STATISTICAL ANALYSIS IN INVERSE PROBLEMS WITH EXPERIMENTAL DATA

ADAM ATTARIAN

14 OCTOBER 2008

In this project we apply hypothesis testing and inverse problem methodology to the spring mass dashpot model with experimentally collected data. We then identify the physical beam parameters and provide confidence intervals for our assertions. In contrast to the last project, we do not have access to truth information and so our uncertainty analysis becomes that much more important.

1. METHODOLOGY

1.1. **Parameter Estimation.**

The spring mass system is given by the differential equation

$$\ddot{x}(t) + C\dot{x}(t) + Kx(t) = 0, \tag{1}$$

with physical parameters $C, K \in \mathbb{R}^+$ and initial conditions

$$x(t_0) = x_0, \qquad \dot{x}(t_0) = v_0.$$

In the lab we collect acceleration data using an accelerometer and an oscilloscope, and so we must derive an expression for acceleration from the mathematical model. If we let $x_1(t) \equiv x(t)$ and $x_2(t) \equiv \dot{x}(t)$, the acceleration of the beam is given by

$$\ddot{x}_1(t) = -Cx_2(t) - Kx_1(t).$$

Both $x_1(t)$ and $x_2(t)$ are obtained by solving (1). As we are using acceleration data, both initial conditions are unknown and must also be treated as parameters to be estimated. In order to estimate $q^T = [x_0, v_0, C, K]$ we formulate the goal function,

$$J(q) = r^T r, \tag{2}$$

where r is an $N \times 1$ vector containing the residuals

$$r_j(q) = \big(-Cx_2(t_j; q) - Kx_1(t_j; q)\big) - z_j, \qquad j = 1, \ldots, N,$$

where z_j are the collected data observations over N time points. This is simply the difference between an evaluation of the model at a certain parameter and the collected data. We then search for the q that minimizes (2) using the Nelder-Mead algorithm. As (2) is merely a sum of squares cost function, $\hat{q}$ is also referred to as the ordinary least squares estimate.

1.2. **Uncertainty Analysis.**

We are estimating the values of C, K, x_0 and v_0, and we wish to provide a quantitative measure for how sure we are of our result.

As the data collected is inherently noisy, it is impossible to estimate exactly the *actual* parameter value q_0 that generates the observations z_j. Therefore, the parameter value that minimizes (2) is only an estimate to q_0 and is denoted $\hat{q}$. We assume the data are corrupted with additive, white gaussian noise with zero mean and variance σ_0^2.

If we treat q_0 and σ_0 as unknowns, we then use their estimates to form the covariance matrix $\hat{\Sigma} \in \text{spd}(\mathbb{R},\text{p})$.

$$\Sigma(\hat{q}) = \hat{\sigma}^2 \left[\chi^T(\hat{q})\chi(\hat{q})\right]^{-1}, \tag{3}$$

where $\hat{q}$ is the estimator from an ordinary least squares process, and χ is the $N \times p$ sensitivity matrix, containing the solutions to the sensitivity equations for p unknown parameters and initial conditions over N longitudinal observations. Given we are estimating sensitivities for two parameters and two initial conditions for two states, we have ten sensitivity equations to solve. These equations form a system of coupled ordinary differential equations and instead of computing these functions explicitly, we use automatic differentiation to compute the needed partial derivatives.

The variance estimator $\hat{\sigma}^2$, is given by

$$\hat{\sigma}^2 = \frac{1}{N-p} r(\hat{q})^T r(\hat{q}). \tag{4}$$

The standard errors to be used in computing confidence intervals are given by

$$SE_k(\hat{q}) = \sqrt{\hat{\Sigma}_{kk}}, \quad k = 1, \ldots, p. \tag{5}$$

The confidence intervals at the $1-\alpha$ level are defined by

$$q_{0,k} \pm t_{1-\alpha/2} SE_k(\hat{q}), \tag{6}$$

where $t_{1-\alpha/2}$ is the critical value pulled from the student's t distribution with $N-p$ degrees of freedom. As $N \to$ large, the t distribution converges to the normal distribution. The critical value is the determined by $P\{T \geq t_{1-\alpha/2}\} = \alpha/2$, and in MATLAB via `tinv(1-alpha/2,N-p)`.

