Proceedings of the 5th WSEAS International Conference on Applied Computer Science, Hangzhou, China, April 16-18, 2006 (pp556-559)

A Univariately Approximate Data Partitioning Algorithm for a Dominantly Multiplicative Multivariate Function

METİN DEMİRALP
Computational Science and Engineering Program, Informatics Institute,
İstanbul Technical University
İTÜ Bilişim Enstitüsü, İTÜ Ayazağa Yerleşkesi, Maslak, 34469, İstanbul, Türkiye (Turkey)
TURKEY (TÜRKİYE)
http://www.be.itu.edu.tr/~demiralp

Abstract: - This paper concerns with the construction of a data partitioning method for a dominantly multiplicative function whose values are given at the points of the Euclid space spanned by the values of its arguments. Method uses the univariate truncation of the HDMR constructed for the logarithm of the function under consideration. The geometry of HDMR is a unit hypercube whose corners are located at the nonnegative parts of the coordinate axes. Weight function in HDMR is assumed to be just 1 for simplicity. The ultimate goal of the partitioning is to produce N (number of independent variables) set of ordered pairs from the unique $(N + 1)$–tuple data set. Although the production is approximate it works well when the function under consideration is dominantly multiplicative. Each set of ordered pairs can be used to determine relevant univariate HDMR component. This means that the N dimensional multivariate interpolation can be reduced to N number of univariate interpolation.

Key-Words: - Data Partitioning, Multivariate Approximation, High Dimensional Model Representation

1 Introduction

Multivariance is perhaps the plague of plethora, in approximating the functions especially when its value is over a certain value, if the approximating procedure is run in a computer. This is basically because of the limited capacities in memory and the execution speed. A function approximating procedure in computer requires a finite number of discrete information due to discrete and limited physical structure of computer although a continous function requires generally denumerably infinite number of information. The discrete data to approximately define a multivariate function which has N independent variable in its argument is a finite set of $(N + 1)$–tuples each of which contains the values of the independent variables and the corresponding value of the investigated function respectively. Therefore the approximation for the considered function is based on this set and obtained by using some of the existing interpolation techniques. That is, approximation is a single multivariate interpolation. Although there exist a lot of multivariate interpolation techniques the knowledge accumulation on the univariate interpolation is greater and greater than the multivariate case. Hence, the univariate interpolation is more preferable. This urges us to somehow convert the multivariate data set to univariate data sets. Univariate interpolation uses a set of ordered pairs whose first and second elements are the values of the independent variable and the value of the related function corresponding to that independent variable value. This means that we need to somehow partition the multivariate data set to sets of univariate data. In other words, we need to use a divide–and–conquer type algorithm. Amongst several possibilities we use the High Dimensional Model Representation (HDMR) Method whose popularity is apparently increasing in last decade. Here we do not use the HDMR of the function under consideration. Instead, we use the HDMR of the natural logarithm of the investigated function after shifting it appropriately to take care of the dominant multiplicative nature in the considered function. Our basic goal here is to formulate the algorithm. The numerical efficiency is beyond the aim of this paper although our future studies will contain various numerical implementations.

Paper is organized as follows. The second section presents the recalling of HDMR while the third section is about the construction of the data partitioning method. Fourth section finalizes the paper via concluding remarks.

2 HDMR

The high dimensional model representation[1-10] of a multivariate function $f(x_1, ..., x_N)$ is given as

$$f(x_1, ..., x_N) \quad = \quad f_0 + \sum_{i_1=1}^{N} f_{i_1}(x_{i_1})$$

Proceedings of the 5th WSEAS International Conference on Applied Computer Science, Hangzhou, China, April 16-18, 2006 (pp556-559)

$$+\sum_{\substack{i_1,i_2=1\\ i_1<i_2}}^{N} f_{i_1,i_2}\left(x_{i_1},x_{i_2}\right)+\cdots \quad (1)$$

where N stands for the number of the independent variables. If we define an Hilbert space over the hyperprism defined by the intervals $a_i \le x_i \le b_i$, (where $1 \le i \le N$ and a_i, b_i are assumed to be given) with the following inner product for two arbitrary square integrable multivariate functions in this space, denoted by $g(x_1,...,x_N)$ and $h(x_1,...,x_N)$ respetively

$$\begin{aligned}(f,g) \quad \equiv \quad & \int_{a_1}^{b_1} dx_1 ... \int_{a_N}^{b_N} dx_N W(x_1,...,x_N) \\ & \times g(x_1,...,x_N)\, h(x_1,...,x_N) \qquad (2)\end{aligned}$$

then we impose the mutual orthogonality amongst the right hand side components of (1). $W(x_1,...,x_N)$ in (2) stands for a product type function, that is,

$$W(x_1,...,x_N) \equiv \prod_{i=1}^{N} W_i(x_i) \qquad (3)$$

where we assume that the integral of each given univariate factors $W_i(x_i)$ $(1 \le i \le N)$ between a_i and b_i is equal to 1 for simplification. These weight factors can be chosen discrete or continuos depending on what we expect from the use of HDMR.

