MIL 780 - Assignment Three

Ryan Balshaw

20 June 2022

Foreword

All code, documents and work relevant to this assignment can be found in my Github repository.

1 Question 1

Consider the following mixture distribution

$$p(x) = a_1 \cdot \mathcal{N}(x|0, 1^2) + a_2 \cdot L(x|5, 2), \tag{1}$$

where $a_1 = 0.4$, $a_2 = 0.6$, $\mathcal{N}(x|0, 1^2)$ is a Gaussian distribution with a mean of zero and a variance of 1^2, and $L(x|5, 2)$ is a Laplacian distribution with a location of 5 and a scaling parameter of 2.

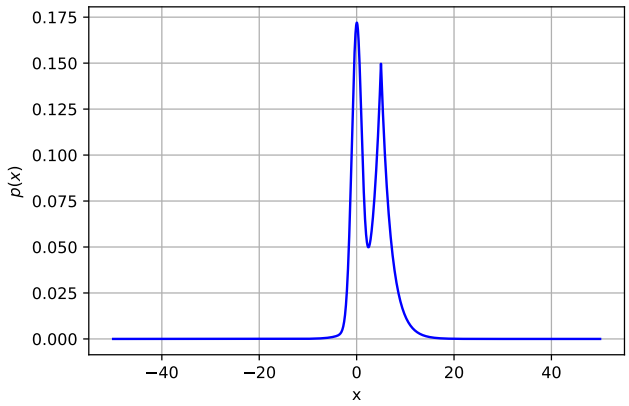


Figure 1: The distribution in Equation (1) visualised over $x \in [-50, 50]$.

To test whether the distribution of interest is valid, we need to consider the properties that constitute a valid distribution. These properties are

1. $p(x) \geq 0$, $\forall x$.
2. $\int_x p(x)dx = 1$.

The first property is resolved as both the Laplacian and Gaussian distributions in Equation (1) are nonnegative $\forall\ x$, and both scale coefficients a_1 and a_2 are non-negative. The second property is resolved as the integral of a linear combination of functions is a linear combination of function integrals. As such, the integral of interest becomes

$$\begin{aligned}\int_x p(x)dx &= \int_x (a_1 \cdot \mathcal{N}(x|0, 1^2) + a_2 \cdot L(x|5, 2))dx \\ &= a_1 \cdot \int_x \mathcal{N}(x|0, 1^2)dx + a_2 \cdot \int_x L(x|5, 2)dx \\ &= a_1 \cdot 1 + a_2 \cdot 1 \\ &= 1.\end{aligned} \tag{2}$$

For this problem, the objective is to use rejection sampling to draw samples from the distribution shown in Equation (1). The goal in rejection sampling is to sample from a complex, possibly unnormalised, distribution $\tilde{p}(x)$ (which may or may not be normalised) using a simpler distribution $q(x)$. This simpler distribution is commonly referred to as the proposal distribution. The rejection sampling procedure is given through

- Initialise a proposal distribution $q(x)$.
- Define/find a scale k such that $k \cdot q(x) > \tilde{p}(x)\ \forall x$.
- Draw a sample $x_i \sim q(x)$.
- Evaluate $q(x_i)$ and $\tilde{p}(x_i)$.
- Sample from a uniform distribution $u_i \sim U[0, 1]$.
- Accept the sample if $\frac{\tilde{p}(x_i)}{k \cdot q(x_i)} > u_i$, reject otherwise (alternatively, if $\tilde{p}(x_i) > u_i \cdot k \cdot q(x_i)$).
- Repeat until a suitable number of samples is drawn.

To proceed with rejection sampling, the goal is to explore the influence of *i)* different proposal distributions,*ii)* different parametrisations of the proposal distributions, and *iii)* the effect of the scale parameter k on the efficiency of the rejection sampling procedure. In the implementation of rejection sampling on this one-dimensional, the parameter k can be resolved automatically by evaluating $k \cdot q(x) > \tilde{p}(x)\ \forall x$ over an x-domain of interest, provided the domain is sufficiently diverse. The considered proposal distributions are the Gaussian distribution, the Laplacian distribution and Student's t distribution. A two degree of freedom t distribution is used for all experiments. The variance of the proposal distribution is of interest in this investigation, as this can greatly affect the efficiency of the rejection sampling process. To absolve any implementation ambiguity, it is assumed that the expectation $\mathbb{E}_{x\sim p(x)}\{x\}$ is known. This is trivial to evaluate as it reduces to $\mathbb{E}_{x\sim p(x)}\{x\} = a_1 \cdot \mu_{gauss} + a_2 \cdot \mu_{laplace} = 3$.

In Figure 2 the results of this investigation is shown. Figure 2 clearly shows that the efficiency of rejection sampling, for the proposal distributions considered, is heavily reliant on the scale parameter k. Regardless of the parametrisation, there is an exponential decay in the efficiency as k is increased. The best performing proposal distribution is the Student's t distribution with a location of 3 and a scale parameter of 5, which had an efficiency just over 30%. In Figure 2, the degradation of the efficiency as a function of an increasing scale parameter is crucial to highlight that distributions that are tight around $p(x)$ are crucial to improved efficiencies. This follows from the rejection sampling algorithm, as if the ratio $\frac{\tilde{p}(x_i)}{k \cdot q(x_i)}$ for any proposal distribution sample $x_i \sim q(x)$ is far from 1 and closer to zero (i.e., the scaled proposal distribution has high sample likelihood while the true distribution does not and vice versa), then more samples will be rejected. Hence, a tight proposal distribution is key to efficient rejection sampling.

In Figure 3, the samples from the proposal distribution with the highest sampling efficiency are shown. It is clear that the rejection sampling process can correctly sample from $p(x)$. Now that samples from $p(x)$ may be drawn, the samples may be used to inspect different aspects of $p(x)$.

The aspects of interest for this problem are:

- $\mathbb{E}_{x\sim p(x)}\{x^2\} = \int_x x^2 p(x) dx$.
- $P(x > 5) = \mathbb{E}_{x\sim p(x)}\{I(x| - \infty, 5)\} = \int_x p(x) I(x|\infty, 5) dx$.
- The 3^{rd} percentile of $p(x)$.

Note that $I(x| - \infty, 5)$ is an indicator function defined through

$$I(x|a, b) = \begin{cases} 1 & \text{if } a \leq x \leq b, \\ 0 & \text{otherwise,} \end{cases}. \tag{3}$$

To estimate the aspects of interest, Monte Carlo integration was used. To compare the performance of the estimates obtained through Monte Carlo integration, numerical integration was used to evaluate the aspects analytically. In Table 1 the results for the two estimation methods are shown. It is clear that the largest disparity in Table 1 was for the $\mathbb{E}_{x\sim p(x)}\{x^2\}$ estimate. The remaining estimates are numerically close, which indicates that the error in the $\mathbb{E}_{x\sim p(x)}\{x^2\}$ estimate may be resolved through sampling $p(x)$ more densely.

To quantify the effect of the number of samples on the accuracy of the aforementioned estimates, Figure 4 was developed by evaluating the estimates as a function of the number of samples. It is clear that as the number of samples increases, the error in the Monte Carlo integration estimates decreases. This correctly suggests that in order to obtain an accurate Monte Carlo estimate, a large number of samples must be used.

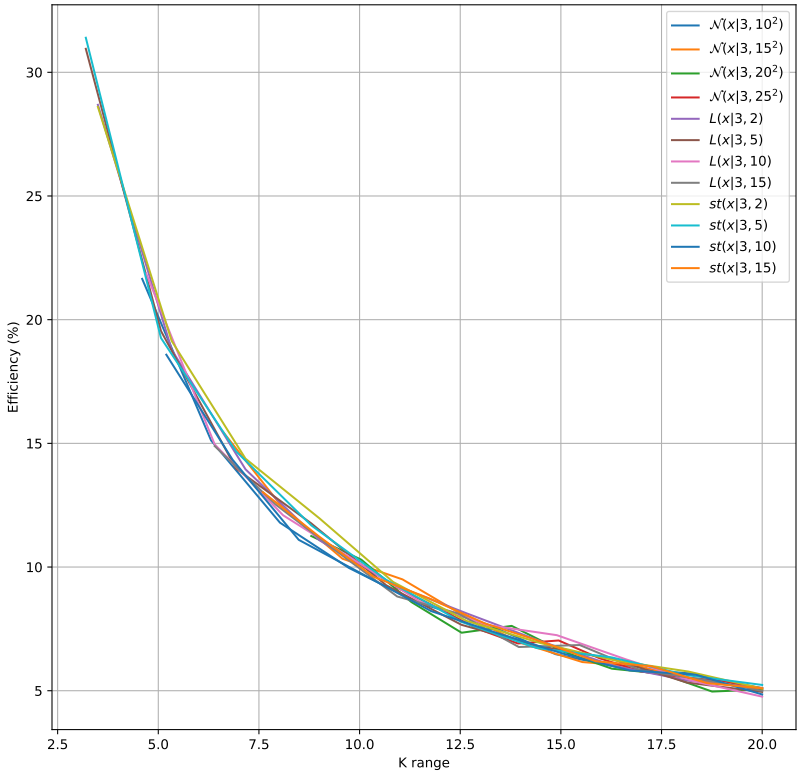


Figure 2: The efficiency of the rejection sampling procedure for 1000 samples under different parametrisations of the proposal distribution $q(x)$. Note that as the scale parameter k was determined automatically, the starting points for each of the different distributions is not consistent.

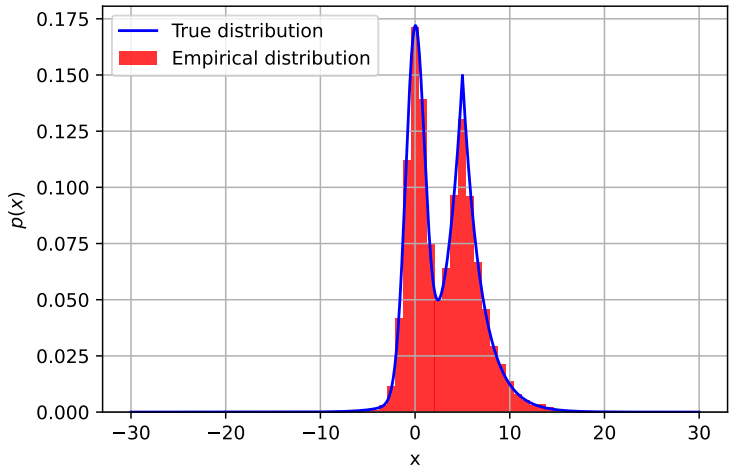
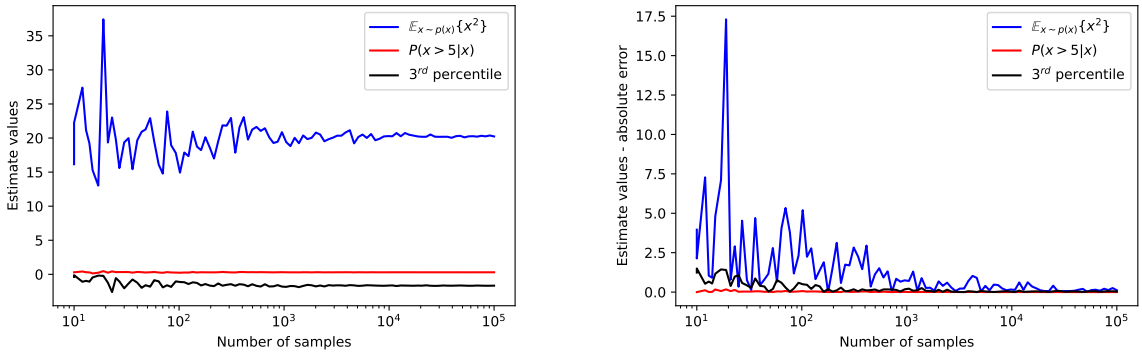


Figure 3: The samples from $p(x)$ using rejection sampling and a proposal distribution $\mathrm{st}_2(x|3, 5)$.

