



A composite Bayesian optimisation framework for material and structural design

R.P. Cardoso Coelho ^{a,b}, A. Francisca Carvalho Alves ^{a,b}, T.M. Nogueira Pires ^a,
F.M. Andrade Pires ^{a,b,*}

^a Faculty of Engineering, University of Porto, Porto, Portugal

^b Institute of Science and Innovation in Mechanical and Industrial Engineering, Porto, Portugal



ARTICLE INFO

Keywords:

Material design
Structural design
Inverse problems
Derivative-free optimisation
Bayesian optimisation

ABSTRACT

In this contribution, a new design framework leveraging Bayesian optimisation is developed to enhance the efficiency and quality of material and structural design processes. The proposed framework comprises two main steps. The first step involves efficiently exploring the design space with a minimum number of sampled points to mitigate computational costs. In the subsequent step, a composite Bayesian optimisation strategy is employed to evaluate the objective function and identify the next candidate for sampling. By building a surrogate model for numerical simulation responses in a fixed-size latent response space and using techniques like Principal Component Analysis for dimensionality reduction, the framework effectively exploits the composition aspect of the objective function. Unlike traditional methods that rely on random sampling across the design space, our Bayesian optimisation approach uses a dynamic, adaptive sampling strategy. This method significantly reduces the number of required experiments while effectively managing uncertainty. We evaluate the framework's performance across various design scenarios and conduct a critical comparative analysis against well-established data-driven approaches. These scenarios include linear and nonlinear material and structural behaviours, addressing multi-objective optimisation and data variability. Our findings demonstrate substantial improvements in performance and quality, particularly in nonlinear settings. This underscores the framework's potential to advance design methodologies in material and structural engineering.

1. Introduction

The rapid advancement of innovative numerical methods and cutting-edge technologies has revolutionised the development of advanced materials and structures tailored for specific applications. In material design, the primary objective is to carefully select the chemical structure, composition, or processing conditions to meet precise design criteria [1]. Typical design requirements include maximising the material's toughness and strength while minimising cost and weight. Although the allure of optimising material and structural design is undeniable, it remains a highly iterative and time-intensive process. The advent of machine learning techniques, increased computational power, and more accurate modelling tools have made simulation-based data-driven design frameworks a powerful alternative for material and structural design. Compared to physical experiments, the reduced costs associated with numerical simulations have further motivated the development of these frameworks. In the numerical community,

* Corresponding author at: Faculty of Engineering, University of Porto, Porto, Portugal.

E-mail addresses: rui.coelho@fe.up.pt (R.P. Cardoso Coelho), a.falves@fe.up.pt (A.F. Carvalho Alves), up202200867@up.pt (T.M. Nogueira Pires), fpires@fe.up.pt (F.M. Andrade Pires).

these challenges are often referred to as *inverse problems*, where the goal is to identify the input parameters that yield a desired output response by solving an optimisation problem. Solving this problem is still challenging, especially when dealing with high-dimensional design spaces and costly numerical simulations. Efficiently selecting candidates for evaluation is crucial to reducing the cost of the optimisation process, making sample efficiency a key priority for exploring the design space.

Solving inverse problems is a cross-disciplinary challenge that has received considerable attention in the literature. In mechanical engineering, notable works include Tanaka and Bui [2], Martins et al. [3], Juang and Sun [4]. Other applications are found in physiology [5], ecology [6] and sedimentation analysis [7]. Inverse problems can be classified into gradient-based or gradient-free optimisation depending on the objective function to optimise. Gradient-based methods are preferred for high-dimensional spaces, but when it is impossible to compute or approximate the derivatives of the objective function, gradient-free methods become the only viable option. This work focuses on problems where the objective function is expensive to evaluate, gradients are unavailable, and observations may be noisy — commonly referred to as *black-box* optimisation problems.

In material modelling, a typical application of inverse problems is identifying material parameters from experimental data, as demonstrated by Pal et al. [8], Pichler et al. [9], Rokonuzzaman and Sakai [10], Andrade-Campos et al. [11], Colak and Cakir [12], Sedighiani et al. [13], Savage et al. [14], Cardoso Coelho et al. [15]. Various optimisation algorithms have been employed to solve these problems, including Nelder-Mead [16,17], genetic [8–14] and evolutionary algorithms [18,19], Bayesian optimisation [20] and the recently proposed LIPO [21] algorithm [22,23]. Additionally, the rising interest in machine learning has driven the development of new methods for optimising hyperparameters, such as learning rates and the number of neurons and layers in neural networks. This context further underscores the need for sample-efficient methods for solving inverse problems with expensive objective functions.

