Journal of Physics: Conference Series

Inverse design under uncertainty with surrogate models

To cite this article: D B Walton *et al* 2024 *J. Phys.: Conf. Ser.* **2647** 212008

Inverse design under uncertainty with surrogate models

D B Walton, C A Featherston, D Kennedy and A Kundu

School of Engineering, Cardiff University, Queen's Buildings, The Parade, Cardiff, CF24 3AA

E-mail: WaltonDB@cardiff.ac.uk, FeatherstonCA@cardiff.ac.uk, KennedyD@cardiff.ac.uk, KunduA2@cardiff.ac.uk

Abstract. In the drive towards net zero the aerospace industry is motivated to develop more efficient aerostructures that can accommodate the next generation of propulsion systems that fall outside of the well understood types that are currently in use. The lack of established standards for such designs means that engineers are faced with an increased level of uncertainty in their design choices before any prototypes are built. Machine learning models are becoming a popular tool for expediting the development of novel designs due to their ability to explore and predict the optimal parameters of large design spaces. It is also possible to quantify and introduce uncertainty into particular models so that practitioners can be made aware of the potential variation in their realised designs. In this paper *Gaussian Process* surrogate models of the performance metrics of the early-stage design of an aircraft wing are created to optimize a subset of design parameters based on some prescribed limits of the intended real system response. This defines the inverse design problem that is solved using *Markov Chain Monte Carlo* sampling. The approach taken requires novel formulation of a Bayesian machine learning framework. In particular, the work investigates the formation of likelihood functions that are flexible given inputs of different scales, can perform marginalisation of stochastic parameters, account for uncertainty in the surrogate model, and optimise the parameters given more than one constraint. A case study is presented in this paper that highlights both a successful implementation of the framework along with a limitation. It is found that the optimization is sensitive to changes in the variances of the likelihoods such that it can be used as a weight to direct the optimization towards a quantity of interest, therefore adjustment of this parameter is used to balance the optimization.

1. Introduction

Finite element models (FEM) provide a robust but expensive way to analyse structural responses and are now widely used in the process of design. From a design optimisation perspective the required number of FEM evaluations necessary to straddle and home in on some design parameter optimum can be limiting due to computational expense and model complexity, and so introduction of surrogate models is becoming more prevalent, particularly in cases where there are multiple parameters being optimised, high fidelity models are needed, and if there are stochastic parameters [1][2]. In [3] an example of surrogate model assisted optimisation under uncertainty can be found where a Gaussian process regression surrogate, trained on a low fidelity FEM model, is introduced into the workflow to optimise the drag coefficient of an airfoil given some uncertainty on the in flight angle of attack, and successfully reduces the expense of evaluating high fidelity FEM models in the process. However, the optimisation process is iterative and an unknown number of evaluations are needed before convergence is reached. In this paper a data driven surrogate model is obtained from an FEM model, and optimisation is performed without any further evaluations of the model.

This study considers the case of an aircraft wing in early stage design. The considered parameters θ are the jig shape and wing stiffness, where the jig twists $\theta_{1,2}$ in figure 1 are the parameters being optimised and describe the shape of the wing (jig shape) with no loads applied, with the aim of letting aerodynamic forces deform the wing in flight to achieve the desired angle of attack α. This deformation also depends on the bending and torsional stiffness (EI, GJ) of the wing at points along the span $\theta_{3,\dots,10}$. The stiffness at these points is considered to be uncertain and is therefore a random variable able to take on any value within some predefined distribution during optimisation. Furthermore, the in flight load response (shear force, bending moment, and torque) along the span is dependent on the parameter values. Therefore, the design optimisation task in this work is to find a distribution over $\theta_{1,2}$ that satisfies some limit on the lift over drag ratio (L/D) while simultaneously satisfying a restraint on the in flight force response. For the purpose of demonstrating the framework used here, the shear force (F) at the root of the wing is targeted, and the L/D is calculated relative to a fixed baseline α when the aircraft is cruising. This problem contains more complexity than [3] in that there are more stochastic parameters, the aerodynamic model is coupled with the structural model, and optimisation of the important parameters is based on two quantities of interest (F and L/D).

