Parameter Optimization of High-Fidelity Simulation Using DOE and Response Surface Methods

IOE 570 Final Project

Bryon J. Sohns

James T. Allison

University of Michigan

Department of Mechanical Engineering

December 15, 2004

Chapter 1

Executive Summary

The goal of this project is to tune the parameters of a high-fidelity simulation of a HUMMWV military vehicle using experimental data from an actual HUMMWV. This model has been developed by RDECOM (at the Detroit Arsenal) to assist them in the design of future HUMMWV platforms.

Initial analysis was performed using an unreplicated full-factorial experimental design and Lenth's method to evaluate factor significance. Response surface methods were then used to determine the simulation parameter settings that resulted in a simulation that most closely matched the experimental data from the actual HUMMWV.

The model predicts the vertical acceleration of the vehicle center of mass in response to a vertical impulse displacement applied to all four wheels simultaneously. The difference between the simulation predictions and experimental data is quantified by an algorithm, AVASIM.

Four model parameters were chosen as factors, and a 2^4 full factorial experiment (without replication) was specified with factor levels at $\pm$ 10% nominal. The model is deterministic, precluding the use of replicates. Applying Lenth's method none of the factors were deemed significant. A response surface method was employed for further investigation of optimal parameter values, and data already obtained for the 2^4 factorial design was used for the required corner points. Additional center and axial points were obtained as required throughout the process. A line search revealed a region with curvature, and a second order model was used to identify a stationary point.

In general the model accuracy response is ill-behaved. It is suggested that the steepest ascent optimization scheme implemented in this study is not adequate—more sophisticated approaches are suggested. However, the steepest ascent results indicate that the accuracy of the model will be greatly improved by setting the model parameters to the stationary point values.

Chapter 2

Introduction

The objective of this project is to use experimental data from an actual vehicle to tune model parameters such that the simulation accurately predicts actual vehicle response. In addition, it is desired to identify which parameters have a significant effect on model accuracy.

The system used for this project is the High Mobility Multipurpose Wheeled Vehicle, or HUMMWV [1]. This vehicle is a mainstay of the Army fleet and is constantly being redesigned to better suit the needs of the Army. This constant redesigning requires that the Army have a high-fidelity simulation model of the vehicle to allow for powerful design techniques to be employed. The simulation model used for this project, developed by RDECOM (at the Detroit Arsenal), is a high-fidelity three-dimensional dynamic model of a HUMMWV with full suspension characteristics. The model is currently configured to parameter values based on an older model vehicle. The Army wishes to parameterize the model to reflect the performance of the latest design in order to use the model in further design studies. In order to do this an actual HUMMWV was set up on a four-poster shaker table and instrumented to record various data. The input to the system is a vertical displacement in the shape of an impulse applied at all four wheels simultaneously. The vertical acceleration of the center of gravity of the vehicle was chosen as the output based on the test setup and input applied. The availability of this experimental data will facilitate the task of finding model parameter values that produce a more accurate simulation. In order to compare the output from the experimental setup with the output from the simulation model an algorithm called AVASIM was employed [5, 6, 7]. This algorithm generates a measure of model accuracy which varies from one to negative infinity, with one representing one hundred percent accuracy and zero representing a user-defined tolerance on the accuracy of the simulation output. This dictates a maximum-the-best optimization strategy.

Four parameters of the simulation model were identified and selected for this study based on discussions with Army personnel. These parameters were the front and rear spring rates and damping coefficients. Access to only one experimental data set was available for this project. Due to this fact and the deterministic nature of the output from a non-heuristic computational simulation, experimental replicates cannot be made. This prompted a full factorial experiment with four factors and no replications for the initial stage of this study. This experimental design was suitable to analyze both the effect of the parameters on model accuracy and the significance of each factor. Lenth's method for unreplicated experiments was employed initially to identify important factors within the simulation model. Unfortunately, this method indicated that none of the factors or interactions were significant.

