Project 1

Bayesian Inference for the Mechanical Model of a Falling Object

Parameter Estimation of Models of Falling Objects

Consider an object that due to the gravitational forces is falling in a medium, such as air. The object is released from a height h at time t_0 as shown in Figure 1. The origin of the coordinate system that measures the position of the object is set at the released position which is assumed to be measured exactly.

Figure 1: Falling object

Accounting also for the forces applied from the air to the object, the equation of motion of the object is given by

$$m\frac{d^2z(t)}{dt^2} = mg - R \tag{1}$$

where $z(t)$ denotes the position of the object, m is the mass of the object, g is the acceleration of gravity, R is the air resistance. A simple and well-known model for the air resistance is $R = mc\upsilon^2(t)$, where $\upsilon(t) = \frac{dz(t)}{dt}$ is the object's velocity and c is the air resistance coefficient which depends on the air friction coefficient C_D, density of the air ρ_{air}, density of the object ρ_{ball} and the geometric characteristics of the object. For a sphere, the air resistance coefficient takes the form $c = \frac{C_D}{2}\frac{3\rho_{air}}{4R_{ball}\rho_{ball}}$. Substituting the air resistance formula into (1), the equation of motion takes the form

$$\frac{d^2z(t)}{dt^2} = g - c\left[\frac{dz(t)}{dt}\right]^2 \tag{2}$$

This is a differential equation which need to be solved with initial conditions $z(t_0) = z_0$ and $\upsilon(t_0) = \upsilon_0$. The values of the initial position and velocity are assumed to be both zero $\left(\upsilon_0 = z_0 = 0\right)$.

The equation can be solved analytically to yield the expression for the velocity

$$\upsilon(t) = \upsilon_\infty \tanh\left[\sqrt{gc}\,(t - t_0)\right] \tag{3}$$

and the position

$$z(t) = \frac{1}{c}\ln\cosh\left[\sqrt{gc}\,(t - t_0)\right] \tag{4}$$

The derivation details are given in the Appendix. The prediction of the velocity and the position of the falling object depends on the model chosen for the air resistance coefficient and the values of the parameters g, c, t_0, z_0 and υ_0 involved in (3) and (4). Unless otherwise stated explicitly, the values of theinitial position and velocity are assumed to be zero.

We are interested in estimating the resistance coefficient c using experimental data $D=[\{\hat{z}_k,t_k\}_{1\to N}]$ of the positions $\hat{z}_k$ of the object at time instances t_k, $k=1,\ldots,N$, taken from the analysis of high-speed camera pictures. The data value at a time instance can be considered to be independent from the data values at other time instances. It is assumed that the model is perfect (no model error), the values of the rest of the variables involved in the model are known, and that the accuracy of the camera equipment and data processing unit for estimating experimentally the objects position has a measurement error which can be adequately modeled as zero-mean Gaussian variable with standard deviation σ. It will be further assumed, for the purposes of this demonstration, that the level of the measurement error is known a priori which means that σ is given. So the model parameter set $\theta=(c)$ includes only a single parameter, the air resistance coefficient to be inferred from the data.

Given the set of independent observations/data, we are interested in updating the uncertainty in the air resistance coefficient c of the model. To infer the value of the single parameter (air friction coefficient) given the data, one needs to set up a model for the prediction error, select the prior, build the expression of the likelihood and use the asymptotic analysis to approximate the posterior by a Gaussian.

Prediction Error Model: Based on the theory, the prediction error equation is

$$\hat{z}_k = z(t_k;c)+e_k \tag{5}$$

where e_k is the prediction error assumed to be a zero-mean Gaussian variable with known variance σ^2, i.e. $e_k \sim N(0,\sigma^2)$.

Prior: For simplicity it is assumed that the prior distribution for the air friction coefficient is uniform, that is,

$$p(c\,|\,I)=\begin{cases}1/(c_{\max}-c_{\min}), & c\in[c_{\min},c_{\max}]\\ 0 & \text{otherwise}\end{cases} \tag{6}$$

where the bounds $c_{\min}$ and $c_{\max}$ are wide so that the parameter inference and uncertainty is not influenced by such choices.

Likelihood: To evaluate the likelihood $p(D\,|\,\theta,I)\equiv p(\{\hat{z}_k\}_{1\to N}\,|\,c,I)$, one uses the prediction error equation, the fact that the measured data at different time instances are independent, and applies successively the product rule of the axiom of probability to finally derive that

$$\begin{aligned} p(D\,|\,c,I) &= p(\{\hat{z}_k\}_{1\to N}\,|\,c,I)=\prod_{k=1}^{N} p(\hat{z}_k\,|\,c,I)\\ &=\prod_{k=1}^{N}\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2\sigma^2}\left[\hat{z}_k-z(t_k;c)\right]^2\right]\end{aligned} \tag{7}$$

Posterior PDF: Substituting in the Bayes formula the expression for the likelihood given in (7) and the uniform prior PDF given in (6), the posterior PDF of the uncertain parameter c takes the form

$$p(c\,|\,D,I)\propto\frac{1}{\sigma^N}\exp\left[-\frac{1}{2\sigma^2}J(c)\right] \tag{8}$$

where the misfit function (or measure of fit function) $J(c)$ is given by

$$J(c) = \sum_{k=1}^{N} \left[\hat{z}_k - z(t_k, c) \right]^2 \tag{9}$$

It can be seen that $J(c)$ is a highly nonlinear function of the model parameter c to be inferred in this exercise.

Most Probable Value or Best Estimate: The function $L(c)$, defined in theory as the minus the logarithm of the posterior PDF, is given by

$$L(c) = -\log p(c \mid D, I) = \frac{1}{2\sigma^2} J(c) + \frac{N}{2} \log \sigma^2 + \text{constant} \tag{10}$$

For constant variance σ^2, the most probable value $\hat{c}$ of c, also known as Maximum-A-Posteriory (MAP) estimate, maximizes the posterior PDF or, equivalently, minimizes $L(c)$ or the misfit function $J(c)$. In contrast to the previous example, the optimization cannot be performed analytically (using $\frac{\partial L(c)}{\partial c} = 0$). Any gradient-based optimization algorithm can be used to find the optimum. These algorithms require that the gradient of the objective function with respect to the parameters is provided. The analytic expression for the gradient of the objective with respect to the air friction coefficient is

$$\frac{\partial L(c)}{\partial c} = \frac{1}{2\sigma^2} \sum_{\kappa=1}^{N} \frac{\partial}{\partial c} [\hat{z}_\kappa - z(t_k, c)]^2 = \frac{1}{\sigma^2} \sum_{\kappa=1}^{N} \left\{ (z(t_k, c) - \hat{z}_\kappa) \frac{\partial z}{\partial c} \right\}$$

where

$$\frac{\partial z}{\partial C} = \frac{(t_\kappa - t_0)\, g \tanh\left[\sqrt{gC}\,(t_\kappa - t_0)\right]}{2C\sqrt{gC}} - \frac{\ln\left\{\cosh\left[\sqrt{gC}\,(t_\kappa - t_0)\right]\right\}}{C^2}$$

Derivation details are given in the Appendix. All results of the optimization in this example are carried out in Matlab using the *fminunc* m-file.

Uncertainty in Model Parameter: The uncertainty in the value of the model parameter c is characterized by the second derivative of the function $L(\mu)$ which is given by (for derivation see Appendix)

$$\left. \frac{d^2 L}{dc^2} \right|_{c=\hat{c}} = \frac{1}{\sigma^2} \sum_{\kappa=1}^{N} \left\{ \left(\frac{\partial z}{\partial c} \right)^2 + (z(t_k, c) - \hat{z}_\kappa) \frac{\partial^2 z}{\partial c^2} \right\}$$

where

$$\frac{\partial^2 z}{\partial C^2} = \frac{g(t_\kappa - t_0)}{2\sqrt{g}} \frac{\frac{\partial T}{\partial C} C^{2/3} - \frac{3}{2} T * C^{1/2}}{C^3} - \frac{\frac{\partial z}{\partial C} * C - z}{C^2}$$

$$T = \tanh\left[\sqrt{gC}\,(t_\kappa - t_0)\right]$$

and

$$\frac{\partial T}{\partial C} = \left(1 - T^2\right)\left(t_\kappa - t_0\right)\frac{g}{2\sqrt{gC}}$$

The measure of the uncertainty, provided by the square root of the inverse of the second derivative of $L(c)$ evaluated at the most probable value, is given by

$$\sigma_{c|D} = \left(\left.\frac{d^2 L}{dc^2}\right|_{c=\hat{c}}\right)^{-1/2} \tag{11}$$

which is obtained in closed form for this example, provided that the MAP value has been estimated numerically.

Given the MAP $\hat{c}$ and the uncertainty index $\sigma_{c|D}$ we can write a measure of the uncertainty interval of c in the form

$$\hat{c} \pm \sigma_{c|D} \tag{12}$$

Asymptotic Posterior PDF: Following the theoretical result and using the MAP $\hat{c}$ and the uncertainty index $\sigma_{c|D}$, the posterior PDF is approximated by the Gaussian distribution

$$p(c \mid D, I) = \frac{1}{\sqrt{2\pi}\sigma_{c|D}} \exp\left[-\frac{1}{2\sigma_{c|D}^2}(c - \hat{c})^2\right] \tag{13}$$

Simulated Experimental Data: In absence of experimental we generate simulated data using the prediction error equation for a nominal value $\underline{\theta}_{nom}$ of the model parameter set $\underline{\theta}$ and a nominal value σ_{nom} of the prediction error parameter σ. The selected time instances can be viewed on figure 2, along with the velocity diagram ($v(t)$) which we obtain from the model. Experimental values are tabulated in Table 1 for $N = 20$ experimental data points and for the following values of the model parameters: $g_{nom} = 9.81$, $c_{nom} = 0.106$, $t_{0,nom} = 1$ sec. The time instances are shown in the first column, while the results in the second third and fourth column are based on nominal values of measurement error given by $\sigma_{nom} = 0.001,\ 0.10,\ 0.30$, respectively. This allows us to study the effect of measurement error on the parameter inference.

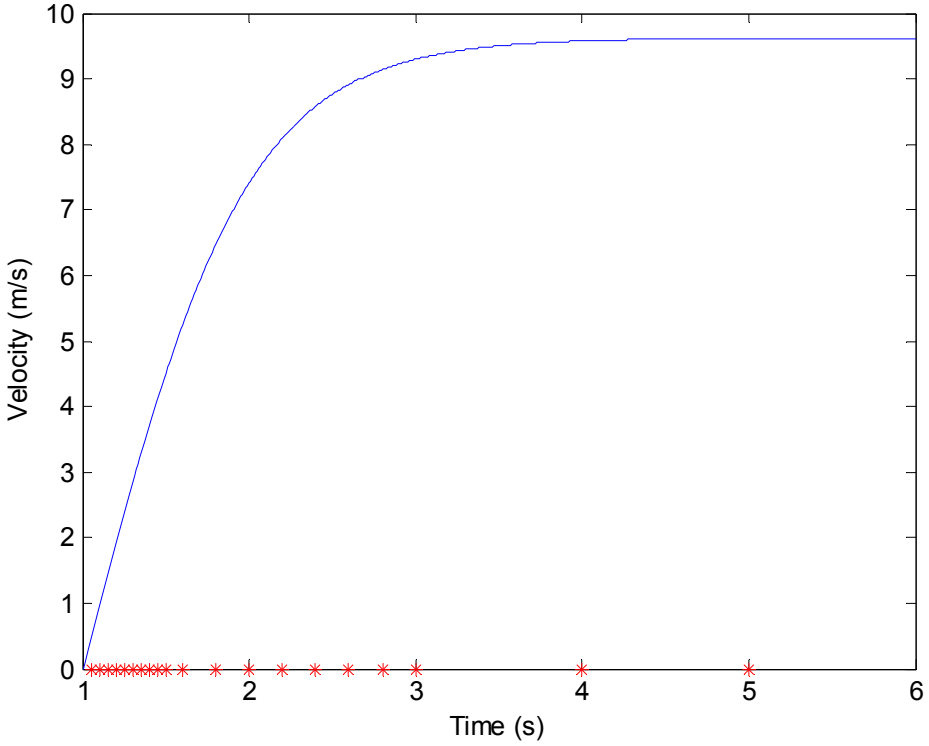
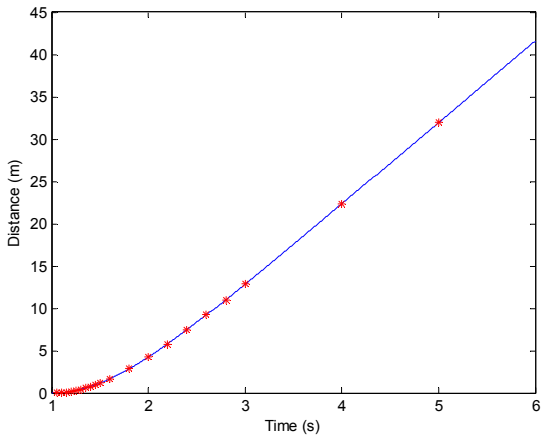


Figure 2: Comparison of the selected time instances with the velocity plot.