The mutual orthogonality of the HDMR components appearing in the right hand side of (1) is sufficient to uniquely determine those components. To facilitate the analysis for the determination of the HDMR components we define the following projection operator

$$\begin{aligned}\mathcal{P}_0 g(x_1,...,x_N) \equiv & \int_{a_1}^{b_1} dx_1 ... \int_{a_N}^{b_N} dx_N W(x_1,...,x_N) \\ & \times g(x_1,...,x_N) \qquad (4)\end{aligned}$$

The orthogonality of all higher than zero order multivariate components to f_0 implies that the integrals of those components over one of their independent variables over the related interval under the corresponding univariate weight function vanish. We call this "vanishing property". If we now consider the action of $\mathcal{P}_0$ on both sides of (1) and then utilize the vanishing properties of the higher than zero variate terms, and the normalized nature of the weight function factors then we can write

$$f_0 = \mathcal{P}_0 f(x_1,...,x_N) \qquad (5)$$

We need to define another integral operator $\mathcal{P}_i$ $(1 \le i \le N)$ for the determination of univariate terms. It is equivalent to $\mathcal{P}_0$'s new form obtained after removing the integration over x_i and discarding the univariate weight function factor $W_i(x_i)$. Its action on an arbitrary square integrable multivariate function produces a univariate function depending on x_i whereas $\mathcal{P}_0$ projects to a constant. The action of this operator on both sides of (1) and the employment of vanishing properties of all HDMR terms except the constant one and the normalization in univariate weight factors enable us to write

$$f_i(x_i) = \mathcal{P}_i f(x_1,...,x_N) - f_0, \qquad 1 \le i \le N \quad (6)$$

The bivariate and the higher multivariate HDMR components can be determined through similar routes although we do not intend to give them explicitly here.

3 Data Partitioning

Let us define the following discrete domain sets for independent variables x_1,...,x_N

$$\mathcal{D}_i \equiv \left\{x_i^{(1)},...,x_i^{(m_1)}\right\}, \qquad 1 \le i \le N \qquad (7)$$

where $a_i \le x_i^{(j)} \le b_i$, $(1 \le j \le m_i)$. We can simplify the analysis by taking all a_i values as 0 and b_i values as 1 without any remarkable loss of generality since certain linear transformations over independent variables convert all intervals to the closed interval between 0 and 1. We can now define the global domain for our HDMR as the cartesian product of $\mathcal{D}_i$ sets. That is,

$$\mathcal{D} \equiv \mathcal{D}_1 \times \cdots \times \mathcal{D}_N \qquad (8)$$

As can be seen easily the set $\mathcal{D}$ contains $m_1...m_N$ N–tuples. If we denote one of these N–tuples by d_j, $(1 \le j \le m_1...m_N)$ then we can write $f(d_j)$ to mean the value of the considered function at the independent variable values appering in d_j. This urges us to define a set of $(N+1)$–tuples composed as $\left(d_j, f\left(d_j\right)\right)$ for j values between 1 and $m_1...m_N$ inclusive. In the light of these discussions we assume that the given data set is explicitly defined as follows

$$\mathcal{F} \equiv \left\{\left(d_j, f\left(d_j\right)\right) \Big| \, 1 \le j \le m_1...m_N\right\} \qquad (9)$$

This enforces us to use the following univariate weight factors in HDMR

$$W_i(x_i) \equiv \sum_{j=1}^{m_i} w_{i,j}\,\delta\left(x_i - x_i^{(j)}\right), \quad 1 \le i \le N \quad (10)$$

where

$$\sum_{j=1}^{m_i} w_{i,j} = 1 \tag{11}$$

and $\delta\left(x_i - x_i^{(j)}\right)$ stands for the Dirac's delta function positioned at $x_i^{(j)}$.

To proceed more with further simplification we can assume that the function values $f\left(d_j\right)$ $(1 \leq j \leq m_1 \times \cdots \times m_N)$ are all nonnegative without any loss of generality since a translation in f gets this nonnegativity.

Now, under the nonnegative function values assumption, we can focus on the HDMR of the natural logarithm of $f(x_1, ..., x_N)$ and write

$$\ln f(x_1, ..., x_N) = \varphi_0 + \sum_{i=1}^{N} \varphi_i(x_i) + \cdots \tag{12}$$

This enables us to write the following result for φ_0 by following the analysis given in the previous section

$$\varphi_0 = \sum_{i_1=1}^{m_1} ... \sum_{i_N=1}^{m_N} w_{1,i_1} ... w_{N,i_N} \ln f\left(x_1^{(i_1)}, ..., x_N^{(i_N)}\right) \tag{13}$$

This formula can be put into the following concise form

$$\varphi_0 = \mathcal{S} \ln f\left(x_1^{(i_1)}, ..., x_N^{(i_N)}\right) \tag{14}$$

if we define the summation operator $\mathcal{S}$ as follows

$$\mathcal{S} \ln f\left(x_1^{(i_1)}, ..., x_N^{(i_N)}\right) = \sum_{i_1=1}^{m_1} ... \sum_{i_N=1}^{m_N} w_{1,i_1} ... w_{N,i_N} \ln f\left(x_1^{(i_1)}, ..., x_N^{(i_N)}\right) \tag{15}$$