After the failure of Lenth's method to identify important factors, it was decided to use a response surface methodology to tune the model parameters with respect to all four factors. The data obtained for the Lenth's method analysis were used in addition to new center points to evaluate the parameter main effects and curvature around the nominal parameter values. After finding an absence of curvature around the nominal values, a line search was performed in the direction of steepest ascent. A change in slope was detected, and a new nominal point was set where that change occurred. A full response surface analysis including the factorial, center and axial points was then used to identify a stationary point and possible optimum parameter value settings. The accuracy at the stationary point was markedly improved over the baseline accuracy.

Chapter 3 provides an overview of the AVASIM algorithm used to evaluate model accuracy, and introduces the design of experiment and response surface methods employed in this study. Chapter 4 presents the results of the investigation, and Chapter 5 summarizes the interpretation of the results and suggests opportunities for future work.

Chapter 3

Simulation and Analysis Methods

This chapter provides some detail on the AVASIM algorithm used to evaluate model accuracy, and provides an overview of the data analysis methods employed in this study.

3.1 AVASIM Overview

The AVASIM algorithm evaluates model accuracy by calculating an Overall Performance Index for a model. To calculate the Overall Performance Index for a specific input and system configuration using AVASIM an output of interest must be first identified. The comparison is made between the model under investigation and the truth. In this project the truth is obtained from experimental data on a real system. However, in other cases it might acquired from a full model. The full model is a more complicated model with enough complexity to provide the most accurate predictions of the systems behavior. For the output of interest, target points can be selected based on the use of the model (Figure 3.1). For example, in a model with a response similar to an ideal second order system, the engineer is usually interested in the overshoot, rise time, steady state value, etc. As many of these points as desired can be defined as targets, in both amplitude and time. For each target point tolerances are then defined.

These tolerances can be either absolute or relative (percentage, etc.) tolerances. Based on intuition and experience relative tolerances are the preferred type of tolerance. Absolute tolerances can be used if the numerical value of a target point is at or near zero (resulting in very large negative performance indices for relatively small errors), or there exists some other problem-specific reason to do so. After target points and tolerances are chosen a performance index at each point is evaluated. This is done using the following formulas:

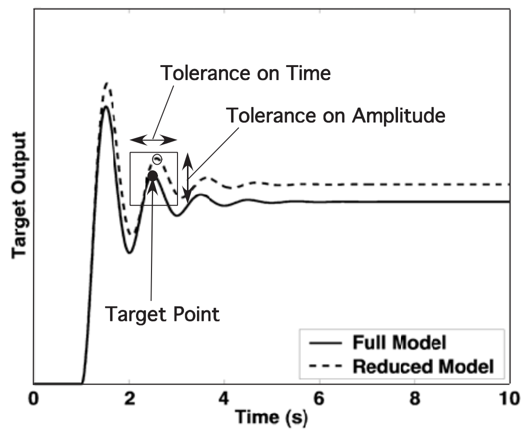

Figure 3.1: Sample Input to AVASIM.

Relative Tolerances:

$$J_{p,i} = 1 - \frac{|y_{i,full} - y_{i,red}|}{y_{i,full} \cdot tolr_i}, \quad i = 1 \dots n$$

Absolute Tolerances:

$$J_{p,i} = 1 - \frac{|y_{i,full} - y_{i,red}|}{tola_i}, \quad i = 1 \dots n$$

It is important to note that a Target Performance Index of 1 at any point corresponds to 100% accuracy. It is also of importance to note that a Target Performance Index of 0 corresponds to the response of the reduced model being at the tolerance for that particular target point. If there are no target points defined, as is the case for this project, this step can be skipped.