2. Results

The most difficult aspect of identifying the beam parameters was choosing a good initial iterate for the optimizer. Simply guessing led the optimizer to consistently converge to the zero solution. While this solution may have minimized cost, it certainly did not reproduce observed dynamics. To aid in this, we compute the spectrum of both the data and the model through an N point FFT. In doing so, we are able to much easily obtain a ball-park estimate by varying K by hand until the frequencies match. The initial condition x_0 is then varied by hand to match the amplitudes between the data spectrum and the model spectrum. This provides a much better initial iterate for the optimizer than just blindly guessing parameter sets.

The optimization is ultimately performed in the time domain, as it is unclear how changing C will affect the spectrum of the model output; additionally if one were to optimize in frequency space the phase shift of the signals would need to be accounted for, adding to the complexity of the goal function. After each $\hat{q}$ is found, the algorithm is restarted using this value as the initial iterate. In most cases, the fit slightly improves.

Data are collected by exciting the beam to two modes, at $f = 6.5$Hz and $f = 55$Hz. The accelerometer's output is volts, and so we are using nonstandard units for acceleration, but this does not affect the model formulation. For each frequency, we test the significance of damping in the model given the collected data.

It should be noted that it is difficult to invert the matrix $\left[\chi^T(\hat{q})\chi(\hat{q})\right]$. The condition number for this matrix are on the order of 10^13 or so, hinting at ill-posednes. This can be attributed to several factors. First, the matrix contains entries of wildly different magnitudes. In the high frequency case, the covariance matrix spans fourteen magnitudes. Secondly, the model is very sensitive to the parameter. A small change in the parameter, especially between K and x_0, can produce very different system outputs. This small change in input producing a large change in output is another hint of ill-posedness. Lastly, the sensitivity analysis is a local analysis – that is, the functions are evaluated at $\hat{q}$, which in turn depends heavily on where and how the data are truncated for analysis. Had we truncated the data differently we would have obtained a possibly very different $\hat{q}$ value, perhaps generating a well-posed $\left[\chi^T(\hat{q})\chi(\hat{q})\right]$.

In the results that follow, in the case of near-rank deficiency, the covariance matrices given are computed using the Moore-Penrose psuedo inverse. Therefore, the reported standard errors and confidence intervals may not be fully reliable given our data set. This is more true for the high frequency data sets.

2.1. **Low Frequency Oscillations.** We first excite the beam at its first fundamental frequency, $f = 6.5$Hz and include damping in the model. Using the collected data, we obtain a parameter estimate

$$\hat{q}^\star_{f=6.5,C\neq 0} = \begin{bmatrix} x_0 \\ v_0 \\ C \\ K \end{bmatrix} = \begin{bmatrix} -0.001975 \\ 0.01247 \\ 0.3245 \\ 1451.6 \end{bmatrix}, \qquad J(\hat{q}^\star_{f=6.5,C\neq 0}) = 9.3671. \tag{7}$$