Table 1: The estimates of interest using the samples of $p(x)$ obtained using rejection sampling.

Method	$\mathbb{E}_{x\sim p(x)}\{x^2\}$	$P(x > 5)$	3^{rd} percentile
Rejection sampling	20.527	0.3037	-1.6204
Numerical evaluation	20.12	0.3	-1.6216



(a) The estimate-sample dependency.

(b) The estimate-sample error.

Figure 4: The effect of the number of samples on the estimates. (a) shows the variation in the estimate and (b) shows the estimate error as a function of the number of samples. Note that the estimates were obtained as a once-off result, and as such this figure does not effectively convey the variation in the estimates for a given number of samples. However, it is clear that increasing the number of samples reduces the error in the estimates.

2 Question 2

Consider the generative model

$$x_n = \theta + \epsilon_n, \tag{4}$$

where ϵ_n is a Laplacian distribution with a known scale parameter of 2 and a location of 0, $\epsilon \sim L(\epsilon|0, 2)$. The population parameter θ is three and not known during the inference process. The following measured data is available for this problem:

$$\begin{aligned}\mathbf{x} =&[4.171206, 3.06754327, 2.16931111, 2.00839214, 4.10014516,\\ &3.51408053, 2.86609886, 6.3492669, 2.90566741, 5.99086803]^T\end{aligned} \tag{5}$$

The objective of this problem is to use Markov Chain Monte Carlo (MCMC) to sample from the posterior distribution over the unknown parameter. The MCMC method in this problem uses the Metropolis-Hastings algorithm, and may be given as

1. Assume the proposal distribution $q(\theta)$, the unnormalised distribution of interest $\tilde{p}(\theta)$, a starting point θ_0, and the number of iterations N.
2. For $n = 1, \cdots, N$:
   - Sample from the proposal distribution $\theta_n \sim q(\theta|\theta_{n-1})$
   - Draw a random sample $u_i \sim U[0, 1]$
   - Accept the θ_n sample if $\min\left(1, \frac{\tilde{p}(\theta_n)q(\theta_{n-1}|\theta_n)}{\tilde{p}(\theta_{n-1})q(\theta_n|\theta_{n-1})}\right) \geq u_i$, reject otherwise
   - If reject: $\theta_n = \theta_{n-1}$
   - If accept: θ_n is kept

Once the samples have been obtained using a MCMC method, the goal is to determine some expectations and probabilities that are of interest for this problem. The elements of interest, which are to be found using Monte Carlo integration, are

- $\mathbb{E}_{\theta\sim p(\theta|\mathbf{x})}\{\theta^2\}$.
- $P(\theta > 3|\mathbf{x})$.
- $P(x > 3|\mathbf{x})$.

It is important to note that $P(x > 3|\mathbf{x})$ is equal to

$$P(x > 3|\mathbf{x}) = 1 - \mathbb{E}_{x\sim p(x|\mathbf{x})}\{I(x| - \infty, 3)\} \tag{6}$$

which requires access to samples from $p(x|\mathbf{x})$. To obtain posterior predictive distribution samples, θ samples may be drawn from the posterior distribution $\theta_i \sim p(\theta|\mathbf{x})$, and a corresponding set of generative model samples $x \sim p(x|\theta_i)$ can then be drawn. The expectation of interest may then be determined for each θ_i sample and the average of these expectations is used to approximate $\mathbb{E}_{x\sim p(x|\mathbf{x})}\{I(x, -\infty, 3)\}$.

The parameters of the proposal distribution are critical to the performance of the MCMC algorithm. To establish how the MCMC algorithm depends on the proposal distribution and the parameters therein, a investigation must occur. As such, a Laplacian and Gaussian proposal distribution will be used and the variance of these distributions will be changed. To determine the effects of the proposal distribution on the MCMC algorithm, the acceptance ratio and the aforementioned estimates will be tracked. The next objective is to describe the distributions of interest.

The prior used in this problem is a Gaussian prior with a mean of 0 and a variance of 25. The distributions of interest here are the prior, the generative model which produces the likelihood function, and the unnormalised posterior. Mathematically, the prior is given as

$$\begin{aligned}p(\theta) &= \mathcal{N}(\theta|\theta_0, \sigma_0^2)\\ &= \frac{1}{\sqrt{2\cdot\pi\cdot\sigma_0^2}}\exp\left(-\frac{1}{2\cdot\sigma_0^2}\cdot(\theta-\theta_0)^2\right),\end{aligned} \tag{7}$$

where $\theta_0 = 0$ and $\sigma_0 = 5$. The generative model

$$p(x|\theta) = L(x|\theta, b) = \frac{1}{2 \cdot b} \cdot \exp\left(-\frac{|x - \theta|}{b}\right), \tag{8}$$

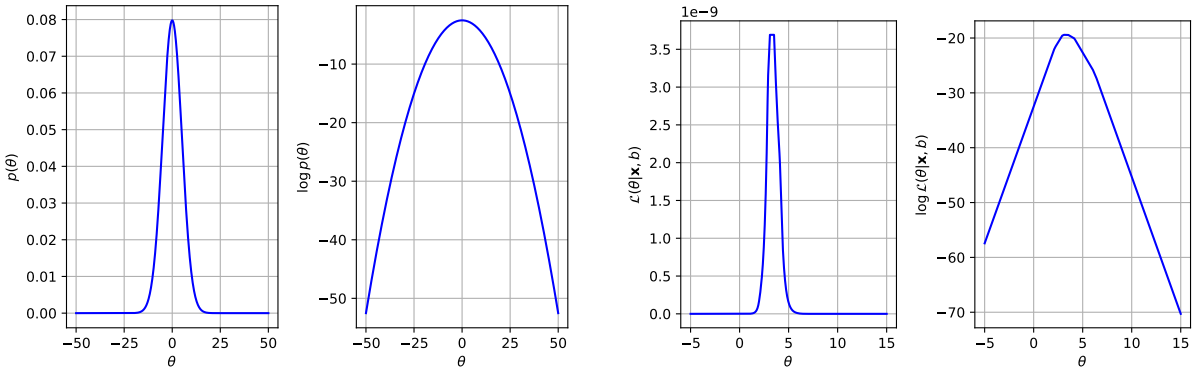
gives rise to the likelihood function

$$\begin{aligned} p(\mathbf{x}|\theta) &= \prod_{n=1}^{N} p(x_n|\theta) \\ \mathcal{L}(\theta|\mathbf{x}, b) &= \prod_{n=1}^{N} \frac{1}{2 \cdot b} \cdot \exp\left(-\frac{|x_n - \theta|}{b}\right), \end{aligned} \tag{9}$$

Finally, the unnormalised posterior is given through

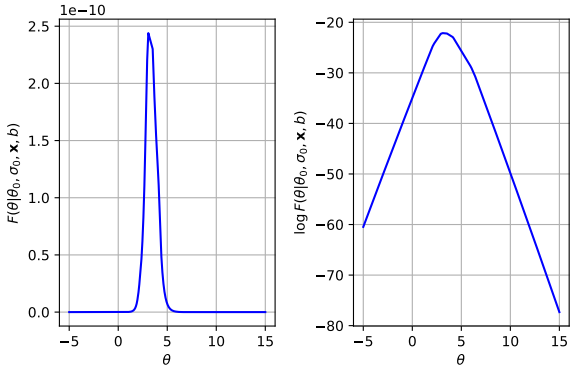
$$\begin{aligned} p(\theta, \mathbf{x}) &= p(\mathbf{x}|\theta)p(\theta) \\ F(\theta|\theta_0, \sigma_0, \mathbf{x}, b) &= \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma_0^2}} \exp\left(-\frac{1}{2 \cdot \sigma_0^2} \cdot (\theta - \theta_0)^2\right) \prod_{n=1}^{N} \frac{1}{2 \cdot b} \cdot \exp\left(-\frac{|x_n - \theta|}{b}\right), \end{aligned} \tag{10}$$

It is important to note here that the goal is to draw samples from the posterior using MCMC. Thus, the distribution of interest $\tilde{p}(\theta)$ is the unnormalised posterior given in Equation (10). In Figure 5, the prior and unnormalised posterior distributions alongside the likelihood function is shown.



(a) The prior.

(b) The likelihood function.



(c) The unnormalised posterior.

Figure 5: The prior, likelihood function and the unnormalised posterior for the second problem.

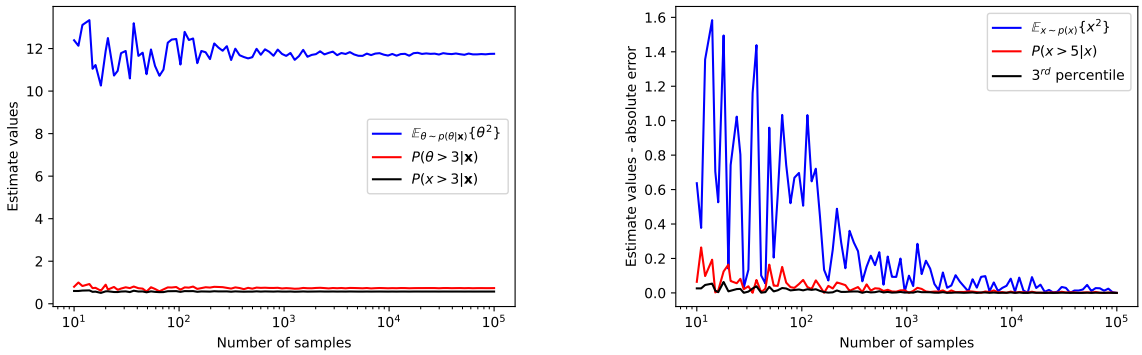
In Table 2, the results of the MCMC method for different proposal distributions is shown. Crucially, it is seen that as the variance of the proposal distribution increased, the acceptance ratio subsequently decreased. This is beneficial to the MCMC method as a highly localised variance biases the method to only accept samples if

the distribution of interest increases in likelihood, whereas a lower variance encourages the model to explore the parameter space and obtain better samples. The clear correlation to the proposal distribution variance indicates that a better change of accepting a new sample if the proposal distribution is highly localised. However, the danger of a high acceptance ratio is that there is no guarantee that you are exploring the space, you only know that you are increasing the likelihood ratio $\frac{\tilde{p}(\theta_n)}{\tilde{p}(\theta_{n-1})}$. Alternatively, a low acceptance ratio may indicate that the exploration has gone off course, and hence you are sampling from regions where $\tilde{p}(\theta)$ does not lie. Thus, there is a natural trade-off between improving the parameter space likelihood improvement and exploration.

Table 2: The acceptance ratio and the estimates of interest for different proposal distributions with different initial parameters. Note that the estimates do not effectively capture the variation in the value of the estimates, and hence are biased as they were determined once off. A tuning period of 50%, a thinning size of 10, and 100 000 iterations were performed for all experiments.

Proposal distribution	Acceptance ratio	$\mathbb{E}_{\theta\sim p(\theta\|\mathbf{x})}\{\theta^2\}$	$P(\theta > 3\|\mathbf{x})$	$P(x > 3\|\mathbf{x})$
Gaussian $(N(\theta\|0, 0.25^2)$	86.43%	11.7337	0.7314	0.5742
Laplacian $L(\theta\|0, 0.25)$	83.41%	11.6734	0.7342	0.5725
Gaussian $(N(\theta\|0, 0.5^2)$	73.31%	11.823	0.731	0.5759
Laplacian $L(\theta\|0, 0.5)$	69.77%	11.7571	0.7368	0.5751
Gaussian $(N(\theta\|0, 2^2)$	33.40%	11.8197	0.7378	0.5754
Laplacian $L(\theta\|0, 2)$	33.2%	11.7278	0.7272	0.5730
Gaussian $(N(\theta\|0, 2^2)$	14.68%	11.6620	0.722	0.5712
Laplacian $L(\theta\|0, 2)$	15.88%	11.6514	0.7314	0.5714

Finally, it is important to quantify the effect of the number of samples on the accuracy of the aforementioned estimates. Figure 4 was developed by evaluating the estimates as a function of the number of samples. It is clear that as the number of samples increases, the error in the Monte Carlo integration step of the MCMC sample estimates decreases. This correctly suggests that in order to obtain an accurate Monte Carlo estimate, a large number of samples must be used.