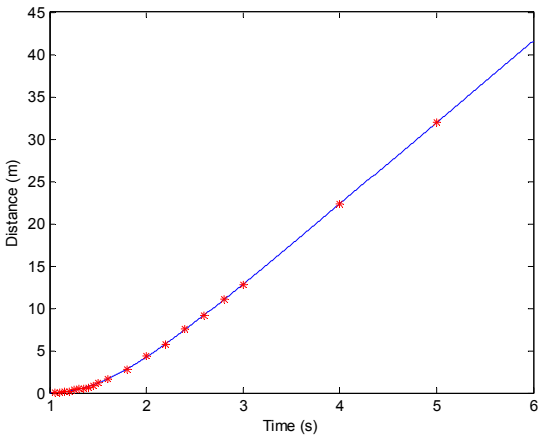
Table 1: Experimental data for different nominal values of σ_{nom}

tκ	σ = 0.001	σ = 0.1	σ = 0.3
1,05	0,010974	0,102239	-0,111934
1,10	0,046636	0,018954	-0,082516
1,15	0,110837	0,212871	0,710952
1,20	0,193019	0,160348	0,480153
1,25	0,303365	0,404578	0,173697
1,30	0,434767	0,497665	0,629414
1,35	0,590750	0,567222	0,480500
1,40	0,763888	0,677387	0,975723
1,45	0,959737	0,855933	1,384999
1,50	1,176771	1,149528	0,695180
1,60	1,665758	1,621698	1,974168
1,80	2,842586	2,801649	3,279906
2,00	4,234222	4,333185	4,249072
2,20	5,786667	5,758120	6,311766
2,40	7,457189	7,571241	7,503489
2,60	9,206042	9,154222	8,836249
2,80	11,013191	11,111480	10,356175
3,00	12,860832	12,807276	12,759479
4,00	22,341639	22,359716	22,556121
5,00	31,944013	32,041227	32,039376

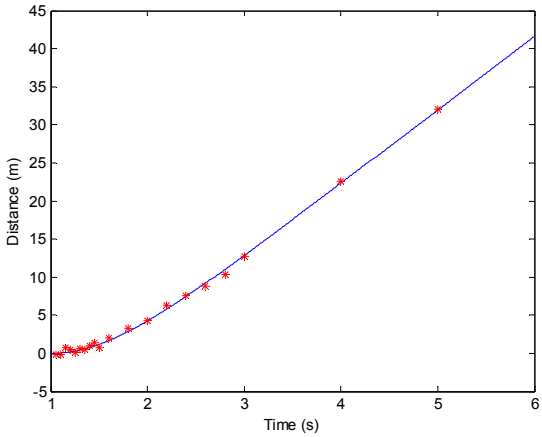
Figure 3 shows the prediction of the position $z(t;\theta_{nom})$ obtained from the model given the nominal values of the model parameters, along with the generated "experimental" data, for various levels of model error. It can be seen that the smaller the σ value, the smaller the error, thus the experimental data fit the model predicted z better.



(a) $\sigma_{nom} = 0.001$



(b) $\sigma_{nom} = 0.1$



(c) $\sigma_{nom} = 0.3$

Figure 3: Comparison of nominal model prediction of position and the experimental measurements for different values of the prediction error (a) $\sigma_{nom} = 0.001$, (b) $\sigma_{nom} = 0.1$, (c) $\sigma_{nom} = 0.3$.

Numerical Results: We will be examining the following experimental cases.

A. Changing the number of data points (N)

- Case 1: N=1
- Case 2: N=2
- Case 3: N=5
- Case 4: N=20

B. Changing the time instances chosen (always for N=2)

We define 3 phases of the drop. Phase 1: the velocity is changing very quickly.
Phase 2: the velocity is changing at a slower pace.
Phase 3: the velocity approximately remains the same.

So we also have

- Case 5: Data taken during phase 2
- Case 6: Data taken during phase 3

Note that case 2 is also for N=2 with data taken during phase 1.

We the experimental data is taken from table 1 and is always the same from table 1. For more visible results we examine the case of $\sigma_{nom} = 0.3$.

The time instances for the different experimental cases (different N values) are reported in Figure 5.

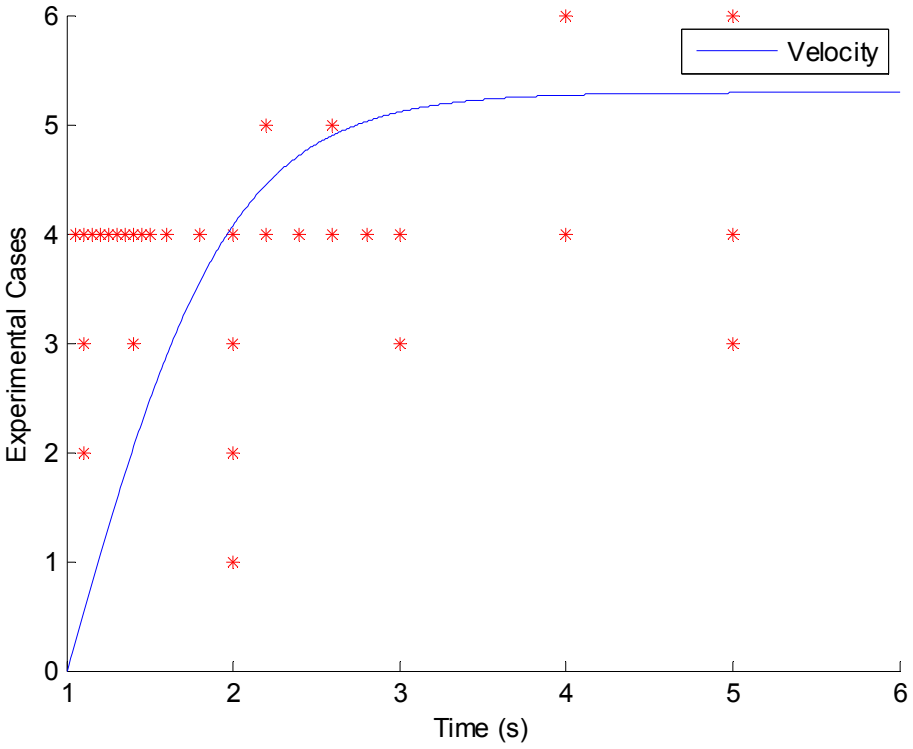


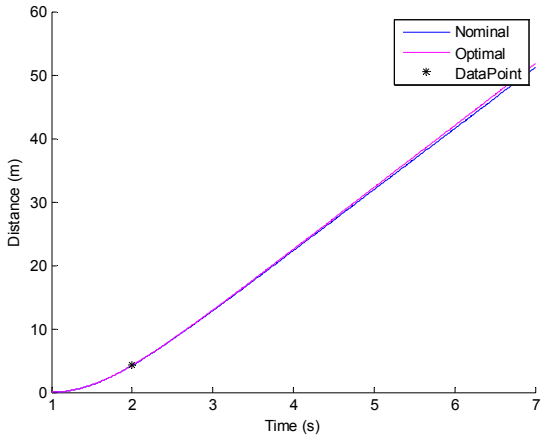
Figure 4: The time instances for the different experimental cases and N values.

Note that data affects the values of $\hat{c}$ and $\sigma_{c|D}$. Figures 5 and 6 compares the curve ($z(t)$) created using the nominal c $(c_{nom} = 0.106)$ with the one created using the optimal c (most probable value, $\hat{c}$) found using the *fminunc* m-file, for the different experimental cases.

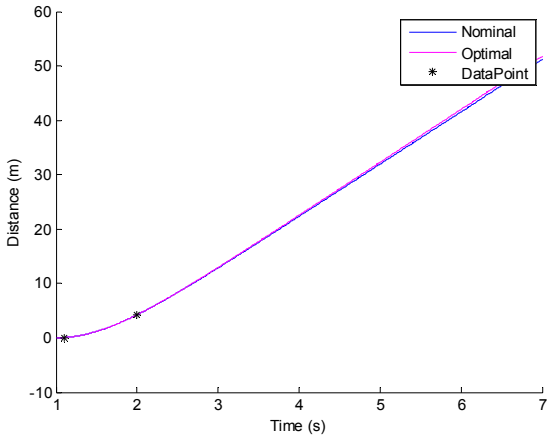
Having in mind that the nominal values were the ones used for generating the "experimental" data of table 1, we can say that the more close the optimal curve is to the nominal one the better it predicts the object's behavior at non measured points.

It can be seen that optimal curve is near the nominal one near given data points. If the given data points are early in motion the optimal curve fails to properly predict the behavior at later times of the motion and diverges from the nominal one.

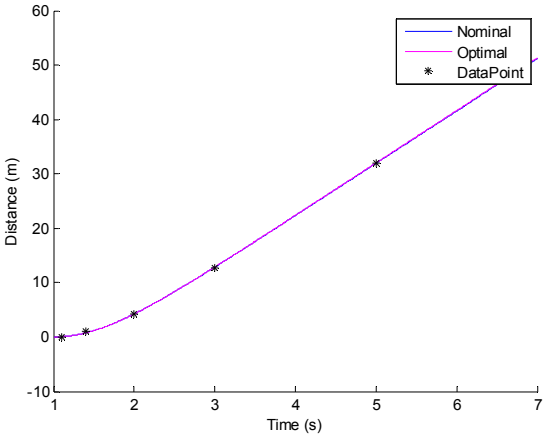
A. Changing the number of data points (N)



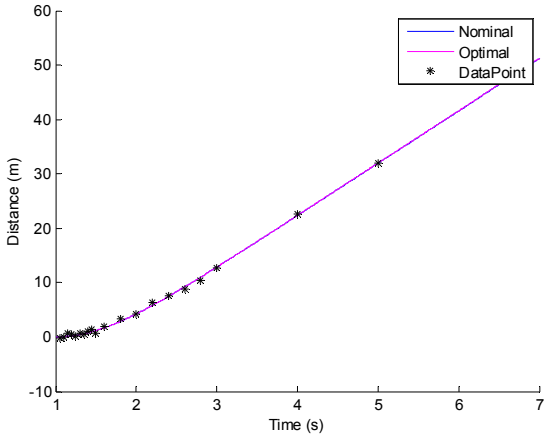
(a) Case 1



(b) Case 2



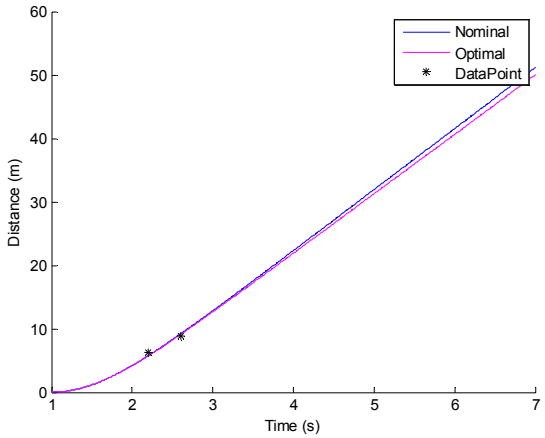
(c) Case 3



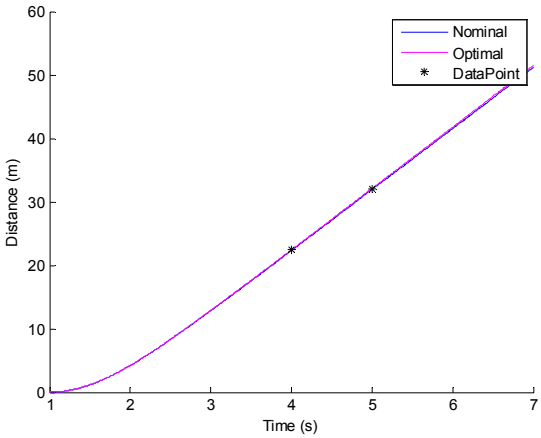
(d) Case 4

Figure 5: Comparison of the curve ($z(t)$) created using the c_{nom} with the one created using the $\hat{c}$ for (a) Case 1: $N = 1$, (b) Case 2: $N = 2$, (c) Case 3: $N = 5$ and (d) Case 4: $N = 20$.

B. Changing the time instances chosen (always for N=2)



(a) Case 5



(b) Case 6

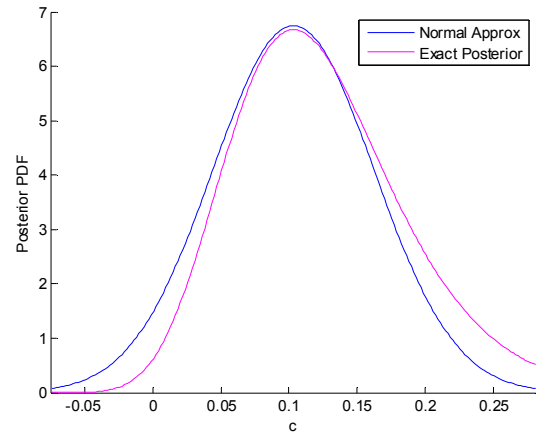
Figure 6: Comparison of the curve ($z(t)$) created using the c_{nom} with the one created using the $\hat{c}$ for $N = 2$ and experimental cases (a) Case 5 and (b) Case 6. The time instances for the different experimental cases are reported in Figure 4.

