White Paper

Parametric Design-Space Analysis:
A New Methodology for Evaluating CAE Projects

The growth in computer power has driven an increase in both size and number of CAE simulations used in engineering and scientific analysis. The increase in number of simulation cases has revealed a gap in current post-processing analysis capabilities. A software tool that can explore the design space of these large data sets, compare results and evaluate overall system performance is needed to fill that gap.

Introduction

The dramatic growth in the use of Computer Aided Engineering (CAE) for engineering design is due, to a large degree, to the equally dramatic increase in computer capabilities. CAE codes solve the complex non-linear partial differential equations (PDEs) that describe such phenomena as fluid flow, structural mechanics, and heat transfer. These solutions require enormous amount of memory and CPU cycles. For example, a high-fidelity solution of the air flow past an airplane will typically require a grid with hundreds of millions of cells. Until recently, the available computer resources were barely sufficient to run a handful of cases. Now it is possible to run hundreds of high-fidelity CAE simulations, making its use in the engineering design process practical.

The role of high-fidelity CAE simulation in engineering design is expanding at the expense of low-fidelity analyses and experimentation. In the aerodynamic design of aircraft, for example, the analysis was typically done with linear potential 'panel' methods or transonic potential methods. These analyses were supplemental to the extensive wind-tunnel testing where the bulk of the force and moment data was acquired.

A scale model of the airplane would be placed in the wind tunnel and force and moment data would be taken over a wide range of flight parameters (speed, altitude, angle-of-attack, and yaw angle) and configurations (control surface positions, flap positions, or any other configuration changes being investigated. In recent years, high-fidelity computational fluid dynamics (CFD) simulations have been supplanting the wind-tunnel as the dominant source of aerodynamic data. The wind tunnel is now viewed primarily as a tool to calibrate the CFD solutions.

The result has been an explosion in the size and number of CAE data sets - one for each set of flight and configuration parameters. Surveys by Tecplot, Inc. indicate that the number of CFD cases run for a particular project now number in the hundreds, or even thousands. Traditional techniques for analyzing the data-verifying quality, exploring the design space, extracting flow features and integrated quantities, and reporting the results, are simply not possible in the time available to the design engineer. As a result, engineers now generally focus on the integrated results (forces and moments) and the detailed flow field is simply ignored. In other words, they are basing their decisions on 0.00001% of the information generated by the CFD simulation and 99.99999% of the valuable information is unused.[1]

Figure 1. Matrix view in Tecplot 360 (Chorus) helps engineers and scientists organize, investigate and compare collections of CFD runs and test data.

---

[1] *Assumes a grid with 100 million cells.*

Needs of the Engineer or Scientist

Simulations are performed to improve engineering design and decision making. The design process begins with a set of requirements that include, among other things, the operating environment of the machine, the desired performance and needed safety factors (some regulated). The designer will develop an array of potential designs, with a set of free parameters, that are evaluated for their ability to meet these requirements. It is the role of simulation, along with experimentation and other analyses, to determine which of the potential designs meets these requirements.

Simulations are also used to analyze existing systems outside of any design process. For example, surface-water or estuary simulations are used to evaluate water management strategies and the flooding potential of rivers. Likewise, ground water and air flow simulations are used in environmental analyses to estimate the propagation of pollutants. These analyses still result in large collections of data sets, representing a range of key input parameters such as rainfall rate and distribution, and various mitigation options such as the amount and timing of excess runoff release from flood-control reservoirs.

The result for each simulation is a set of metadata, like forces and moments, and one or more data files containing the field data. Engineers need to manage this field-data and metadata for a large collection of cases, verify the quality (convergence and accuracy) of the solution, identify trends and anomalies in the metadata, identify the root cause of these trends and anomalies by examination of the field-data, compare data from different sources, and collaborate with others by sharing the data.

Components of Parametric Analysis

Data Management

In the related collections of data sets, the combined size and number of runs is growing with Moore's law. The metadata from each case (input parameters and scalar results for each simulation) is generally stored in a database or spreadsheet. Field data is generally stored in binary output files on file servers with high-bandwidth connections to the compute server.

As the size of the collections grows, it has become increasingly difficult to track and manage the data. In a presentation at the Society of Petroleum Engineers, Jim Crompton said that reservoir simulation "engineers spend 30% of their time looking for data, verifying data accuracy and formatting data. They work on their personal space, so it can get lost when they leave."[2] For this reason, management of the simulation data is a critical component of parametric design-space analysis software.