(a) The estimate-sample dependency.

(b) The estimate-sample error.

Figure 6: The effect of the number of samples on the MCMC-based estimates. (a) shows the variation in the estimate and (b) shows the estimate error as a function of the number of samples. Note that the estimates were obtained as a once-off result, and as such this figure does not effectively convey the variation in the estimates for a given number of samples. However, it is clear that increasing the number of samples reduces the error in the estimates. A Gaussian distribution with a mean of 0 and a variance of 2^2 was used for all iterations.

3 Question 3

The objective of the third problem is to perform Bayesian inference on a linear elastic-linear hardening material model. The model of interest is given by

$$\sigma_{meas} = \sigma_l(\epsilon, \boldsymbol{\theta}) + \nu, \tag{11}$$

where ϵ is strain, $\boldsymbol{\theta} = [E, \sigma_{y_0}, H]^T$, E is the Young's Modulus, σ_{y_0} is the yield strength and H is the plastic modulus. The additive noise ν is obtained from a zero-mean Gaussian with an unknown variance η^2. The material model $\sigma_l(\eta, \boldsymbol{\theta})$ is given by

$$\sigma_l(\epsilon, \boldsymbol{\theta}) = E \cdot \epsilon \cdot \left(1 - h(\epsilon - \frac{\sigma_{y_0}}{E})\right) + \left(\sigma_{y_0} + \frac{H \cdot E}{H + E} \cdot \left[\epsilon - \frac{\sigma_{y_0}}{E}\right]\right) \cdot h(\epsilon - \frac{\sigma_{y_0}}{E}), \tag{12}$$

where $h(\cdot)$ is the heavyside function

$$h(x) = \begin{cases} 1 & \text{if } x \geq 0, \\ 0 & \text{if } x \leq 0 \end{cases}. \tag{13}$$

To ensure that the model does not succumb to the units of the different parameters, the parameters are scaled as follows:

- $E = 10^{11} \cdot \tilde{E}$
- $\sigma_{y_0} = 10^{11} \cdot \tilde{\sigma}_{y_0}$
- $H = 10^{11} \cdot \tilde{H}$

where $\tilde{\boldsymbol{\theta}} = \left[\tilde{E}, \tilde{\sigma}_{y_0}, \tilde{H}\right]$. The prior over the scaled parameters $\{\tilde{\boldsymbol{\theta}}, \eta\}$ is given as a factorial Gamma prior

$$p(\tilde{\boldsymbol{\theta}}, \eta) = \text{Ga}\left(\tilde{E}, \alpha = 3, \beta = 1\right) \cdot \text{Ga}\left(\tilde{\sigma}_{y_0}, \alpha = 3, \beta = 1\right) \cdot \text{Ga}\left(\tilde{H}, \alpha = 3, \beta = 1\right) \cdot \text{Ga}\left(\eta, \alpha = 3, \beta = 1\right). \tag{14}$$

The goal is to use PyMC3 to sample from the posterior distribution, and then compare the posterior to a pre-defined Gaussian posterior distribution. The goal here is to ask questions about the material model given some observed data. The observed data in this question is shown in Figure 7 superimposed over two initialisations of the material model with zero noise.

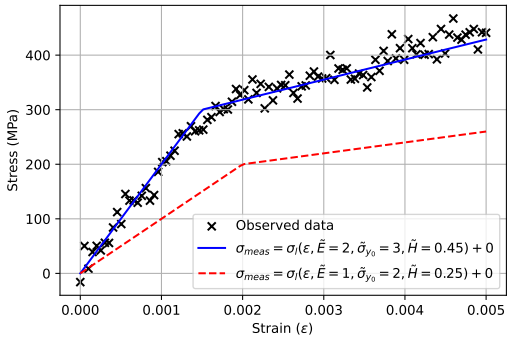


Figure 7: The observed material data for the third problem.

3.1 PyMC3 analysis

In the first part of this problem, the objective is to use `PyMC3` to sample from the posterior distribution. In Figure 8 the samples from the prior and posterior distributions $p(\tilde{\boldsymbol{\theta}}, \eta)$ and $p(\tilde{\boldsymbol{\theta}}, \eta | \boldsymbol{\sigma}_{meas}, \boldsymbol{\epsilon})$ are shown. The change in the parameters is distinct, as the posterior distributions are all less skewed. The MAP estimate of the model parameters was found to be $\{\boldsymbol{\theta}, \eta\}_{MAP} = \{E = 189.844GPa, \sigma_{y_0} = 300.997MPa, H = 54.188GPa, \eta = 16.002\}$. In Figure 9 the prior and posterior distribution samples are used to draw samples from the prior and posterior predictive

distributions. In Figure 10, the `PyMC3` posterior predictive samples is shown. I suspect that PyMC3 follows the sample process to sample from the prior and posterior predictive distributions as was done in this work.

Using the posterior samples the posterior mean and covariance was determined. This is given as

$$\boldsymbol{\mu}_{posterior} = [190400.157MPa, 299.925MPa, 54681.867MPa, 16.39]^T, \tag{15}$$

$$\boldsymbol{\Sigma}_{post} = \begin{bmatrix} 1.501e^7 & -1.228e^4 & 1.778e^6 & 0 \\ -1.228e^4 & 4.193e^1 & -1.946e^3 & 0 \\ 1.778e^6 & -1.946e^4 & 1.309e^7 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{16}$$

Finally, an investigation was performed into the effect of the prior hyper-parameters. In Figure 11 the influence of the α and β terms in the priors on the posterior MAP estimate is shown. It is clear that model parameters are only sensitive to hyper-parameters where $\alpha \geq 9$ or $\beta \leq 0.5$. After these ranges, there is little noticeable variation in the MAP parameters.

3.2 Application and inference

We can now use the results from `PyMC3` to answer questions about the model in this problem. To compare to the `PyMC3` results, a pre-defined posterior distribution is given as

$$p(\boldsymbol{\theta}, \eta | \boldsymbol{\sigma}_{meas}, \boldsymbol{\epsilon}) = \mathcal{N}\left(\begin{bmatrix} E \\ \sigma_{y_0} \\ H \\ eta \end{bmatrix} | \begin{bmatrix} 190000 \\ 300 \\ 55000 \\ 18 \end{bmatrix}, \begin{bmatrix} 2 \cdot e^7 & -1e^{-4} & 1e^6 & 0 \\ -1e^{-4} & 30 & -1e^{-4} & 0 \\ 1e^6 & -1e^4 & 1e^7 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \right). \tag{17}$$

Firstly, given an initial estimate of the model parameters $E = 200GPa, \sigma_{y_0} = 300MPa, H = 45GPa$, we can determine the log-likelihood function as a function of η. This is possible as the generative model was assumed to be a Gaussian distribution

$$p(\sigma_{meas} | \sigma_l(\boldsymbol{\theta}, \epsilon), \eta) = \frac{1}{\sqrt{2 \cdot \pi \cdot \eta^2}} \cdot \exp\left(-\frac{1}{2} \cdot \frac{(\sigma_{meas} - \sigma_l(\boldsymbol{\theta}, \epsilon))^2}{\eta^2} \right). \tag{18}$$

Hence, the log-likelihood function, given some observed data $\{\boldsymbol{\sigma}_{meas}, \boldsymbol{\epsilon}\}$, is given by

$$LL(\boldsymbol{\theta}, \eta | \boldsymbol{\sigma}_{meas}, \boldsymbol{\epsilon}) = -\frac{N}{2} \cdot \log(2 \cdot \pi) - \frac{N}{2} \cdot \log(\eta^2) - \frac{1}{2 \cdot \eta^2} \sum_{n=1}^{N} \left[\sigma_{meas,n} - \sigma_l(\boldsymbol{\theta}, \epsilon_n) \right]^2. \tag{19}$$

If the assumption is made that the noise η is unknown, the resulting likelihood function is shown in Figure 12. The parameter η that maximised the likelihood function is $\eta = 19.023$. If the noise is known ($\eta = 18$), then the likelihood function evaluates to $LL(\boldsymbol{\theta}, \eta | \boldsymbol{\sigma}_{meas}, \boldsymbol{\epsilon}) = -437.307$ and the likelihood function evaluates to $L(\boldsymbol{\theta}, \eta | \boldsymbol{\sigma}_{meas}, \boldsymbol{\epsilon}) = 1.2022e^{-190}$.

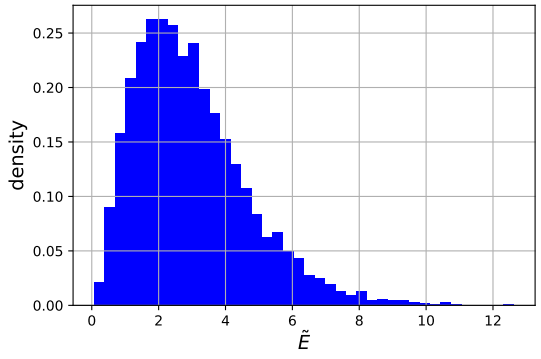
The next question of interest is: *what is the* 95% *credible interval for the posterior yield stress?* In this setting, the objective is to evaluate

$$\nu_{\alpha/2} \leq \sigma_{y_0} | \boldsymbol{\sigma}_{meas}, \boldsymbol{\epsilon} \leq \nu_{1-\alpha/2}, \tag{20}$$

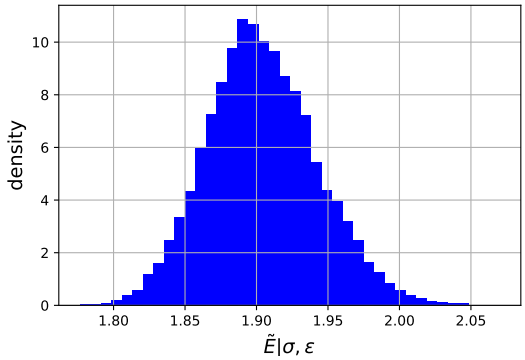
where α is the interval probability and ν_k is the k^{th} percentile. This interval was found using samples from the posterior using the posterior given in Equation (17) and `PyMC3`. These are given as

$$\text{Pre-defined posterior samples: } 289.274 \text{ MPa} \leq \sigma_{y_0} | \boldsymbol{\sigma}_{meas}, \boldsymbol{\epsilon} \leq 310.702 \text{ MPa}, \tag{21}$$

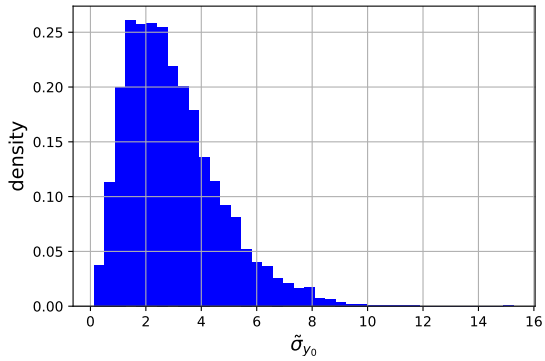
$$\texttt{PyMC3} \text{ posterior samples: } 286.801 \text{ MPa} \leq \sigma_{y_0} | \boldsymbol{\sigma}_{meas}, \boldsymbol{\epsilon} \leq 312.087 \text{ MPa}, \tag{22}$$



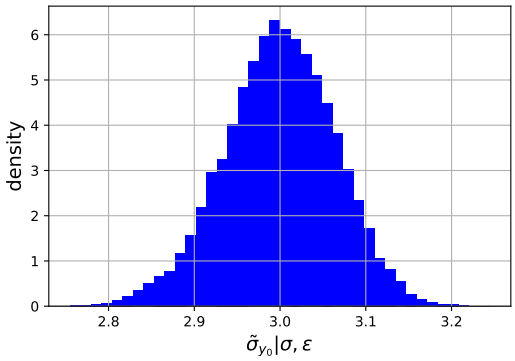
(a) Prior $\tilde{E}$ samples.