Figure's 7 and 8 show the evolution of the posterior PDF $p(c \mid D, I)$ (the posterior uncertainty in c) . The asymptotic results for the posterior PDF are compared also with the exact results using the posterior PDF in (3).

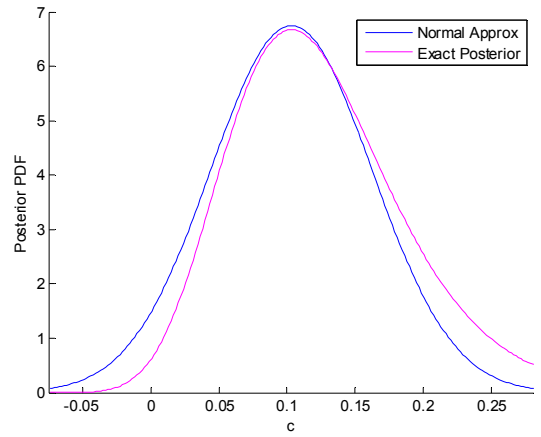
It can be seen that as the number of data increases the Gaussian posterior PDF converges to the exact posterior PDF. Also, as the number of data increases the uncertainty in the parameter c decreases.

Comparing the cases with N=2 (that is Case 1, Case 2 and Case 3) why see that the later the measurements the more close is the exact posterior PDF to it's Gaussian approximation. Also, as the time of data collection increases the uncertainty in the parameter c decreases.

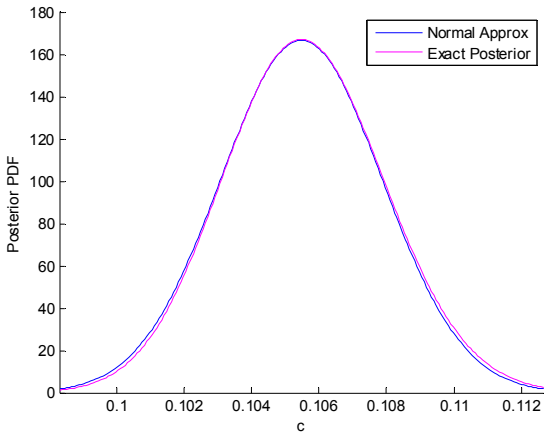
A. Changing the number of data points (N)



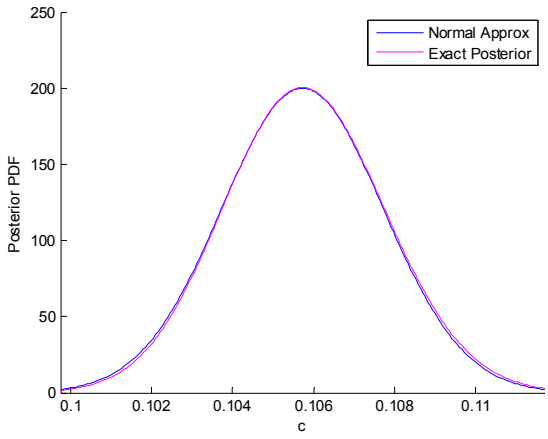
(a) Case 1



(b) Case 2



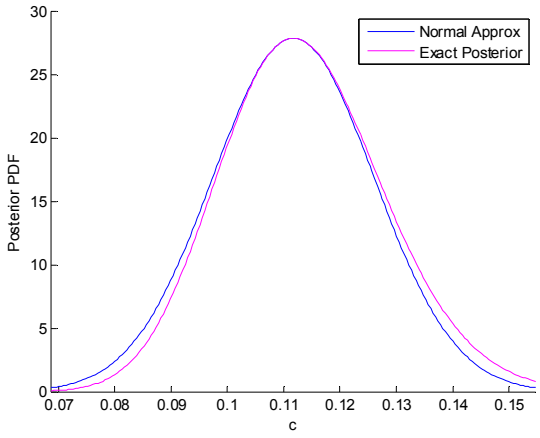
(c) Case 3



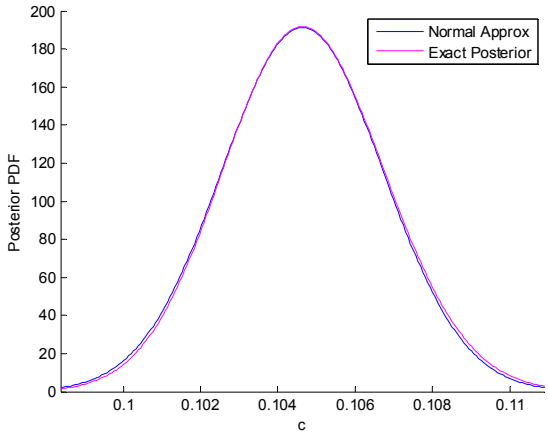
(d) Case 4

Figure 7: Comparison of approximate Gaussian posterior PDF with exact posterior PDF for (a) Case 1: $N = 1$, (b) Case 2: $N = 2$, (c) Case 3: $N = 5$ and (d) Case 4: $N = 20$.

B. Changing the time instances chosen (always for N=2)



(a) Case 5



(b) Case 6

Figure 8: Comparison of approximate Gaussian posterior PDF with exact posterior PDF for $N = 2$ and experimental cases (a) Case 5 and (b) Case 6. The time instances for the different experimental cases are reported in Figure 4.

Example 2: Parameter Estimation of Models of Falling Objects

Consider the falling object problem introduced in Example 1. The interest in this case lies in estimating two models parameters: the air resistance coefficient c, and the initial time t_0 that the object was released, given the set of independent observations/data (see Table 1). The other model parameters are considered known and the size of the error is due to measurement only and thus it can be considered to be known as well. The analysis is based on the same data and assumption introduced in Example 1. The prediction error equation is given by (5). The prior for both parameters is assumed uniform, given by (6) for the parameter c and by a similar expression for the parameter t_0 involving wide enough bounds $t_{0,\min}$ and $t_{0,\max}$ so that the inference is not affected by such bounds. The likelihood function, the posterior PDF and the function $L(\underline{\theta})$ remain the same as the ones given in (7), (8) and (10), respectively, with the parameter set $\underline{\theta} = (c, t_0)$ involving two parameters instead of one.

Most Probable Value: For constant variance σ^2, the most probable value or the MAP value $\underline{\hat{\theta}} = (\hat{c}, \hat{t}_0)$ is obtained by minimizing $L(\underline{\theta})$ or the misfit function $J(\underline{\theta})$. Similar to the one parameter case, the optimization can not be performed analytically. A gradient-based optimization algorithm is used to find the optimum. The analytic expressions for the gradient of the objective function, required in gradient-based optimization algorithms, with respect to the air resistance coefficient and the initial time are given by

$$\frac{dL(\underline{\theta})}{dc} = \frac{1}{\sigma^2}\sum_{\kappa=1}^{\mathrm{N}}\left\{(z(t_k,\underline{\theta}) - \hat{\mathrm{z}}_\kappa)\frac{\partial z}{\partial c}\right\}$$

$$\frac{dL(\underline{\theta})}{dt_0} = \frac{1}{\sigma^2}\sum_{\kappa=1}^{\mathrm{N}}\left\{(z(t_k,\underline{\theta}) - \hat{\mathrm{z}}_\kappa)\frac{\partial z}{\partial t_0}\right\}$$

Derivation details can be found in the Appendix. All results of the optimization in this example are carried out in Matlab using the *fminunc* m-file.

Uncertainty in Model Parameter: The uncertainty in the value of the model parameters is characterized by the Hessian of the function $L(\underline{\theta})$. The components of this Hessian are given by (for derivation see Appendix)

$$H_{cc}(\underline{\hat{\theta}}) = \left.\frac{\partial^2 L}{\partial c^2}\right|_{\underline{\theta}=\underline{\hat{\theta}}} = \frac{1}{\sigma^2}\sum_{\kappa=1}^{\mathrm{N}}\left\{\left(\frac{\partial z}{\partial c}\right)^2 + (z(t_k,\underline{\theta}) - \hat{\mathrm{z}}_\kappa)\frac{\partial^2 z}{\partial c^2}\right\}$$

$$H_{ct_0}(\underline{\hat{\theta}}) = \left.\frac{\partial^2 L}{\partial c \partial t_0}\right|_{\underline{\theta}=\underline{\hat{\theta}}} = \frac{1}{\sigma^2}\sum_{\kappa=1}^{\mathrm{N}}\left\{\frac{\partial z}{\partial c}\frac{\partial z}{\partial t_0} + (z(t_k,\underline{\theta}) - \hat{\mathrm{z}}_\kappa)\frac{\partial^2 z}{\partial c \partial t_0}\right\}$$

$$H_{t_0t_0}(\underline{\hat{\theta}}) = \left.\frac{\partial^2 L}{\partial t_0^2}\right|_{\underline{\theta}=\underline{\hat{\theta}}} = \frac{1}{\sigma^2}\sum_{\kappa=1}^{\mathrm{N}}\left\{\left(\frac{\partial z}{\partial t_0}\right)^2 + (z(t_k,\underline{\theta}) - \hat{\mathrm{z}}_\kappa)\frac{\partial^2 z}{\partial {t_0}^2}\right\}$$

Where

$$\frac{\partial z}{\partial g} = \frac{(t_\kappa - t_0)\tanh\left[\sqrt{gC}\,(t_\kappa - t_0)\right]}{2\sqrt{gC}}$$

$$\frac{\partial z}{\partial C} = \frac{(t_\kappa - t_0)\, g \tanh\left[\sqrt{gC}\,(t_\kappa - t_0)\right]}{2C\sqrt{gC}} - \frac{\ln\left\{\cosh\left[\sqrt{gC}\,(t_\kappa - t_0)\right]\right\}}{C^2}$$

$$\frac{\partial z}{\partial t_0} = \frac{-\sqrt{gC}\tanh\left[\sqrt{gC}\,(t_\kappa - t_0)\right]}{C}$$

$$\frac{\partial^2 z}{\partial C^2} = \frac{g(t_\kappa - t_0)}{2\sqrt{g}}\frac{\frac{\partial T}{\partial C}C^{2/3} - \frac{3}{2}T * C^{1/2}}{C^3} - \frac{\frac{\partial z}{\partial C} * C - z}{C^2}$$

$$\frac{\partial^2 z}{\partial C \partial t_0} = \frac{g}{2C\sqrt{gC}}\left[-T + (t_\kappa - t_0)\frac{\partial T}{\partial t_0}\right] - \frac{1}{C}\frac{\partial z}{\partial t_0}$$

$$\frac{\partial^2 z}{\partial t_0^{\,2}} = -\frac{\sqrt{gC}}{C}\frac{\partial T}{\partial t_0}$$

$$T = \tanh\left[\sqrt{gC}\,(t_\kappa - t_0)\right]$$

$$\frac{\partial T}{\partial C} = \left(1 - T^2\right)(t_\kappa - t_0)\frac{g}{2\sqrt{gC}}$$

$$\frac{\partial T}{\partial t_0} = -\left(1 - T^2\right)\sqrt{gC}$$

Following the theoretical development, the uncertainty in the two-dimensional parameter space is completely characterized in the neighborhood of the MAP value by the eigenvalues and the eigenvectors of the Hessian matrix.

Asymptotic Posterior PDF: Following the theoretical result and using the MAP estimate $\hat{\underline{\theta}}$ and the Hessian matrix $H(\hat{\underline{\theta}})$, the posterior PDF is approximated by the two-variable Gaussian distribution

$$p(\underline{\theta} \mid D, I) = \frac{\sqrt{\det[H(\hat{\underline{\theta}})]}}{\left(\sqrt{2\pi}\right)^n}\exp\left[-\frac{1}{2}(\underline{\theta} - \hat{\underline{\theta}})^T H(\hat{\underline{\theta}})(\underline{\theta} - \hat{\underline{\theta}})\right] \qquad (14)$$

where $n = 2$ in this case. Note that the covariance matrix $C_{\underline{\theta}|D} = H^{-1}(\hat{\underline{\theta}})$ of the Gaussian distribution can also be written in the form

$$C_{\underline{\theta}|D} = \begin{bmatrix} \sigma^2_{cc|D} & C_{ct_0|D} \\ C_{ct_0|D} & \sigma^2_{t_0t_0|D} \end{bmatrix} = \begin{bmatrix} \sigma^2_{cc|D} & \rho_{ct_0|D}\sigma_{cc|D}\sigma_{t_0t_0|D} \\ \rho_{ct_0|D}\sigma_{cc|D}\sigma_{t_0t_0|D} & \sigma^2_{t_0t_0|D} \end{bmatrix} \tag{15}$$

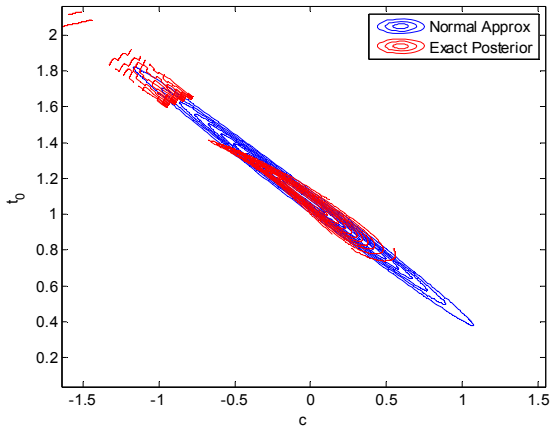
where $\rho_{ct_0|D} = C_{ct_0|D} / (\sigma_{cc|D}\sigma_{t_0t_0|D})$ is the correlation coefficient between the two uncertain parameters c and t_0 given the data. Note that $\rho_{ct_0|D}$ is bound by $-1 \le \rho_{ct_0|D} \le 1$, with values close to zero implying uncorrelated parameters, while values close to one implying fully correlated parameters.

