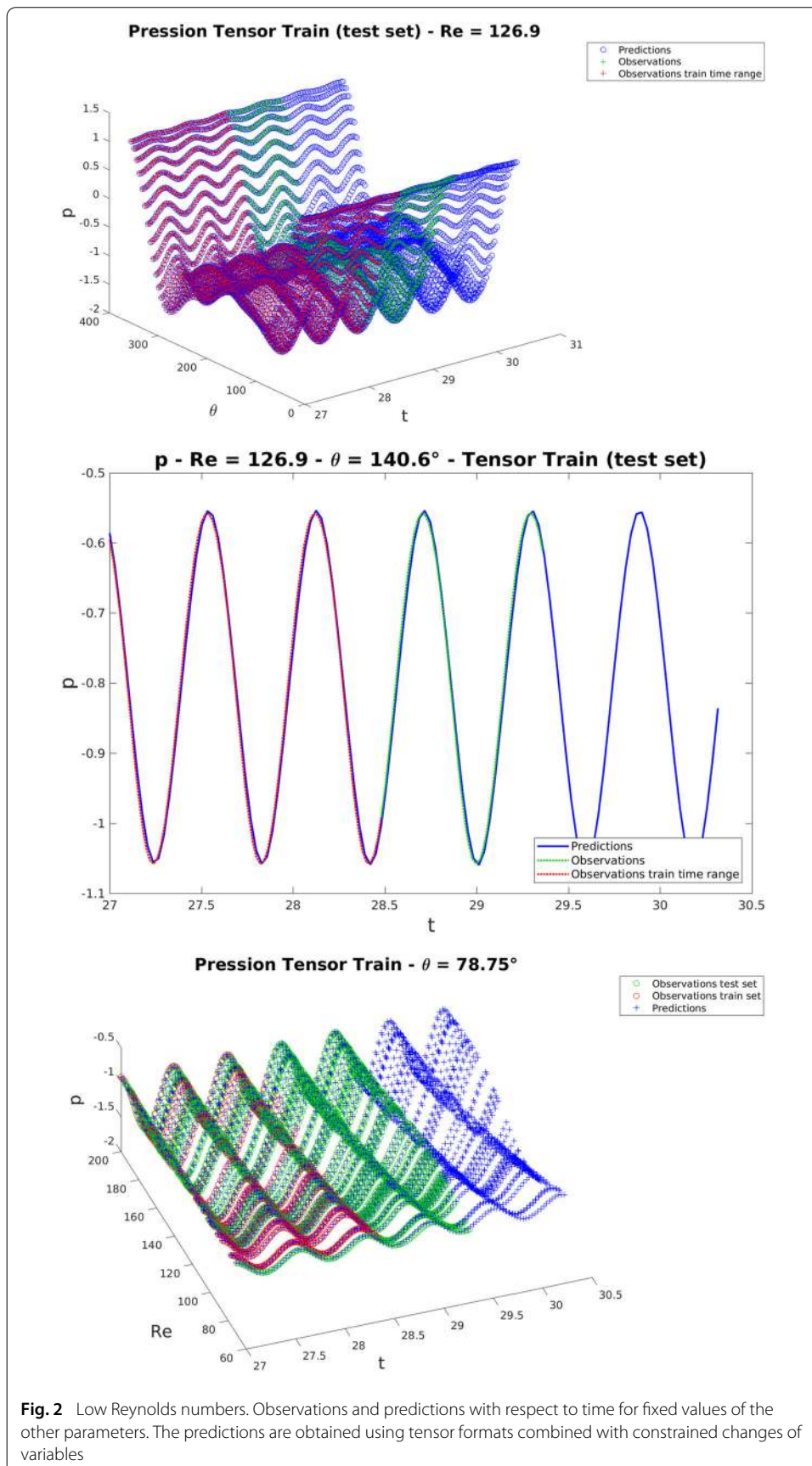- training sample S using 20 simulations of Re and considering only the 181 first time steps.