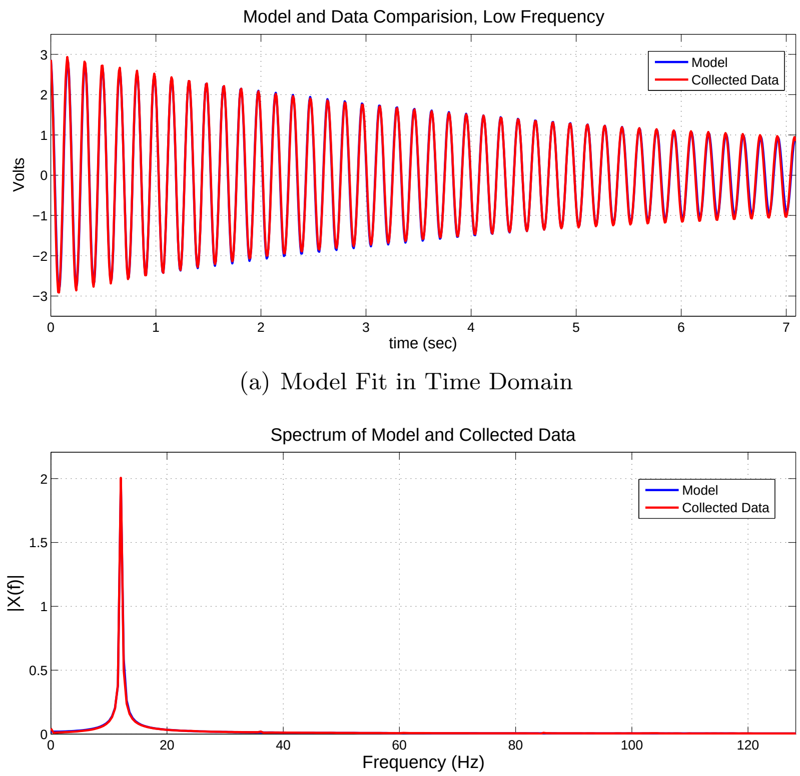
The model versus data fit, in both time and frequency domains, is given in Figure 1. The covariance matrix for this estimate is given in (8).

$$\hat{\Sigma}(q^\star_{f=6.5,C\neq 0}) = \begin{bmatrix} 0.087648 & -0.0074888 & -0.00018885 & 0.0013824 \\ -0.0074888 & 127.88 & -0.0011254 & -0.2752 \\ -0.00018885 & -0.0011254 & 6.928\times 10^{-7} & -6.1361\times 10^{-7} \\ 0.0013824 & -0.2752 & -6.1361\times 10^{-7} & 0.001015 \end{bmatrix}. \tag{8}$$

The standard errors and 95% confidence levels are then given in Table 1.

Parameter	Standard Error	95% CI
C	0.29605	0.3245 ± 0.58103
K	11.309	1451.6 ± 22.194
x_0	8.324×10^{-4}	-0.001975 ± 0.001633
v_0	0.03186	0.01247 ± 0.06253

TABLE 1. Standard Errors and Confidence Bounds for $f = 6.5$Hz and $C \neq 0$.



(a) Model Fit in Time Domain

(b) Model Fit in Frequency Domain

FIGURE 1. Model Fit for the Low Frequency, $C \neq 0$

To test the significance of damping in the model, we set $C = 0$ and recompute $\hat{q}_{f=6.5}$, obtaining the best fit to the data with no damping. We obtain the parameter estimate

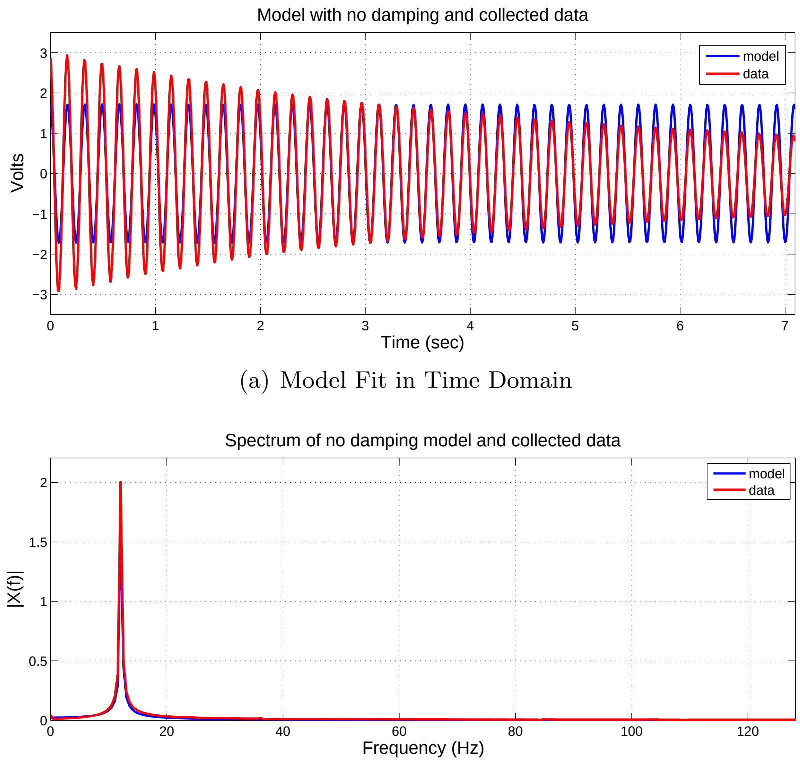
$$\hat{q}^{\star}_{f=6.5,C=0} = \begin{bmatrix} x_0 \\ v_0 \\ K \end{bmatrix} = \begin{bmatrix} -0.001173 \\ 0.005999 \\ 1452.9 \end{bmatrix}, \qquad J(\hat{q}^{\star}_{f=6.5,C=0}) = 153.13. \tag{9}$$