The framework, based on work by [4], requires generating data from the FEM analysis of the wing at a fixed number of design of experiments (DOE) points that uniformly cover the design space. The data is used to train two separate Gaussian process surrogate emulators as mappings $f_A : \theta \mapsto L/D$ and $f_L : \theta \mapsto F$. Bayesian inference [5] and Markov Chain Monte Carlo (MCMC) [6] is used to optimise θ, where maps are evaluated in the likelihood function and points are accepted based on the posterior probability of them meeting the given constraints equation 1

$$\pi(\theta|f, \mathcal{D}, \mathcal{M}, \tau) \sim p(\theta|\mathcal{M})\mathcal{L}(f(\theta)|\mathcal{M}, \tau) \tag{1}$$

where $\mathcal{D}$, $\mathcal{M}$, and τ are the data, model, and targets respectively. $\pi(.)$, $p(.)$, $\mathcal{L}(.)$ are the posterior, prior, and likelihood of Bayes theorem.

2. Statistical Methods

2.1. Gaussian Process Emulator

The coupled CFD and CSM model is emulated with two separate Gaussian possesses [7] so they can be queried separately at new design points during MCMC sampling. This method gives an estimate of the output as a normally distributed random variable and can be expressed as follows. The emulator is a map f from the multi-dimensional input space to the output

$$f : \mathbb{R}^d \mapsto \mathbb{R} \tag{2}$$

$$f(\theta|\mathcal{D}) \sim \mathcal{GP}(M(\theta), K(\theta, \theta')) \tag{3}$$

$$\mathcal{GP}(M(\theta_i), K(\theta, \theta_i')) \sim \mathcal{N}(\mu_i, \sigma_i^2) \tag{4}$$

2.2. Differential Evolution Markov Chain

An adaptive Metropolis Hastings algorithm is used as the MCMC sampler due to the complexity of finding the necessary gradients for the state of the art Hamiltonian Monte Carlo based sampling algorithms such as the *No U-turn Sampler* [8]. The *Differential Evolution Markov*

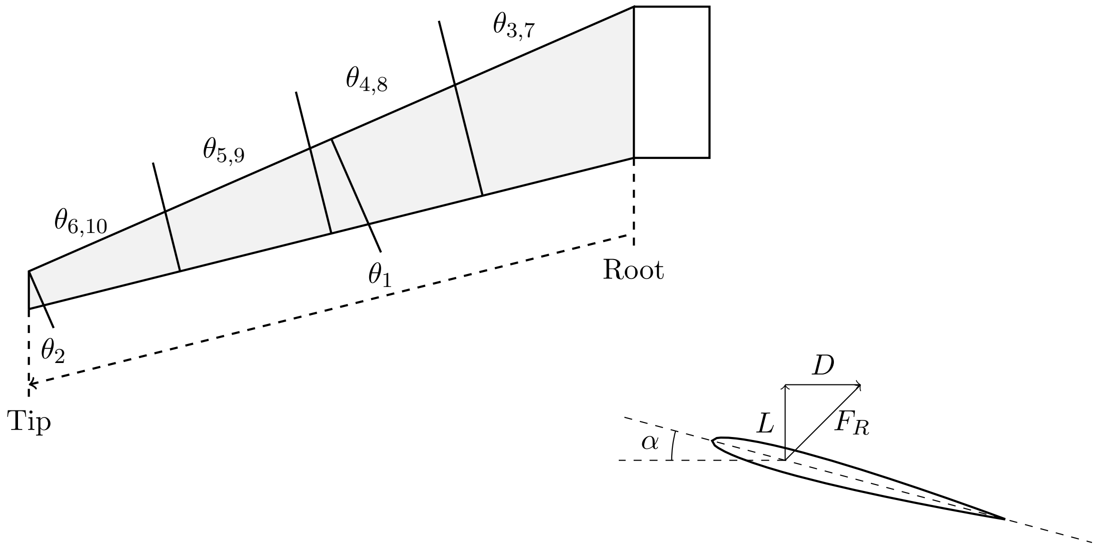


Figure 1: Figure shows approximate location of design parameters on aircraft wing. $\theta_{1,2}$ are located at a point midway between the root and the tip, and at the tip itself. $\theta_{3,...,10}$ are parameter values between the the root, three separating lines, and the tip. In the bottom right hand corner is a standard depiction of an airfoil with applied aerodynamic forces (L, D, F_R) and angle of attack α for reference when reading the text.

