Data Analytics for Simulation Repositories in Industry

Rodrigo Iza Teran[1] and Jochen Garcke[1,2]

[1]Fraunhofer-Institut SCAI
Numerical Data-Driven Prediction
Schloss Birlinghoven
53754 Sankt Augustin
rodrigo.iza-teran@scai.fraunhofer.de

[2]Universität Bonn
Institute for Numerical Simulation
Wegelerstr. 6
53115 Bonn, Germany
garcke@ins.uni-bonn.de

Abstract: Simulations are used intensively in the developing process of new industrial products and have achieved a high degree of detail. In that workflow often thousand finite element model variants, representing different product configurations, are simulated within a few days or even overnight. Currently the decision process for finding the optimal product parameters involves the comparative evaluation of large finite element simulation bundles by post-processing each one of those results using 3D visualisation software. This time consuming process creates a severe bottleneck in the product design and evaluation workflow.

To handle these data we investigate an analysis approach based on nonlinear dimensionality reduction to find a low dimensional parameterisation of the dataset. In such a reduced representation, similar model variants are organised in clusters and the influence of the input variables can be analysed along such a parameterisation. We demonstrate the application of this approach to a realistic and relevant industrial example for robustness analysis of the bumper location in a frontal crash simulation. The approach has the potential to considerably speed up the virtual product development process by allowing an intuitive and interactive simultaneous evaluation of many product designs.

1 Introduction

Finite element simulations of physical behaviour are an essential tool in the development and improvement of industrial products for many years now [Mey07]. Here, an approximate representation of the object by a mesh is employed and, based on a mathematical model of the physical process, a numerical simulation is performed. For an accurate description such numerical approximations typically need to be very fine, in the order of million of unknowns in our case, with larger simulations needed for some other applications. The size of the resulting numerical data are huge, one simulation in time generates number of unknowns multiplied by the number of (saved) time steps numerical values. Furthermore, during the research and development process engineers easily generate several hundred variants of a specific finite element model that simulates different operating conditions of the product, multiplying the data once more. Post-processing software tools are readily available to display the 3D geometrical information of such a model and the results of one numerical simulation. Engineers analyse and compare the different simulations using their own engineering knowledge, although this is generally limited to the simultaneous analysis of only a few simulations at a time. The complex structure of the data and its sheer size, the required 3D visualisation of the geometry and the needed inspection of the associated design variables of each configuration prohibits a detailed comparative analysis of more than a few simulations by hand. There is need for a more efficient product development process which overcomes the current limitations.

The best known approach to tackle this data analysis challenge are Principal Component Analysis (PCA) to identify variation modes and its counterpart classical Multidimensional Scaling (CMDS) for finding a low dimensional embedding. For several thousand simulations, corresponding to a specific product development phase, the PCA can recover the principal variations in just a few components. The approach has already been successfully used for the analysis of numerical simulation data, see [AGHH08, MT08, TNNC10]. But in spite of their success so far, there are many situations where PCA or CMDS are not adequate. Being linear methodologies, they cannot cope with a presence of nonlinear correlations in the data, for example when dealing with time dependent highly nonlinear deformations. Approaches that attempt to overcome these limitations have been proposed in recent years, see [BGIT+13, Iza14]. In line of this research we use a nonlinear dimensionality reduction method called diffusion maps, which is a machine learning method known to be able to recover the so called intrinsic geometrical structure of the data in the presence of nonlinearities. Assuming the intrinsic geometry is low dimensional, recovering such nonlinear structures can be very effective for analysing a number of simulations simultaneously. The contribution of this work in comparison to [BGIT+13, Iza14] consists in constructing and using a realistic large scale industrial scenario to proof the effectiveness of nonlinear dimensionality reduction methods in the virtual product development process.

In section 2 we will introduce the applied data workflow, followed by section 3 about the essential basics of nonlinear dimensionality reduction. The application of the approach for the crash simulation bundle is presented in section 4. We finally give an overview of the potential application of the methodology and discuss also further efforts in this area.

2 A Workflow for Analysing Numerical Simulation Data

We describe a general workflow that allows for the analysis of high dimensional data from bundles of large finite element simulations. Four steps are involved in it. The first two steps in the analysis workflow involve costly data intensive computations. Preferably these take place in a parallel server infrastructure where the bulky data is stored, therefore avoiding transfer of the big data and exploiting the parallel HPC resources.