The posterior marginal distribution $p(c \mid D, I)$ of the parameter c is Gaussian with mean $\hat{c}$ and variance $\sigma^2_{cc|D}$, while the marginal distribution $p(t_0 \mid D, I)$ of the parameter t_0 is Gaussian with mean $\hat{t}_0$ and variance $\sigma^2_{t_0t_0|D}$. In particular, the marginal distribution of c is Gaussian, i.e. $c \sim N(\hat{c}, \sigma^2_{cc|D})$, with both $\hat{c}$ and $\sigma^2_{cc|D}$ obtained from the two-parameter inference problem and should be different from the corresponding values obtained from the single-parameter inference problem in Example 1.

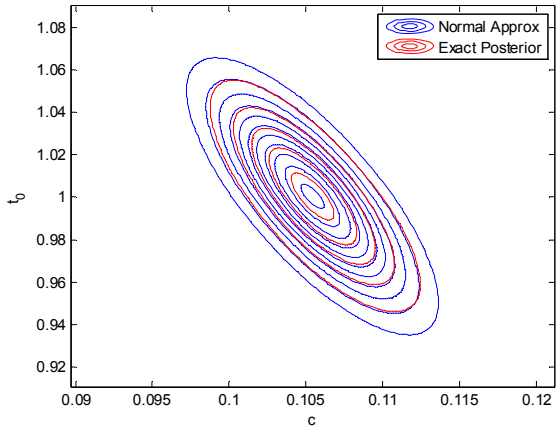
Numerical Results: Parameter inference results are obtained using the simulated experimental data in Table 1. We will be using the same experimental cases, with the exception of Case 1. We do not examine the case of $N = 1$ because it's impossible to evaluate two parameters with only one data point. The Matlab code itself gives an error that the $H(\hat{\underline{\theta}})$ is not a positive definite matrix.

A. Changing the number of data points (N)

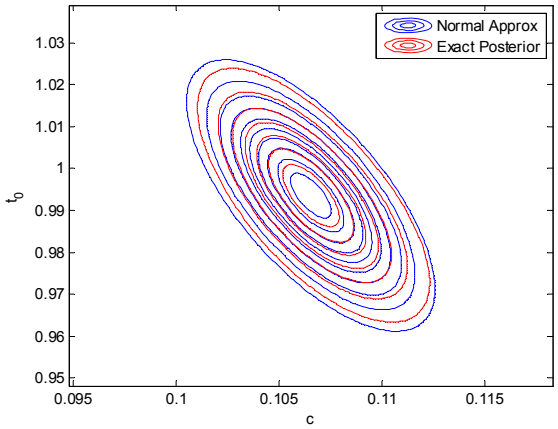
Figure 6 shows the contour plots of the posterior PDF $p(c, t_0 \mid D, I)$ (the posterior uncertainty in c and t_0) as a function of the number of data N for $N =$ 2, 5, 20 (experimental cases 2, 3 and 4). Note that data affects the MAP values $\hat{c}$ and $\hat{t}_0$ as well as the uncertainty in these values. The asymptotic results for the posterior PDF are compared also with the exact results using the posterior PDF in (8).Comparing cases 2, 3 and 4, it can be seen that as the number of data increases the Gaussian posterior PDF converges to the exact posterior PDF. Also, as the number of data increases the uncertainty in the parameters c and t_0 decreases.



(a) $N = 2$,



(b) $N = 5$

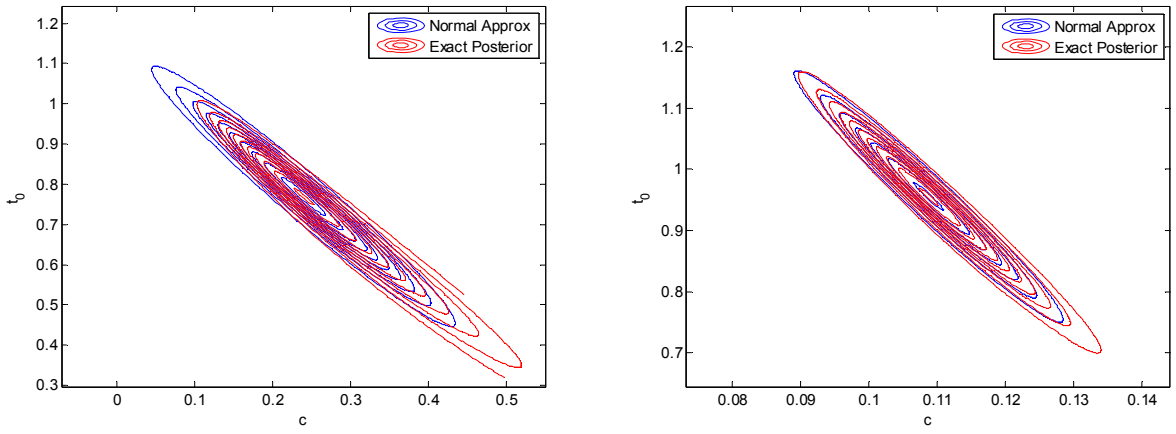


(c) $N = 20$

Figure 9: Comparison of contour plots of approximate Gaussian posterior PDF with exact posterior PDF for (a) Case 2: $N = 2$, (b) Case 3: $N = 5$ and (c) Case 4: $N = 20$. The time instances for the different experimental cases are reported in Figure 4.

B. Changing the time instances chosen (always for N=2)

The effect of the time instances on the parameter inference is investigated. Results for the contour plots of the Gaussian posterior PDF of the model parameter are shown in Figure 8 for the experimental cases 5 and 6. The contours of the Gaussian posterior PDF is compared with the contours of the exact posterior PDF. The values of the correlation coefficients for each experimental case are given in Table 2. Comparing the cases for N=2 (these are cases 2,5 and 6) It can be seen that as we take data in later moments of the motion the Gaussian posterior PDF converges more to the exact posterior PDF. Also, as the number of data is taken in the later phases of the motion the uncertainty in the parameters c and t_0 decreases.



(a) Case 5 (b) Case 6

Figure 10: Comparison of approximate Gaussian posterior PDF with exact posterior PDF for $N = 2$ and experimental cases (a) Case 5 and (b) Case 6. The time instances for the different experimental cases are reported in Figure 4.

The optimal values (Maximum-A-Posteriory (MAP)) and the values of the correlation coefficients $\rho_{ct_0|D}$ for each experimental case are given in Table 2. It can be seen (by observing cases 2,3 and 4) that the correlation coefficient decreases when the number of experimental data (N) increases. And by observing cases 2, 5 and 6 it can be seen the drop phase during which the data taken does not affect the correlation coefficient.

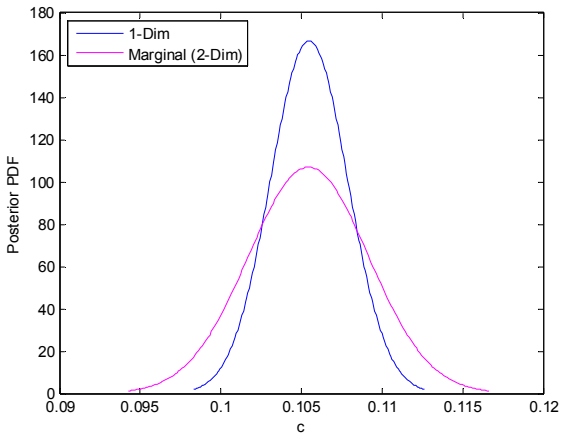
Table 2: MAP values and Correlation coefficient $\rho_{ct_0|D}$ for each experimental case

Case2	c_{opt}	t_{0opt}	*Corr. Coef.*
2	-0.0471	1.1000	0.9961
3	0.1054	1.0004	0.7673
4	0.1065	0.9936	0.6886
5	0.2393	0.7699	0.9742
6	0.1087	0.9559	0.9763

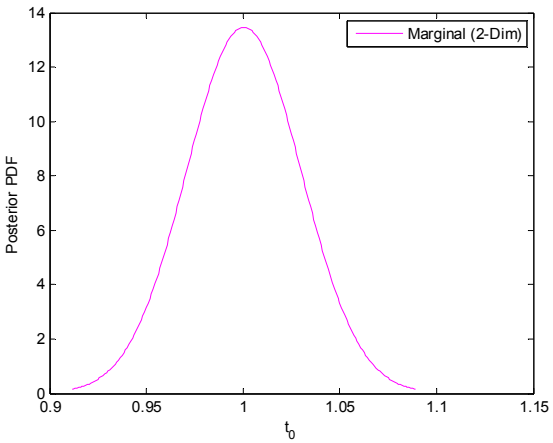
Using the posterior approximation as a Gaussian distribution we can find the marginal distributions for the parameters c and t_0. That is shown in Figures 7 (left) and 7 (right), respectively, for N=5 and N=20. In Figure 7 (left) the uncertainty in c obtained from the one-parameter inference problem in Example 1 is also shown. We use N=5 and N=20 (experimental cases 3 and 4) because that is the case in which the Gaussian posterior PDF converges to the exact posterior PDF the most.

It can be seen that the larger the amount of experimental data (N), the smaller the covariance.

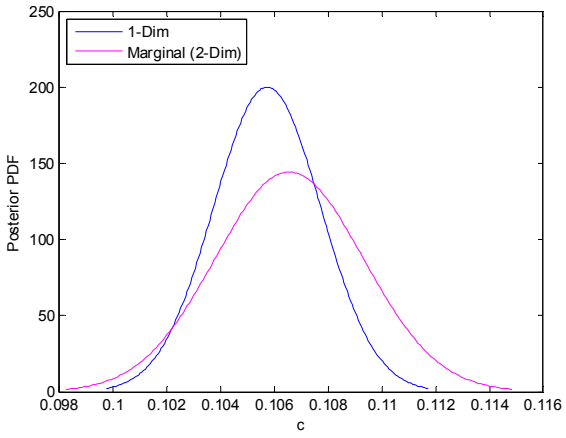
Also from the figures on the left we can see that the optimal values of c (MAP values $\hat{c}$ or c_{opt}) are not the same for the 1-parameter problem and the 2-parameter problem neither are the covariances.



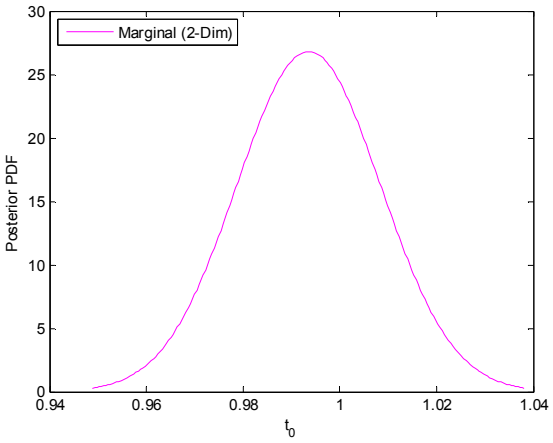
(a) Left Figure



(a) Right Figure



(b) Left Figure



(b) Right Figure

Figure 7: Posterior marginal distributions Posterior for the parameters c (left) and t_0 (right) for (a) Case 3 (N=5) and (b) Case 4 (N=20).

Example 3: Parameter Estimation of Models of Falling Objects

Return to the falling object problem introduced in Example 1. We are interested in estimating all three model parameters (acceleration of gravity g, air resistance coefficient c, initial time t_0) and the prediction error parameter σ. The analysis is based on the same data and assumption introduced in Example 1. The prediction error equation, the likelihood function, the posterior PDF and the function $L(\underline{\theta})$ remain the same as in example 1, with the parameter set $\underline{\theta} = (g, c, t_0, \sigma)$ involving four parameters instead of one. The prior for all four parameters is assumed uniform, given by (6) for the parameter c and by similar expressions for the other three parameters, using wide enough bounds so that the inference is not affected by such bounds.