Chain algorithm [9] uses 5 to propose the next step in the chain and they are accepted or rejected using the usual Metropolis ratio

$$x^* = x_i + \gamma(x_{R1} - x_{R2}) + e \tag{5}$$

where $(x_{R1} - x_{R2})$ is a differential term between two random previous points in the chain. The γ term is a scaling factor that can be tuned for a given number of steps to balance exploration and exploitation of the parameter space. e is a noise term that is drawn from a random distribution (usually normal) centered around x_i, its standard deviation can also be tuned. It is clear that this stepping method means that the Markov chains are not strictly memory less, but the stochastic differential term and noise term can ensure that the auto-covariance of x_i tends to zero as the chain evolves.

2.3. Likelihood Functions

Two types of optimisation are presented here, constrained and bound, therefore, the likelihood functions take on different forms. The constrained optimisation requires a target τ for which the likelihood is in the form of a standard normal distribution PDF

$$\mathcal{L}(f(\theta) = \tau) = \frac{1}{\sigma\sqrt{2\pi}} \exp \frac{-((\tau - f(\theta)) - \mu)^2}{2\sigma^2} \tag{6}$$

Taking the natural log of equation 6 and assuming independence gives the final constrained likelihood $\mathcal{L}_c$ to be maximised by the following form, where the subscripts L and A refer to the aerodynamic and load responses respectively

$$\ln \mathcal{L} = \ln \mathcal{L}_L + \ln \mathcal{L}_A \tag{7}$$

For the bound optimisation the standard normal CDF in equation 8 applies an upper bound, and $1 - CDF$ is used to apply a lower bound (equation 9), where $t = (\tau - f(\theta))$. The final joint likelihood is also calculated using equation 7.

$$\mathcal{L}(f(\theta) \leq \tau) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{f(\theta)} \exp \frac{-(t-\mu)^2}{2\sigma^2} dt \tag{8}$$

$$\mathcal{L}(f(\theta) \geq \tau) = 1 - \left\{ \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{f(\theta)} \exp \frac{-(t-\mu)^2}{2\sigma^2} dt \right\} \tag{9}$$

2.3.1. Marginalisation In order to evaluate the likelihoods as a function of just the important parameters $\theta_{1,2}$ it is necessary to marginalise out the effect the stiffness parameters have on $f(\theta_i)$ at each step in the chain. This is done with a precalculated $n \times d_s$ Latin Hypercube sample S^{LHS}, and an $n \times d_t$ matrix T^{θ_i} that has $T_n^{\theta_i} = \theta_i, \forall n \in \{1, 2, \ldots, n\}$. Concatenation then gives $\hat{\theta} = [T^{\theta_i}, S^{LHS}]$ which is the $n \times d$ argument of $f(\theta_i)$ in the chain that produces an $n \times 1$ dimensional output $\mathbf{y}_i$. The final marginalisation step requires taking the mean of $\mathbf{y}_i$ thus giving the approximated $f(\theta_{1,2})$

$$f(\theta_{1,2}) \approx \frac{1}{n} \sum_{n=1}^{n} f(\hat{\theta}) \tag{10}$$

2.3.2. Normalisation To limit bias in the likelihood function induced by the scale of the variance between the two outputs, a normalisation step before evaluation but after $f(\theta_{1,2})_i$ is obtained in the Markov chain is used giving $\hat{f}(\theta_{1,2})$. This ensures that both variances are ≈ 1 while retaining the relationship between points. The normalisation is of a standard score form, except the sample covariance between the outputs Σ is used in the denominator of equation 11

$$\hat{f}(\theta_{1,2}) = (f(\theta_{1,2}) - \mu_{\mathcal{D}})((\Sigma_{\mathcal{D}})^{-0.5})^{-1} \tag{11}$$

where $\mu_{\mathcal{D}}$ is the vector containing the mean of each output dimension. Both $\mu_{\mathcal{D}}$ and $\Sigma_{\mathcal{D}}$ are pre-calculated from the data before sampling begins, along with the normalised target values, where $f(.)$ in equation 11 is replaced with $[\tau_L, \tau_A]$.