From a mathematical standpoint, optimisation can be solved using various methods that only require observations of the objective function, but evaluating the objective value remains a complex task. Numerical tools used for solving the physical response of systems necessitate multiple steps for each function call: setting up the numerical model, solving governing equations, extracting desired responses, and computing the objective value. Similarly, experimentally-driven optimisations require conducting an experiment for each function call. Several strategies have been proposed to address these challenges to streamline the optimisation process and leverage the problem structure. Some strategies have been proposed to alleviate the burden of the optimisation process and to exploit this problem structure. For instance, Bessa and colleagues [24] introduced a computational framework for material and structural design, comprising three main steps: (i) exploring the design space using Design of Experiments (DoE) methods, (ii) efficiently simulating each design sample computationally, and (iii) constructing a surrogate model using machine learning techniques. In material design contexts, this approach approximates the optimisation of the original objective function through optimisation of a less costly surrogate model. This surrogate model, as demonstrated Bessa and Pellegrino [25], also facilitates the introduction of uncertainty quantification and multi-objective optimisation into the design process. Naturally, constructing this low-fidelity surrogate model involves creating a dataset and subsequently training a machine learning model. Examples of predictive models trained on computational data can be found in recent studies [26–29]. Additionally, Bessa et al. [24] advocate for using reduced-order models to accelerate further dataset evaluation, such as the self-consistent clustering analysis (SCA) [30].

Inspired by the work in Bessa et al. [24], numerous authors have advanced techniques for solving optimisation problems using machine learning. Abueidda et al. [31] and [32] proposed an optimisation framework based on neural networks optimised with a Genetic algorithm. Abueidda et al. [31] employed this framework to identify optimal microstructural descriptors of a two-dimensional checkerboard composite, while [32] used it to optimise the post-buckling of composite stiffened panels under compression loads. Sridhara et al. [1], applied a neural network model with automatic differentiation for gradient-based optimisation in microstructural design. Huang et al. [33] introduced a six-step mechanistic data science framework for materials design, encompassing (i) multimodal data generation and collection, (ii) mechanistic feature extraction, (iii) knowledge-driven dimension reduction, (iv) reduced-order modelling, (v) mechanistic learning through regression and classification, and (vi) system and design. The dimension reduction in step (iii) is achieved using principal component analysis (PCA). This framework successfully predicted composite material systems for targeted properties using a knowledge database.

Despite the excellent results achieved in these studies, Zhang et al. [34] argued that optimising surrogate models is inefficient if the design objective is well-defined. Their approach avoids choosing sampling points close to optimal locations, instead selecting random points across the entire design space. In contrast, Bayesian Optimisation (BO) provides a more efficient, adaptive sampling paradigm [35]. Wang and Dowling [36] demonstrated that a Bayesian Optimisation-driven inverse design framework requires significantly fewer experiments compared to the traditional Edisonian approach.

Bayesian optimisation is one of the most popular methods for solving expensive black-box optimisation problems, and has seen numerous applications over the last years [37–42]. It is widely recognised for its sample efficiency in exploring the design space to optimising the objective function. The core idea behind this method is constructing a probabilistic surrogate model for the objective function, which guides the search for the optimal solution by solving cheaper inner optimisation problems. For instance, Cardoso Coelho et al. [15] compared several optimisation algorithms for the parameter identification of various constitutive models, demonstrating that Bayesian optimisation is one of the best options for this task. Additionally, their work proposes a composite Bayesian optimisation framework that performs orders of magnitude better than standard Bayesian optimisation in some instances. This strategy builds on the work of Astudillo and Frazier [43], leveraging the compositional nature of the objective function to improve the quality of the inner surrogate model dramatically. However, this method has been proposed for the curve fitting problem, and its application to design problems still needs to be explored.