Once each Target Point Performance Index has been calculated a Response Performance Index is generated. This is done through the creation of a threshold model. This is simply amplitude scaled and time shifted version of the full model of the form:

$$y_{thr} = a \cdot y_{full}(t + b)$$

The scaling factors a and b are chosen to be as large as possible while insuring that the tolerances at all target points are met. If no target points have been defined then a percentage tolerance can be defined directly for a and b. This leads to a maximum perturbed model that still has acceptable accuracy. Once this metamodel output has been generated, the residual sums between the reduced and the full model and the threshold and the full model are calculated using the following formulas:

$$RS_{red} = \int_0^T |y_{full}(t) - y_{red}(t)| dt$$

$$RS_{thr} = \int_0^T |y_{full}(t) - y_{thr}(t)| dt$$

After these values are obtained the Response Performance Index in obtained in a fashion similar to the Target Performance Indices using the following formula:

$$J_r = 1 - \frac{RS_{red}}{RS_{thr}}$$

Finally, the Overall Performance Index for the reduced model for a specific input and system configuration is calculated as an average of the Target Point and Response Performance Indices using the following formula:

$$PI = \frac{1}{2}\left[\left(\frac{1}{n}\sum_{i=1}^{n} J_{p,i}\right) + J_r\right]$$

If there are no target points defined then this value is equal to the Response Performance Index. Once this value has been calculated there have been two proposed methods of assessing model validity called the liberal and conservative criterion. The conservative criterion states that all the individual Target Point Performance Indices ($J_{p,i}$) and the Response Performance Index (J_r) must be positive in order for the model to have sufficient accuracy and therefore be valid. The liberal criteria states that it is only necessary for the Overall Performance Index (PI) to be positive (meaning that some Target Point or Response Indices can be negative, or have insufficient accuracy) for the model to be valid. When there are no target points defined these criterion are equal.

The response of model accuracy evaluation tools is notoriously ill-behaved. The response surface is typically highly non-linear, sometimes discontinuous, multimodal, and usually numerically noisy. These properties pose substantial challenges with regard to finding an unconstrained optimum, discussed in the next section.

3.2 Problem Development

The nominal vehicle parameters are set to:

$$\mathbf{x_0} = (A, B, C, D)^T = (0.770, 0.617, 94.020, 98.277)^T$$

where A = front spring rate, B = rear spring rate, C = front damping rate, and D = rear damping rate. The factorial levels were set to $\pm$ 10% of these nominal values. AVASIM was used to evaluate model accuracy at all 2^4 factorial points. Since no replicates were available, Lenth's method (individual error rate approach) was selected to check factor significance. With Lenth's method the pseudo standard error (PSE) can be used to estimate the standard error [8]. The full factorial experiment also provides insight into the response of the model accuracy evaluation.

Response surface methods provide a generally efficient method for performing unconstrained optimization. The well known steepest ascent method [2, 8] uses the principal that the gradient of a function points in the direction of steepest ascent. A line search in this direction is performed until a maximum in the gradient direction is found. The gradient at that new point is calculated, and used as an updated line search direction. This process can be repeated until the gradient is determined to be the zero vector, at which point necessary conditions for optimality are satisfied (this point is called a stationary point $\mathbf{x_s}$). It can be shown that the search directions for adjacent iterates are orthogonal. If the response function is highly elliptic the steepest ascent method can converge very slowly. Alternate methods such Newton's method or Quasi-Newton methods such as BFGS update are more efficient in this case.

Instead of iterating until the gradient is zero, a second order model can be fit to the data around the current point. This of course only makes sense if curvature is present in the region, which can be checked by either comparing the average value of factorial corner points to the average of center points, or calculating the p-value of the aggregate quadratic term. If curvature is present, a second order model is then fit to the data (requiring additional axial points). The stationary point of this second order model is then used to estimate the location of stationary point[1]. If a second order model approximates the response well, this is a reasonable approach. Care should be exercised if the stationary point is determined to be far outside the range of sampled data.

[1]This is the essence of Newton's method for unconstrained optimization, except that a second order model is iteratively fit to every new estimate for the stationary point.

Chapter 4

Results

The model accuracy response corresponding to the 2^4 full factorial points described in §3.2 are displayed in Table 4.1.