- Extraction - the raw data for the analysis are obtained directly from the simulation, these variables can be of different types such as scalars, vectors or tensors defined on nodes or elements of a finite element mesh.
- Preprocessing - this step is usually necessary to cope with the huge data size (millions of nodes and elements) and its complex nature. A usual preprocessing step

consists in the use of only subsets of the datasets in areas of interest (for example the supporting frontal beams of a car in a frontal crash), also sub-sampling and clustering can be employed [BGIT$^+$13]. The data can also be transformed in different ways in order to obtain a compact representation; for example using principal component analysis (PCA).

- Dimensionality Reduction - in this step a low dimensional representation is obtained from the dataset that parameterises the information. We use here diffusion maps as a dimensionality reduction framework, but other dimensionality reduction methods can be used as well.
- Exploration - based on the low dimensional representation found after the dimensionality reduction. The simulation variables are organised in the low dimensional embedding space, i.e. each numerical simulation is represented as a point in the obtained low dimensional space. Due to the reduced dimensionality, the datasets can be efficiently explored.

3 Nonlinear Dimensionality Reduction with Diffusion Maps

Nonlinear dimensionality reduction is a very active area of research in machine learning in recent years, often known as manifold learning to emphasise the actual objective of such methods, namely the identification of low dimensional structures to represent high-dimensional information. Several methods have been introduced and we refer to [LV07, Wan11] for details about each of this methods. In this work we concentrate on so called kernel methods, which are based on the construction of a similarity matrix with coefficients calculated using a kernel function of the type $e^{-d(x,y)}$, where $x \in R^n$ and n can be very large (e.g. up to the total number of nodes of the mesh in our application). A specific distance function d can be used in the method like Euclidean or Geodesic (approximated by the graph distance on the mesh for our application). After building the similarity matrix a SVD decomposition is used to extract the eigenvectors corresponding to the largest eigenvalues. In this research work we use a specific variant of a kernel method called diffusion map, for details about the method see [LL06, CL06, Wan11]; for its use for the analysis of numerical simulation data see [Iza14].

Notice that in our analysis we use simulation data on a finite element mesh directly. A finite element mesh contains nodes and elements and we assume that the mesh connectivity is the same and use only the values defined at the nodes on the mesh as data set. If the mesh connectivity is different a reference mesh can be used to map the simulation values from the slightly different numerical simulation meshes to it.

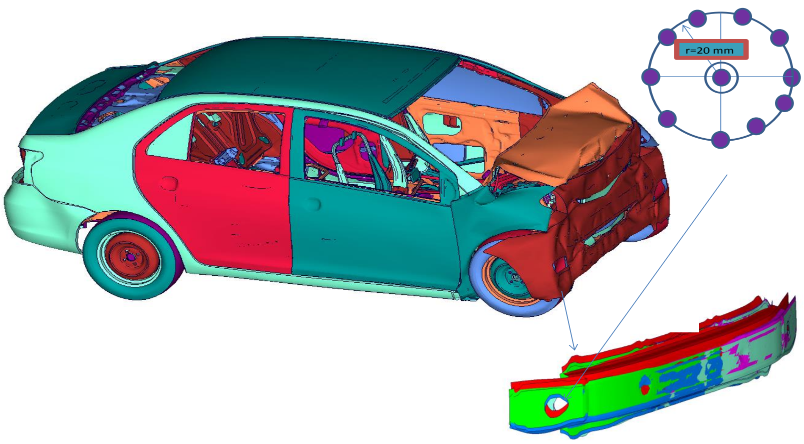


Figure 1: Variations of the bumper position along a circle.

4 Robustness Analysis in Frontal Crash

For this case study we investigate a current Toyota model with 998,000 nodes and corresponding elements, which is publically available from the NCAC[1]. As a realistic and difficult engineering problem, the position of the bumper was changed along a circle, see Figure 1, which results in observably different crash behaviour during a frontal impact with a barrier. A total of $M = 243$ numerical simulations where performed using LS-DYNA[2] and 26 intermediate time steps were saved for each run. The training dataset is saved in files in a binary format. So called post-processor software, can be used to read this data and extract (parts of) it. We use the software Animator[3] to extract specific components of a car or structure. The components we choose are the ones that are critical for the engineer during the analysis of the structural behaviour under crash (see Figure 2). Note that in [BGIT$^+$13, Iza14] car crash data was analysed stemming from a simpler model from that repository, with more than one order of magnitude less grid points, to study the general analysis procedure. In this work we analyse simulation data in a size currently being used from a difficult engineering task, and further proceed to an interpretation of the results from an engineering perspective.