In this work, we extend the composite Bayesian optimisation strategy to address the response design problem, enabling a more efficient solution for material and structural design challenges. This extension requires generalising the composition proposed

by Cardoso Coelho et al. [15] to optimise arbitrary responses, making our approach a superset of the previous work and allowing the curve fitting problem to be recovered as a particular case. The composite variant offers several compelling advantages over standard Bayesian optimisation:

- *Comprehensive Response Analysis*: The optimiser observes the entire response instead of a single scalar objective value. This reduces information loss during the computation of the objective function, allowing the optimiser to make more informed decisions.
- *Embedded Reduction Function*: The reduction function that computes the scalar objective value from the system's response is directly embedded within the optimisation framework. This eliminates the need for the surrogate model to learn the shape of the reduction function, streamlining the optimisation process.
- *Improved Predictive Ability*: The composite strategy employs multi-dimensional Gaussian processes, enhancing the predictive capability of the surrogate model for the objective function.

Additionally, we explore Principal Component Analysis (PCA) as a dimensionality reduction technique to lower the computational cost of the method and, as a side benefit, enhance optimisation quality. Furthermore, we extend the method to multi-objective optimisation, which can be seamlessly integrated with the composite setting. To address the hyperparameters that arise from this formulation, we discuss heuristics to guide their selection. We then present a comparative analysis of our method's performance against classical Bayesian optimisation, the data-driven framework developed by [24] for both single- and multi-objective optimisation problems, and, as a baseline, simple random search. Several applications are considered in this comparison, including the optimisation of an analytical response, the maximisation of the toughness of a particle-reinforced composite, the design of a compression-torsion coupling metamaterial under both single- and multi-objective optimisation, and the cross-section design of a metallic foam beam under bending. These examples demonstrate the broad applicability of the proposed method and its potential to solve complex design problems more efficiently than state-of-the-art methods.

This paper is organised to provide a comprehensive overview and detailed analysis of the proposed method. In Section 2, we begin by outlining the optimisation problem and thoroughly detailing the formulation of the composite Bayesian optimisation strategy, addressing both single- and multi-objective scenarios. Section 3 delves into practical heuristics for selecting the hyperparameters of the method, offering guidance on how to fine-tune the approach effectively. Section 4 provides an in-depth comparative analysis, evaluating the performance of the proposed method against current state-of-the-art approaches across a range of design problems, highlighting its advantages and areas of improvement. Finally, Section 5 summarises the key findings of the study, presents conclusions drawn from the analysis, and discusses potential avenues for future research to enhance further and apply the proposed method.

2. A data-driven approach to the inverse problem

Material and structural design problems require finding a system's optimal geometry, chemical composition, or other relevant physical attributes to achieve a desired performance. In this context, one must parametrise the system into a set of input design variables in a given domain or parameter space. Examples of design variables span material phase volume fractions, material properties, manufacturing parameters, and geometric features. The response of the system is strongly dependent on this set of parameters. Finding the optimal solution requires searching the design space to maximise or minimise a given objective function. In the most general case, computing the objective function requires the evaluation of the system for a given set of input design variables — the system's response can only be observed via *sampling* and no additional information is available.

Numerous ways can be employed to evaluate the system's response. These comprise analytical models, numerical simulations, and experimental tests. While analytical models are generally more informative,¹ they are often limited to simple geometries and material models. On the other hand, experimental tests are a high-fidelity representation of the system's response, but they are often expensive, noisy and time-consuming. Within this context, numerical simulations are a popular choice for evaluating the system's response, allowing it to solve the problem at an arbitrary level of complexity. With the advent of high-performance computing, these tools have become increasingly popular in the field of material and structural design, enabling the application of *data-driven* approaches to the inverse problem.

Since this is a highly iterative process, automated unsupervised optimisation techniques are required to explore the design space efficiently. To this end, we explore a novel composite Bayesian optimisation framework for the structural and material design inverse problem. This strategy has been implemented in the *piglot* package, an in-house open-source Python package for the derivative-free optimisation of numerical responses [44]. As depicted in Fig. 1, the package connects the numerical solver with the optimisation algorithm, providing a seamless interface for the design problem.

In this section, we explore the theoretical foundations of the data-driven approach to the inverse problem. Firstly, the inverse problem is posed in Section 2.1, where the objective function is defined as a black-box function that depends on the output of a numerical simulation. Then, the classical Bayesian optimisation framework is described in Section 2.2. Subsequently, the composite Bayesian optimisation framework is introduced in Section 2.3. Finally, the multi-objective problem is presented in Section 2.4, along with its solution using Bayesian optimisation.