Table 4.1: 2^4 full factorial experiment results.

A	B	C	D	PI
−	−	−	−	−14.037
−	−	−	+	−12.761
−	−	+	−	−13.223
−	−	+	+	−12.693
−	+	−	−	−12.671
−	+	−	+	−12.772
−	+	+	−	−11.666
−	+	+	+	−11.829
+	−	−	−	−15.036
+	−	−	+	−12.453
+	−	+	−	−13.531
+	−	+	+	−12.951
+	+	−	−	−14.225
+	+	−	+	−12.808
+	+	+	−	−13.156
+	+	+	+	−12.446

This data was used to calculate the main effects and interactions, and the PSE required for Lenth's method. The cutoff value for significance was found to be $IER_{5\%,15} = 2.16$, and the pseudo standard error was $PSE = 0.457$. Table 4.2 displays all of the effects and t test values for each effect. It was discovered that none of the factorial effects were significant according to Lenth's method.

The data gathered from the original 2^4 experiment was used to begin the response surface process for finding the optimum model parameter settings. Seven center points, generated by small (within 1%) perturbations around the nominal values, were used to perform a curvature check. MINITAB was used to fit a first order model (with an additional aggregate quadratic term) to the data, and the p-value for the aggregate quadratic term was found to be 0.971, indicating an exceedingly small presence of any curvature. It was decided to

Table 4.2: Lenth's method calculations.

effect	$\lvert\overline{y}\rvert$	t
A	0.619	1.354
B	0.639	1.397
C	0.658	1.439
D	0.854	1.867
AB	0.305	0.667
AC	0.049	0.107
AD	0.468	1.024
BC	0.186	0.407
BD	0.389	0.849
CD	0.440	0.961
ABC	0.080	0.175
ABD	0.129	0.282
ACD	0.238	0.520
BCD	0.248	0.541
ABCD	0.077	0.168

perform a line search in the steepest ascent direction. The scaled coefficients of the linear model provide the search direction:

$$\boldsymbol{\Delta} = (-0.725, 0.748, 0.771, 1.000)^T$$

Because steps of 2Δ would quickly bring the parameter values out of the feasible range, higher resolution 1Δ step sizes were used. The line search produced the following sequence of PI values:

$$\{-12.0407,\ -11.1058,\ -11.0782,\ -10.8952,\ -11.0702,\ -11.0899,\ -11.4695\}$$

A change of slope can be observed around the 4Δ response, and the investigation was directed to that point. The nominal values were updated to the 4Δ parameter values. Another 2^4 experiment was designed, centered at the new nominal values. The factorial points were set to $\pm$ 10% of the updated nominal values, and seven zero points were specified using small perturbations of the nominal values. Again MINITAB was used to fit a first order model (with an additional aggregate quadratic term) to the data. The p-value for the aggregate quadratic term in this case was found to be 0.004, indicating a strong presence of curvature. Since a stationary point may be near, a second order model (equation 4.1) was fit after eight axial points were specified and evaluated (using a value of $\alpha = 2.0$).

$$PI = \beta_0 + \mathbf{x}^T\mathbf{b} + \mathbf{x}^T\mathbf{B}\mathbf{x} \tag{4.1}$$

The values of β_0, $\mathbf{b}$, and $\mathbf{B}$ were found to be:

$$\beta_0 = -11.0199$$

$$\mathbf{b} = (-0.1266, -0.4241, 0.2548, -0.0009)^T$$

$$\mathbf{B} = \begin{bmatrix} -0.0128 & -0.0047 & -0.0082 & 0.0094 \\ -0.0047 & -0.0953 & -0.0278 & 0.0261 \\ -0.0082 & -0.0278 & -0.0369 & -0.0217 \\ 0.0094 & 0.0261 & -0.0217 & -0.0010 \end{bmatrix}$$