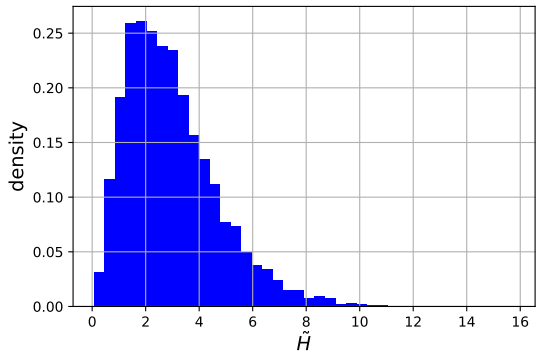
(b) Posterior $\tilde{E}$ samples.



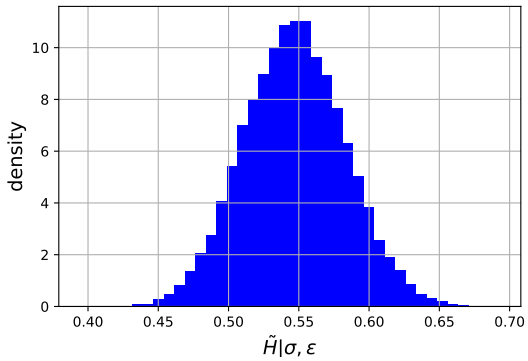
(c) Prior $\tilde{\sigma}_{y_0}$ samples.



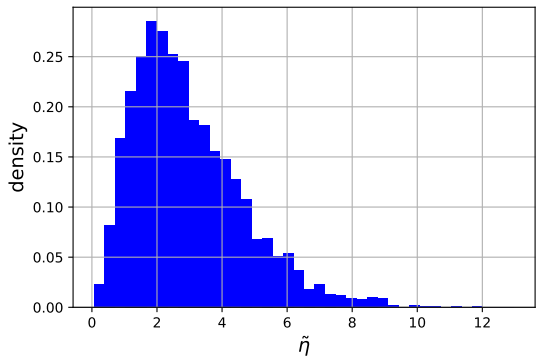
(d) Posterior $\tilde{\sigma}_{y_0}$ samples.



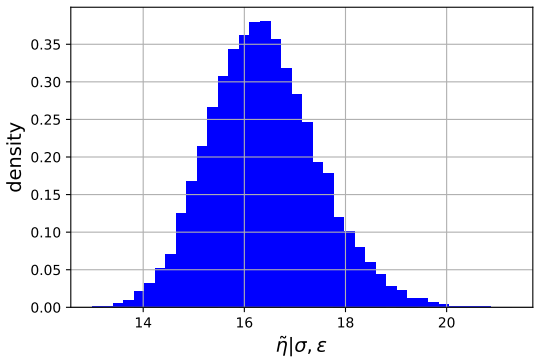
(e) Prior $\tilde{H}$ samples.



(f) Posterior $\tilde{H}$ samples.

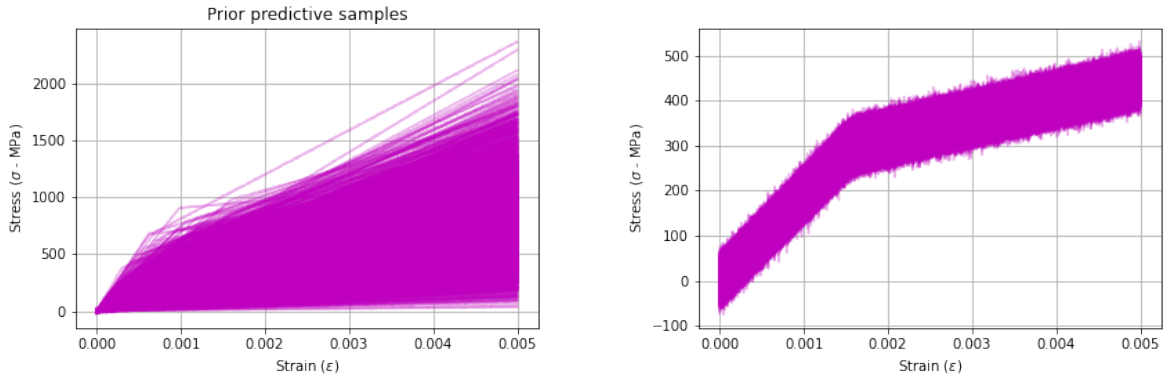


(g) Prior η samples.



(h) Posterior η samples.

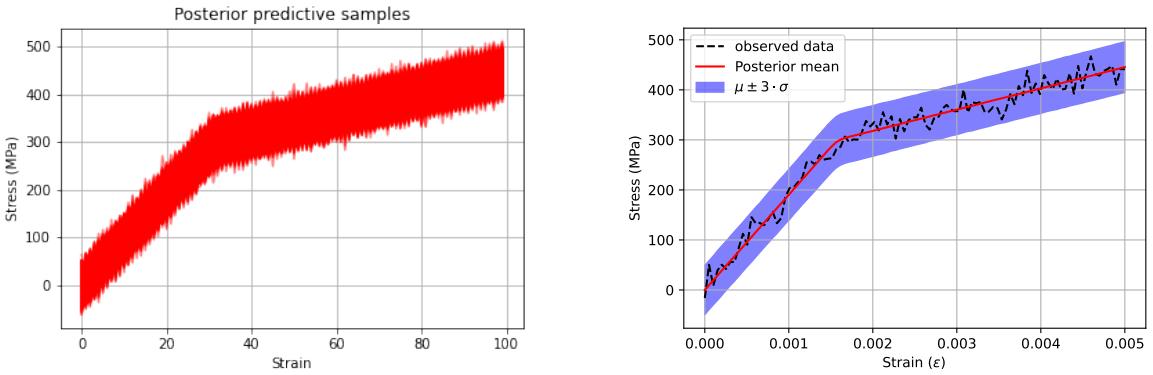
Figure 8: Samples from the prior and posterior distributions over $\{\tilde{\boldsymbol{\theta}}, \eta\}$.



(a) Prior predictive distribution samples.

(b) Posterior predictive distribution samples.

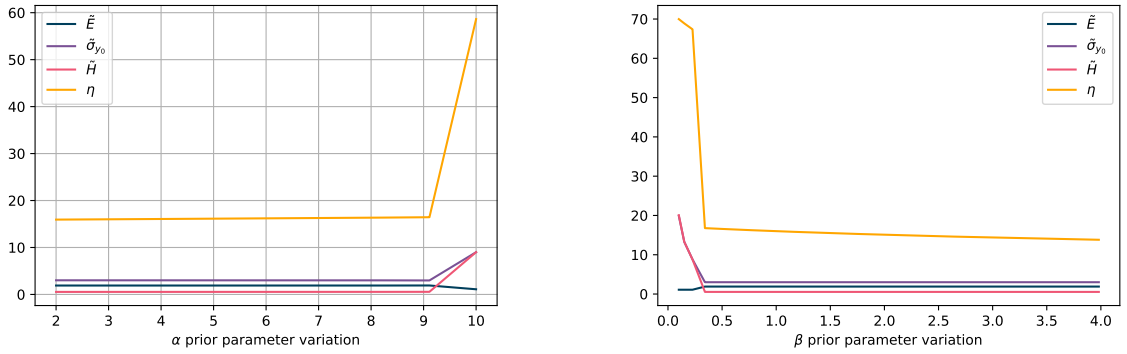
Figure 9: Samples from the prior and posterior predictive distributions for the linear elasticity-linear hardening material model using the prior and posterior distribution `PyMC3` samples.



(a) Prior predictive distribution samples.

(b) Posterior predictive distribution samples.

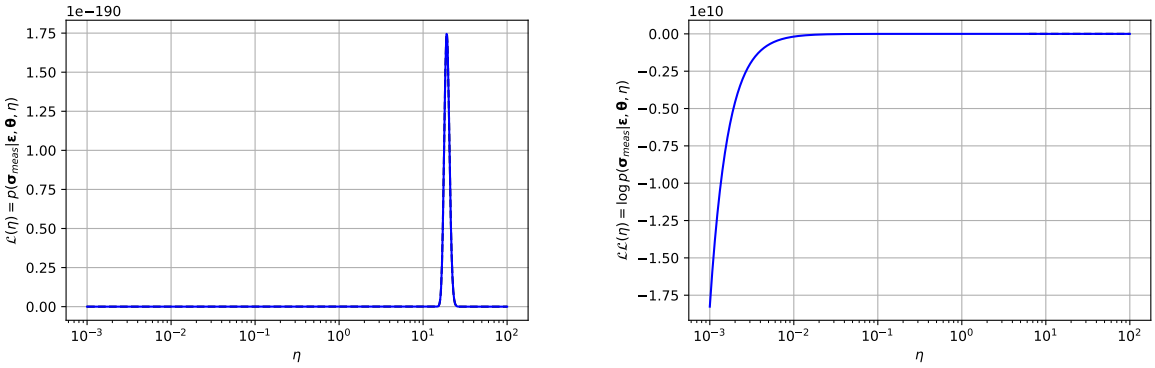
Figure 10: Samples from the prior and posterior predictive distributions for the linear elasticity-linear hardening material model obtained using `PyMC3`.



(a) α variation.

(b) β variation.

Figure 11: The effect of the prior hyper-parameters on the MAP estimates of the model hyper-parameters.



(a) Likelihood function over η.

(b) Log-likelihood function over η.

Figure 12: The likelihood and log-likelihood function for $E = 200GPa, \sigma_{y_0} = 300MPa, H = 45GPa$ and an unknown noise.

The third question is: *what is the expectation* $\mathbb{E}\{H|\boldsymbol{\sigma}_{meas}, \boldsymbol{\epsilon}\}$*?* This can be found using Monte Carlo integration for samples from the posterior marginal distribution $p(H|\boldsymbol{\sigma}_{meas}, \boldsymbol{\epsilon})$. Using the predefined posterior and the `PyMC3` samples, this evaluates to

- $\mathbb{E}\{H|\boldsymbol{\sigma}_{meas}, \boldsymbol{\epsilon}\}_{pre-defined} = 54.987$ GPa
- $\mathbb{E}\{H|\boldsymbol{\sigma}_{meas}, \boldsymbol{\epsilon}\}_{\texttt{PyMC3}} = 54.592$ GPa

Additionally, the MAP estimate of the H parameter is given as

- MAP estimate (pre-defined posterior): 55 GPa
- MAP estimate (`PyMC3` posterior): 54.188 GPa

The fourth question of interest is: *What is the probability that the material will yield if the maximum stress is 250 MPa?* The can be answered by evaluating the marginal yield stress probability $p(\sigma_{y_0} < 250|\boldsymbol{\sigma}_{meas}, \boldsymbol{\epsilon}) = \mathbb{E}_{\sigma_{y_0} \sim p(\sigma_{y_0}|\boldsymbol{\sigma}_{meas}, \boldsymbol{\epsilon})}\{I(\sigma_{y_0}| - \infty, 250)\}$. Using samples from the pre-defined posterior and by using `PyMC3`, this probability evaluates to $\mathbb{E}_{\sigma_{y_0} \sim p(\sigma_{y_0}|\boldsymbol{\sigma}_{meas}, \boldsymbol{\epsilon})}\{I(\sigma_{y_0}| - \infty, 250)\} = 0$. To establish why this occurs, please consider Figure 13. It is clear from Figure 13 that the maximum stress of 250 is far below the posterior samples, and hence it is highly unlikely that the material will fail under a maximum stress of 250 MPa.

