Capturing Scientists' Insight for DDDAS

Paul Reynolds, David Brogan, Joseph Carnahan, Yannick Loitière, and Michael Spiegel

Computer Science Department
University of Virginia

Abstract. One of the intended consequences of utilizing simulations in dynamic, data-driven application systems is that the simulations will adjust to new data as it arrives. These adjustments will be difficult because of the unpredictable nature of the world and because simulations are so carefully tuned to model specific operating conditions. Accommodating new data may require adapting or replacing numerical methods, simulation parameters, or the analytical scientific models from which the simulation is derived. In this research, we emphasize the important role a scientist's insight can play in facilitating the runtime adaptation of a simulation to accurately utilize new data. We present the tools that serve to capture and apply a scientist's insight about opportunities for, and limitations of, simulation adaptation. Additionaly, we report on the two ongoing collaborations that serve to guide and evaluate our research.

1 Introduction

In dynamic, data-driven application systems (DDDAS), we have observed that scientists are regularly confronted with the challenge of creating simulations capable of adapting to unanticipated runtime conditions. Runtime conditions may trigger adjustment or replacement of the analytical scientific models from which the system is derived, the numerical methods that implement those models, or the computational infrastructure that executes the numerical methods [1]. How, for example, should a weather simulation respond to newly acquired data that invalidates its predictions? Can some simulation parameters be adjusted automatically by a Kalman filter or must the Kalman filter itself be reparameterized? Perhaps an entirely different underlying model is required to appropriately simulate the new portion of state space exposed by the new data. Because so many aspects of DDDAS are candidates for change, effective automated adaptation to runtime conditions often requires leveraging subject matter expert (SME) insight to guide and constrain the adaptation.

For SME insight to be used in automatic adaptation, the insight must be captured and represented in a way that an automated simulation adaptation system can interpret and understand [2]. To enable this, we are developing formal programming language constructs for describing properties of candidate simulation adaptations. By succinctly encoding a SME's insight about potential simulation adaptations, we seek methods for specifying and automatically exploring simulation expansion opportunities subject to the identified constraints. Extremely

International Conference on Computational Science, pp. 1–8, 2006.

simple examples of such functionality come from the C++ type-conversion procedures where the programmer can specify constraints using typecasts while relegating opportunities for automatic conversion to the compiler. The semantics of the interacting components of DDDAS are much more complex than simple data types. However, we believe it is worthwhile to seek to capture the SME's insight 1) so that SMEs don't have to address particulars early on, 2) so future simulation users can take advantage of early SME insights, 3) so the adaptation process can be conducted in a semi-automated manner, and 4) so future incarnations of a simulation can have a high likelihood of being as efficient as possible.

In this paper, we present ongoing work to encode SME insight for supporting semi-automated program verification and automatic simulation adaptation. Our research is guided by two collaborative simulation development efforts. We report on the role of SME insight for modeling the strong interactions between quark and gluons in hadronic physics simulations and modeling thermoacoustic combustion instabilities in simulations of lean, premixed gas turbine engines.

2 Flexible Points

Opportunities for adapting and fine tuning appear throughout typical simulations because simulations typically include a large number of assumptions with acceptable alternatives. Within simulations themselves we call these opportunities for adaptation *flexible points* [3]. We have addressed a number of issues related to the discovery and use of flexible points, including

- understanding why flexible points work and how they can best be exploited in simulation adaptation,
- studying the relationship between flexible points in a simulation and the simulation's underlying model,
- identifying the limitations of using SME insight to identify assumptions in simulations, and
- mapping out the variety of different flexible points and their possible uses.

To use flexible points more effectively, we have evaluated different simulation optimization techniques for manipulating them. Depending on whether the goal is to develop SME insight about potential adaptations or to exploit SME insight to find the ideal adaptation for a specific problem, different techniques may be more appropriate. Unlike most optimization problems, we are interested in techniques that can be monitored and interrupted by the SME when new insight has been gained or when the optimization appears to be homing in on an unacceptable result [4].

To understand why flexible points work, we have studied a domain we call *coercible software*, which is distinguished by the existence of model abstraction opportunities, where decisions must be made about the level of abstraction to use in simulating phenomena. Because simulations rely heavily on choices of

the abstractions for the phenomena that they represent, most simulations are examples of coercible software [5].

To chart the range of flexible points that our language constructs will need to capture, we have developed an evolving taxonomy of flexible points [2, 6]. By distinguishing different types of flexible points, we can develop a toolkit of language constructs that are powerful enough to capture SME insight about a wide variety of simulation adaptations. As we continue to work towards effective tools for capturing and applying SME insight to simulation adaptation, several important challenges remain. These include: analyzing the ways that flexible points can interact with each other, developing requirements for flexible point language constructs, and prototyping and evaluating language constructs for capturing SME insight as flexible points.