3. Results and Discussion

The results presented here are for an extreme case when trying to optimise $\theta_{1,2}$ for a region in the output space that is hard to reach in the bounded case, due to most of the space falling outside the bound. It is also a case where the bound in one dimension is more restrictive than the other. Figure 2 shows the bounds of the entire output space as a dotted line before marginalisation of the stiffness parameters. Inside this domain are samples from the subspace of reachable points after marginalisation. The dashed lines show the bounds extending from the point 'x' which is the target for the constrained case also. The bound target puts a minimum on the L/D and a maximum on F. It is clear that given the subspace there is a relatively small area of acceptable points that meet the target restrictions. Considering the non-marginalised full space, the proportion of reachable points in the L/D dimension relative to the unreachable points is greater than those in the F dimension., which gives rise to the possibility of the optimisation favouring one dimension over the other.

The raw data has magnitudes of $\times 10^{-2}$ and $\times 10^{5}$ in the L/D and F dimensions respectively which also gives rise to an imbalance in the likelihood functions. Therefore, normalisation of the data and targets is necessary before the likelihoods are evaluated, see section 2.3.2. Figures 3a, 3b, 4a, and 4b show the log likelihoods evaluated across the subspace sample points before normalisation. It can be seen that there is a difference in variance between the two dimensions

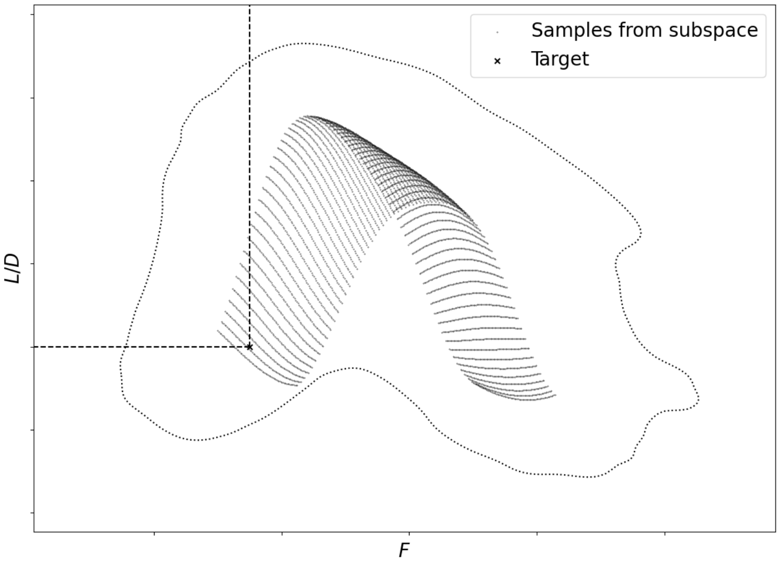
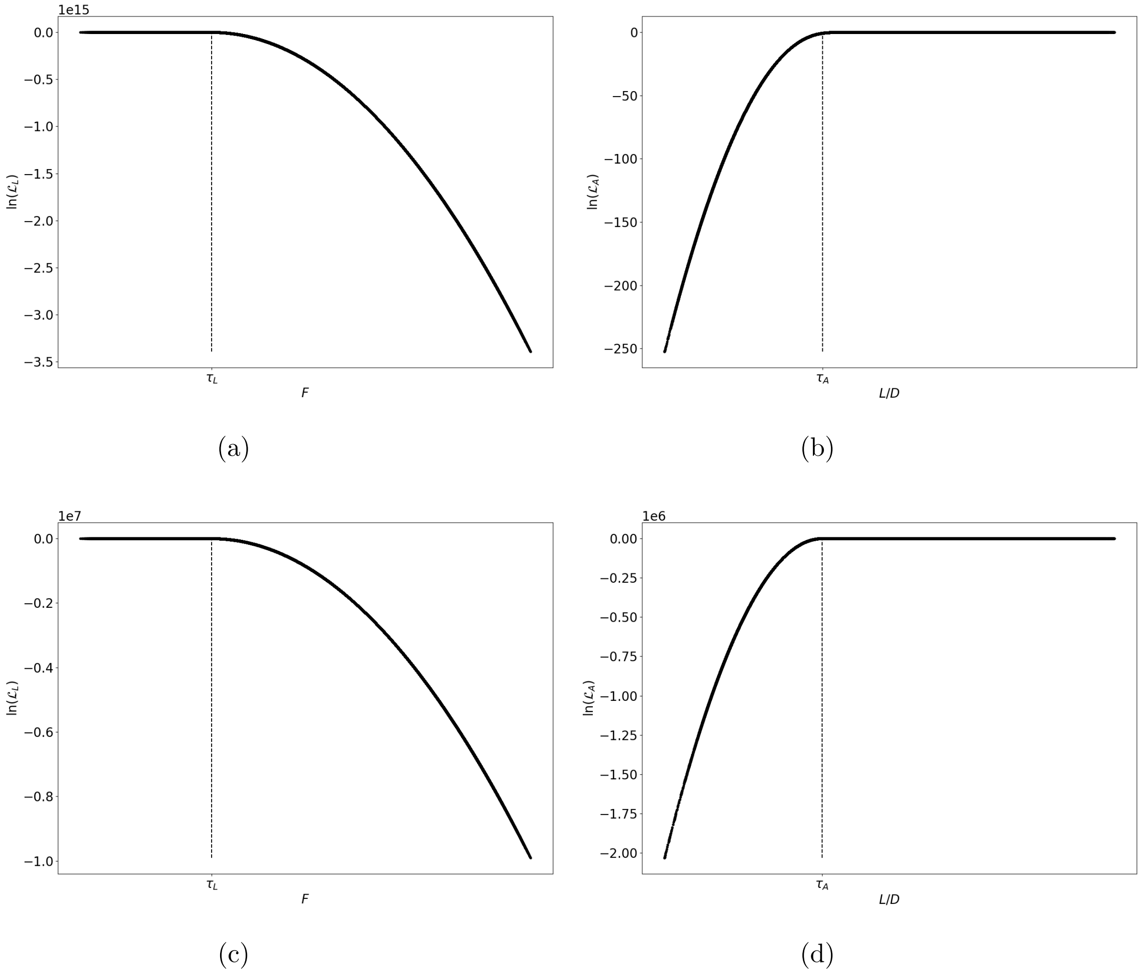