The algorithm provided an approximation with dimension $M = 6 : X = (g_1(\xi), \ldots, g_4(\xi), \xi)$ and TT-ranks $r = [1\ 3\ 4\ 4\ 2\ 1]$. The model error was estimated using two different samples:
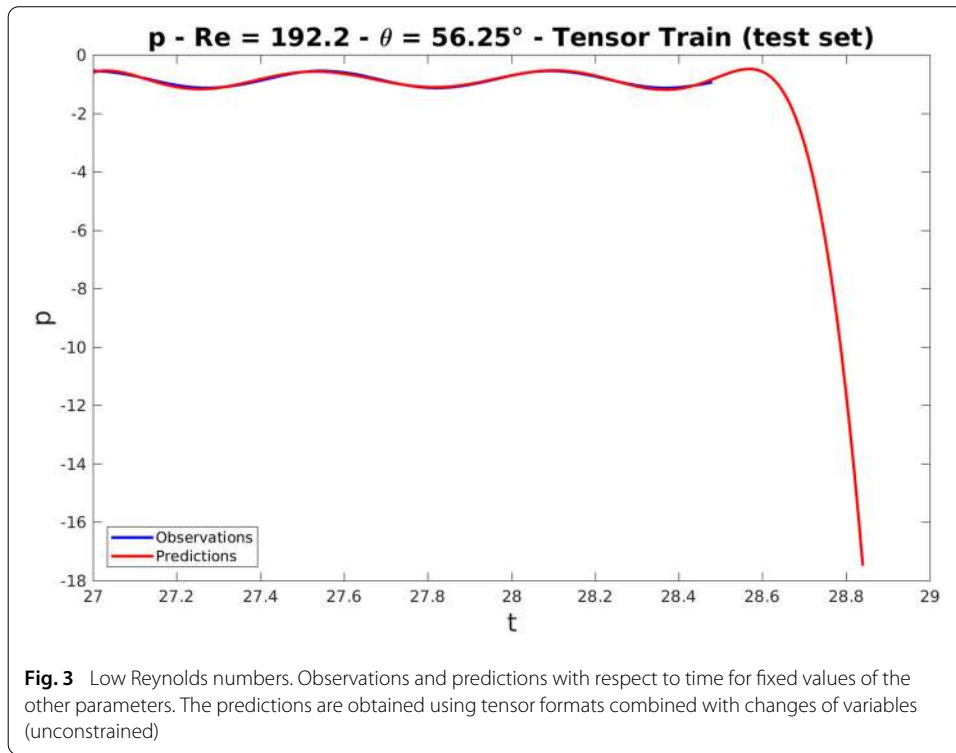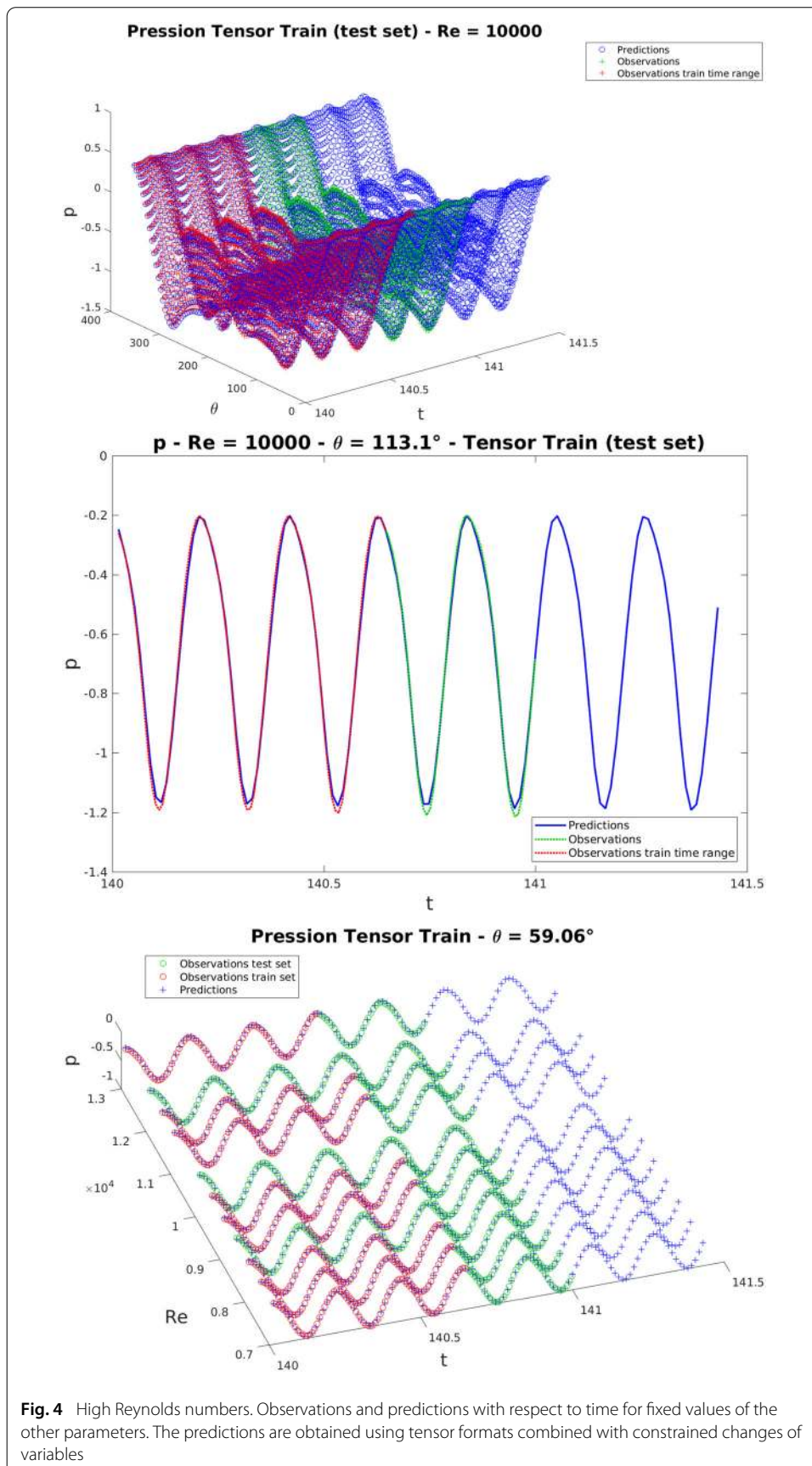
- estimation of the approximation error on the sample S_{test} with 30 simulations of Re on the train time range (consisting of the 181 first time steps): $e_{\text{test}}^2 = 1.14\%$,
- estimation of the extrapolation error on the sample S_{extra} with 30 simulations of Re on the whole time range (consisting of the 300 available time steps): $e_{\text{extra}}^2 = 1.19\%$.