To analyze how flexible points interact with each other, we are exploring different assumptions about model abstractions. Starting from the basis established in compositional modeling [7], our goal is to identify those properties that make it possible to automatically manipulate model abstractions without conflicting or unexpected effects. Then, given a set of working assumptions about model abstractions and a taxonomy of flexible points, the requirements for flexible point constructs must be formalized. Lastly, language constructs that meet these requirements must be prototyped and evaluated for use in data-driven simulation adaptation scenarios.

To ensure that our theoretical work on flexible points stays true to the needs of application experts, we have maintained a tightly knit group of researchers who coordinate on end-to-end issues. What good are flexible points if application experts simply require improved visualization tools? We are convinced that flexible points and the technology that supports them are clearly required –a conclusion confirmed by experience gained through our collaborations. In the following sections we describe the application work that is the forge in which our theorizing is tested.

3 Hadronic Physics

When researchers in elementary particle physics utilize traditional methodologies, models are provided in a functional form with free parameters that are adjusted to fit empirical observations. With the acquisition of new data, these free parameters are obvious flexible point candidates for any necessary simulation adaptations. While many methods can optimally tune these flexible points to fit experimental data, there is always some uncertainty as to whether suboptimal tuning results are due to inadequacies in the optimization process or if the model itself is in need of refinements. When data arrives in infrequent batches, the physicist can use domain insight to determine if adjustments to the model are required. In a DDDAS setting, however, data will arrive so frequently that the physicist must have a better understanding of the limitations of the flexible points and the origins of any errors in simulation behavior.

The challenge of identifying and understanding flexible points has become more relevant in recent years because physicists have become increasingly receptive to computational methods derived from the field of artificial intelligence. Many more tunable parameters are being included in these new simulations and computers are taking a more active role in the very development of the mathematical models. It has become acceptable, for example, to replace quadratics with neural-network-based function approximators [8]. Some argue this is a step forward because the traditional quadratic equations were an artifact of outdated techniques that inserted theoretical bias into the simulations. Others question the utility of neural networks where the voluminous parameters have little intuitive meaning and there are unknown consequences of such underlying biases as the selected network topology and threshold function.

Thus, physicists are concerned with understanding the two types of bias, theoretical and systematic, potentially initiated by the flexible points of their simulations. Theoretical bias is the bias introduced by researchers in the form of the precise structure of the models they use, which invariably constrains the form of the solutions. Systematic bias is the bias introduced by algorithms, such as optimization algorithms, which due to the internal operation of the algorithm may favor some results in ways that are not justified by their objective functions. For example, an optimization algorithm may return a parameterization as its final result without revealing that the returned result is only marginally better than several local minima. The physicist must understand and articulate the impact such behavior, or bias, of the optimization algorithm has on the rest of the simulation.

3.1 The SOMPDF collaboration: context and research directions

Parton Distribution Functions (PDF) are the distribution of quark and gluon momenta measured during a collision of protons and/or atomic nuclei that are accelerated to relativistic speeds [9]. Finding a functional parameterization of PDFs constitutes a major research effort in elementary particle physics. Physicists design models that provide a quark/gluon distribution, PDF, at a specific energy scale. The simulated PDFs at multiple energy scales are then combined to produce the proton structure functions (observables) that can be matched with the experimental data obtained from supercolliders. To match experimental data well, a global fitting procedure is applied to the parameters of the PDF models. Some physicists have begun to question the global fitting procedure because the $\chi 2$ results of the fit are likely to underestimate both the systematic/theoretical bias and experimental errors from the various data sets. In order to reduce the impact of theoretical assumptions the usage of neural network methods was proposed [8].

To replace the systematic bias injected by a global parameter fitting process, we provide an interactive fitting tool that helps the physicist control the systematic and theoretical bias present in the fitting process, resulting in a PDF model that is better understood. We extend our previous work based on the

Self-Organizing Map (SOM) algorithm [6] to create a SOM approach to creating PDFs. Our SOMPDF method is an iterative search process in which the expert interactively delineates the boundary between acceptable and unacceptable results. The SOMPDF method samples the parameter space to generate the results of multiple candidate parameterizations. These results are clustered into a SOM and judged by the expert. A statistical analysis of the user-selected PDF parameterizations permits the creation of a new set of similar exploratory parameterizations that will be tested in the next iteration of the SOMPDF fitting. This method capitalizes on the strengths of clustering algorithms because the clusters will provide finer-grained statistical distributions than if the data were treated as a monolithic whole. Potentially fruitful pockets of state space can be extracted and explored. Furthermore, the visual properties of SOMs lend themselves particularly well to user interaction. The SOMs are easy to visualize because they are two-dimensional projections of the nonlinear, high-dimensional state space. Additionally, there is a topological ordering over the SOM that ensures similar data from the high-dimensional space will map to nearby datapoints on the two-dimensional SOM. Using intuitive notions of proximity and a simple point-and-click interface, the user will be able to quickly partition the PDF state space.