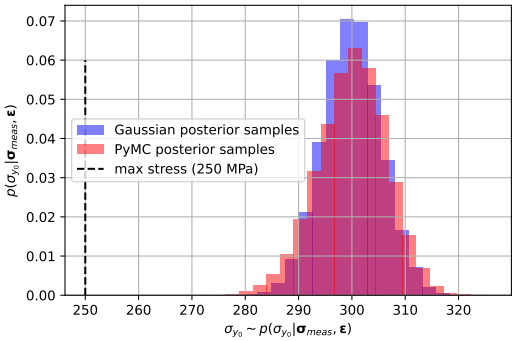


Figure 13: Samples from the marginal distribution $p(\sigma_{y_0}|\boldsymbol{\sigma}_{meas}, \boldsymbol{\epsilon})$ over the yield stress.

The fifth question of interest is: *given the historical data, what is the probability that the measured stress* σ_{meas} *will be more than 200 MPa for a strain of 0.001?* This question can be re-written as: what is the probability $p(\sigma > 200|\epsilon = 0.001, \boldsymbol{\sigma}_{meas}, \boldsymbol{\epsilon})$? This can be expanded through

$$\begin{aligned} p(\sigma > 200|\epsilon = 0.001, \boldsymbol{\sigma}_{meas}, \boldsymbol{\epsilon}) &= 1 - p(\sigma \leq 200\epsilon = 0.001, \boldsymbol{\sigma}_{meas}, \boldsymbol{\epsilon}) \\ &= 1 - \mathbb{E}_{\sigma_{meas} \sim p(\sigma_{meas}|\epsilon=0.001, \boldsymbol{\sigma}_{meas}, \boldsymbol{\epsilon})}\{I(\sigma_{meas}| - \infty, 200)\}. \end{aligned} \tag{23}$$

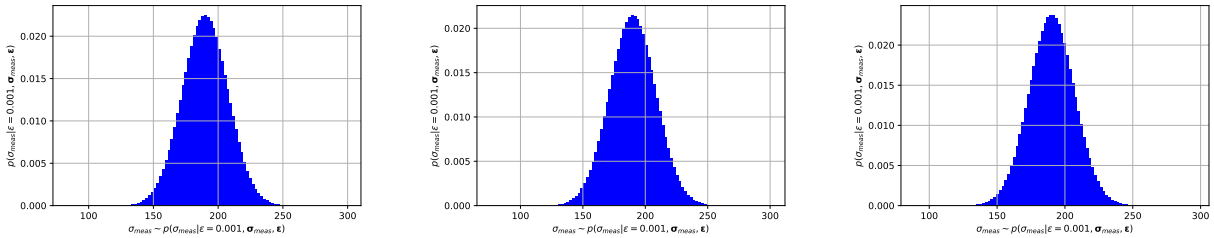
This probability requires samples from the posterior predictive distribution. As such, the sampling procedure to generate posterior predictive samples is as follows:

1. Sample the posterior $\{\boldsymbol{\theta}_i, \eta_i\} \sim p(\boldsymbol{\theta}, \eta|\boldsymbol{\sigma}_{meas}, \boldsymbol{\epsilon})$.
2. Sample the generative model $\sigma_{meas} \sim p(\sigma_{meas}|\sigma_l(\boldsymbol{\theta}_i, \epsilon = 0.001), \eta_i)$.
3. Repeat

This procedure ensures that we can draw samples from the posterior predictive distribution for $\epsilon = 0.001$. As initially the noise η was unknown, a maximum likelihood estimate of $\eta = 17.149$ for $\boldsymbol{\theta} = [190000, 300, 55000]^T$ MPa. Using the MLE noise, the pre-defined noise ($\eta = 18$), and posterior samples using `PyMC3`, the probability of interest was found to be

1. $p(\sigma > 200|\epsilon = 0.001, \boldsymbol{\sigma}_{meas}, \boldsymbol{\epsilon})_{pre-defined, \eta=17.149} = 0.2871$.
2. $p(\sigma > 200|\epsilon = 0.001, \boldsymbol{\sigma}_{meas}, \boldsymbol{\epsilon})_{pre-defined, \eta=18} = 0.2956$.
3. $p(\sigma > 200|\epsilon = 0.001, \boldsymbol{\sigma}_{meas}, \boldsymbol{\epsilon})_{\texttt{PyMC3}} = 0.2835$.

Thus, we can conclude that there is a 30% chance that the measured stress will be greater than 200MPa at a strain of 0.001. In Figure 14, the three empirical posterior predictive distributions used for this question are shown.



(a) $\eta = 17.149$. (b) $\eta = 18$. (c) `PyMC3`.

Figure 14: The empirical posterior predictive distribution where $\epsilon = 0.001$ for three different variations of the posterior predictive distribution.

4 Question 4

My chosen paper was:

- Diamond DH, Heyns PS, Oberholster AJ (2016) Online shaft encoder geometry compensation for arbitrary shaft speed profiles using Bayesian regression. Mech. Syst. Signal Process. 81:402–418

The reason here was simple: I have used the method in this paper in both my Masters and my PhD, and I believe that it is applicable and important to the field. The approach proposed in [1], hereafter referred to as Bayesian Geometry Compensation (`BGC`), is based on cubic-spline polynomial regression implementation with a Bayesian regression posterior distribution update step to encode some prior information into the problem when solving for the regression coefficients. [1] proceeds with an alternate derivation to the process detailed in this write-up, but ultimately the estimation problem is underdetermined and thus Bayesian inference is used to ensure that at least one possible solution is obtained, given some prior knowledge of the problem. I suppose the goal of this question was to go through the Bayesian inference process for shaft profile compensation, to make explicit reference to the steps taken and draw parallels to the content of the MIL course. However, I chose to explore whether the system of equations is indeed underdetermined, and established a procedure which produces a unique solution. I then conducted a comparison to the methodology proposed by [1]. Let's begin.

4.1 System derivation

Consider an application where there are N shaft sections over one rotation, and M shaft rotations. The model of interest, for the shaft rotation $\theta(t)$ is discretised over time into sectional domains. This discretisation is facilitated through cubic splines [1] and is given by a peicewise function of the form

$$\theta(t) = \begin{cases} \theta_1(t) & t_1 \leq x \leq t_2, \\ \theta_2(t) & t_2 \leq x \leq t_3, \\ \vdots & \\ \theta_K(t) & t_{K-1} \leq x \leq t_K, \end{cases}, \tag{24}$$

where K is the number of sections over the whole domain of interest, $\theta_k(t), k = 1, \cdots, K$ are polynomials, the time points t_i refer to the zero-crossing time (i.e., the time where one section ends and another begins), and $t \in [t_1, t_K]$ is the time over the whole domain. In this application, we introduce notation to distinguish between the K polynomials. This notation utilises the problem-specific information that the domain of interest has a fixed number of rotations M and a fixed number of sections per rotation M. Hence, the number of piecewise polynomials K becomes $K = NM$. For the n^{th} section in the m^{th} shaft rotation ($n = 1, \cdots N$, $m = 1 \cdots, M$), the nm^{th} polynomial is given as

$$\theta_{nm}(t) = a_{nm} + b_{nm} \cdot (t - t_{nm}) + c_{nm} \cdot (t - t_{nm})^2 + d_{nm} \cdot (t - t_{nm})^3, \tag{25}$$

where the time over the nm^{th} polynomial is $t \in [t_{nm}, t_{(n+1)m}]$. A naive approach is to assume that θ_{nm} is constant and set to $\frac{2 \cdot \pi}{N}$. In this naive setting, each increment is equispaced over the domain. However, as discussed in [1], this is too naive for practical purposes. Hence, we assume that the domain is broken up into sections and we exploit the physics of the problem to try and solve for the equations. For the considered problem, each piecewise polynomial has 4 unknowns, and a result there are $4NM$ unknowns over the whole domain. Furthermore, let the change in time over one cubic spline index be

$$T_{nm} = t_{(n+1)m} - t_{nm}, \tag{26}$$

where t_{nm} and $t_{(n+1)m}$ are the start and endpoints respectively of the considered spline. T_{nm} represents the time period between two consecutive zero-crossing times. The first and second derivatives of the $\theta_{nm}(t)$ polynomial is

$$\dot{\theta}_{nm}(t) = b_{nm} + 2 \cdot c_{nm} \cdot (t - t_{nm}) + 3 \cdot d_{nm} \cdot (t - t_{nm})^2, \tag{27}$$

$$\ddot{\theta}_{nm}(t) = 2 \cdot c_{nm} + 6 \cdot d_{nm} \cdot (t - t_{nm}), \tag{28}$$

The issue here, is that the conventional approach to applying cubic splines cannot be used, as we do not know any function value information of the shaft rotation. However, I believe that one can utilise domain knowledge to ensure that a linear system of equations is found. Our assumptions are as follows:

1. The shaft speed is initialised at $\theta(0) = 0$ and ends at $\theta(T_{(N+1)M}) = M \cdot 2\pi$.
2. The discretisation is a continuous function through time.
3. Each shaft section distance remains constant for all M rotations.
4. The sum of all N sections add up to one revolution, or $2 \cdot \pi$ radians.

Using these assumptions, let's try to obtain a system of equations that can we can solve. The first, and perhaps most obvious relationship is that the angular rotation between one segment and the next must be equal. This is done to ensure that the angular displacement is consistent between each polynomial. This is given, generally, by

$$\begin{aligned}
&\theta_{nm}(t_{(n+1)m}) = \theta_{(n+1)m}(t_{(n+1)m}) \\
&a_{nm} + b_{nm} \cdot T_{nm} + c_{nm} \cdot T_{nm}^2 + d_{nm} \cdot T_{nm}^3 = a_{(n+1)m} \\
&a_{nm} - a_{(n+1)m} + b_{nm} \cdot T_{nm} + c_{nm} \cdot T_{nm}^2 + d_{nm} \cdot T_{nm}^3 = 0.
\end{aligned} \tag{29}$$

The angular rotation relationship gives $NM - 1$ equations. The second relationship is the angular velocity (shaft speed) consistency relationship. This enforces that the shaft speed from one section to the next must be piecewise continuous. This is given through

$$\begin{aligned}
&\dot{\theta}_{nm}(t_{(n+1)m}) = \dot{\theta}_{(n+1)m}(t_{(n+1)m}) \\
&b_{nm} + 2 \cdot c_{nm} \cdot T_{nm} + 3 \cdot d_{nm} \cdot T_{nm}^2 = b_{(n+1)m} \\
&b_{nm} - b_{(n+1)m} + 2 \cdot c_{nm} \cdot T_{nm} + 3 \cdot d_{nm} \cdot T_{nm}^2 = 0
\end{aligned} \tag{30}$$

The angular speed consistency relationship gives $NM - 1$ equations. The third relationship is the angular acceleration consistency relationship, which requires the angular acceleration from one section to the next to be piecewise continuous. This is given through

$$\begin{aligned}
\ddot{\theta}_{nm}(t_{(n+1)m}) &= \ddot{\theta}_{(n+1)m}(t_{(n+1)m}) \\
2 \cdot c_{nm} + 6 \cdot d_{nm} \cdot T_{nm} &= c_{(n+1)m} \\
2 \cdot c_{nm} - c_{(n+1)m} + 6 \cdot d_{nm} \cdot T_{nm} &= 0
\end{aligned} \tag{31}$$