The univariate function $\varphi_j\left(x_j\right)$ can not be evaluated analytically. Instead its values at $x_j = x_j^{(1)}$,..., $x_j = x_j^{(m_j)}$ can be determined. To this end we define a new summation operator $\mathcal{S}_j$ which can be derived from $\mathcal{S}$ by removing the sum over x_j and discarding the factor w_{j,i_j}. This enables us to express the result as follows

$$\varphi_j\left(x_j^{(i_j)}\right) = \left(\mathcal{S}_j - \mathcal{S}\right) \ln f\left(x_1^{(i_1)}, ..., x_N^{(i_N)}\right) \qquad 1 \leq j \leq N, \qquad 1 \leq i_j \leq m_j \tag{16}$$

This set of equations define N sets of ordered pairs such that the ordered pairs of the j–th set are defined by the general term $\left(x_j^{i_j}, \varphi_j\left(x_j^{i_j}\right)\right)$. Each of these N set of ordered pairs can be used to interpolate the corresponding univariate component analytically. The resulting function will have different structures depending on the type of the interpolation scheme used.

Now (12) can be rewritten in the following approximation form

$$f(x_1, ..., x_N) \approx e^{\varphi_0} \left[\prod_{i=1}^{N} e^{\varphi_i(x_i)}\right] \cdots \tag{17}$$

(17) is the ultimate multivariate interpolation formula we sought.

4 Concluding Remarks

The main goal of this paper was to construct a multivariate approximate interpolation formula to approximate a function from a discrete data. Data was assumed to be given at all nodes of a hyperprismatic mesh. That is, we have dealt with the uniform data. We have used HDMR to approximate the logarithm of the function under consideration instead of itself since the logarithm converts the multiplicativity to additivity. The additive nature of HDMR was taken to scene for this reason. The dominancy in the multiplicativity enabled us to truncate HDMR at univariate level. All these allowed us to construct (17).

The approach presented here can be extended by increasing the truncations order to get better results.

If the function under consideration is dominantly additive then the function's itself can be expanded to HDMR and the univariate truncation of the result gives another multivariate interpolation formula.

All these mean that a lot of fruitful application possibilities are arising in the horizon for future applications although (17) seems to be a milestone in multivariate interpolation of uniform data at the nodes of a hyperprismatic mesh.

Acknowledgement:
Author is grateful to State Planning Organization (DPT) for its generous support and to Turkish Academy of Sciences for its enthusiasm increasing stimulation.

References:

[1] I.M. Sobol, Sensitivity Estimates for Nonlinear Mathematical Models, *Mathematical Modeling and Computational Experiments*, 1993, VOL. 1, pp 407-414.

[2] J.A. Shorter, P.C. Ip, and H. Rabitz, An Efficient Chemical Kinetics Solver Using High Dimensional Model Representation, txitext J. Phys. Chem. A, 1999, VOL 103, pp 7192-7198.

[3] Ö. Alış and H. Rabitz, Efficient Implementation of High Dimensional Model Representations, *Journal of Mathematical Chemistry*, 2001, VOL 29, pp 127-142.

[4] G. Li, S. W. Wang, H. Rabitz, S. Wang and P. Jaffe, Global uncertainty assessments by high dimensional model representation (HDMR). *Chemical Engineering Science*, 2002, VOL 57, pp 4445-4460.

[5] G. Li, M. Artamonov, H. Rabitz, S. Wang, P.G. Georgopoulos and M. Demiralp, High Dimensional Model Representations Generated from Low Order Terms -lp-RS-HDMR, Journal of Computational Chemistry, 2003, VOL 24, pp 647-656.

[6] İ. Yaman and M. Demiralp, High Dimensional Model Representation Approximation of an Evolution Operator with a First Order Partial Differential Operator Argument, *Applied and Numerical Analysis and Computational Mathematics*, 2003, VOL 1, pp 287-296.

[7] M.A. Tunga and M. Demiralp, A Factorized High Dimensional Model Representation on the Partitioned Random Discrete Data, *Applied and Numerical Analysis and Computational Mathematics*, 2003, VOL 1, pp 233-243.

[8] M.A. Tunga and M. Demiralp, A factorized high dimensional model representation on the nodes of a finite hyperprismatic regular grid, *Applied Mathematics and Computation*, 2005, VOL 164, pp 865-883.

[9] M.A. Tunga and M. Demiralp, Hybrid high dimensional model representation (HHDMR) on the partitioned data, *Journal of Computational and Applied Mathematics*, 2006, VOL 185, pp 107-132.

[10] M. Demiralp, High Dimensional Model Representation and its application varieties, *Tools for Mathematical Methods, Mathematical Research, St-Petersburg*, 2003, VOL 9, pp 146-159.