The model fit with no damping and data are given in Figure 2. The covariance matrix is given in (10), and the standard errors and 95% confidence levels are given in Table 2.

$$\hat{\Sigma}(q^{\star}_{f=6.5,C=0}) = \begin{bmatrix} 6.61 & -4.6647 \times 10^{-5} & -0.013352 \\ -4.6647 \times 10^{-5} & 6.9949 \times 10^{-9} & 9.4016 \times 10^{-8} \\ -0.013352 & 9.4016 \times 10^{-8} & 3.669 \times 10^{-5} \end{bmatrix} \tag{10}$$

In order to test the significance of the damping term, we form the null hypothesis H_0 that damping does *not* have a statistically signifiant role in the model. The test statistic $\hat{U}$ is defined as

$$\hat{U} = N \frac{J(q^{\star}_{f=6.5,C=0}) - J(q^{\star}_{f=6.5,C\neq 0})}{J(q^{\star}_{f=6.5,C\neq 0})} \sim \chi^2(1). \tag{11}$$



(a) Model Fit in Time Domain

(b) Model Fit in Frequency Domain

FIGURE 2. Model Fit for the Low Frequency, $C = 0$

Parameter	Standard Error	95% CI
K	2.571	1452.9 ± 5.0458
x_0	8.3636×10^{-5}	$-0.001173 \pm 1.641 \times 10^{-4}$
v_0	0.0006058	0.005999 ± 0.003026

TABLE 2. Standard Errors and Confidence Bounds for $f = 6.5$Hz and $C = 0$.

Equation (11) is analogous to the relative error between the two goal function values. We can use this distribution to test our hypothesis: if the test statistic, $\hat{U} > \tau$, where τ is a critical value associated with some specific $1 - \alpha$ confidence level, then we *reject* H_0.

Let us choose to use a quite restrictive α level of .001, representing 99.9% confidence. The τ critical value for $\alpha = .001$ is $\tau = 10.83$. For the low frequency data, $\hat{U} = 13951$. Obviously since $\hat{U} \gg \tau$, we reject H_0, and thus *damping plays a statistically significant role*. This is true for nearly any α level given the magnitude of $\hat{U}$.

The minimum value α^* at which H_0 can be rejected is the p–value. Smaller p-values are better: this says that the evidence is stronger in favor of rejecting the null hypothesis. The p–value can be computed by evaluating 1 minus the cumulative distribution function at $\hat{U}$, which is defined as $1 - P\{x < \hat{U}\}, x \sim \chi^2(1)$, or in MATLAB as `1-chi2cdf(`$\hat{U}$`,1)`.

Given the magnitude of $\hat{U}$, it is very far out in the right tail of the distribution and so the associated p–value is identically zero. Therefore we assert that with absolute certainty, given the data, damping plays a significant role in the spring mass dashpot model.

2.2. **High Frequency Oscillations.** The experiments of the previous section are repeated at $f = 55$Hz. For the sake of brevity, the results will be summarized. The methodology is identical to the the previous section.