Setting the gradient of the second order model to zero, the stationary point was determined to be:

$$\mathbf{x}_* = (0.4127, -0.2484, -0.4515, 6.8110)^T$$

The values of PI at the stationary point $\mathbf{x}_*$ calculated with the second order model and the simulation were respectively:

$$PI(\mathbf{x}_*) = -11.0541 \quad \text{(using second order model)}$$

$$PI(\mathbf{x}_*) = -11.4523 \quad \text{(using simulation)}$$

The second order model was only roughly accurate in this case. To determine the nature of the stationary point the eigenvectors of the $\mathbf{B}$ matrix were found:

$$\lambda_{\mathbf{B}} = (-0.1098, -0.0444, -0.0150, 0.0233)^T$$

The $\mathbf{B}$ matrix is therefore indefinite, indicating a saddlepoint. However, note from the $\mathbf{b}$ vector that the effect from factor D is not very significant. If the factor D is ignored, the resulting 3×3 $\mathbf{B}$ matrix is negative definite, indicating a local maximum. Setting D to its nominal value and running the simulation again with the other factors at stationary point values, we find that $PI = -11.1459$, which is in fact better than the simulation response at the stationary point.

The accuracy of the model at the baseline parameter values and at the stationary point can be depicted graphically. Figure 4.1 illustrates how the dynamic time response predicted by the model using baseline parameter values roughly agrees with the experimental data. Figure 4.2 shows that the model with the stationary point parameter values predicts a dynamic time response that more closely follows the experimental data.

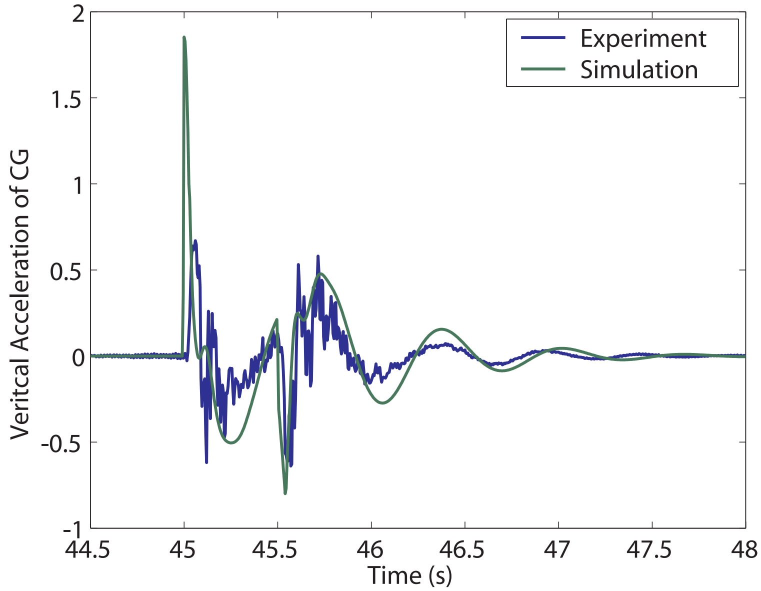

Figure 4.1: Dynamic time response: experimental response vs. predicted response with initial model parameter settings.