In this section, we have described how physicists are expanding their PDF simulations to take greater advantage of empirical data and automatic parameterization techniques. The SOMPDF we are developing serves to integrate the physicist's insight with the model parameterization process. Early observations are that the SOMPDF challenges some of the physicist's insights and refines or reinforces others. Not only does the physicist develop new ideas about relevant and related parameters, but the physicist's evaluation functions are adjusted as well. The next section describes another collaboration with SMEs who confront a more challenging case of having to change underlying scientific models.

4 Thermoacoustic Instability

Continuous combustion processes are central to the application of industrial burners, steam and gas turbines, waste generators, and jet and ramjet engines. Under certain conditions, the heat release rate of the combustion process and the dynamic gas pressure of the combustion chamber can become coupled. This coupling will lead to the growth of large-amplitude fluctuations known as thermoacoustic instabilities [10].

Modeling the thermoacoustic problem involves an accurate description of chemical reactions, fluid-dynamics, and acoustic mechanisms of the system. The chemical reactions proceed at time scales of $10^{-2} - 10^{-8}$ seconds, while the other mechanisms proceed at time scales on the order of $10^{-2} - 10^{-4}$ seconds. This is an instance of multiresolution modeling, which can be identified by the need to simulate a unified phenomenon, given several submodels with different levels of temporal or spatial resolution [11]. It remains a hard problem to maintain a consistent representation between different levels of resolution [12]. Two opera-

tors are needed to aggregate a set of attributes to a lower level of resolution and to disaggregate an attribute to a higher level of resolution. These operators are often not relatively inverse functions.

A detailed model of chemical kinetics in $CH_4/O_2/N_2$ combustion consists of 17 chemical species in 39 elementary reactions. The detailed model is considered highly accurate because it matches experiments in studies of gas combustion [13]. But the detailed model is impractical in the study of thermoacoustic coupling. The extremely fast reactions transpire on time scales that are several orders of magnitude higher than the fluid dynamics and pressure acoustics. A one-step kinetic model uses a single global reaction to capture the combustion of methane. The one-step model is an ad hoc kinetic model, meaning the parameters of the model must be tuned to a particular application. The parameters of the model are derived from the Arrhenius form of the reaction rate ($\omega = [CH_4]^{\alpha}[O_2]^{\beta}Ae^{-\frac{E_a}{RT}}$). The reaction rate (ω) is calculated using four parameters: the CH_4 and O_2 reaction orders (α, β), the pre-exponential factor (A), and the activation energy (E_a). Unfortunately it has been shown that the one-step model is inadequate for simulating the thermoacoustic coupling phenomenon [14].

A one-step model is insufficient because it requires a high activation energy (E_a). The high activation energy causes an amplification of acoustic pressure fluctuations where the detailed model shows a constant magnitude [14]. It is possible that a two-step ad hoc model would behave correctly. In a two-step model, one of the reactions can serve as the rate-limiting reaction. The rate-limiting reaction will prevent the amplification of pressure fluctuations. We are searching the space of two-step models to find an appropriate set of reaction parameters. The two-step model has twice as many degrees of freedom as the one-step model, so there are eight independent parameters. There are many candidate global optimization techniques for minimizing the relative error of the two-step model and detailed model: branch and bound, simulated annealing, stochastic tunneling, genetic algorithms, etc. An additional difficulty is encountered while defining the valid parameter space for the ad hoc two-step model. Most random combinations of the parameter space lead to nonconvergent differential equations in the chemical kinetics simulation [15]. It is impossible to determine a priori which parameter sets do not converge. It is also undecidable whether the chemical kinetics simulation will converge and terminate, given an arbitrary large amount of time.

We have developed a heuristic that searches for converging parameter sets. These parameter sets are subsequently used in the global search techniques for the two-step model. The heuristic begins with an arbitrary pair of (α, β) values and a fixed value for E_a. The space of pre-exponential values is then searched by order-of-magnitude approximation to find those values which quickly converge in the simulation. Once a converging pre-exponential factor is discovered, the space of neighbors is searched to find a value that minimizes the relative error of the models.

We are searching for an appropriate two-step model using the parameter set heuristic. By varying the search domain of (α, β, E_a) different regions of the search space can be studied. In addition there is a set of possible reactions available in the two-step ad hoc model. Whereas the one-step global model is unique, the choice of reactions used in the two-step model is non-unique. We have not yet found an adequate two-step model. Further analysis is needed to show an existence proof, or a nonexistence proof, for an adequate two-step model.