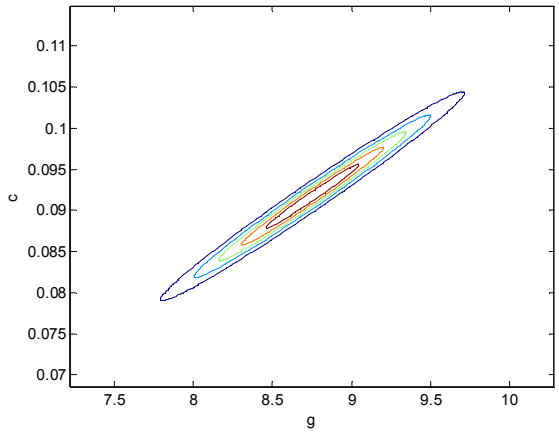
Most Probable Value: The most probable value or the MAP value $\hat{\underline{\theta}} = (\hat{g}, \hat{c}, \hat{t}_0, \hat{\sigma})$ is obtained by minimizing $L(\underline{\theta})$. The optimization is performed numerically using a gradient-based optimization algorithm. The analytic expressions for the gradient of the objective function, required in gradient-based optimization algorithms, with respect to all four parameters are given in the Appendix. All results of the optimization in this example are carried out in Matlab using the *fminunc* m-file.

Uncertainty in Model Parameter: The uncertainty in the value of the model parameters is characterized by the 4x4 Hessian matrix of the function $L(\underline{\theta})$. The components of this Hessian are also given in the Appendix. Following the theoretical development, the uncertainty in the four-dimensional parameter space is completely characterized in the neighborhood of the MAP value by the eigenvalues and the eigenvectors of the Hessian matrix.

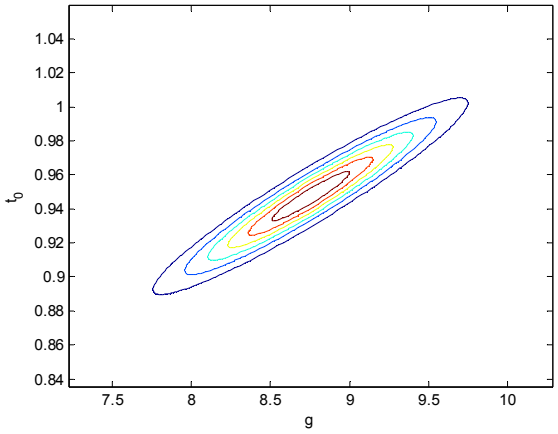
Asymptotic Posterior PDF: Following the theoretical result and using the MAP estimate $\hat{\underline{\theta}}$ and the Hessian matrix $H(\hat{\underline{\theta}})$, the posterior PDF is approximated by the multi-variable Gaussian distribution (14). Note that the covariance matrix $C_{\underline{\theta}|D} = H^{-1}(\hat{\underline{\theta}})$ of the Gaussian distribution can also be written in a 4x4 matrix similar to the 2x2 matrix in (15) with the correlation coefficients $\rho_{gc|D} = C_{gc|D} / (\sigma_{gg|D}\sigma_{cc|D})$, $\rho_{gt_0|D}$, $\rho_{g\sigma|D}$, $\rho_{ct_0|D}$, $\rho_{c\sigma|D}$ and $\rho_{t_0\sigma|D}$ quantifying the relative correlation between the parameter pair involved in the subscripts.

From this multi-variable Gaussian distribution, the marginal distribution of one or two or three parameters are Gaussian and can easily be recovered from the elements of the mean vector $\hat{\underline{\theta}}$ and the covariance matrix $C_{\underline{\theta}|D}$. For example, the joint marginal distribution of the parameter set (c, t_0) is Gaussian with mean $(\hat{c}, \hat{t}_0)$ and covariance matrix $C_{(c,t_0)|D} = \begin{bmatrix} \sigma^2_{cc|D} & c_{ct_0|D} \\ c_{ct_0|D} & \sigma^2_{t_0t_0|D} \end{bmatrix}$, both obtained from the four-parameter inference problem and should be different from the corresponding values obtained from the two-parameter inference problem in Example 2. The marginal distribution of c is also Gaussian, i.e. $c \sim N(\hat{c}, \sigma^2_{cc|D})$, with both $\hat{c}$ and $\sigma^2_{cc|D}$ obtained from the four-parameter inference problem and should be different from the corresponding values obtained from the single-parameter inference problem in Example 1 or the parameter inference problem in Example 2.

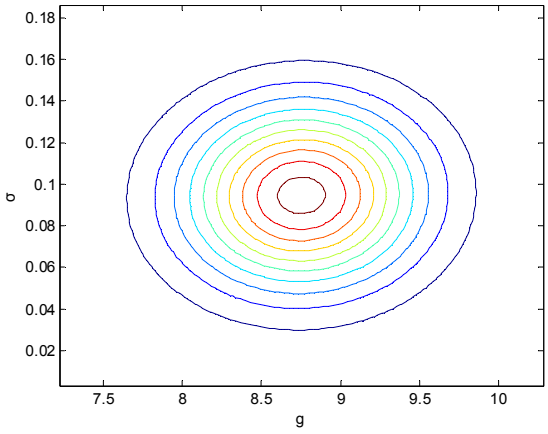
Numerical Results: Parameter inference results for the four-parameter case are obtained using the simulated data in Table 1. Figures 12 shows contour plots of the marginal distributions of a pair of parameters in two-dimensional parameter spaces for number of data $N = 5$.



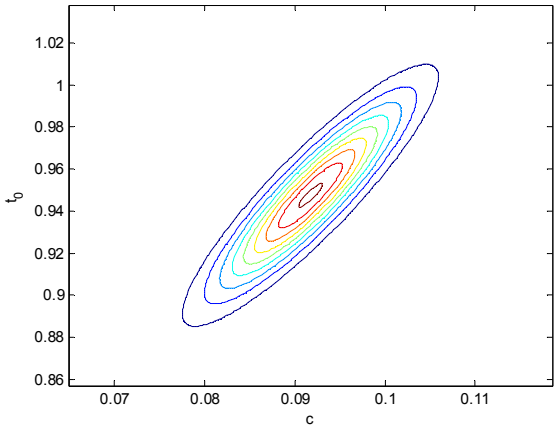
(a) $p(g,c \mid D,I)$



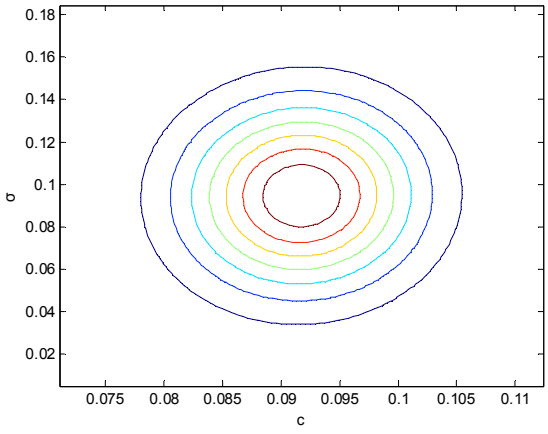
(b) $p(g,t_0 \mid D,I)$



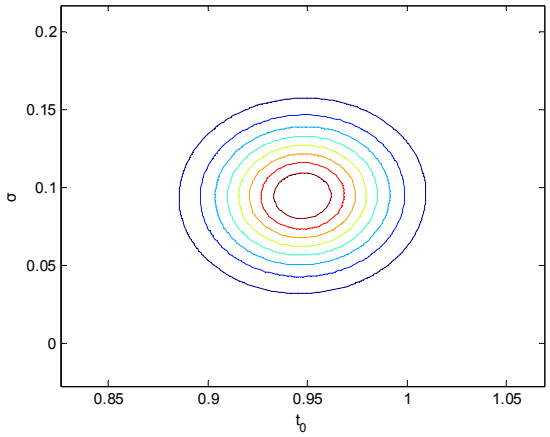
(c) $p(g,\sigma \mid D,I)$



(d) $p(c,t_0 \mid D,I)$



(e) $p(c,\sigma \mid D,I)$



(f) $p(t_0,\sigma \mid D,I)$

Figure 12: Marginal posterior uncertainty for N=5 and for parameter pairs: (a) $p(g,c \mid D,I)$, (b) $p(g,t_0 \mid D,I)$, (c) $p(g,\sigma \mid D,I)$,(d) $p(c,t_0 \mid D,I)$, (e) $p(c,\sigma \mid D,I)$ and (f) $p(t_0,\sigma \mid D,I)$

Table 3: Optimal values found ($\hat{g}, \hat{c}, \hat{t}_0, \hat{\sigma}$) for N=5

N	g_{opt}	c_{opt}	t_{0opt}	σ_{opt}
5	8.7529	0.0917	0.9475	0.0946

In figure 13 the asymptotic results for the marginal posterior PDF of parameters c and t_0 (figure 12(d)) are compared with the approximate and exact results of the posterior PDF obtained using the two-parameter inference case in Example 2.

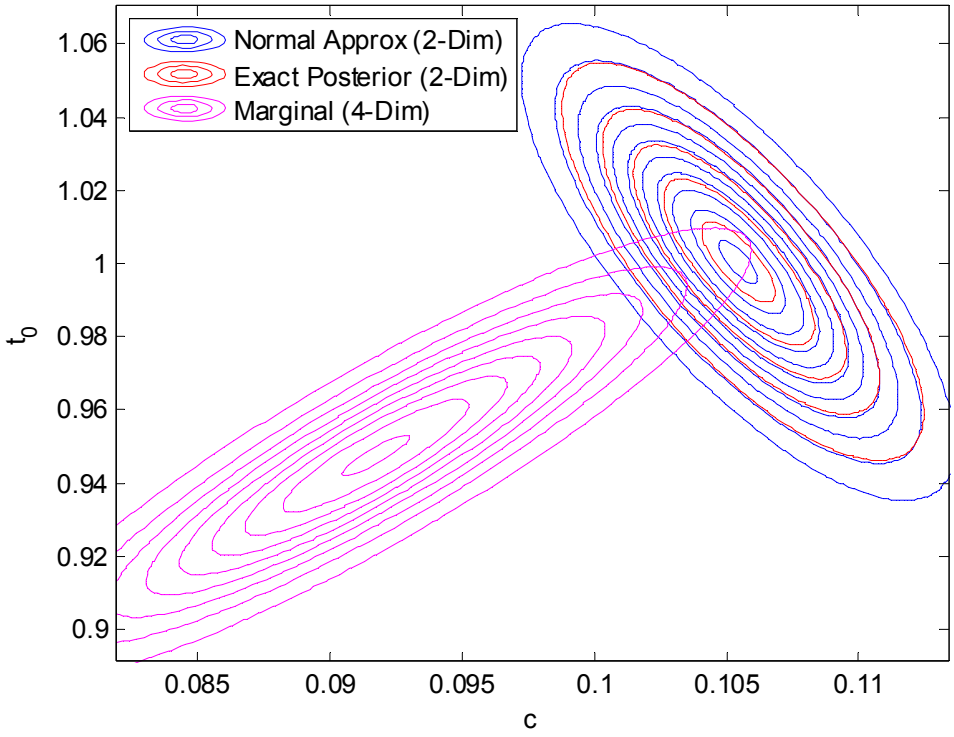


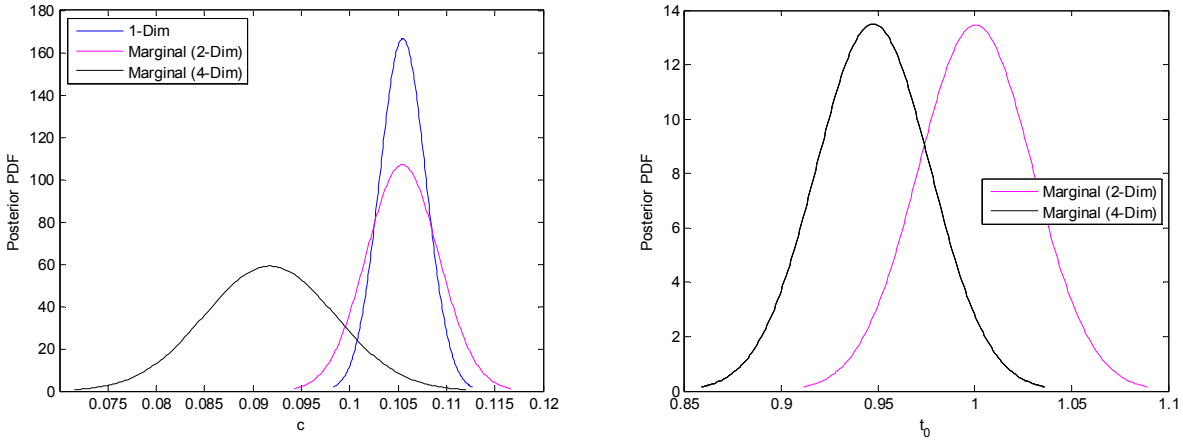
Figure 13: Comparison between the asymptotic results for the marginal posterior PDF of parameters c and t_0 (figure 13(d)) and the approximate and exact results of the posterior PDF obtained using the two-parameter inference case in Example 2, for N=5.