¹ For instance, derivatives can be extracted to accelerate optimisation routines.

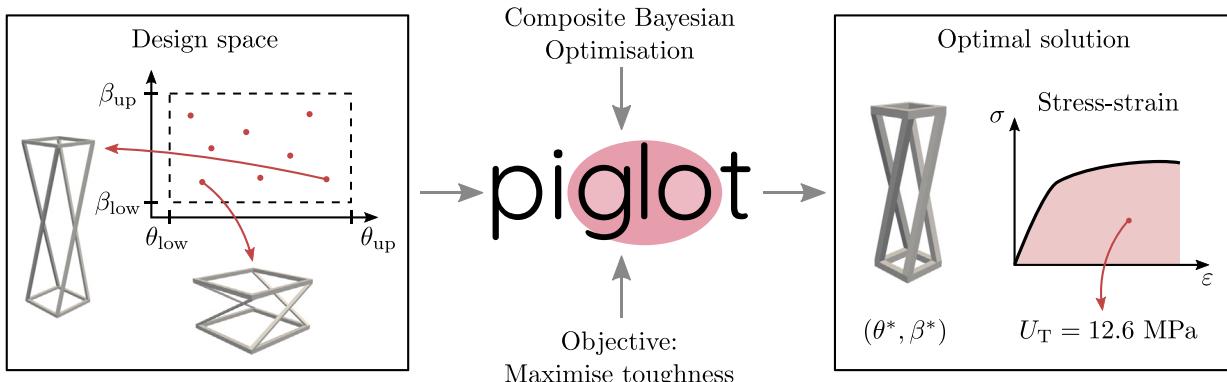


Fig. 1. Schematic representation of the design problem solved with **piglot**. In this case, we optimise two geometrical parameters θ and β of a metamaterial to maximise the toughness of the structure.

2.1. Inverse problem

Within the setting of this work, the inverse problem is mathematically defined as follows: given a set of input parameters θ in the parameter space \mathcal{B} , the goal is to find the optimal set of parameters θ^* that maximises a given objective function $J(\theta)$,

$$\theta^* = \arg \max_{\theta \in \mathcal{B}} J(\theta). \quad (1)$$

The objective function $J(\theta)$ is a deterministic scalar-valued function evaluated through the solution of a physical model or a numerical simulation. Without loss of generality, this function is assumed to be determined from a reduction operation $\mathcal{L}(\theta, \cdot)$ applied to the output of a numerical simulation $S[\theta]$, that is,

$$J(\theta) = \mathcal{L}(\theta, S[\theta]). \quad (2)$$

The reduction operation $\mathcal{L}(\cdot)$ is responsible for computing a scalar value for the quantity of interest from the output of the numerical simulation. On the other hand, the functional $S[\theta]$ represents the numerical simulation that solves the physical model for a given set of input parameters θ . In practice, the output quantities are generally discrete and time-dependent for numerical simulations. Thus, a discrete-time grid $t(\theta)$ and the output of the numerical simulation $Y(\theta)$ can be obtained from the simulation functional, that is,

$$S[\theta] \Rightarrow t(\theta), Y(\theta). \quad (3)$$

The time is a vector-valued quantity and, for completeness, is assumed to be dependent on the input parameters.² In contrast, the matrix-valued term $Y(\theta)$ can represent *any* output of the numerical simulation, including both scalar quantities such as forces or pointwise stresses, but also full-field measurements such as displacement or stress fields. This allows recasting the objective function into the discretised variant,

$$J(\theta) = \mathcal{L}(\theta, t(\theta), Y(\theta)). \quad (4)$$

With this expression, the compositional nature of the objective function is explicitly stated, where the reduction function \mathcal{L} is applied to the time grid $t(\theta)$ and the output of the numerical simulation $Y(\theta)$. This aspect is crucial for the proposed methodology, and is further explored in Section 2.3.

It is assumed that the fields $t(\theta)$ and $Y(\theta)$ are expensive to evaluate, and that their gradients are not available. Therefore, the objective function $J(\theta)$ is treated as a *black-box* from which we can only extract information via sampling on the parameter space. This assumption places strong constraints on the optimisation procedure: gradient-based methods are not feasible and, to reduce the computational cost, the number of evaluations of the objective function must be minimised.