Figure 2: Figure shows the outline of the full output space with dotted line. Inside this space are samples from the subspace that exists after marginalising out the stiffness parameters. The 'x' and the dashed lines indicate the location of the constrained target point and the target area bounds.

of order $\times 10^{15}$. Given the additive nature of equation 7, this causes the algorithm to give preference to one output. In figures 3c, 3d, 4c, and 4d it can be seen that after normalisation of the data the difference in variance is greatly reduced and they are of similar scale while the shapes of the functions are preserved. It is also clear from their shapes that points further from the target are given likelihoods that approach a minimum, while points that are closer, or exceed the target, tend to a maximum. The bell shape of a normal PDF in the constrained likelihoods can clearly be seen in figures 4c and 4d, however the sigmoid shape of the normal CDF is not observed due to $\ln \mathcal{L}$ tending to $-inf$ as points move further away from the target.

To evaluate the results of the optimisation, the posterior distribution over $f(\theta_{1,2})$ given the accepted points in the Markov chains $\pi(f(\theta_{1,2})|\mathcal{M}, \tau)$ is needed. These points are re-evaluated using equation 10 together with their log likelihoods and plotted in figures 5a and 5b. For the constrained case in figure 5b we observe a cluster of points around the target which is almost Gaussian distributed. It can be seen that the posterior distribution does not exceed the bounds of the marginalised subspace and that the points of maximum log likelihood are located on or close to the target, indicating that the optimisation was successful and our confidence in an accepted point can be quantified by its likelihood or by fitting a kernel density estimation to the data. In the case of the bounded optimisation in figure 5a, the posterior distribution is not within the target area. Notably, the target on L/D is completely satisfied as all points are above the target, but the bound on F is not satisfied. This indicates a limitation in the algorithm, and suggests that it could be favouring optimisation of L/D given τ_A. In addition to the observed locations of the accepted points, it can be seen in the colour bar that none of the points have the theoretical maximum log likelihood of zero, also indicating that there is a need for reformulation of the likelihood so that the MCMC sampling can find and converge around



(a)