On Fig. 2 are plotted the predictions with blue crosses and the observations with circles, the red ones were used for learning the approximation and the green ones for estimating the model error. We observe a very good match of the predictions with the observations even beyond the train time range.

The approximation constructed using the proposed change of variable is able to extrapolate the wall pressure beyond the time range used for training the approximation. This extrapolation was made possible by introducing the constraint (24). As an illustration, Fig. 3 shows the observations (in blue) versus the predictions (in red) on a longer time interval obtained without constraint. The approximation obviously looses its periodicity.

Table 1 summarizes the results obtained with the proposed approach and those obtained using the change of variable without using the tensor train format. One can observe that tensor formats enable to break the curse of dimensionality and thus to ease the learning of the approximation based on observations. Indeed the storage complexity of tensor train formats is $\mathcal{O}(MnR^2)$ where n is the order of the dimension of the representation space in each dimension and R is the order of the rank. That is, it grows only linearly with the dimension M and quadratically with the rank whereas the storage complexity without using tensor formats, *i.e.* on the polynomial chaos, grows factorially or exponentially with the dimension M. Exploiting low complexity representations as low rank structures is necessary to address the problem when the dimension M increases with the definition of new variables.

Fig. 2 Low Reynolds numbers. Observations and predictions with respect to time for fixed values of the other parameters. The predictions are obtained using tensor formats combined with constrained changes of variables

Fig. 3 Low Reynolds numbers. Observations and predictions with respect to time for fixed values of the other parameters. The predictions are obtained using tensor formats combined with changes of variables (unconstrained)

Table 1 Storage complexity and test error of the approximations obtained with the change of variable and with and without using Tensor Train (TT) format

	Storage complexity	Test error e^2_{test} (%)
TT format + change of variables	594	1.14
Polynomial chaos + change of variables	8482	5.42

High Reynolds numbers

We now consider the approximation of the wall pressure for higher Reynolds numbers for which the data is given for 11 values of the Reynolds number in $\mathcal{X}_{Re} = [7000, 13000]$. The approximation is constructed with the following setting:

- for g_i: polynomial bases with a maximal degree of 1 for t and 2 for θ and Re,
- for h: polynomial bases of degree 20 for θ and 6 for Re,
- training sample S using 8 simulations of Re and considering only the 1000 first time steps.

The algorithm provided an approximation with dimension $M = 7 : X = (g_1(\xi), \ldots, g_5(\xi), \xi)$ and TT-ranks $r = [1\ 3\ 5\ 5\ 5\ 5\ 1\ 1]$. The model error was estimated using two different samples:

- estimation of the approximation error on the sample S_{test} with 3 simulations of Re on the train time range (consisting of the 1000 first time steps): $e^2_{test} = 2.97\%$,
- estimation of the extrapolation error on the sample S_{extra} with 3 simulations of Re on the whole time range (consisting of the 1500 available time steps): $e^2_{extra} = 3.02\%$.