There are two components of data management:

- Safe storage of the data
- Rapid understanding of what data is available

In general, the metadata should be stored in a database like any critical company information. On the other hand, the field data doesn't work efficiently with most database formats so the database may just contain links to traditional data files stored in a hierarchical file system.

Visual representations of the data sets give the user a quick understanding of the available results and how they relate to one another. One way to do this is with scatter plots (Figure 2) where symbols show the values of the independent variables (input parameters to the simulation) for all of the solutions in one image. For highly dimensional data, the dependence on the independent variables not displayed can be determined through interactive filtering, depth cues and/or arrays of scatter plots. These techniques are discussed more in the next section.

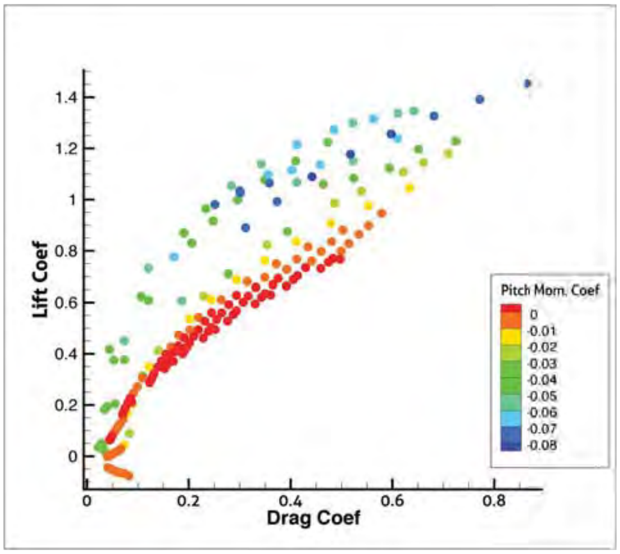

Figure 2. Scatter plot in Tecplot 360, a suite of design space analysis tools.

---

[2] *Jim Crompton, "Putting the Focus on Data," lecture at the Pacific Northwest Section of the SPE September Meeting, September 14, 2010.*

Visualization and Analysis of High-Dimensional Metadata

In general, the metadata is a function of several independent parameters, each of which represents a separate dimension for visualization and analysis. In aerodynamic analysis of an airplane, for example, the independent parameters might be:

- Flight conditions
- Angle-of- attack
- Yaw angle
- Speed
- Altitude
- Configuration parameters
- Positions of control surface like ailerons, elevator and rudder.

This is at least seven parameters (dimensions). The engineer wishes to analyze the relationship between the integrated quantities such as lift, drag, pitching moment and the seven independent parameters.

The engineer may create this aerodynamic database for one or more of the following purposes:

- Analysis or simulation of the behavior of the vehicle
- Inclusion in a control or guidance system
- Optimization of a configuration parameter
- Analysis of vehicle sensitivity to input parameters
- Analysis of uncertainty

In any case, the first step is usually to understand the relationships between the dependent and independent parameters.

Visualization of seven dimensional data is difficult. Humans see the world in three dimensions (four if you include motion over time) and computers only have two dimensional screens (three if you include animations). One option is to do an array of 21 conventional two-dimensional plots, each of which displays the variation of a dependent variable over two of the dimensions. Another option is to use techniques from business analytics, such as parallel coordinate plots (Figure 3) and box plots (Figure 4).

Statistical techniques are critical in the analysis of high-dimensional metadata. Interactive exploration using filters and plots such as scatter plots is a common technique to better understand the data. Adjusting filters allows the user to explore the dependence of the data on non-displayed dimensions. Other techniques, such as box plots, reduce dimensionality by summarizing the behavior over the non-displayed dimensions.

Perhaps the most important analytical tools are surrogate models (Figure 5), which estimate the variation of the functional relationship between a dependent variable and the independent variables. This functional form may be used in visualization, estimation of optimal configurations, sensitivity analyses, and as a substitute for the full simulation in subsequent analysis.

In visualization, surrogate models are especially useful when the data are sparse. They are used to: create:

- Line plots
- Carpet plots
- Iso-surfaces

In optimization, they provide the functional form that may be solved or searched for maxima and minima.