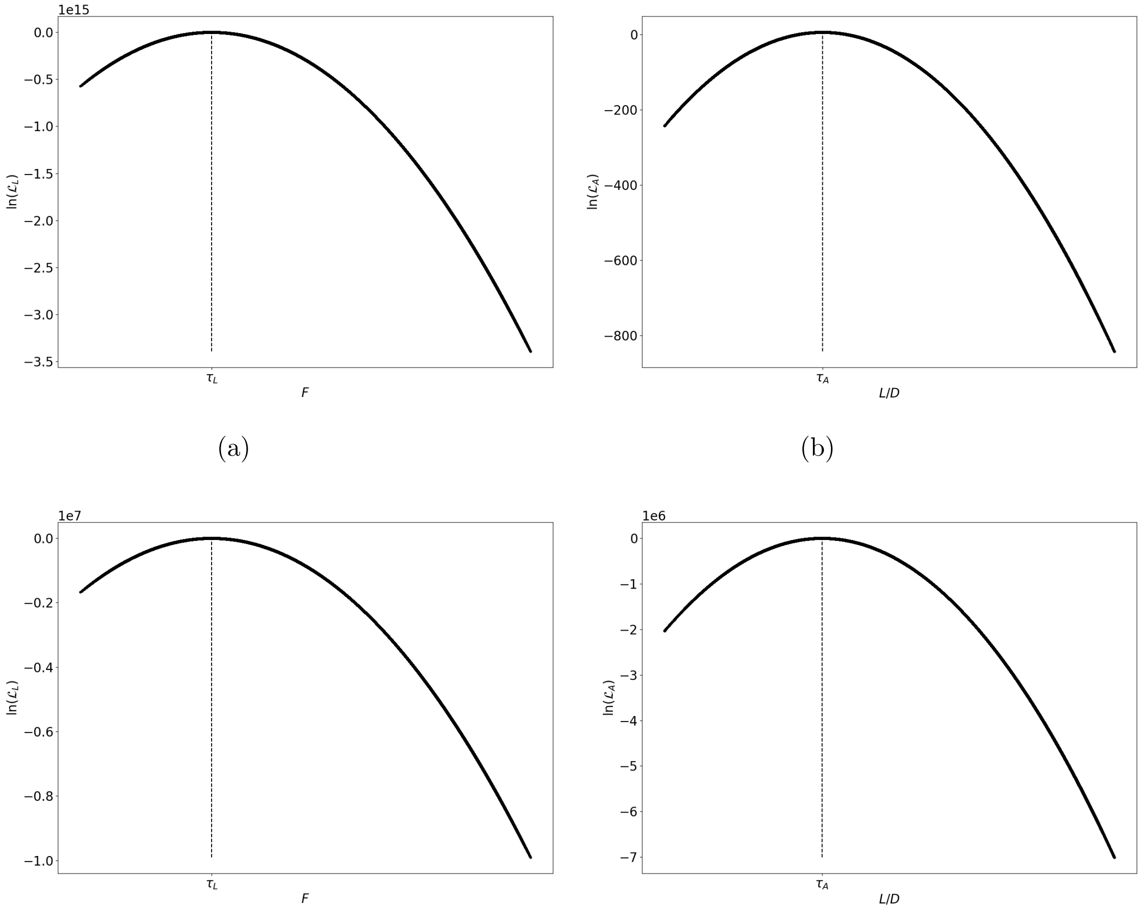
(b)

(c)

(d)

Figure 3: Plot of likelihood function for the bounded case. 3a and 3b show the likelihood evaluated *before* normalisation of the data. 3c and 3d show the likelihood evaluated *after* normalisation of the data. The dotted line shows the location of the upper bound of F in 3a and 3c, with the lower bound of L/D in 3b and 3d.

the maximum posterior given both bounds.



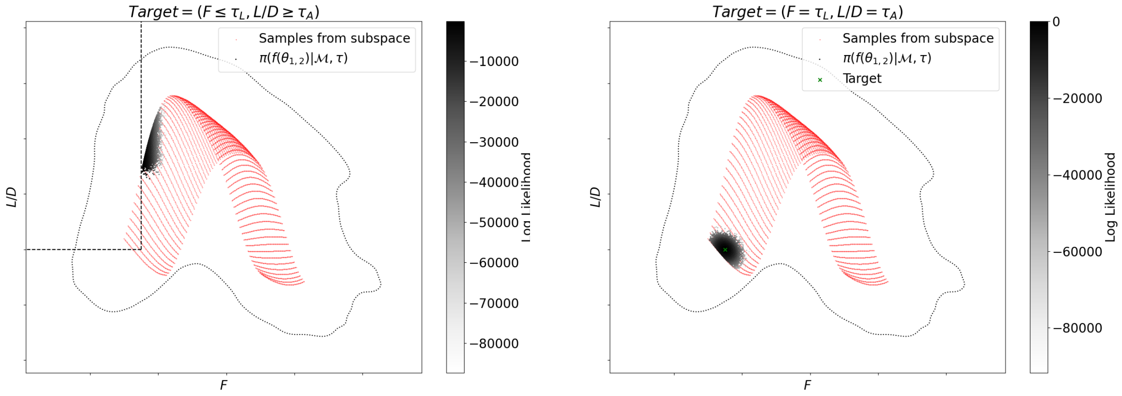
(a)