As a final note, we limit the scope of the work to the optimisation of the time-dependent responses. By adopting a reparameterisation of the time axis, we can directly optimise quantities computed from force-displacement or stress-strain curves, which are common in the field of material design. By considering the decomposition of the objective function into \mathcal{L} and S , this approach can be easily generalisable for optimising any quantity computable from the numerical response.

2.2. Bayesian optimisation

Bayesian optimisation is a widely recognised technique for optimising costly-to-evaluate black-box functions with unknown derivatives [45]. The fundamental principle involves constructing a surrogate model based on a dataset of known function

² In the most common cases, the time is independent of the parameters. This dependency is introduced for the completeness of the formulation.

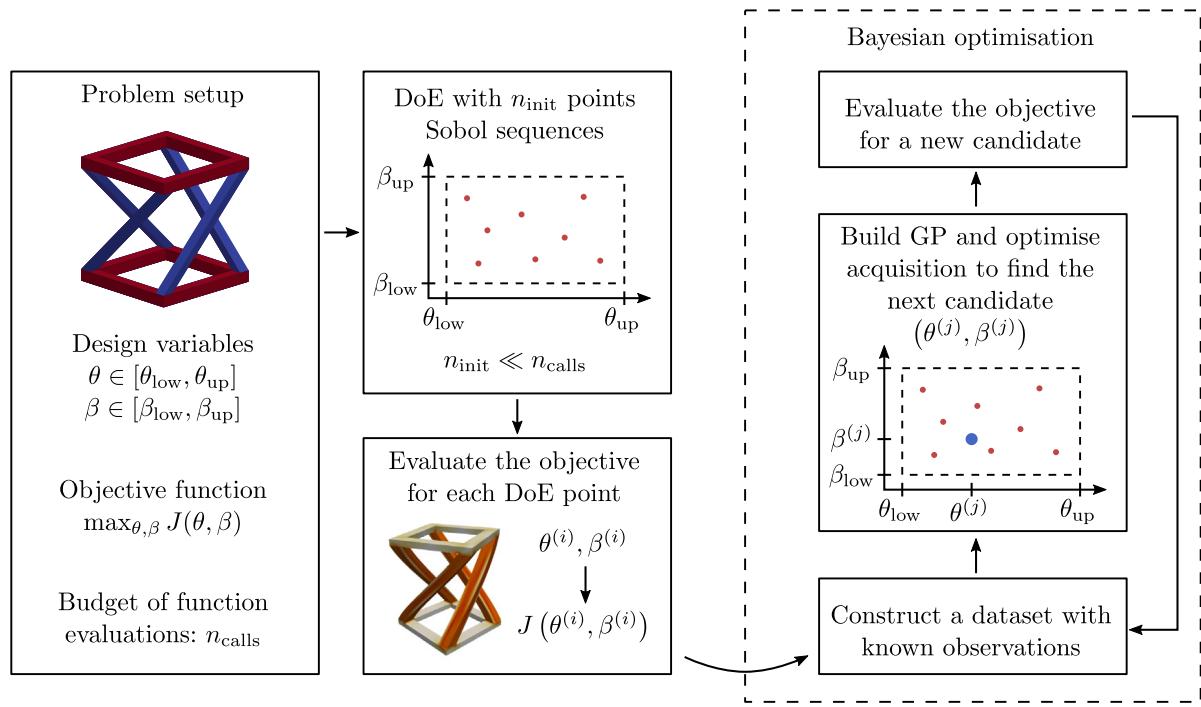


Fig. 2. Bayesian optimisation framework for the design problem.

evaluations. Subsequently, a collection of utility functions is utilised to identify the next candidate(s) sampling point(s). The basic structure of a Bayesian optimisation loop is outlined in Algorithm 1. In this work, we use the BoTorch package [46] to implement the Bayesian optimisation algorithm.

Algorithm 1 Overview of the Bayesian optimisation method (adapted from Frazier [45]).