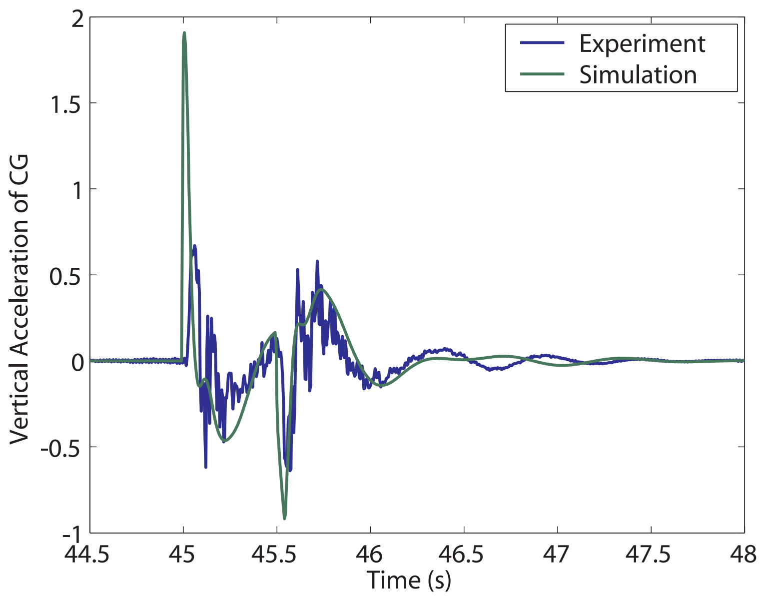

Figure 4.2: Dynamic time response: experimental response vs. predicted response with stationary point parameter settings.

Chapter 5

Conclusion

The AVASIM algorithm was used to quantify the accuracy of a computer simulation of a HUMMWV Army vehicle with respect to actual experimental data. The objective of this study was to find parameter values that maximized this accuracy. A 2^4 experimental design without replication was used for initial analysis. This indicated a lack of curvature in the region around the baseline constraints, requiring a line search to find a region with curvature so that a stationary point and optimum parameter set could be found. A steepest ascent response surface method was employed, and a stationary point with marked improvement over the baseline performance.

The functional nature of PI merits some discussion. As mentioned earlier, it is known to be highly non-linear, sometimes discontinuous, and somewhat noisy. This makes response surface methods difficult to implement. Second order models do not fit the response well. This was demonstrated by the discrepancy between the second order response and the simulation response at the same parameter value settings presented in Chapter 4.

Other optimization approaches are better suited for this task. One suggested approach is to use a gradient-free algorithm, such as DIRECT[1] [3] to find a region in the model parameter space likely to have the global optimum, followed up with a more sophisticated gradient-based algorithm, such as Sequential Quadratic Programming [4]. This approach is likely to find the global optimum more efficiently.

[1]A method of systematic search over the design space, acronym for DIvided RECTangles.

Bibliography

[1] J. Aardema. Failure analysis of the lower rear ball joint on the high-mobility multipurpose wheeled vehicle (HUMMWV). Technical Report Technical Report No. 13337, TACOM, 1988.

[2] Mokhtar S. Bazaraa, Hanif D. Sherali, and C. M. Shetty. *Nonlinear Programming: Theory and Algorithms*. John Wiley and Sons, Inc., second edition, 1993.

[3] Donald R. Jones. DIRECT. In *Encyclopedia of Optimization*. Kluwer Academic Publishers, 1999.

[4] M.J.D. Powell. Algoritms for nonliear constraints that use lagrangian functions. *Mathematical Programming*, 14:224–248, 1978.

[5] P. Sendur, J.L. Stein, L.S. Louca, and H. Peng. A model accuracy and validation algorithm. In *Proceedings of the 2002 ASME International Mechanical Engineering Congress and Exposition*, June 1995. New Orleans, LA., Published by American Society of Mechanical Engineers, ISBN 0-7918-1692-3, New York, NY.

[6] P. Sendur, J.L. Stein, L.S. Louca, and H. Peng. An algorithm for the assessment of reduced dynamic system models for design. In *Proceedings of the International Conference on Simulation and Multimedia in Engineering Education*, pages 92–101, 2003. Orlando, FL., Published by the Society for Modeling and Simulation International Simulation, ISBN 1-56555-261-1, San Diego, CA.

[7] P. Sendur, J.L. Stein, L.S. Louca, and H. Peng. An algorithm for the selection of physical system model order based on desired state accuracy and computational efficiency. In *Proceedings of the 2003 ASME IMECE Conference*, 2003.

[8] C.F. Wu and Michael Hamada. *Experiments: Planning, Analysis, and Parameter Design Optimization*. John Wiley and Sons, Inc., 2000.