(b)

(c)

(d)

Figure 4: Plot of likelihood function for the constrained case. 4a and 4b show the likelihood evaluated *before* normalisation of the data. 4c and 4d show the likelihood evaluated *after* normalisation of the data. The dotted lines show the location of the targeted point in both the F and L/D dimensions.



(a) (b)

Figure 5: Results of optimisation. Plotted over the space and subspace from figure 2 is the distribution $f(\theta_{1,2})$ evaluated at the accepted points in the Markov chain. The colour of the points indicates their likelihood. 5a shows the results of the bounded case, where the bounded area is above and left of the dotted lines. 5b shows the results of the constrained case with the targeted point shown as a green 'x'.

4. Conclusion

The case study in this paper presents two types of optimisation using the data driven Bayesian framework under development for optimising a set of design parameters. The use of a Gaussian process emulator, trained on data collected from high fidelity simulations, allows for fast prediction on new unseen input vectors, thus enabling the evaluation of likelihood functions for an MCMC sampling based optimisation algorithm. However, the complexity of the likelihood is increased due to marginalisation of the stochastic parameters, and limits the choice of MCMC sampling methods. For the case where the output is targeted on a specific point (constrained), the algorithm worked well and produced a posterior distribution over the outputs given the accepted points in the Markov chain that cluster tightly around the target. For the case where the target is a bounded area it is observed that the proposed approach under performed, highlighting a limitation in the current framework, and requires further investigation.

Further work is being done to investigate the formation of the bounded likelihood function, including: using different probability distributions with thinner tails that produce steeper gradients in the CDF and therefore better define the bounded area; adjusting the variance parameter in the distributions so that the log likelihoods in both dimensions have exactly the same variance, minimum, and maximum when evaluated across the entire space; introducing a penalty term in the likelihood to force the MCMC algorithm to reject points that produce outputs outside the bound. This paper also focused on just one area of the reachable space, so current work also involves testing the optimisation at different locations across the output space.

References

[1] Arregui-Mena J D, Margetts L and Mummery P M 2014 *Archives of Computational Methods in Engineering 2014 23:1* **23**(1) 171–190 ISSN 1886-1784 URL `https://link.springer.com/article/10.1007/s11831-014-9139-3`

[2] Chatterjee T, Chakraborty S and Chowdhury R 2019 *Archives of Computational Methods in Engineering* **26**(1) 245–274 ISSN 18861784 URL `https://link.springer.com/article/10.1007/s11831-017-9240-5`

[3] Morales E, Korondi Z, Quagliarella D, Tognaccini R, Marchi M, Parussini L and Poloni C 2021 *Space Technology Proceedings* **8** 35–53 URL `https://link.springer.com/chapter/10.1007/978-3-030-80542-5`$_3$

[4] Kundu A, Matthies H and Friswell M 2018 *Computer Methods in Applied Mechanics and Engineering* **337** 281 – 304 ISSN 0045-7825 URL `http://www.sciencedirect.com/science/article/pii/S0045782518301658`

[5] Gelman A, Carlin J B, Stern H S and Rubin D B 1995 *Bayesian Data Analysis* URL `https://www.taylorfrancis.com/books/mono/10.1201/9780429258411`

[6] Brooks S, Gelman A, Jones G L and Meng X L 2011 *Handbook of Markov Chain Monte Carlo* 1–592

[7] Rasmussen C E 2006 *Gaussian processes for machine learning* Adaptive computation and machine learning (Cambridge, Mass.: MIT Press) ISBN 978-0-262-18253-9

[8] Hoffman M D and Gelman A 2011 *Journal of Machine Learning Research* **15** 1593–1623 ISSN 15337928 URL `https://arxiv.org/abs/1111.4246v1`

[9] Braak C J T and Vrugt J A 2008 *Statistics and Computing* **18**(4) 435–446 ISSN 09603174 URL `https://link.springer.com/article/10.1007/s11222-008-9104-9`