The angular speed consistency relationship gives $NM - 1$ equations. Before I continue with other relationships, let's take a step back and evaluate what we have done and where we are going. The equations that have been developed detail how to enforce consistency in between the polynomials, but they do not enforce consistency at the start or end points. To ensure that we have starting point consistency, it is enforced that

$$\theta_{1,1}(t_{11}) = 0, \tag{32}$$

Then, borrowing a decision from cubic spline literature, it is enforced that

$$\begin{aligned}
\ddot{\theta}_{11}(t_{11}) &= 0 \\
2 \cdot c_{11} &= 0
\end{aligned} \tag{33}$$

$$\begin{aligned}
\ddot{\theta}_{NM}(t_{(N+1)M}) &= 0 \\
2 \cdot c_{NM} + 6 \cdot d_{NM} T_{NM} &= 0
\end{aligned} \tag{34}$$

which is a natural cubic spline approximation (i.e., the second derivatives at the endpoints are zero). At this point, the process has given us $3NM$ equations that can be used. To finalise the set of equations, I exploit points three and four of the assumptions made. These are used to ensure that the angular sectional increment distance is consistent with the known properties of the system. The first constraint is

$$\begin{aligned}
\int_{t_{nm}}^{t_{(n+1)m}} \dot{\theta}_{nm} dt &= \int_{t_{n(m+1)}}^{t_{(n+1)(m+1)}} \dot{\theta}_{n(m+1)} dt \\
b_{nm} \cdot T_{nm} + c_{nm} \cdot T_{nm}^2 + d_{nm} \cdot T_{nm}^3 &= b_{n(m+1)} \cdot T_{n(m+1)} + c_{n(m+1)} \cdot T_{n(m+1)}^2 + d_{n(m+1)} \cdot T_{n(m+1)}^3
\end{aligned} \tag{35}$$

where this constraint is applied to each of the N sections $M-1$ times, which produces $MN-N$ equations. To resolve the final set of constraints, I chose to enforce that the amount of distance travelled in the first rotation be proportional to the time ratio for one rotation. This is achieved through

$$\begin{aligned} \int_{t_{nm}}^{t_{(n+1)m}} \dot{\theta}_{nm} dt &= 2 \cdot \pi \cdot \frac{T_{nm}}{\sum_{n=1}^{N} T_{nm}} \\ b_{nm} \cdot T_{nm} + c_{nm} \cdot T_{nm}^2 + d_{nm} \cdot T_{nm}^3 &= 2 \cdot \pi \cdot \frac{T_{nm}}{\sum_{n=1}^{N} T_{nm}}. \end{aligned} \tag{36}$$

where this equation need only be applied to each of the N sections (i.e., N times) in one rotation ($m = 1$) as the consistent angular increment constraint ensures that all other sections are constrained in the same manner. Hence, we should be able to resolve a system of equations that is uniquely solvable.

To estimate the shaft speed, the total angular increment $\theta_n = b_{nm} \cdot T_{nm} + c_{nm} \cdot T_{nm}^2 + d_{nm} \cdot T_{nm}^3$ is used through

$$\omega_n = \frac{\theta_n}{T_{nm}}. \tag{37}$$

I realise that this notation may be confusing, as all that has happened is that I have dropped the subscript m. However, [1] uses this notation, and for consistency this notation is preserved. To quantify how this approach performs, three investigations are performed. The first is the simple process of estimating the shaft speed for a system with constant operating conditions. In the second investigation, a comparison is conducted on the examples from [1]. Finally, the performance of the `BGC` approach is compared to the approach with a unique solution. Henceforth, the unique solution case will be referred to as the `Uniquely solvable` approach.

4.2 Investigation 1

Consider an application where the pulse points for the m^{th} rotation are given at

$$t_{pulse}(m) = [m, m + \frac{1}{4}, m + \frac{3}{4}, 2 \cdot m]. \tag{38}$$

In this setting, the pulse points describe a rotating system where the shaft speed is constant as they are immutable and hence there is no change in velocity. Such a system is seen in Figure15(a). The solution for the `Uniquely solvable` process is seen in Figure 15(a). It is clear that under an such a rotating system, the `Uniquely solvable` process can estimate the shaft speed exactly.

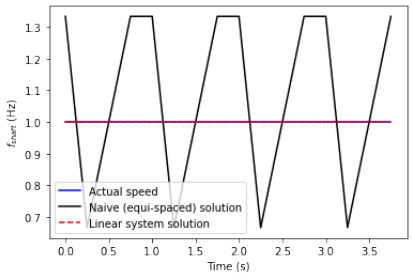


Figure 15: The results for the naive speed estimation problem (equispaced increments) versus the unique solution case.

4.3 Investigation 2

For the second investigation, the `BGC` methodology was compared to the `Uniquely solvable` process. The analysis procedure followed in [1] is used to facilitate this comparison. This requires two steps, *i)* generation of the underlying pulse increments, and *ii)* a set of trail shaft speed profiles. The underlying pulse increments is retained through each of the trail speed profiles, to simulate an asset under different operating conditions. As the shaft speed profiles are parametrised with numerical functions, it is trivial to obtain the zero-crossing times through numerical integration.

In Figure 16 the angular increment sizes for the problem considered are shown. The increment size was obtained by sampling $N-1$ points from a Gaussian distribution

$$\theta_n \sim \mathcal{N}(\frac{2\cdot\pi}{N}, \frac{1}{N^2}), \tag{39}$$

and then using the final point to ensure that $\sum_{n=1}^{N}\theta_n = 2\cdot\pi$. As such, the final increment size is found through

$$\theta_N = 2\cdot\pi - \sum_{n}^{N-1}\theta_n. \tag{40}$$

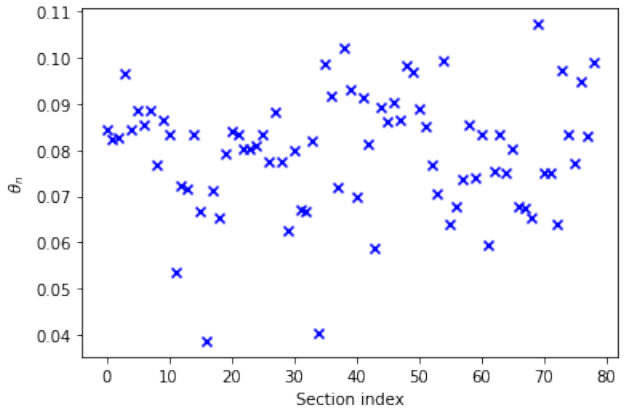


Figure 16: The actual angular increment sizes used for the speed estimation comparison problem.

For the second step of the comparison function, four trial speed profiles are considered. These are given through

$$\omega_1(t) = 100 + 10\cdot t, \tag{41}$$

$$\omega_2(t) = 100 + 5\cdot\sin\left(2\cdot\pi\cdot 30\cdot t\right), \tag{42}$$

$$\omega_3(t) = 100 + 10\cdot t + 5\cdot\sin\left(2\cdot\pi\cdot 30\cdot t\right), \tag{43}$$

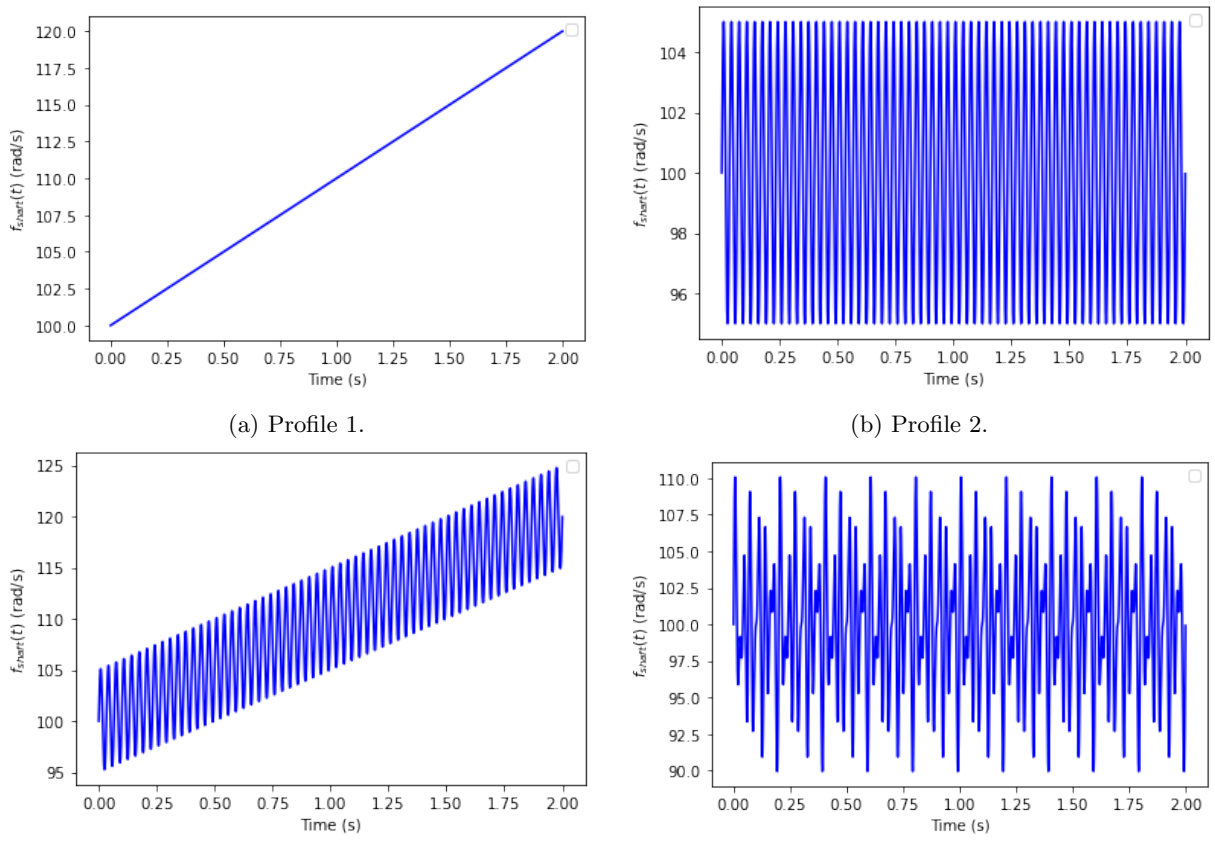
$$\omega_4(t) = 100 + 5\cdot\sin\left(2\cdot\pi\cdot 30\cdot t\right) + 4\cdot\sin\left(2\cdot\pi\cdot 45\cdot t\right) + 2\cdot\sin\left(2\cdot\pi\cdot 20\cdot t\right). \tag{44}$$

Finally, to compare between the `BGC` results to the `Uniquely solvable` results, the sum of squared errors is used. I chose to use the sum as the increment size is already small, and hence the mean squared error is not representative enough of the difference in performance. The squared error between the actual angular increments and the estimated increments is given through

$$E = \sum_{n=1}^{N}\left(\theta_n^{\text{actual}} - \theta_n^{\text{approximated}}\right)^2. \tag{45}$$

In Figure 17 the true shaft profiles for the considered problem are shown. In Figure 18 the speed profiles found through Equation (37) and the true angular increments is shown. Notice how the speed plotting defined through Equation (37) produces some noise. In Figure 19 the shaft profiles found through the `Uniquely solvable` procedure are shown. In Figure 20 the shaft profiles found through the `BGC` procedure are shown. In the comparison between Figures 19 and 20, it is clear that both procedures cannot accurately estimate the true underlying shaft speed. There are obvious deviations between the true and estimated shaft profiles. This is attributed to the estimation error that is present in both procedures, and the nature of the estimation comparison process, where the latter relies on Equation (37) to visualise the estimated shaft speed. Perhaps, for visualisation purposes, it would have been better to use the cubic spline polynomials as all of the unknowns were solved for.