The values of the correlation coefficients $\rho_{\theta_i\theta_j|D}$ are given in Table 5. It can be seen that the two.

Table 5: Correlation coefficient $\rho_{ct_0|D}$ for the experimental case 3 (N=5)

$\rho_{gc\|D}$	$\rho_{gt_0\|D}$	$\rho_{g\sigma\|D}$	$\rho_{ct_0\|D}$	$\rho_{c\sigma\|D}$	$\rho_{t_0\sigma\|D}$
0.9876	0.9497	0.0206	0.9004	0.0203	0.0197

The marginal distributions for the parameters c and t_0 are shown in Figures 14 (left) and 14 (right), respectively. In Figure 14 (left) the uncertainty in c, $p(c \mid D, I)$, obtained from the one-parameter inference problem in Example 1, and the marginal posterior PDF $p(c \mid D, I)$ derived from the two-parameter inference problem in Example 2 are also shown. In Figure 11 (right) the uncertainty in t_0 obtained from the marginal posterior PDF $p(t_0 \mid D, I)$ derived from the two-parameter inference problem in Example 2 is also shown.



Left Figure Right Figure

Figure 14: Comparison of the posterior marginal distributions for the parameters c (left) and t_0 (right) for the experimental case 3. The time instances for the experimental case 3 are reported in Figure 4.

Table 6 summarizes the results of the MAP values $\hat{c}$ of c computed for $N = 5$ based on the marginal distribution for the parameter obtained for the following inference cases (IF): IF1 involving a single parameter (see Example 1), IF2 involving two parameters (see Example 2), IF3 involving four parameters (see Example 3). It can be clearly seen that the MAP values are affected by the parameter inference case considered. It should also be noted that the MAP values $\hat{c}$ of c corresponding to the four-dimensional exact posterior PDF $p(g, c, t_0, \sigma \mid D, I)$, two-dimensional exact marginal posterior PDF $p(c, t_0 \mid D, I)$, and the one-dimensional exact marginal posterior PDF $p(c \mid D, I)$ are expected to be different. Estimation of this is quite challenging since the due to the challenge in estimating the marginal posterior PDFs from the exact four-dimensional posterior PDF. Approximating the exact posterior PDF by a Gaussian posterior PDF has an effect of estimating the same MAP value $\hat{c}$ independently of the used posterior PDFs or marginal posterior PDFs.

Table 6 : MAP values $\hat{c}$ of c computed for $N = 5$ and for the inference cases IF1 (one parameter - Example 2), IF2 (two parameters - Example 2), and IF3 (four parameters - Example 3).

IF	IF1	IF2	IF3
MAP $\hat{c}$	0.1054	0.1054	0.0917

Example 5: Optimal Experimental Design (OED) for Falling Objects

Considering the falling object problem, the objective of the optimal experimental design is to estimate the optimal time instances to measure the position of the object so that the information that the corresponding data contain for estimating the values of the model parameters is maximized. This is achieved by minimizing the information entropy associated with the posterior distribution of the model parameters. The design variables in the minimization problem are the time instances.

The asymptotic approximation of the information entropy, valid for large number of data ($N \to \infty$), is used for optimal experimental design. This asymptotic estimate of the information entropy is given for this particular problem in the form

$$I\left(\underline{\tau};D\right) \approx \tilde{I}(\underline{\tau};\underline{\theta}_0,\Sigma) = \frac{1}{2} N_\theta \ln(2\pi) - \frac{1}{2}\ln[\det Q(\underline{\tau};\underline{\theta}_0,\Sigma)] \qquad (16)$$

Change the symbol Q to H

where $\underline{\tau} = (t_1,\ldots,t_N)$ contains the design variables, $\underline{\theta}_0 \equiv \underline{\hat{\theta}}(\underline{\tau},\Sigma,D)$ is the optimal value of the parameter set $\underline{\theta}$ that minimizes the measure of fit $J(\underline{\theta};\underline{\tau},D)$ given in ??, and $Q(\underline{\tau};\underline{\theta},\Sigma)$ is an $N_\theta \times N_\theta$ semi-positive definite matrix defined as $\nabla_{\underline{\theta}}\nabla_{\underline{\theta}}^T J(\underline{\theta};\Sigma,D)$ and asymptotically approximated by

$$Q(\underline{\tau};\underline{\theta},\Sigma) = \sum_{k=1}^{N} [\nabla_{\underline{\theta}} \underline{z}(t_k;\underline{\theta})]^T \Sigma^{-1} [\nabla_{\underline{\theta}} \underline{z}(t_k;\underline{\theta})] \qquad (17)$$

in which $\nabla_{\underline{\theta}} = [\partial/\partial\theta_1,\cdots,\partial/\partial\theta_{N_\theta}]^T$ is the usual gradient vector with respect to the parameter set $\underline{\theta}$. The matrix $Q(L,\underline{\theta},\Sigma)$ is a semi-positive definite matrix, known as the Fisher information matrix (FIM), containing the information about the uncertainty in the values of the parameters $\underline{\theta}$ based on the data from all measured positions specified in $\underline{\tau}$.

In the initial stage of designing the experiment, the data and consequently the values of the optimal model parameters $\underline{\hat{\theta}}$ and the form of the prediction error covariance matrix Σ are not available. In practice, useful designs can be obtained by taking the optimal model parameters $\underline{\hat{\theta}}$ and prediction error covariance Σ to have some nominal values $\underline{\theta}_0$ and Σ to arise from a correlation function such as (20), chosen by the designer to be representative of the system and the expected model and measurement errors.

Prediction Error Model: An analysis of the prediction error correlation models is crucial and is presented next. The prediction error $\underline{e}_k = \underline{e}_{k,meas} + \underline{e}_{k,\mathrm{mod}}$ in ?? is due to a term, $\underline{e}_{k,meas}$, accounting for the measurement error and a term, $\underline{e}_{k,\mathrm{mod}el}$, accounting for the model error. Assuming independence between the measurement error and model error, the covariance Σ_t of the total prediction error is given in the form

$$\Sigma = \Sigma^{meas} + \Sigma^{\mathrm{mod}el} \qquad (18)$$

where Σ^{meas} and $\Sigma^{\mathrm{mod}el}$ are the covariance matrices of the measurement and model errors, respectively.

The designer has to assume values for the individual covariance matrices in (18). Such assumptions may depend on the nature of the problem analyzed. One reasonable choice is to

assume that the measurement error is independent of the location of sensors so that the covariance Σ^{meas} takes the diagonal form $\Sigma^{meas} = \sigma_1^2 I$, where I is the identity matrix.

For the model errors, a certain degree of correlation should be expected for the model errors between two neighborhood locations arising from the underlining model dynamics. This correlation can be taken into account by selecting a non-diagonal covariance matrix Σ^{model}. Specifically, the correlation Σ_{ij}^{model} between the predictions errors $\varepsilon_{i,\mathrm{mod}}$ and $\varepsilon_{j,\mathrm{mod}}$ at DOFs i and j, respectively, is selected in this work to be

$$\Sigma_{ij}^{model} = E[\varepsilon_{i,\mathrm{mod}}\varepsilon_{j,\mathrm{mod}}] = \sqrt{\Sigma_{ii}^{model}\Sigma_{jj}^{model}}\,R(\delta_{ij}) \tag{19}$$

that accounts for the time distance $\delta_{ij} = |t_i - t_j|$ between the time instances t_i and t_j, where $R(\delta_{ij})$ is a correlation function satisfying $R(0) = 1$. In general, the covariance matrix should be consistent with the actual errors and correlations as observed from measurements. However, in an experimental design stage such measurements are not available to guide the selection of the correlation between prediction errors. Instead a correlation function should be postulated to proceed with the design of the optimal sensor locations. Several correlation functions can be explored. For demonstration purposes, the following exponential correlation function is assumed:

$$R(\delta) = \exp[-\delta/\lambda] \tag{20}$$

where λ is a measure of the temporal correlation length. However, the formulation presented is general and does not depend on the choice of the correlation model.

Assuming the following model for the variance $\Sigma_{kk}^{model} = \sigma_2^2 \hat{z}_k^2$ of the prediction error at a time instant t_k, where σ_2^2 is independent of the time instant, the covariance matrix of the prediction error in (19) takes the form

$$\Sigma_{ij}^{model} = E[\varepsilon_{i,\mathrm{mod}}\varepsilon_{j,\mathrm{mod}}] = \sigma_2^2 \hat{z}_i\,\hat{z}_j R(\delta_{ij})$$

Substituting in (18), the (i, j) element Σ_{ij} of the covariance matrix Σ is given by

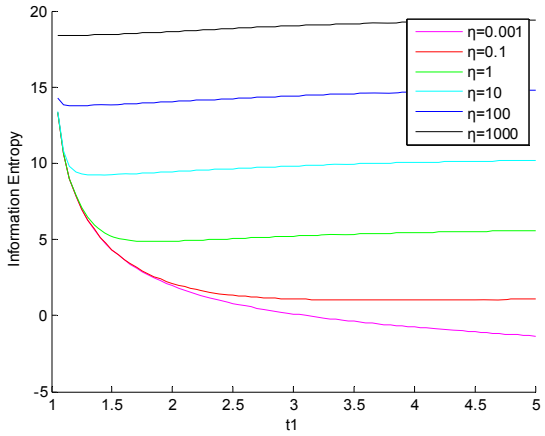
$$\Sigma_{ij} = \sigma_1^2 + \sigma_2^2 \hat{z}_i\,\hat{z}_j R(\delta_{ij}) = \sigma_1^2(1 + \eta^2 \hat{z}_i\,\hat{z}_j R(\delta_{ij})) \tag{21}$$

where $\eta = \sigma_2/\sigma_1$ is a measure of the size of the model error in relation to the measurement error. Values of $\eta \approx 0$ imply that the measurement error dominates the modeling error in the prediction error equation, while values of $\eta >> 1$ imply that the measurement error is negligible compared to the model error. Substituting (21) into (17) and then (16), the optimal design based on the information entropy depends only on the ratio $\eta = \sigma_2/\sigma_1$ and is independent of the value of σ_1. To perform the optimal design one has to replace the measured values in (21) by the nominal model predicted values, i.e. $\hat{z}_k \approx z(t_k; \underline{\theta}_0)$.

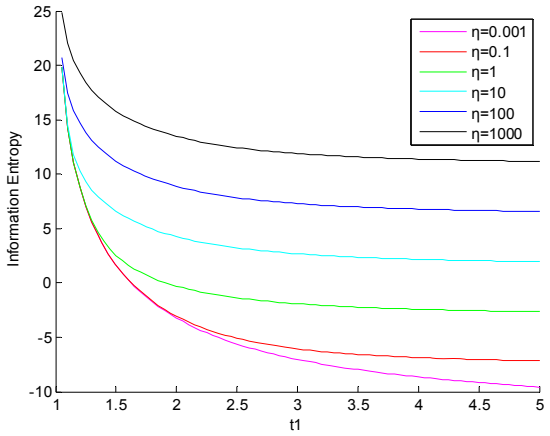
Numerical Results:

OED Case $\underline{\theta} = (c)$, $N = 1$: Consider the problem of a single parameter g, c or t_0. The information entropy as a function of the time we take the measurement for a single time instant is given in Figure 15 for different values of c. The selection of the λ value makes no difference on

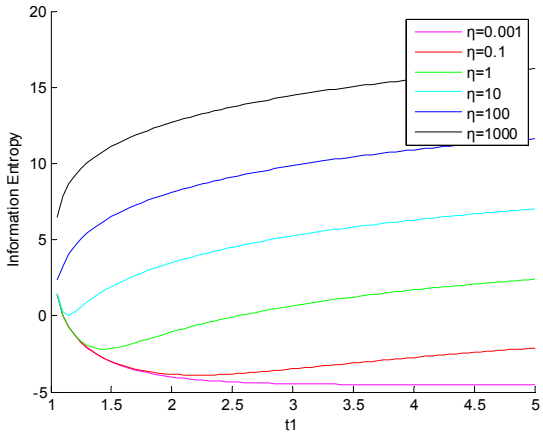
the results when N=1. It can be seen that the later the measurement the lower the information entropy and thus the uncertainty for the parameter c.