Often, extensive parameter sweeps are necessary to find satisfactory solutions – in our case converging parameter sets – and are practiced widely by SMEs. In most cases sweeps are employed when SMEs have a high degree of confidence in their models and believe that the ideal solution is certain to be found with the right bindings of parameter values. However, as is often the case, they eventually grow weary of parameter sweeps and begin to explore alternatives. The tension between persisting with sweeps and revisiting model design is often dictated by the depth of tradition associated with a model. In some communities models have existed for decades, and to consider modifying them borders on blasphemy.

In our thermoacoustic coupling work we have learned firsthand that provisions for revisiting assumptions about model abstractions, carried along with a model, would provide an expert with more readily accessible information about available alternatives. Rather than spending months, or years, on parameter sweeps, and then finally turning to question the model itself, an expert could routinely review all of the options available – alternative bindings to decisions made during model design – in a manner consistent with the model designer's intentions.

5 Conclusion

For the thermoacoustic instability model, extensive parameter sweeps did not reveal a valid set of model parameters for the two-step version of the model. In the end, only the SMEs could say whether any alternatives to the sets of equations employed could even be considered As a result, no automatic system could explore possible forms for the two-step model. Our experience has demonstrated what the model lacks: design-time capture of flexible point information. On the other hand, the search for a valid model for the Parton distribution functions has benefited considerably from tools that use self-organizing maps to support abstraction and visualization of simulation behavior. With better visualization tools, the SMEs have been able to direct the search for a better parameterizations with far more success than could have been achieved through brute-force parameter sweeps.

Our current belief is that far too much time is spent on parameter sweeps in the application communities. We do not see a clear path to a general approach to DDDAS without resolving this problem. Our investigations of SOMs for parameter identification, and flexible points for formal capture of critical alternatives in selection of model design abstractions, are meant to address the challenges simulationists currently face when seeking to adapt their models to meet desired

objectives. We find our work on flexible points and SOMs for parameter identification on the mark for resolving issues that make adaptation slow, and so we are pursuing the ideas aggressively.

Acknowledgments. The authors gratefully acknowledge the support of the NSF under grant 0426971.

References

1. Douglas, C., Deshmukh, A.: Dynamic data-driven application systems: creating a dynamic and symbiotic coupling of application/simulations with measurements/experiments. In: NSF Sponsored Workshop on Dynamic Data Driven Application Systems. (2000)
2. Carnahan, J.C., Reynolds, P.F., Brogan, D.C.: Language constructs for identifying flexible points in coercible simulations. In: Proceedings of the Fall Simulation Interoperability Workshop. (2004)
3. Carnahan, J.C., Reynolds, P.F., Brogan, D.C.: Visualizing coercible simulations. In: Proceedings of the Winter Simulation Conference. (2004) 411–420
4. Waziruddin, S., Brogan, D.C., Reynolds, P.F.: Coercion through optimization: A classification of optimization techniques. In: Proceedings of the Fall Simulation Interoperability Workshop. (2004)
5. Carnahan, J.C., Reynolds, P.F., Brogan, D.C.: Simulation-specific properties and software reuse. In: Proceedings of the Winter Simulation Conference. (2005) 2492–2499
6. Brogan, D.C., Reynolds, P.F., Bartholet, R.G., Carnahan, J.C., Loitière, Y.: Semi-automated simulation transformation for DDDAS. In: International Conference on Computational Science. (2005) 721–728
7. Nayak, P.P.: Causal approximations. Artificial Intelligence **70** (1994) 277–334
8. Del Debbio, L., Forte, S., et al.: Unbiased determination of the proton structure function f_p^2 with faithful uncertainty estimation. In: hep-ph/0501067. (2005)
9. Feynman, R.: Photon-Hadron Interactions. W. A. Benjamin, Inc. (1972)
10. Hathout, J.P.: Thermoacoustic instability. In Ghoniem, A.F., ed.: Fundamentals and Modeling in Combustion. Volume 2. (1999)
11. Davis, P.K., Bigelow, J.H.: Experiments in multiresolution modeling. RAND Monograph (1998) MR-104.
12. Reynolds, P.F., Srinivasan, S., Natrajan, A.: Consistency maintenance in multiresolution simulation. ACM Transactions on Modeling and Computer Simulation **7** (1997) 368–392
13. Peters, N.: Flame calculations with reduced mechanisms an outline. In Peters, N., Rogg, B., eds.: Reduced kinetic mechanisms for applications in combustion system s. Volume m15 of Lecture Notes in Physics. Springer Verlag (1993) 224240
14. Zambon, A.C.: Modeling of Thermoacoustic Instabilities in Counterflow Flames. PhD thesis, University of Virginia (2005) Department of Mechanical and Aerospace Engineering.
15. Kee, R., Rupley, F., Miller, J.: Chemkin II: A fortran chemical kinetics package for the analysis of gas-phase chemical kinetics. Technical report, Sandia National Laboratories (1989) Sandia Report SAND89-8009.