In Figure 19, it is clear that in the first portion, for all of the estimated speed profiles, is constant. This is the crucial flaw in the `Uniquely solvable` solution and is a direct consequence of the assumption made in Equation



(a) Profile 1.

(b) Profile 2.

(c) Profile 3.

(d) Profile 4.

Figure 17: The actual speed profiles for the considered problem.

Table 3: The total squared error between the actual section angular increment size and the estimated size.

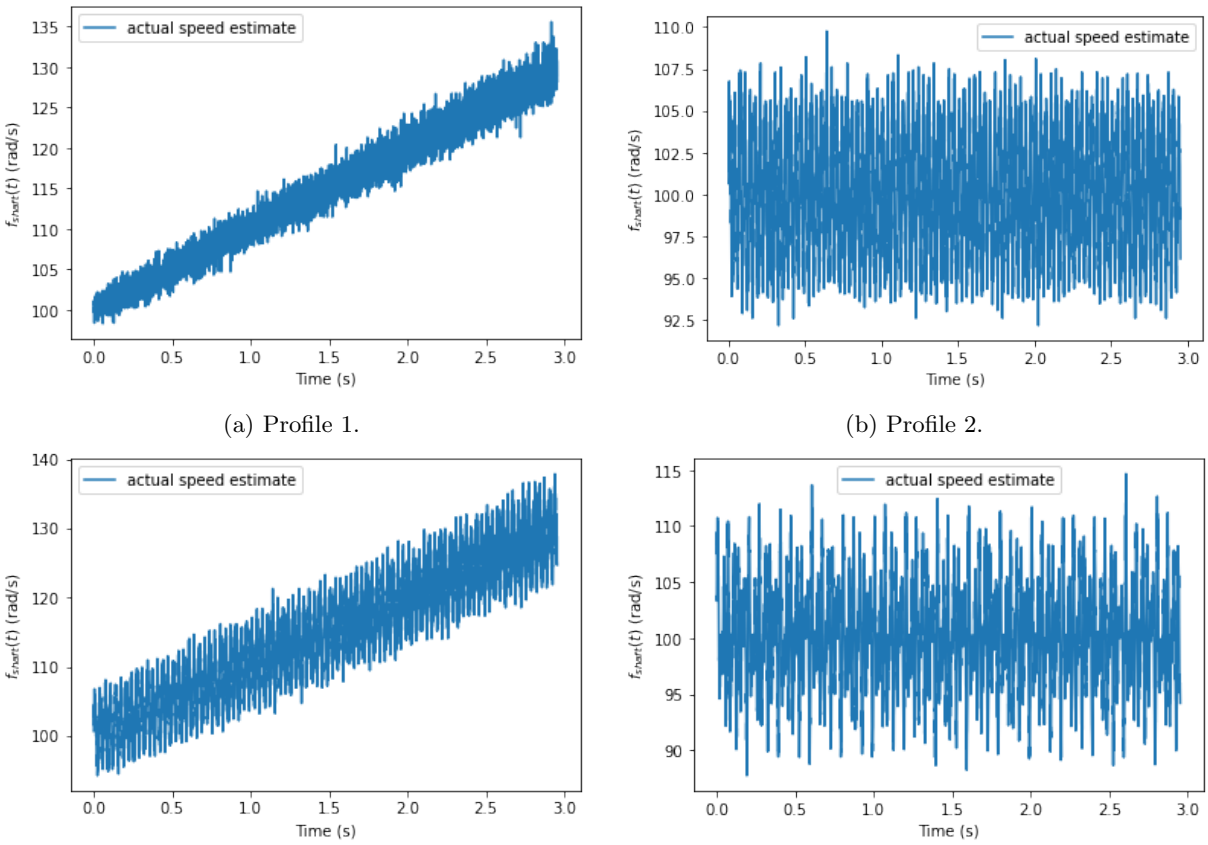
	Speed profile 1	Speed profile 2	Speed profile 3	Speed profile 4
My implementation	0.8247	0.7658	0.8969	0.7561
Diamond implementation	0.81220	0.7267	0.8137	0.7190

(36). To elaborate, consider the relationship between Equations (36) and (37). The shaft speed, for the first shaft rotation (i.e., $m = 1$) , is given through

$$\begin{aligned} \omega_n &= \frac{2 \cdot \pi \cdot T_{nm}}{T_{nm} \cdot \sum_{n=1}^{N} T_{nm}} \\ &= \frac{2 \cdot \pi}{\sum_{n=1}^{N} T_{nm}}, \end{aligned} \tag{46}$$

which makes the flaws of the assumption required for Equation (36) clear. To the detriment of the `Uniquely solvable` solution, the assumption in Equation (36) enforces that . One may wonder why the angular sectional distance for the remaining sections do not resemble the first rotation. This, at the time of writing this part of the report, escapes me a little.

In Figure 16 the true and approximated angular increments are shown. It is clear that for all of the considered speed profiles that the `Uniquely solvable` solution is not as accurate as `BGC`. To further emphasise this point, consider the sum-of-square-errors shown in Table 3 for each speed profile. It is clear that the `Uniquely solvable` solution always has the greater error.



(a) Profile 1.

(b) Profile 2.

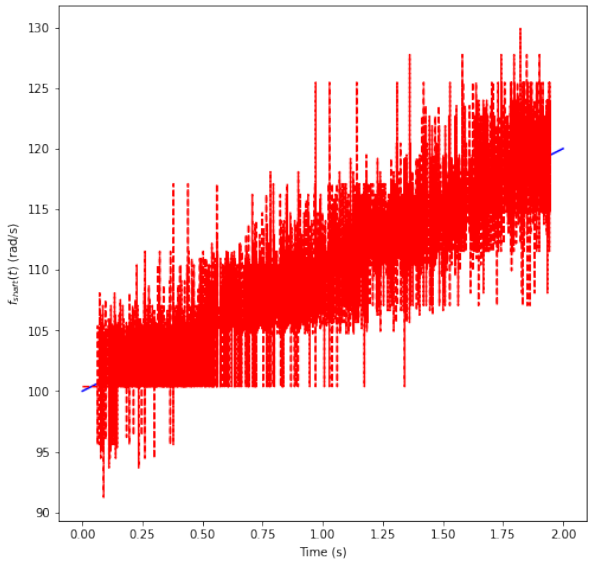
(c) Profile 3.

(d) Profile 4.

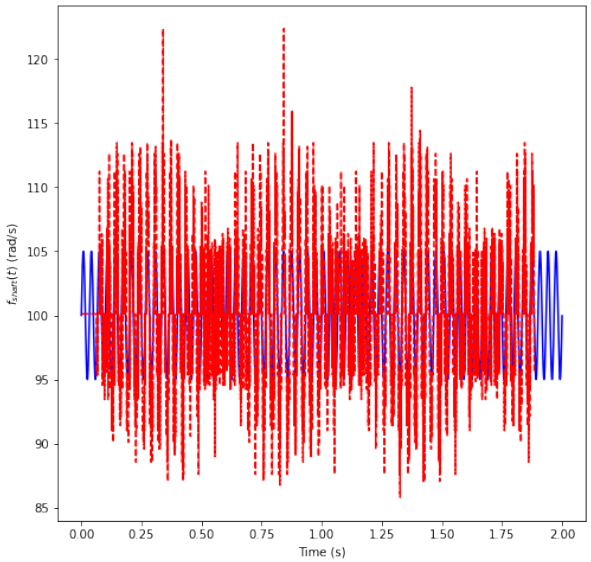
Figure 18: The actual speed profiles for the considered problem when using Equation (37) and the true angular increments.

4.4 Investigation 3

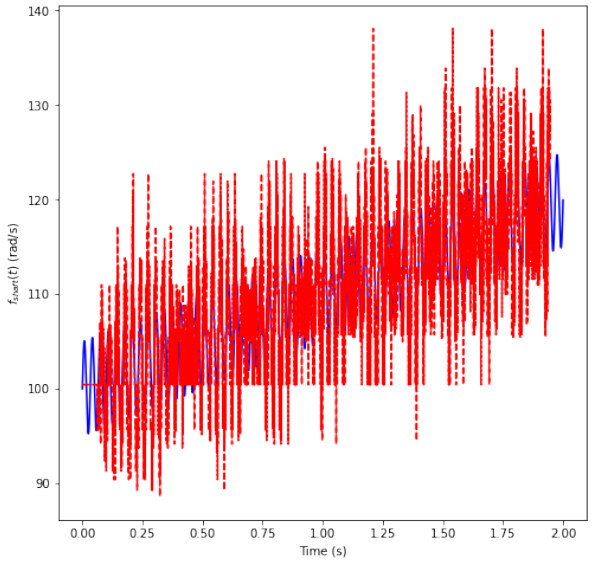
In the final investigation, a tachometer signal from the C-AIM gearbox dataset was used to compare the two methods. Figure 22 shows the results for the two estimation procedures. It is clear that both the methods perform similarly, but the `uniquely solvable` solution still produces a linear angular velocity in the first shaft rotation section. Hence, it is blatantly clear `BGC` is superior here, and the assumption responsible for Equation (36) must be changed. Perhaps extending the assumption to a constant acceleration based on the change in time, or something of that manner, will be beneficial. Unfortunately I could not come up with any ideas, but it was fun to undergo this analysis process.



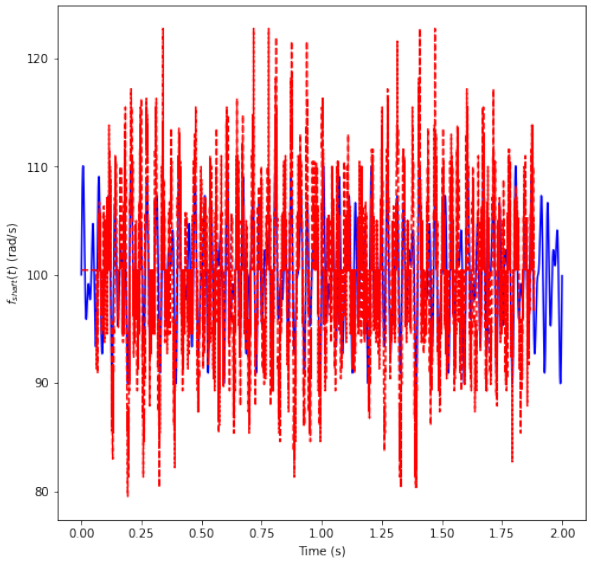
(a) Estimated profile 1.



(b) Estimated profile 2.

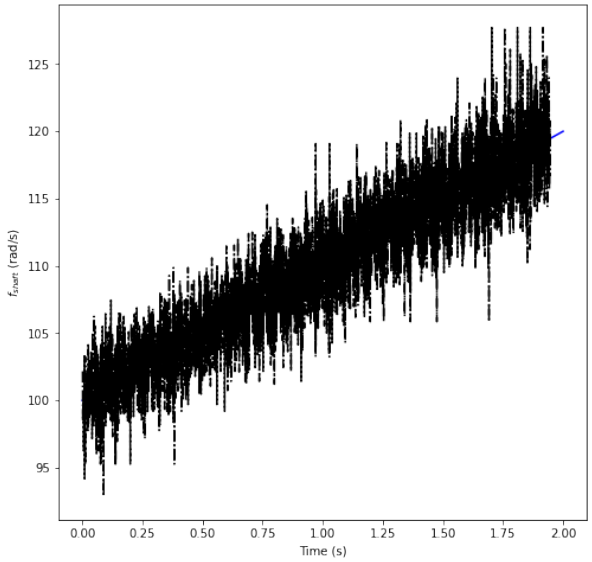


(c) Estimated profile 3.