For $C \neq 0$, we find

$$\hat{q}^{\star}_{f=55,C\neq 0} = \begin{bmatrix} x_0 \\ v_0 \\ C \\ K \end{bmatrix} = \begin{bmatrix} -1.3545 \times 10^{-5} \\ -2.9692 \times 10^{-6} \\ 1.8202 \\ 1.092 \times 10^{5} \end{bmatrix}, \qquad J(\hat{q}^{\star}_{f=55,C\neq 0}) = 2.8884.$$

The high frequency data and the associated model fit, in both time and frequency, are given in Figure 3. The associated uncertainty analysis are given in (12) and Table 3.

$$(12) \quad \hat{\Sigma}(q^{\star}_{f=55,C\neq 0}) = \begin{bmatrix} 19.804 & 13.558 & -7.1498 \times 10^{-5} & 1.0551 \times 10^{-4} \\ 13.558 & 2.1781 \times 10^{6} & 1.2001 \times 10^{-4} & -7.919 \\ -7.1498 \times 10^{-5} & 1.2001 \times 10^{-4} & 4.5077 \times 10^{-10} & -1.1833 \times 10^{-9} \\ 1.0551 \times 10^{-4} & -7.919 & -1.1833 \times 10^{-9} & 4.9994 \times 10^{-5} \end{bmatrix}.$$

Parameter	Standard Error	95% CI
C	4.4501	1.8202 ± 8.7429
K	1475.8	$1.092 \times 10^{5} \pm 2899.5$
x_0	2.1231×10^{-5}	$-1.3545 \times 10^{-5} \pm 4.1712 \times 10^{-5}$
v_0	0.0070707	$-2.9692 \times 10^{-6} \pm 0.013891$

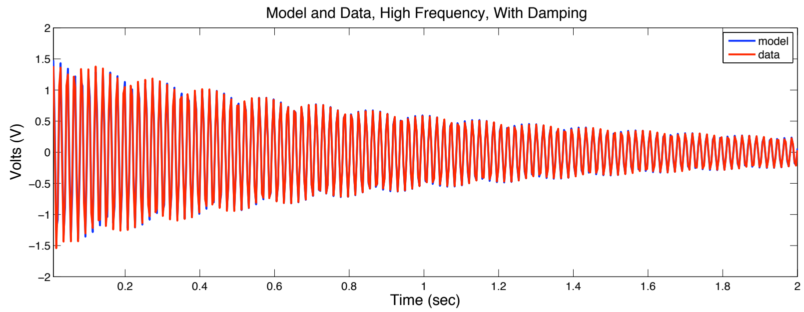
Table 3. Standard Errors and Confidence Bounds for $f = 55$Hz and $C \neq 0$.

For $C = 0$, we obtain the parameter estimate $q^{\star}_{f=55,C=0}$,

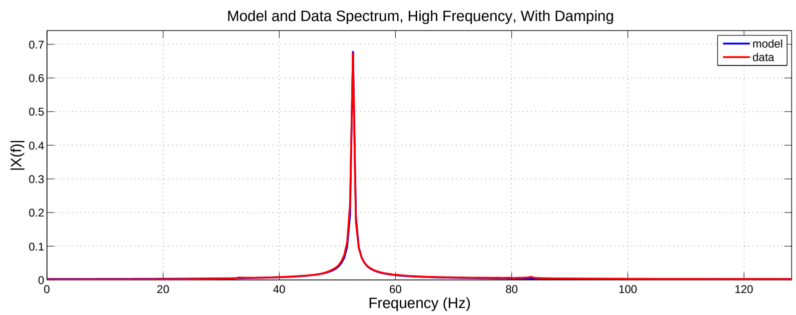
$$q^{\star}_{f=55,C=0} = \begin{bmatrix} x_0 \\ v_0 \\ K \end{bmatrix} = \begin{bmatrix} -6.222 \times 10^{-6} \\ 3.3957 \times 10^{-4} \\ 1.0907 \times 10^{5} \end{bmatrix}.$$

The model fit with no damping, in both time and frequency, is given in Figure 4, and the covariance matrix for this estimate is given in (13). The standard errors and confidence regions are given in Table 4.

$$(13) \qquad \hat{\Sigma}(q^{\star}_{f=55,C=0}) = \begin{bmatrix} 2.0289 \times 10^{5} & -2.9629 \times 10^{-3} & -0.58824 \\ -2.9629 \times 10^{-3} & 6.5408 \times 10^{-12} & 8.5766 \times^{-10} \\ -0.58824 & 8.5766 \times^{-10} & 2.3723 \times 10^{-6} \end{bmatrix}.$$



