An Efficient Algorithm for the Discovery of Complex Material Formulations

George Brell

October 5, 2007

1 Introduction

The discovery of new material formulations is a difficult problem in many fields of science and engineering, including plasma physics, due to the need for complex materials with specific properties. Even in complex materials comprised of only three or four components, the number of possible fractional combinations far exceeds the resources of most laboratories to investigate. Using Random Sampling-High Dimensional Model Representation (RS-HDMR), we can significantly reduce the cost of designing multi-component formulations. Used in conjunction with advanced searching techniques, such as evolutionary-algorithms, RS-HDMR appears capable of dramatically increasing the efficiency of discovering new material formulations.

2 Approach

RS-HDMR is an algorithmic approach to high dimensional, non-linear data analysis and interpolation that expresses a model function output, f(x) as a finite, hierarchical correlated expansion in terms of the n input variables, $(x_1, \ldots, x_n)$ (Li, et al., 2007):

$$f(x) = f_0 + \sum_{i=1}^{n} f_i(x_i) + \sum_{1 \le i < j \le n} f_{ij}(x_i, x_j) + \cdots + \sum_{1 \le i_1 < \cdots < i_l \le n} f_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots x_{i_l}) + \cdots + f_{12\ldots n}(x_1, x_2, \ldots x_n)$$

where f_0 is a constant representing the mean of $f(x)$, $f_i(x_i)$ represents the individual contribution of the i-th variable to the output and the further component functions represent higher order interactions amongst the input variables. Each of these is approximated via optimal weighted orthonormal polynomials.

In the case of complex material formulations, x_i is the mole fraction of the i-th material component. This relation constrains the system being considered as $\sum_{i=1}^{n} x_i = 1$ (i.e. the mole fractions must sum to unity). This seemingly simple constraint significantly reduces the volume of the space under investigation , which appears to have an important impact on the success of RS-HDMR in this context.

Typically, many of the component functions will not be significant to the output function and these terms may be identified via a statistical F-test and excluded from the approximation. In practice, only first, second and third order functions are often necessary. The individual component functions are then determined sequentially from lower to higher order, which allows for $f(x)$ to be represented in terms of the independent and correlated contributions of the material inputs. Furthermore, we can order the functions in terms of their contribution to $f(x)$ by analysis of the magnitude of the individual component contributions.

3 Results and Applications

The primary focus of my research has thus far been in modeling arbitrarily constructed data sets to test the limits of HDMR in terms of both its accuracy and determining the minimal number of random sampled complex materials necessary to provide a reliable model for eventual application to the discovery of material formulations. With this goal in mind, we generated a large array of functions $f(x)$ of differing complexity and number of variables and thoroughly examined the quality of the RS-HDMR maps.

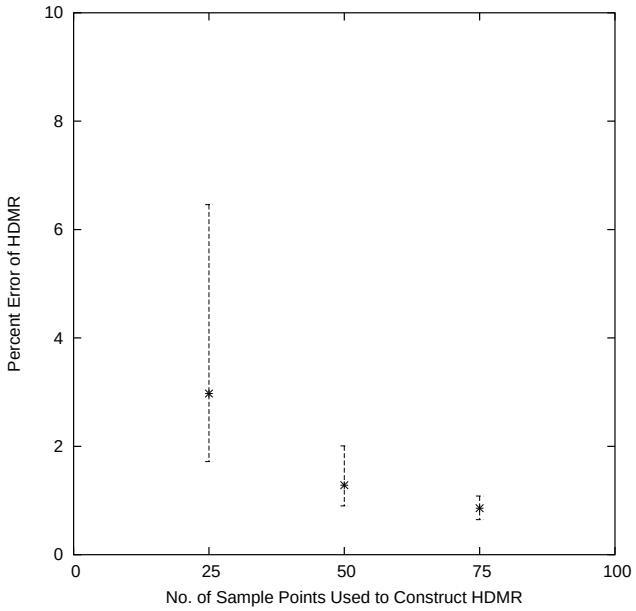

Figure 1. Plots the average percent error of the HDMR versus the number of sample points used in its construction. The error bars present are the standard deviations. The percent error is referenced to 2 which is the middle of the dynamic range (1 to 3) of $f(x)$ in this example.

As a typical illustration, the figure above presents the error present in an HDMR representing a material of four components whose observable property is given by the the polynomial $f(x) = 3x_1 + x_2 + x_3 + x_4 + 3x_1x_2 + x_1x_3 + x_2x_3$. One of the primary variables affecting the accuracy of the HDMR is the number of input data points used in its construction. Since the individual material samples utilized can have a significant effect on the accuracy of the HDMR-based model, multiple HDMR's were constructed for analysis using different sample sizes. As the figure shows at 75 points, the average error present in the HDMR representation of $f(x)$ is less than 1% of the function's midrange value. Even at 25 sample points for HDMR construction, the average error is only around 3%. This level of accuracy is incredibly high considering the very low number of points used to construct the HDMR. Importantly, the behavior above was found to be typical of a vast number of functions up to dimension $n = 10$, although the number of necessary samples modestly increases with n.

Based on these results, HDMR appears to be a valuable tool for efficiently seeking complex materials with many components. Utilizing HDMR after making a minimal number of random experimental compositions and measuring their properties should allow for targeting particularly rich areas of the composition space. The search for effective compositions could be led by an evolutionary algorithm guided by the experimentally-generated HDMR. The low number of sample points required to generate an accurate HDMR can significantly reduce the total number of experiments required to formulate a specific material.

4 Further Investigation

I am currently conducting further research into the applications of HDMR towards complex material formulations. Outside of the models utilized and mentioned previously, I am testing more complex formulations. These include further-constrained systems, i.e. materials with specifically limited components, and samples with random errors. I am also considering alternative sampling methodology as a method to increase accuracy. I also plan to explore the underlying reasons for why RS-HDMR appears to be so efficient; the present hypothesis is that the functional constraint $\sum_{i=1}^{n} x_i = 1$ greatly reduces the effective sample volume in the component space.