(a) g is the only parameter



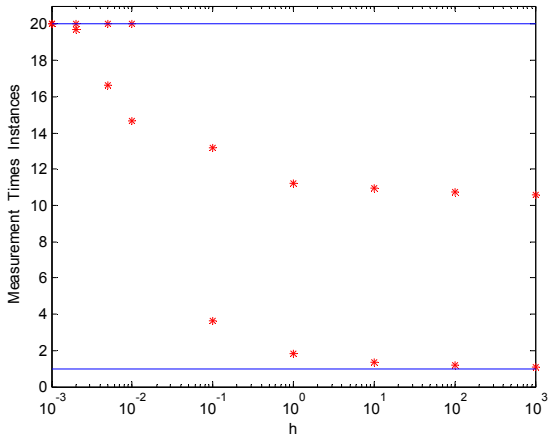
(b) c is the only parameter



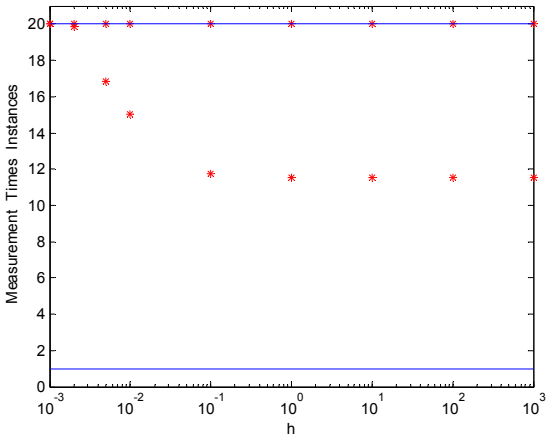
(c) t_0 is the only parameter

Figure 15: Information entropy as a function of time instance that a measurement is taken. We selected 6 η values: $\eta = 0.001, 0.1, 1, 10, 100, 1000$. (a) g is the only parameter, (b) c is the only parameter, (c) t_0 is the only parameter.

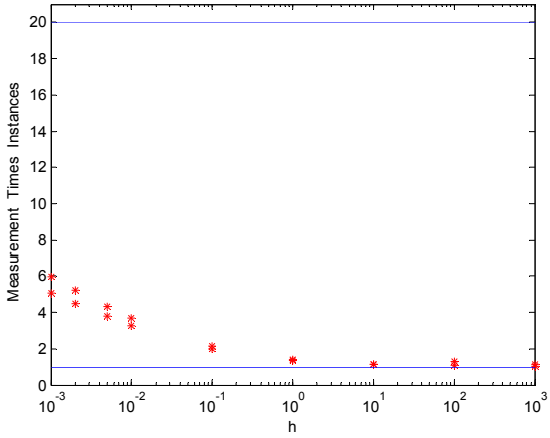
OED Case $\underline{\theta} = (c)$**,** $N = 2$: Consider the problem of a single parameter that is g, c or t_0. The optimal time instances for two measurements t_1, t_2 ($t_1, t_2 \in [1, 20]$) as a function of η is given in Figure 16 for correlation time $\lambda = 2$ sec. It can be seen that for very small values of h the two time instances approach each other, but for bigger h values they have a almost constant difference of $\approx$ 8.5 sec. For small values of h ($h \to 0$) we can say that there is no model error, in that case it is wrong to assume asymptotic approximation and the CMA gives as the best solution one that has N times the same time instance.



(a) g is the only parameter



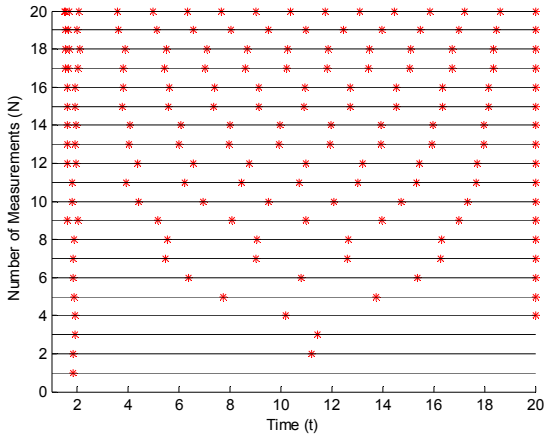
(b) c is the only parameter

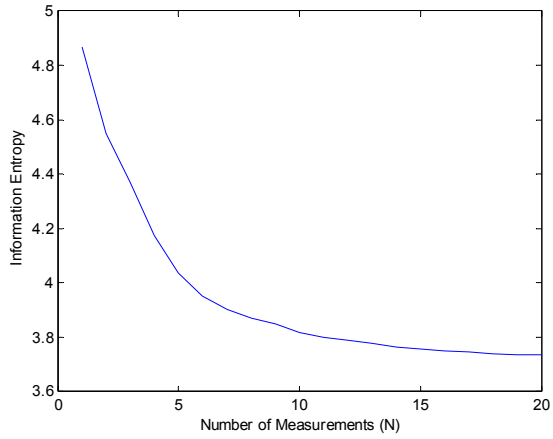


(c) t_0 is the only parameter

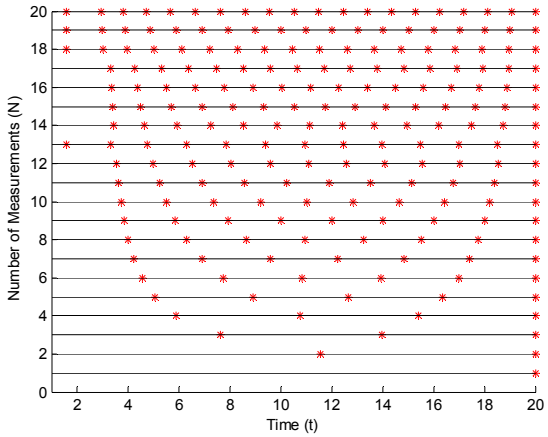
Figure 16: Optimal time instances t_1,t_2 (for $t_1,t_2 \in [1,20]$) that measurements are taken as a function of the ratio of model to measurement errors η. The correlation time is set to $\lambda = 2$ sec. We chose $\eta \in [0.001,\ 0.002,\ 0.005,\ 0.01,\ 0.1,\ 1,\ 10,\ 100,\ 1000]$. The graphs are in semi-logarithmic scale

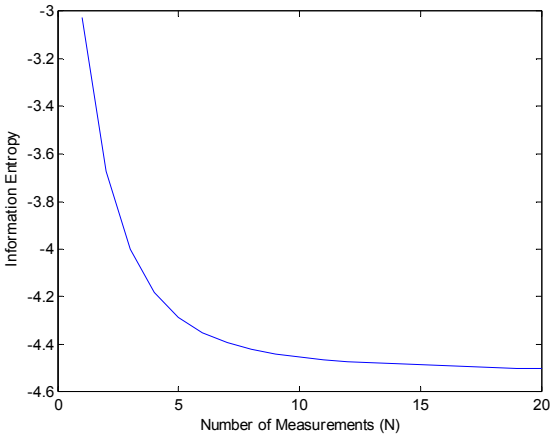
OED Case $\underline{\theta} = (c)$**,** $N =$ **1:10**: Consider the problem of a single parameter g, c or t_0. The optimal time instances for 1 to N measurements is given in Figure 17 (left) for fixed ratio of model to measurement errors $\eta = 1$ and for correlation time $\lambda = 2$ sec. The minimum information entropy computed at the optimal time instances as a function of the number of measurements is given in Figure 17 (right).



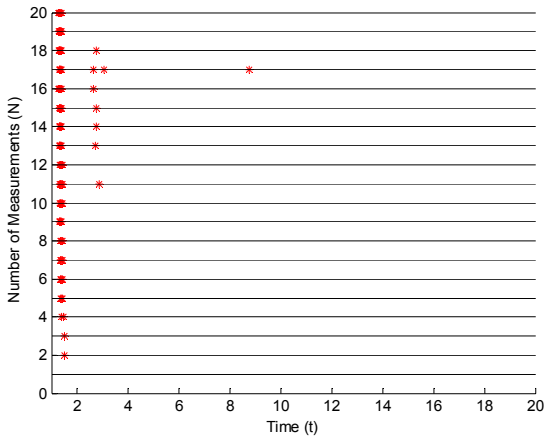


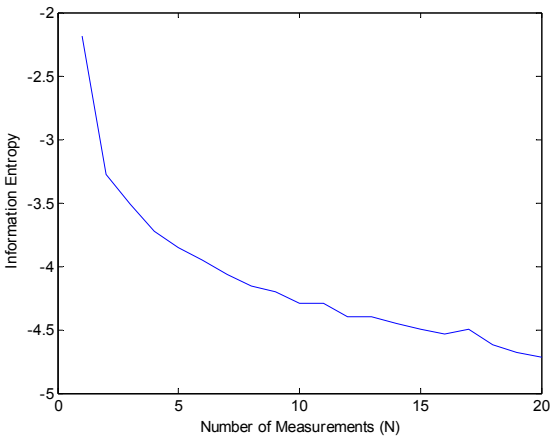
(a) g is the only parameter





(b) c is the only parameter





(c) t_0 is the only parameter

Figure 17: (Left) optimal time instances for 1 to N measurements. (Right) Minimum information entropy as a function of the number of measurements. $\eta = 1$ and $\lambda = 2$ sec.

Exercise: Inference of Resistance Coefficients for Falling Object on Liquids

Consider a sphere of diameter d_s and density ρ_s that due to the gravitational forces is falling in a fluid with density ρ and dynamic viscosity μ (note that the kinematic viscosity is $\nu = \mu / \rho$). The terminal velocity (U_0) is acquired when the force of buoyancy together with the fluid's resistance to the movement, become equal with the weight of the sphere. We can approximate the fluid's resistance from Stokes law, and for Reynold number $\mathrm{Re} = U_0 d_\sigma / \mathrm{v} < 0.12$ we get that

$$U_0 = \frac{{d_\sigma}^2 (\rho_\sigma - \rho) g}{18 \mu}$$

We are interested in estimating the dynamic viscosity μ of the fluid, using experimental data $D = [\{\mathrm{U}_{0_k}, \rho_{s_k}\}_{1 \to N}]$ of the terminal velocities U_{0_k} of the object with density ρ_{s_k}, $k = 1, \ldots, N$, taken from measuring the travelling time of the sphere for certain fall length, during which the sphere has it's terminal velocity. The data value at a certain density ρ_{s_k} can be considered to be independent from the data values at densities. It is assumed that the models used are perfect (no model error), the values of the rest of the variables involved in the model are known, and that the accuracy of the time measurements has a measurement error which can be adequately modeled as zero-mean Gaussian variable with standard deviation σ. It will be further assumed, for the purposes of this demonstration, that the level of the measurement error is known a priori which means that σ is given. So the model parameter set $\theta = (\mu)$ includes only a single parameter, the dynamic viscosity of the fluid to be inferred from the data.

Given the set of independent observations/data, we are interested in updating the uncertainty in the dynamic viscosity μ of the model. To infer the value of the single parameter (dynamic viscosity) given the data, one needs to set up a model for the prediction error, select the prior, build the expression of the likelihood and use the asymptotic analysis to approximate the posterior by a Gaussian.

Prediction Error Model: Based on the theory, the prediction error equation is

$$U_{0_k} = U_0(\rho_{\sigma_k}, \mu) + e_k \qquad \text{where}$$

e_k is the prediction error assumed to be a zero-mean Gaussian variable with known variance σ^2, i.e. $e_k \sim N(0, \sigma^2)$.

Prior: For simplicity it is assumed that the prior distribution for the air friction coefficient is uniform, that is,

$$p(\mu \mid I) = \begin{cases} 1/(\mu_{\max} - \mu_{\min}), & \mu \in [\mu_{\min}, \mu_{\max}] \\ 0 & \text{otherwise} \end{cases}$$

where the bounds $\mu_{\min}$ and $\mu_{\max}$ are wide so that the parameter inference and uncertainty is not influenced by such choices.

Likelihood: To evaluate the likelihood $p(D \mid \mu) \equiv p(\{\mathrm{U}_{0_k}\}_{1 \to N} \mid \mu)$, one uses the prediction error equation, the fact that the measured data at different time instances are independent, and applies successively the product rule of the axiom of probability to finally derive that

$$p(D \mid \mu) = p(\{\mathrm{U}_{0_k}\}_{1\to N} \mid \mu) = \prod_{k=1}^{N} p(\mathrm{U}_{0_k} \mid \mu)$$
$$= \prod_{k=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2} \left[U_{0_k} - U_0(\rho_{s_k}, \mu) \right]^2 \right] \tag{22}$$

Posterior PDF: Substituting in the Bayes formula the expression for the likelihood and the uniform prior PDF, the posterior PDF of the uncertain parameter c takes the form

$$p(\mu \mid D) \propto \frac{1}{\sigma^N} \exp\left[-\frac{1}{2\sigma^2} J(\mu) \right]$$

where the misfit function (or measure of fit function) $J(\mu)$ is given by

$$J(\mu) = \sum_{k=1}^{N} \left[U_{0_k} - U_0(\rho_{s_k}, \mu) \right]^2$$

It can be seen that $J(\mu)$ is a highly nonlinear function of the model parameter μ to be inferred in this exercise.

Appendix 1.1: Solution of equation of motion for the falling object

Introducing the velocity of the object, the set of equations describing the motion is the following

$$\frac{d\upsilon}{dt} = g - c\upsilon^2 \qquad (1)$$

$$\frac{dz}{dt} = \upsilon \qquad (2)$$

with initial conditions $z(t_0) = z_0$ and $\upsilon(t_0) = \upsilon_0$.