On Fig. 4 are plotted the predictions with blue crosses and the observations with circles, the red ones were used for learning the approximation and the green ones for estimating the

Fig. 4 High Reynolds numbers. Observations and predictions with respect to time for fixed values of the other parameters. The predictions are obtained using tensor formats combined with constrained changes of variables

model error. Again we observe a very good match of the predictions with the observations even beyond the train time range. The TT-rank of the approximation is low making possible the approximation using few samples of the Reynolds number R_e.

Conclusion

This paper presents a new strategy to approximate multivariate functions with periodicity. It gives the principles of the method based on the combination of functions $h(g_1(\xi, t), \ldots, g_m(\xi, t), \xi)$ chosen in appropriate classes of functions. The functions $g_i(\xi, t)$ define new variables of the multivariate functions h which is here represented in the class of rank structured functions. Algorithms are proposed for constructing the approximation based on observations of the function, a constraint is added for the definition of the new parameters to promote periodicity of the representation. The numerical simulations yield good results. An analysis on the convergence of the approximation is to be studied.

Declarations

Competing interests
The authors declare that they have no competing interests.

Author details
[1]GeM UMR CNRS 6183, Université de Nantes, Nantes, France, [2]LOMC UMR CNRS 6294, Université Le Havre Normandie, Normandie Université, Le Havre, France.

References
1. Bass RF, Grochenig K. Random sampling of multivariate trigonometric polynomials. SIAM J. Math. Anal. 2006;36(3):773–95. https://doi.org/10.1137/S0036141003432316.
2. Kammerer L, Potts D, Volkmer T. Approximation of multivariate periodic functions by trigonometric polynomials based on rank-1 lattice sampling. Journal of Complexity. 2015;31(4):543–76. https://doi.org/10.1016/j.jco.2015.02.004.
3. Briand T. Trigonometric polynomial interpolation of images. Image Processing On Line. 2019. https://doi.org/10.5201/ipol.2019.273.
4. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. : Springer; 2009.
5. Tipireddy R, Ghanem R. Basis adaptation in homogeneous chaos spaces. Journal of Computational Physics. 2014;259:304–17.
6. Constantine PG, Dow E, Wang Q. Active subspace methods in theory and practice: applications to kriging surfaces. SIAM Journal on Scientific Computing. 2014;36(4):1500–24.
7. Cortesi A, Constantine P, Magin T, Congedo PM. Forward and backward uncertainty quantification with active subspaces: application to hypersonic flows around a cylinder. INRIA: Technical report; 2017.
8. Friedman JH, Stuetzle W. Projection pursuit regression. Journal of the American statistical Association. 1981;76(376):817–23.
9. Hastie T, Tibshirani R, Wainwright M. Statistical Learning with Sparsity: the Lasso and Generalizations. : CRC Press; 2015.
10. Kolda TG, Bader BW. Tensor decompositions and applications. SIAM Review. 2009;51(3):455–500. https://doi.org/10.1137/07070111X.
11. Chevreuil M, Lebrun R, Nouy A, Rai P. A least-squares method for sparse low rank approximation of multivariate functions. SIAM/ASA Journal on Uncertainty Quantification. 2015;3(1):897–921. https://doi.org/10.1137/13091899X.
12. Grelier E, Nouy A, Chevreuil M. Learning with tree-based tensor formats; 2019. arXiv:1811.04455.

13. Cawley GC, Talbot NLC. Fast exact leave-one-out cross-validation of sparse least-squares support vector machines. Neural Networks. 2004;17:1467–75.
14. Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. 1996;58(1):267–88.
15. Chen SS, Donoho DL, Saunders MA. Atomic decomposition by basis pursuit. SIAM Journal on Scientific Computing. 1999;20:33–61.
16. Efron B, Hastie T, Johnstone I, Tibshirani R. Least angle regression. The Annals of Statistics. 2004;32(2):407–99.
17. Nouy A, Grelier E, Giraldi L. Approximationtoolbox. Zenodo. 2020. https://doi.org/10.5281/zenodo.3653970.
18. Grelier E Learning with tree-based tensor formats : Application to uncertainty quantification in vibroacoustics. PhD thesis, Centrale Nantes; 2019.
19. Falcó A, Hackbusch W, Nouy A. Tree-based tensor formats. SeMA Journal; 2018.
20. Cichocki A, Lee N, Oseledets I, Phan A-H, Zhao Q, Mandic D. Tensor networks for dimensionality reduction and large-scale optimization: Part 1 low-rank tensor decompositions. Foundations and Trends in ® Machine Learning. 2016;9(4–5):249–429.
21. Oseledets I, Tyrtyshnikov E. Recursive decomposition of multidimensional tensors. Doklady Math; 2009.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.