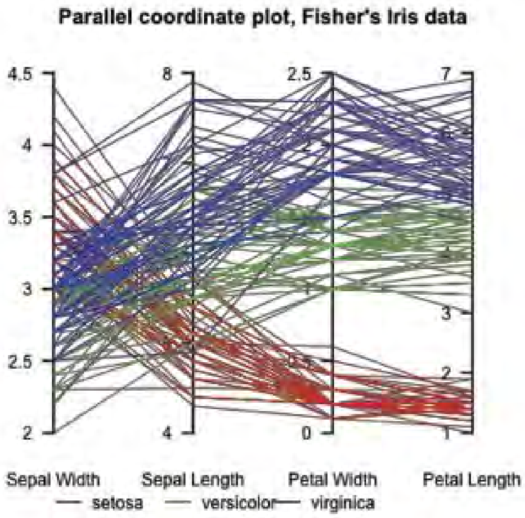

Figure 3: Example parallel coordinate plots for 4 dimensional data

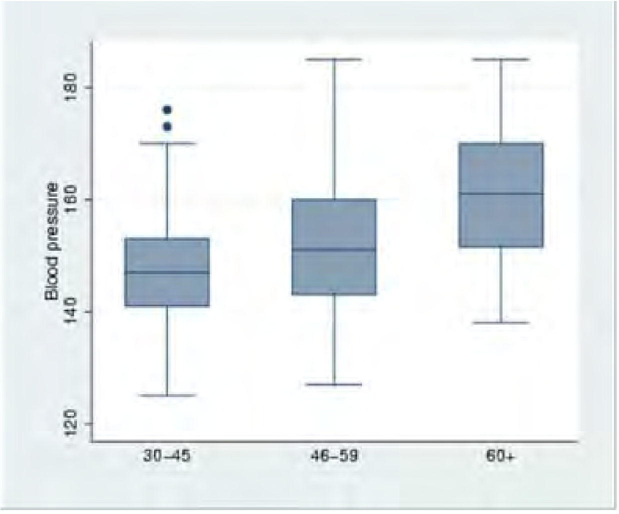

Figure 4: Box plot: center line is median, box extends from 25th to 75th percentile, and whisker extends from 5th to 95th percentile.

For sensitivity analysis, where the variation of dependent variables with independent variables would be difficult to estimate without a surrogate model, this functional form becomes critical.

Finally, subsequent analyses such as Monte-Carlo simulations, flight simulators, and control and guidance system design would be virtually impossible if the full simulation had to be run every time the dependent variables needed to be evaluated as a new set of independent variables.

The key difference between parametric design-space analysis and more traditional analytics like business analytics is that field data is available for each simulation case – each point in the metadata.

These large data sets contain the detailed variation of the field variables like pressure throughout the computational domain (space and, perhaps, time). This data is very valuable. For example, it can be analyzed to find the root cause of anomalies in a ways that are not possible, or at least extremely difficult, with measured data.

For example if the engineer notices an unexpected inflection in the lift versus angle-of-attack plot she could visualize the flow field near the aircraft wing to search for a cause. It might be a region of boundary layer separation caused by interactions between the engine nacelle and the wing, and the engineer would obtain a valuable and timely insight that could be used to improve the vehicle design. A parametric analysis tool must be able to make this "deep dive" easy.

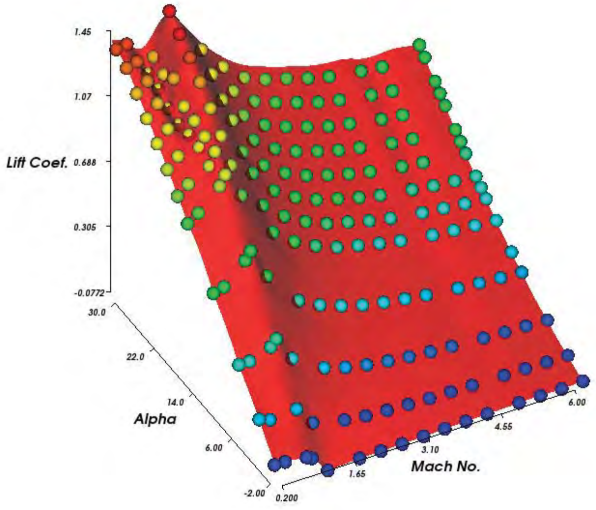

Figure 5. Surrogate model in Tecplot 360's Chorus design-space analysis tool.

Visualization and Analysis of Field Data

Computer aided engineering (CAE) simulations generally solve a set of partial differential equations (PDEs) to get the distribution of scalar, vector and tensor variables on a grid filling the computational domain. The grid is either an IJK-ordered grid (a mapping of a rectangular grid to a non-rectangular domain) or a finite-element grid where the domain is subdivided into elementary shapes like tetrahedra and hexahedra. The PDEs are solved using iterative techniques to get the distribution of the field variables on the grid. The field variables are physical quantities like pressure, temperature, velocity or stress. The field data are integrated to get the metadata variables described in the previous section.