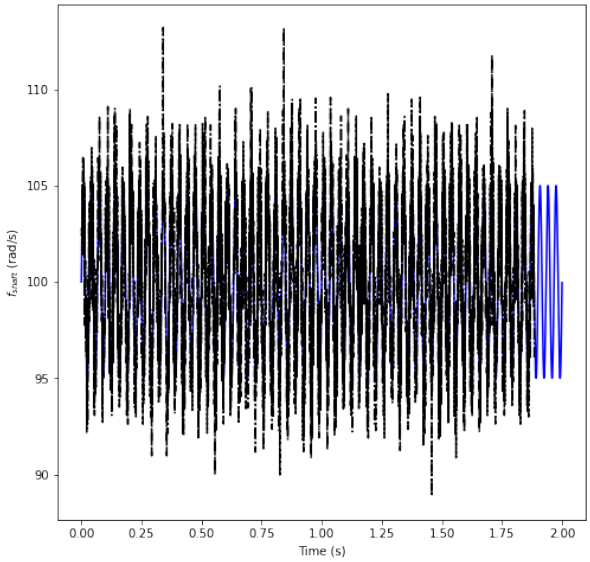


(d) Estimated profile 4.

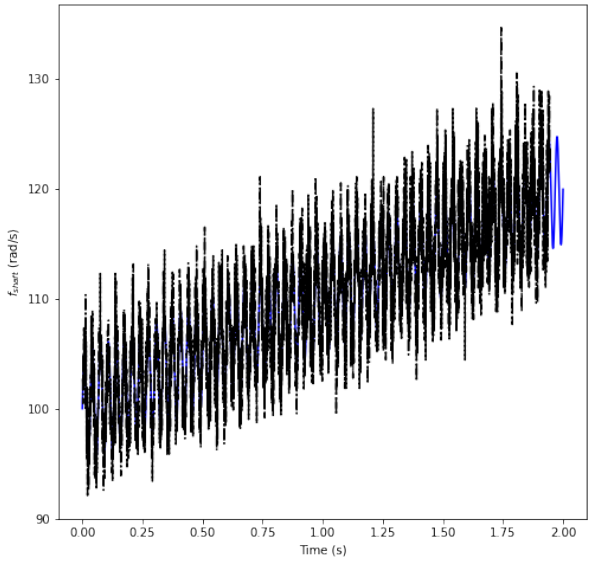
Figure 19: The estimated shaft profiles for the process I attempted. Notice how each profile has a section in the beginning that is constant. This is a result of the assumption in Equation (36).



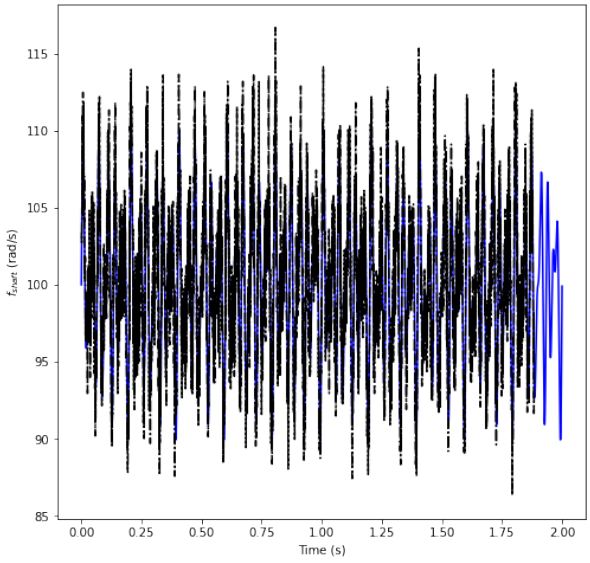
(a) Estimated profile 1.



(b) Estimated profile 2.

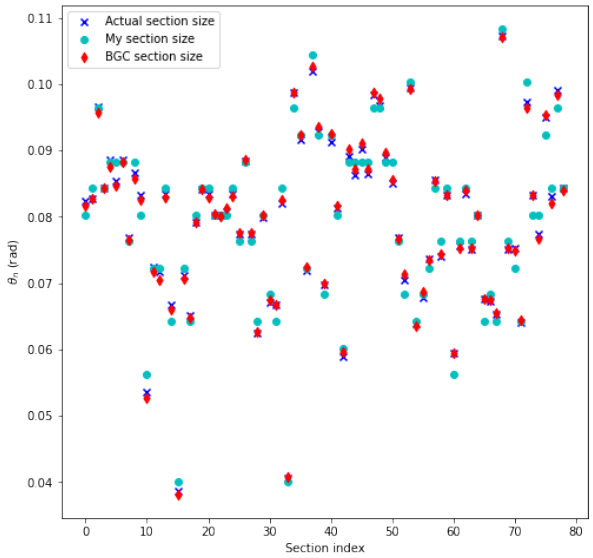


(c) Estimated profile 3.

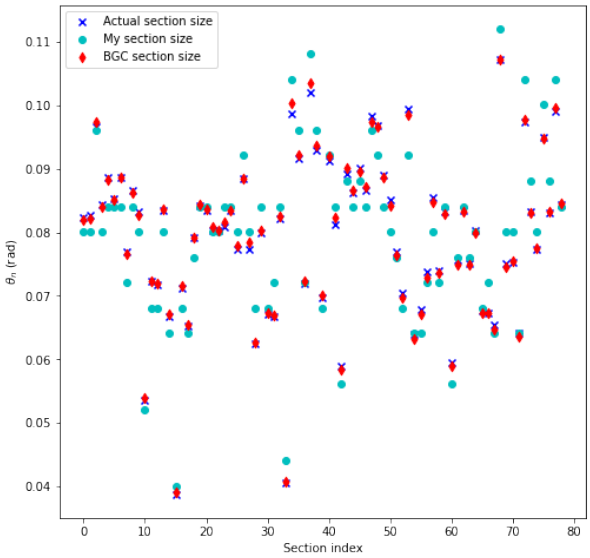


(d) Estimated profile 4.

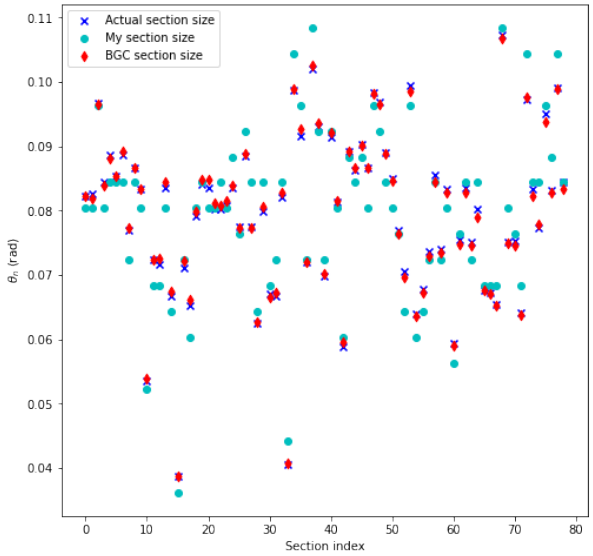
Figure 20: The estimated shaft profiles for the process presented in [1].



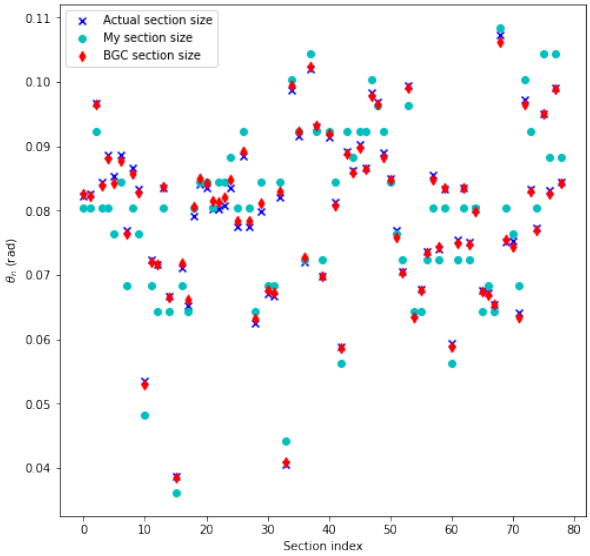
(a) Profile 1 increments.



(b) Profile 2 increments.

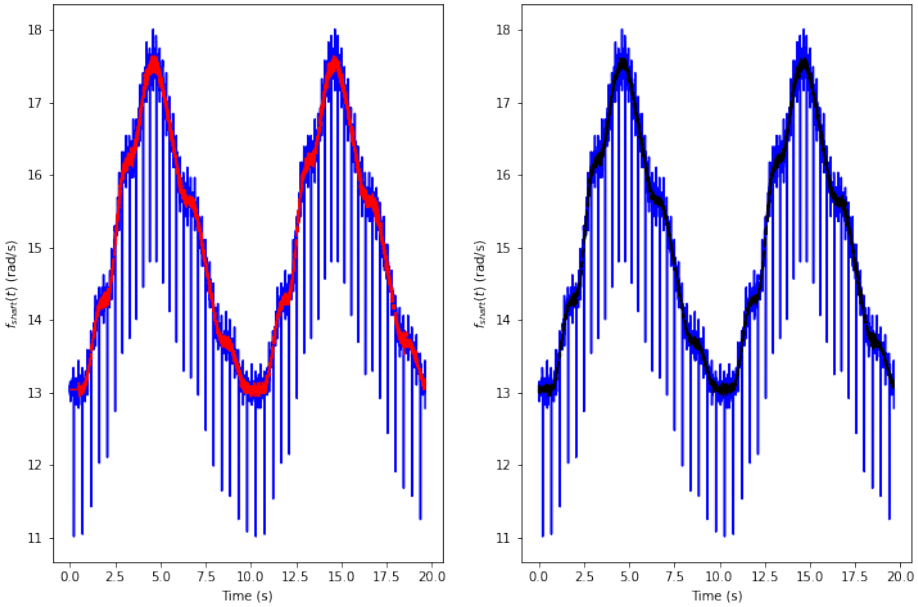


(c) Profile 3 increments.

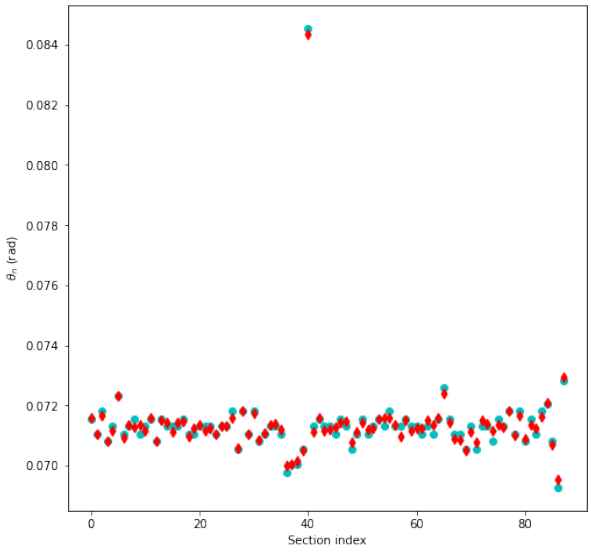


(d) Profile 4 increments.

Figure 21: The section angular increment sizes for the four profiles analysed.



(a) Estimated profile through the `uniquely solvable` (left) and `BGC` (right) procedures.



(b) Estimated angular sectional increments.

Figure 22: The estimated speed profiles and angular increments for the `uniquely solvable` and `BGC` procedures. Note that the naive speed profile, obtained using the assumption that each of the sections are equi-spaced, is plotted in blue in both of the sub-plots in (a).

References

[1] D. H. Diamond, P. S. Heyns, and A. J. Oberholster, "Online shaft encoder geometry compensation for arbitrary shaft speed profiles using Bayesian regression," *Mechanical Systems and Signal Processing*, vol. 81, pp. 402–418, 2016, ISSN: 10961216. DOI: 10.1016/j.ymssp.2016.02.060. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0888327016001114.