Journal of Computational Physics 374 (2018) 1219–1238

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

Gradient-based optimization for regression in the functional tensor-train format

Alex A. Gorodetsky [a,*,1], John D. Jakeman [b]

[a] University of Michigan, 3053 FXB, 1320 Beal Avenue, Ann Arbor, MI, 48109, USA
[b] Optimization and Uncertainty Quantification, Sandia National Laboratories, Albuquerque, NM, 87123, USA

ARTICLE INFO

Article history:
Received 11 January 2018
Received in revised form 5 July 2018
Accepted 2 August 2018
Available online 14 August 2018

Keywords:
Tensors
Regression
Function approximation
Uncertainty quantification
Alternating least squares
Stochastic gradient descent

ABSTRACT

Predictive analysis of complex computational models, such as uncertainty quantification (UQ), must often rely on using an existing database of simulation runs. In this paper we consider the task of performing low-multilinear-rank regression on such a database. Specifically we develop and analyze an efficient gradient computation that enables gradient-based optimization procedures, including stochastic gradient descent and quasi-Newton methods, for learning the parameters of a functional tensor-train (FT). We compare our algorithms with 22 other nonparametric and parametric regression methods on 10 real-world data sets and show that for many physical systems, exploiting low-rank structure facilitates efficient construction of surrogate models. We use a number of synthetic functions to build insight into behavior of our algorithms, including the rank adaptation and group-sparsity regularization procedures that we developed to reduce overfitting. Finally we conclude the paper by building a surrogate of a physical model of a propulsion plant on a naval vessel.

1. Introduction

Assessment of uncertainty in a computational model is essential to increasing the credibility of simulation-based knowledge discovery, prediction, and design. Sources of model uncertainty must be identified and the effect of these uncertainties on the model output (prediction) quantified. The accuracy to which uncertainty can be quantified is limited by the computational resources available. Many applications require vast amounts of computational effort, thereby limiting the number model evaluations that can be used to interrogate the uncertainty in the system behavior. Consequently a significant portion of methods developed for uncertainty quantification (UQ) in recent years have focused on constructing surrogates of expensive simulation models using only a limited number of model evaluations.

Within the computational science community, both parametric and non-parametric function approximation methods have been extensively used for Uncertainty Quantification (UQ). Nonparametric Gaussian process models (GP) [46,41] and parametric Polynomial Chaos Expansions (PCE) [18,56] are arguably the two most popular methods used. Gaussian process regression can be interpreted as a Bayesian method for function approximation, providing a posterior probability distribution over functions. Maximum likelihood estimation and Markov Chain Monte Carlo sampling are the two most popular methods used to characterize the posterior distribution of the GP. Polynomial chaos expansions represent a response surface as

* Corresponding author.
E-mail address: goroda@umich.edu (A.A. Gorodetsky).
[1] This work was performed while the author was at Sandia National Laboratories.

https://doi.org/10.1016/j.jcp.2018.08.010
0021-9991/

a linear combination of orthonormal multivariate polynomials. The choice of the orthonormal polynomials is related to the distribution of the uncertain model variables. Various approaches have been adopted to compute the coefficients of the PCE basis, and these include, pseudo-spectral projection [8,9], sparse grid interpolation [17,39], and regression using ℓ_1-minimization [3,13,37]. For a comparison between nonparametric GP methods and parametric PCE methods see [20], and for an attempt to combine the benefits of both approaches see [50].

High-dimensional approximation problems, such as regression, pose challenges for both parametric and nonparametric representation formats. Parametric approaches, for example those using a PCE representation, are limited by their expressivity. For example, if the multivariate basis functions, which describe a parametric representation, are tensor products of univariate polynomials, then the curse of dimensionality is obtained because the number of combinations of basis functions grows exponentially with dimension. Nonparametric methods, for example Gaussian process regression, have great expressive capabilities. However, they also encounter the curse of dimensionality since their excess risk grows exponentially with dimension [25].

To counteract these computational burdens for both types of methods, attention has focused on constraining the functional representation to limit the curse of dimensionality. One popular constraint is limiting the model to that of additive separable forms [27,38,48]

$$f(x) = f_1(x_1) + f_2(x_2) + \dots f_d(x_d). \tag{1}$$

One can also use second order interactions, e.g., $f_{12}(x_1, x_2), f_{13}(x_1, x_3), \dots$, to increase expressivity while maintaining tractability [31]. However, further increasing the number of interaction terms in the ANOVA model [15]

$$f(x) = \sum_i f_i(x_i) + \sum_{i,j} f_{ij}(x_i, x_j) + \sum_{i,j,k} f_{ijk}(x_i, x_j, x_k) + \cdots$$

will still naturally encounter the curse of dimensionality unless adaptive methods that identify the order of interaction interactively are utilized [17,35,16,30,29].

In this paper, we propose algorithms to improve regression in a functional representation that takes advantage of *low-rank* structure to mitigate the curse of dimensionality while maintaining high expressivity. Low-rank functional representations are parametric models that enable a wide variety of interactions between variables and can generate high order representations. More specifically, they are continuous analogues of tensor decompositions and exploit *separability*, i.e., that a function can be represented by the sum of products of univariate functions. One example is the canonical polyadic (CP) [5] representation consisting of a sum of products of univariate functions $f(x) = \sum_{i=1}^{R} f_1^{(i)}(x_1) \dots f_d^{(i)}(x_d)$, and the number of free parameters of such a representation scales linearly with dimension. Instead of the CP format, we use a continuous analogue of the discrete tensor train (TT) decomposition [42] called the functional tensor-train (FT) [43,22] to allow for a greater number of interactions between variables.

Low-rank functional tensor decompositions have been used for regression previously. Existing approaches [14,36,6,47] implicitly make two simplifying assumptions to facilitate the use of certain algorithms and data structures from the low-rank tensor decomposition literature. Specifically, they assume linear, with respect to the parameters of $f_k^{(i)}$,[2] and identical basis expansions for each univariate function of a particular dimension. These approaches convert the problem from one of determining a low-rank *function* to one of representing the coefficients of a tensor-product basis as a low-rank *tensor*. Following this conversion, many of these techniques use alternating minimization to determine the coefficients of the FT. Alternating minimization, such as alternating least squares, transforms a nonlinear optimization problem for fitting the parameters of each univariate function to data into a linear problem by only considering a single dimension at a time. Existing approaches either use efficient linear algebra routines to solve the linear system at each iteration [14] or sparsity inducing methods such as the LASSO [36]. Recently, iterative thresholding has also been used to find low rank coefficients; however, such an approach has been limited to problems with low-dimensionality [47].

In this paper we take a different approach: we use gradient-based optimization procedures such as quasi-Newton methods and stochastic gradient descent, and we do not restrict ourselves by the assumptions that lead to the consideration of a tensor of coefficients. Overall, our contributions include:

1. Derivation and computational complexity analysis of a gradient-based fitting procedure that yields more accurate approximations than alternating minimization;
2. Usage of both linear and nonlinear parameterizations of univariate function in each mode to facilitate a larger class of functions than is possible when using solely linear representations; and
3. Creation of rank-adaptation and regularization schemes to reduce overfitting.

We demonstrate the efficacy of our approach on both synthetic functions and on several real-world data sets used in existing literature to test other regression algorithms [31]. Specifically, we show that gradient-based optimization provides

[2] Note, this causes the final multivariate function output to be *multilinear* with respect to its parameters.