One of the primary goals of parametric design-space analysis is to verify the quality of the CAE solutions. There are three main sources of error in the simulations:

- Violation of the assumptions in the equations
- Insufficient convergence
- Truncation error

The partial differential equations solved by the CAE code are based on simplifying assumptions. Structural dynamics codes usually assume small displacements so that the equations can be linearized. If the computed displacements are too large this assumption can lead to substantial errors. Computational fluid dynamics codes which solve the Navier-Stokes equations nearly always model the effects of turbulence rather than computing all turbulent eddies. These models are imperfect and portions of the flow are frequently inaccurate.

These, and many other assumptions, need to be tested to ensure the accuracy of the results.

Additional assumptions are generally made in the application of boundary conditions. For example, the actual flow domain for external aerodynamics is extremely large but it is always truncated for CFD grids. If the outer boundary is too close to the vehicle, it may alter the results. Also, it is critical that the boundary conditions be well posed (for example, there is no inflow on an outflow boundary) or the solution may not give meaningful results.

Many of the CAE codes solve non-linear PDEs using iterative techniques. These solvers must be iterated to convergence or there will be errors in the solution. This converge is usually verified using convergence history plots. These are line plots versus iteration number of either residuals, which must go to zero, or they are integrated quantities, which must reach a constant value.

Truncation error is estimated using a grid convergence study. This requires the solution at the same input values (independent parameters), but with a series of increasingly fine grids. The truncation error varies with the square of the grid spacing (for second-order schemes) so the difference between the coarse-grid solution and the fine-grid solution will give an estimate of the error on the coarse grid.

A primary goal of the "deep dive" is to investigate the root cause of anomalies in metadata (example in Figure 6). For CFD solutions, these anomalies are typically caused by fluid-dynamic phenomena like boundary-layer separation or vortices.

Another required capability of the "deep dive" is comparison of the field-data solutions. This can either be a visual comparison or a numerical comparison. The software should show both solutions and some representation of the change between the solutions.

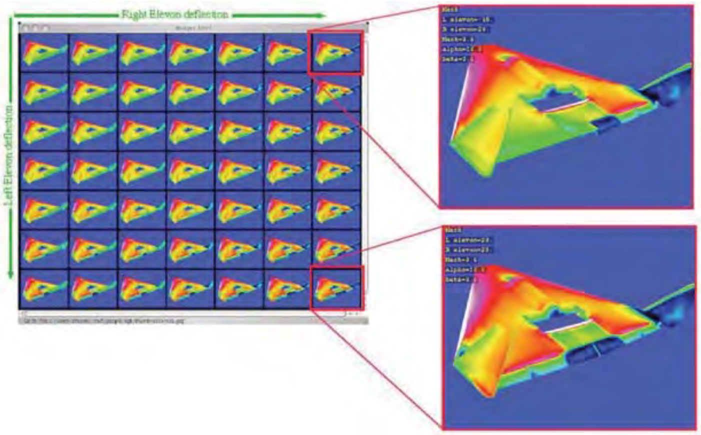

Figure 6. Investigating the root cause of anomalies or trends in metadata. (Image from KrisnaKumar, K., et. al., "Intelligent Control of the Bees Flyer," AIAA 2004-6274)

Data Mining

The dependent variables in the metadata are scalar descriptive data that are computed from the field data. Typically they are integrated quantities like forces and moments, but other quantities are possible. For example, the maximum temperature within the field would be a useful descriptive quantity in heat transfer computations. Other quantities would require more detailed feature-extraction capabilities, like the percentage of the surface with boundary layer separation.

In addition to the scalar descriptive data, images of field data visualizations are commonly extracted. These provide a quick, but limited, method of viewing and comparing the field data solutions. Figure 6 is an example of an array of images.

Future of Parametric Design-Space Analysis

The number of CAE cases for each project is also expected to grow. As computer power expands following Moore's law, a portion of the addition capability will be used to increase the number of cells in each case, to reduce the truncation error and model more complicated geometries, and the remainder will be used to run additional cases. The product of grid size and number of cases will increase at Moore's law – the number of cases will increase more slowly than Moore's law.

Tecplot 360. Tecplot 360's suite of visualization & analysis tools integrates CFD post-processing, field and parametric data management, and powerful analytics into a single environment. An engineer using Tecplot 360 can manage and analyze collections of CFD simulations, and compare them in a single environment while evaluating overall system performance.

Learn about using Tecplot 360 for parametric CFD analysis:

https://www.tecplot.com/360