(a) Model Fit in Time Domain



(b) Model Fit in Frequency Domain

FIGURE 3. Model Fit for the High Frequency, $C \neq 0$

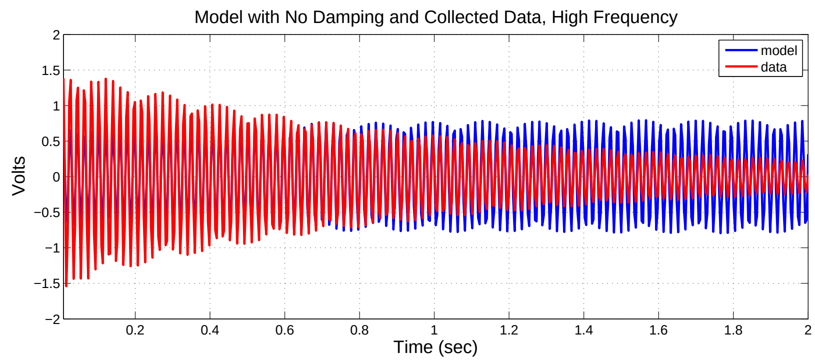
Parameter	Standard Error	95% CI
K	450.44	$1.0907 \times 10^5 \pm 884.95$
x_0	2.5575×10^{-6}	$-6.222 \times 10^{-6} \pm 5.0246 \times 10^{-6}$
v_0	0.0015402	$3.3957 \times 10^{-4} \pm 0.003026$

TABLE 4. Standard Errors and Confidence Bounds for $f = 55$Hz and $C = 0$.

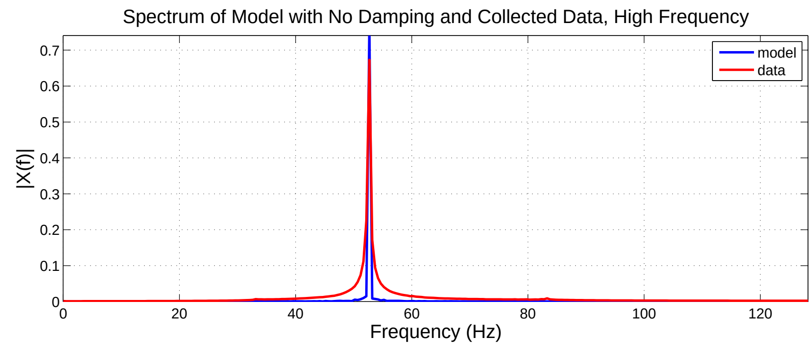
To test the significance of the damping, we form the test statisitic $\hat{U}$ as before. For this frequency, $\hat{U} = 7842.6$. As before, $\hat{U} \gg \tau\ \forall \alpha$, and so we reject the hypothesis that damping does not play a significant role with near absolute certainty. As with the low frequency data, we have our p–value to be identically zero.

3. Conclusion and Analysis

In performing the experiments and the analysis we find that the initial iterate for the optimizer plays a crucial role in obtaining sensible parameter estimates. Through using Fourier analysis and matching frequencies, a good initial iterate is obtained from which a time domain optimization can be performed. There is some error intrinsic in these results, either from the numerical difficulty associated with computing the inverse of $\chi^T \chi$, but also in the experiment setup. The accelerometer is attached to the end of the beam. This weight is nontrivial and is not accounted for in the model, so the data may have extra noise that may not be modeled with simple white gaussian noise.



(a) Model Fit in Time Domain



(b) Model Fit in Frequency Domain

Figure 4. Model Fit for the High Frequency, $C = 0$

Most of the results match up with expectations. The K parameters have relatively small confidence intervals, agreeing with the sensitivity of the cost function to K. There is also a very tight tolerance on x_0, again agreeing with the sensitivity experienced when attempting to find a suitable initial iterate. Lastly, we conclude that damping plays a statistically significant role at both frequencies.