We apply the workflow explained in section 2 to all simulations, and extract in the preprocessing step as relevant structural parts the firewall and the structural beams, see Figure 2. We perform the data analysis at time step 14, when most of the crash affecting the frontal structure took place. For the entries of the feature vector we employ for each mesh point the distance to the corresponding mesh point of a chosen reference model, this gives a vector in $\mathbb{R}^n$, where n is the total number of nodes of the selected parts. The dimension-

[1] Finite Element Model Archive, `http://www.ncac.gwu.edu/vml/models.htm`
[2] Livermore Software Technology `www.lstc.com/products/ls-dyna`
[3] GNS mbH `http://gns-mbh.com/animator.html`

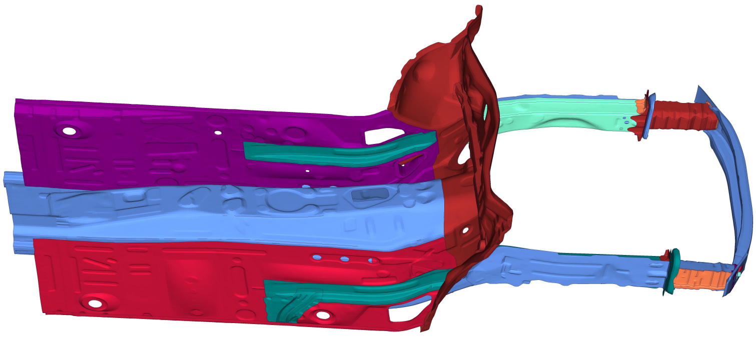


Figure 2: Selected structural parts relevant for the engineering analysis.

ality reduction is then performed with these feature vectors. For the exploration step of the workflow, each of the simulations is now represented as a point in a 3D plot which therefore gives a parametric representation of the 243 simulations in 3D.

To have a comparison with a standard linear approach we first performed a CMDS analysis for the same dataset using the first spectral coordinates as embedding for the exploration step, which is shown in Figure 4(a). Although simulations using a similar angle are arranged nearby, no distinct structures can be identified.

In Figure 3 the low dimensional embedding with diffusion maps is shown. It can clearly be seen that the simulations are organised according to a certain type of deformation mode. The colour of the deformed structure corresponds to the difference to the chosen reference model and the colour of the points in the 3D plot corresponds to a specific angle for the bumper location in degrees.

From the embedding plot for diffusion maps in Figure 3, it can be seen that at least three deformation modes can be associated to a range of positioning angles of the bumper. We extracted for clarity a beam from each mode in order to show the typical deformation in each mode and the angle dependence, see Figure 4.

The computational time for data loading and preprocessing in an off-line-phase takes less than an hour, which is mostly due to the data extraction, and needs to take place once for each data set. The actual data exploration afterwards can be done interactively. The selection of the corresponding simulations and the generation of the 3D pictures shown in Figure 4 took less than an hour. In comparison an engineer in charge of this project normally needs to analyse each of the deformation modes by hand and then classify them. A task that for that amount of simulations can require several days, if not weeks.

5 Overview and Perspectives

We studied a methodology for analyzing finite element simulations arising in the virtual product development process on a specific realistic industrial application. We could show