From (1) we get

$$\int_{\upsilon_0}^{\upsilon} \frac{1}{g - c\upsilon^2} d\upsilon = \int_{t_0}^{t} dt \qquad (3)$$

The integrand can be written as

$$\frac{1}{g - c\upsilon^2} = \frac{1}{2\sqrt{g}}\left(\frac{1}{\sqrt{g} - \sqrt{c}\upsilon} + \frac{1}{\sqrt{g} + \sqrt{c}\upsilon}\right)$$

Using this

$$\int_{\upsilon_0}^{\upsilon} \frac{1}{g - c\upsilon^2} d\upsilon = \frac{1}{2\sqrt{g}}\left(\int_{\upsilon_0}^{\upsilon} \frac{1}{\sqrt{g} - \sqrt{c}\upsilon} d\upsilon + \int_{\upsilon_0}^{\upsilon} \frac{1}{\sqrt{g} + \sqrt{c}\upsilon} d\upsilon\right) = \frac{1}{2\sqrt{g}}(I_1 + I_2) \qquad (4)$$

Calculating the integrals I_1 and I_2 we get that

$$I_1 = \int_{\upsilon_0}^{\upsilon} \frac{1}{\sqrt{g} - \sqrt{c}\upsilon} d\upsilon = -\frac{1}{\sqrt{c}}\left[\ln\left|\sqrt{g} - \sqrt{c}\upsilon\right| - \ln\left|\sqrt{g} - \sqrt{c}\upsilon_0\right|\right] = -\frac{1}{\sqrt{c}} \ln\left|\frac{\sqrt{g} - \sqrt{c}\upsilon}{\sqrt{g} - \sqrt{c}\upsilon_0}\right|$$

$$I_2 = \int_{\upsilon_0}^{\upsilon} \frac{1}{\sqrt{g} + \sqrt{c}\upsilon} d\upsilon = \frac{1}{\sqrt{c}}\left[\ln\left|\sqrt{g} + \sqrt{c}\upsilon\right| - \ln\left|\sqrt{g} + \sqrt{c}\upsilon_0\right|\right] = \frac{1}{\sqrt{c}} \ln\left|\frac{\sqrt{g} + \sqrt{c}\upsilon}{\sqrt{g} + \sqrt{c}\upsilon_0}\right|$$

Using (3) and (4)

$$\frac{1}{2\sqrt{gc}}\left[-\frac{1}{\sqrt{c}} \ln\left|\frac{\sqrt{g} - \sqrt{c}\upsilon}{\sqrt{g} - \sqrt{c}\upsilon_0}\right| + \frac{1}{\sqrt{c}} \ln\left|\frac{\sqrt{g} + \sqrt{c}\upsilon}{\sqrt{g} + \sqrt{c}\upsilon_0}\right|\right] = t - t_0$$

After manipulating this expression one derives

$$\ln\left|\frac{\left(\sqrt{g} + \sqrt{c}\upsilon\right)\left(\sqrt{g} - \sqrt{c}\upsilon_0\right)}{\left(\sqrt{g} + \sqrt{c}\upsilon_0\right)\left(\sqrt{g} - \sqrt{c}\upsilon\right)}\right| = 2\sqrt{gc}(t - t_0)$$

Assuming that $\upsilon(t_0) = \upsilon_0 = 0$ we get

$$\ln\left|\frac{\left(\sqrt{g}+\sqrt{c}\upsilon\right)}{\left(\sqrt{g}-\sqrt{c}\upsilon\right)}\right| = 2\sqrt{gc}(t-t_0)$$

Introducing $\upsilon_\infty = \sqrt{\frac{g}{c}}$ the expression takes the form

$$\frac{\upsilon_\infty + \upsilon}{\upsilon_\infty - \upsilon} = \exp\left[2\sqrt{gc}\left(t-t_0\right)\right]$$

Manipulating the expression we result in

$$\upsilon(t) = \upsilon_\infty \tanh\left[\sqrt{gc}\left(t-t_0\right)\right]$$

From (2) by integration we get

$$\int_{z_0}^{z} dz = \int_{t_0}^{t} \upsilon_\infty \tanh\left[\sqrt{gc}\left(t-t_0\right)\right]dt$$

Using the known resultthat $\int \tanh x dx = \ln\cosh x + \text{const}$. we can get

$$z - z_0 = \frac{1}{c}\left[\ln\cosh\left[\sqrt{gc}\left(t-t_0\right)\right]\right]_{t_0}^{t} = \frac{1}{c}\ln\cosh\left[\sqrt{gc}\left(t-t_0\right)\right]$$

Assuming that $z(t_0) = z_0 = 0$ we finally derive that

$$z(t) = \frac{1}{c}\ln\cosh\left[\sqrt{gc}\left(t-t_0\right)\right]$$

Appendix 1.2 Analytical Expressions for Derivatives and Hessian of $L(\underline{\theta})$

Assuming uniform prior distribution, from Bayes rule we have

$$p(\underline{\theta}\,|\,\underline{\hat{z}}) = Const * p(\underline{\hat{z}}\,|\,\underline{\theta}) = \frac{Const}{\left(\sqrt{2\pi}\right)^{N}\sqrt{\det\Sigma}}\exp\left\{-\frac{1}{2}\left[\underline{\hat{z}}-\underline{z}(\underline{\theta})\right]^{\mathrm{T}}\Sigma^{-1}\left[\underline{\hat{z}}-\underline{z}(\underline{\theta})\right]\right\}$$

The function to be minimized is

$$L(\underline{\theta}) = -\ln[\text{'posterior'}] = -\ln[\,p(\underline{\theta}\,|\,\underline{\hat{z}})]$$

from which we get

$$L(\underline{\theta},\underline{\varphi}) = \frac{1}{2}\left[\underline{\hat{z}}-\underline{z}(\underline{\theta})\right]^{\mathrm{T}}\Sigma^{-1}(\underline{\varphi})\left[\underline{\hat{z}}-\underline{z}(\underline{\theta})\right] + \frac{1}{2}\ln\left(\det\Sigma(\underline{\varphi})\right) + Const.$$

where we separated the parameters in model parameters $\underline{\theta}$ and prediction error parameters $\underline{\varphi}$. For uncorrelated measurement only errors we have $\Sigma = \sigma^2\mathrm{I}$, and thus $\underline{\varphi} \equiv \sigma$.

Assuming that σ is also a parameter (not fixed) we get

$$L(\underline{\theta},\sigma) = \frac{1}{2\sigma^2}\sum_{\kappa=1}^{N}[\hat{z}_\kappa - z(t_k,\underline{\theta})]^2 + N\ln\sigma + Const.$$

The derivatives of $L(\underline{\theta},\sigma)$ with respect to $\underline{\theta},\sigma$ are obtained analytically in the form

$$\frac{\partial L}{\partial\theta_i} = \frac{1}{2\sigma^2}\sum_{\kappa=1}^{\mathrm{N}}\frac{\partial}{\partial\theta_i}[\hat{z}_\kappa - z(t_k,\underline{\theta})]^2 = \frac{1}{\sigma^2}\sum_{\kappa=1}^{\mathrm{N}}\left\{(z(t_k,\underline{\theta})-\hat{z}_\kappa)\frac{\partial z}{\partial\theta_i}\right\}$$

and

$$\frac{\partial L}{\partial\sigma} = \frac{1}{2}\sum_{\kappa=1}^{\mathrm{N}}[\hat{z}_\kappa - z(t_k,\underline{\theta})]^2\frac{\partial}{\partial\sigma}\left(\frac{1}{\sigma^2}\right) + \mathrm{N}\frac{\partial}{\partial\sigma}(\ln\sigma) = -\frac{1}{\sigma^3}\sum_{\kappa=1}^{\mathrm{N}}[\hat{z}_\kappa - z(t_k,\underline{\theta})]^2 + \frac{\mathrm{N}}{\sigma}$$

where $\dfrac{\partial z}{\partial\theta_i}$ depends on the parameter used in θ_i.

The elements H_{ij} of the hessian matrix of $L(\underline{\theta},\sigma)$ are

$$\frac{\partial^2 L}{\partial\theta_i\partial\theta_j} = \frac{1}{\sigma^2}\sum_{\kappa=1}^{\mathrm{N}}\left\{\frac{\partial z}{\partial\theta_j}\frac{\partial z}{\partial\theta_i} + (z(t_k,\underline{\theta})-\hat{z}_\kappa)\frac{\partial^2 z}{\partial\theta_i\partial\theta_j}\right\}$$

$$\frac{\partial^2 L}{\partial\sigma^2} = \frac{3}{\sigma^4}\sum_{\kappa=1}^{\mathrm{N}}[\hat{z}_\kappa - z(t_k,\underline{\theta})]^2 - \frac{\mathrm{N}}{\sigma^2}$$

$$\frac{\partial^2 L}{\partial\theta_i\partial\sigma} = -\frac{2}{\sigma^3}\sum_{\kappa=1}^{\mathrm{N}}\left\{(z(t_k,\underline{\theta})-\hat{z}_\kappa)\frac{\partial z}{\partial\theta_i}\right\}$$

In the falling object problem we got that

$$z_\kappa = z(\underline{\theta}, t_\kappa) = \frac{1}{C}\ln\left\{\cosh\left[\sqrt{gC}\,(t_\kappa - t_0)\right]\right\}$$

and that

$$\underline{z} = \begin{Bmatrix} z(t_1, \underline{\theta}) \\ z(t_2, \underline{\theta}) \\ \cdots \\ z(t_N, \underline{\theta}) \end{Bmatrix} = \begin{Bmatrix} z_1 \\ z_2 \\ \cdots \\ z_N \end{Bmatrix},$$

where $\underline{\theta} = \{g, C, t_0\}$.

The first derivatives of z are

$$\frac{\partial z}{\partial g} = \frac{(t_\kappa - t_0)\tanh\left[\sqrt{gC}\,(t_\kappa - t_0)\right]}{2\sqrt{gC}}$$

$$\frac{\partial z}{\partial C} = \frac{(t_\kappa - t_0)\,g\tanh\left[\sqrt{gC}\,(t_\kappa - t_0)\right]}{2C\sqrt{gC}} - \frac{\ln\left\{\cosh\left[\sqrt{gC}\,(t_\kappa - t_0)\right]\right\}}{C^2}$$

$$\frac{\partial z}{\partial t_0} = \frac{-\sqrt{gC}\tanh\left[\sqrt{gC}\,(t_\kappa - t_0)\right]}{C}$$

Introducing $T = \tanh\left[\sqrt{gC}\,(t_\kappa - t_0)\right]$, we get

$$\frac{\partial z}{\partial g} = \frac{(t_\kappa - t_0) * T}{2\sqrt{gC}}$$

$$\frac{\partial z}{\partial C} = \frac{(t_\kappa - t_0)\,g * T}{2C\sqrt{gC}} - \frac{z}{C}$$

$$\frac{\partial z}{\partial t_0} = \frac{-\sqrt{gC} * T}{C}$$

while the derivatives of T with respect top the parameters are given by

$$\frac{\partial T}{\partial g} = \left(1 - T^2\right)(t_\kappa - t_0)\frac{C}{2\sqrt{gC}}$$

$$\frac{\partial T}{\partial C} = \left(1 - T^2\right)(t_\kappa - t_0)\frac{g}{2\sqrt{gC}}$$

$$\frac{\partial T}{\partial t_0} = -\left(1 - T^2\right)\sqrt{gC}$$

Finally, the second derivatives of z are obtained in the form

$$\frac{\partial^2 z}{\partial g^2} = \left(t_\kappa - t_0\right)\frac{2\sqrt{gC}\,\frac{\partial T}{\partial g} - T\,\frac{C}{\sqrt{gC}}}{4gC}$$

$$\frac{\partial^2 z}{\partial g \partial C} = \left(t_\kappa - t_0\right)\frac{2\sqrt{gC}\,\frac{\partial T}{\partial C} - T\,\frac{g}{\sqrt{gC}}}{4gC}$$

$$\frac{\partial^2 z}{\partial g \partial t_0} = \frac{-T + \left(t_\kappa - t_0\right)\frac{\partial T}{\partial t_0}}{2\sqrt{gC}}$$

$$\frac{\partial^2 z}{\partial C^2} = \frac{g\left(t_\kappa - t_0\right)}{2\sqrt{g}}\,\frac{\frac{\partial T}{\partial C}C^{2/3} - \frac{3}{2}T * C^{1/2}}{C^3} - \frac{\frac{\partial z}{\partial C} * C - z}{C^2}$$

$$\frac{\partial^2 z}{\partial C \partial t_0} = \frac{g}{2C\sqrt{gC}}\left[-T + \left(t_\kappa - t_0\right)\frac{\partial T}{\partial t_0}\right] - \frac{1}{C}\frac{\partial z}{\partial t_0}$$

$$\frac{\partial^2 z}{\partial t_0^{\,2}} = -\frac{\sqrt{gC}}{C}\frac{\partial T}{\partial t_0}$$