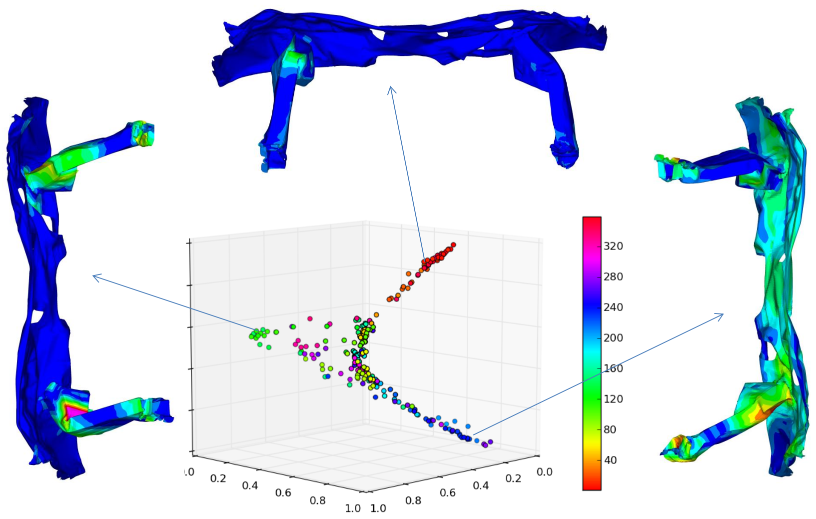
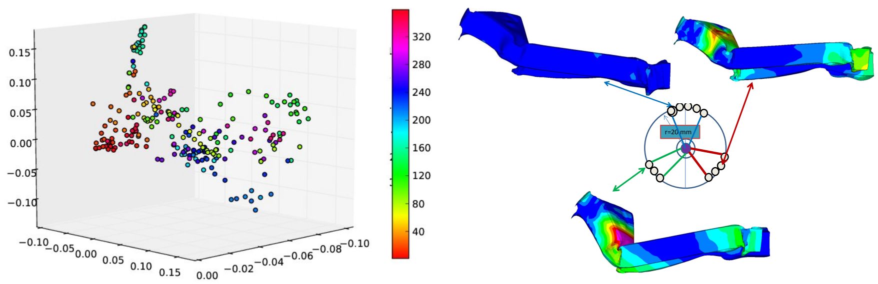


Figure 3: Embedding obtained by diffusion maps. Each simulation run is colour coded with the angle of the bumper variation.



(a) CMDS embedding

(b) deformation modes

Figure 4: We show the CMDS embedding, where each simulation run is colour coded with the angle of the bumper variation, and the bumper deformation modes derived from the diffusion maps embedding from Figure 3.

that organising the data along low dimensional structures allows the simultaneous analysis of many simulations. A realistic example in crash simulation, whereby the position of a bumper is changed, shows that different deformation modes in the structural beams of the car can be easily identified in the low dimensional embedding. Using nonlinear techniques reduces the complexity and time for investigating such large bundles of huge numerical simulation data.

To deploy such an approach for the analysis of large scale simulation data in real-life environments, it will need to handle efficient data storage (including compression) which will take place on data servers in the future instead of workstations as nowadays is the practice in industry. It will also need efficient transfer of the (relevant) data between server and client, as well as efficient data processing for analysis and visualisation procedures, like the one outlined in this work, on the server and the client.

Acknowledgements

The variations of the bumper positions were designed by Claudius Schoene of GNS mbH in the course of the KMU-innovativ project FEMMINER supported by the BMBF.

References

[AGHH08] S. Ackermann, L. Gaul, M. Hanss, and T. Hambrecht. PCA for Detection of Globally Important Input Parameters in Nonlinear Finite Element Analysis. In *Weimar Optimization and Stochastic Days 5.0*, 2008.

[BGIT+13] B. Bohn, J Garcke, R. Iza Teran, A. Paprotny, B. Peherstorfer, U. Schepsmeier, and C.-A. Thole. Analysis of car crash simulation data with nonlinear machine learning methods. In *Proceedings of the ICCS 2013*, volume 18 of *Procedia Computer Science*, pages 621–630. Elsevier, 2013.

[CL06] R. Coifman and S. Lafon. Diffusion Maps. *Applied and Computational Harmonic Analysis*, 21(1):5–30, 2006.

[Iza14] R. Iza Teran. Enabling the Analysis of Finite Element Simulation Bundles. *Internat. Jour. for Uncertainty Quantification*, 4(2):95–110, 2014.

[LL06] S. Lafon and A. B. Lee. Diffusion Maps and Coarse-Graining: A Unified Framework for Dimensionality Reduction, Graph Partitioning, and Data Set Parameterization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(9):1393–1403, 2006.

[LV07] J. Lee and M Verleysen. *Nonlinear Dimensionality Reduction*. Springer, 2007.

[Mey07] M. Meywerk. *CAE-Methoden in der Fahrzeugtechnik*. Springer, 2007.

[MT08] L. Mei and C.-A. Thole. Data analysis for parallel car-crash simulation results and model optimization. *Sim. Modelling Practice and Theory*, 16(3):329–337, 2008.

[TNNC10] C.-A. Thole, L. Nikitina, I. Nikitin, and T. Clees. Advanced mode analysis for crash simulation results. In *Proc. 9th LS-DYNA Forum*, 2010.

[Wan11] J. Wang. *Geometric Structure of High-Dimensional Data and Dimensionality Reduction*. Springer, 2011.