- 1: Design n_0 initial experiments and observe f at each point θ
 - 2: Build the initial dataset and set $n = n_0$
 - 3: Assign the number of iterations N
 - 4: **while** $n \leq N$ **do**
 - 5: Build a surrogate model on existing observations θ (Section 2.2.1)
 - 6: Define an acquisition function α and find the optimum $\theta_n = \arg \max_{\theta \in \mathcal{B}} \alpha(\theta)$ (Section 2.2.2)
 - 7: Observe $f(\theta_n)$ and update the dataset
 - 8: Increment n
 - 9: **end while**
 - 10: Return observed maximiser of $f(\theta)$
-

A schematic representation of the Bayesian optimisation framework for the design problem is depicted in Fig. 2. The first step is the problem setup, where the design variables, the objective function and the budget of evaluations n_{calls} are defined for the numerical problem at hand. Then, an initial Design of Experiments (DoE) is performed to sample the design space, using n_{init} points, where the objective function is evaluated at each point. It is assumed that this dataset is only built to start the process and that the number of DoE points is much smaller than the total budget of evaluations, that is, $n_{\text{init}} \ll n_{\text{calls}}$. Finally, the Bayesian optimisation loop is initiated, where the surrogate model is built from the initial dataset and the acquisition function is optimised to select the next sampling point. This is repeated until the budget of evaluations is exhausted, at which point the algorithm returns the best-observed point.

2.2.1. Surrogate model

In the Bayesian optimisation context, the most frequently adopted surrogate model is a Gaussian process (GP) [47]. A GP is a probabilistic model that defines a distribution over functions, that is,

$$f \sim \mathcal{GP}(\mu(\theta), k(\theta, \theta')), \quad (5)$$

where $\mu(\theta)$ and $k(\theta, \theta')$ are the prior mean and covariance functions. Without loss of generality, the prior mean is usually set to zero, and the covariance function is defined by a kernel function $k(\theta, \theta')$ that accounts for the correlation between the input parameters,

usually through the Euclidean distance between θ and θ' . These quantities constitute the *prior* of the GP, which is then conditioned with the observed data to obtain the *posterior* distribution. After n observations, the posterior distribution of the GP is given by

$$f(\theta) \sim \mathcal{N}(\mu_n(\theta), \sigma_n^2(\theta)), \quad (6)$$

where the posterior mean $\mu_n(\theta)$ and variance $\sigma_n^2(\theta)$ can be computed by the expressions in Rasmussen and Williams [47, Section 2.2] and Frazier [45, Section 3]. In other words, the value of the function f at any point θ is assumed as a Gaussian distribution with mean $\mu_n(\theta)$ and variance $\sigma_n^2(\theta)$.

While the GP is, by construction, a hyperparameter-free model, the definition of the kernel requires estimating a set of hyperparameters for the covariance function. Consider a Matérn kernel, given by,

$$k_{\text{Matern}}(\theta, \theta') = \alpha \frac{2^{1-v}}{\Gamma(v)} \left(\sqrt{2v} (\theta - \theta')^T \mathbf{L}^{-1} (\theta - \theta') \right)^v K_v \left(\sqrt{2v} (\theta - \theta')^T \mathbf{L}^{-1} (\theta - \theta') \right), \quad (7)$$

where Γ is the gamma function, K_v is the modified Bessel function, and v is the smoothness parameter. In this work, the latter is fixed as $v = 2.5$. Furthermore, α is the output scale parameter and \mathbf{L} is a diagonal matrix with the lengthscale hyperparameters, which must be estimated according to the training data. In addition, the observations may contain a certain level of noise σ_ϵ , that is,

$$y = f(\theta) + \epsilon, \quad \text{with } \epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2), \quad (8)$$

which must be accounted for in the GP model. If the observation noise is not known for each data point, this quantity must also be estimated from the training data. The estimation of these hyperparameters is performed by maximising the marginal log-likelihood (MLL) of the GP model. For these models, the integrated marginalised term is Gaussian which, in turn, allows the MLL to be expressed analytically. In this work, we use the BoTorch package to estimate the hyperparameters of the GP model via a built-in L-BFGS-B optimiser.

2.2.2. Acquisition functions

After constructing the probabilistic surrogate model for the objective function, the next step is to define an acquisition function that guides the selection of the next sampling point. An acquisition function quantifies the utility of sampling a given point in the design space. Classical acquisition functions include [48]:

Probability of Improvement (PI): Taking into account the posterior probability distribution, P , this acquisition computes the probability that a point θ improves the current optimum $f^* = f(\theta^*)$, that is [49],

$$\alpha^{\text{PI}}(\theta) = P(f(\theta) \geq f^*). \quad (9)$$

It should be remarked that this function returns the probability of *any* improvement of the function, regardless of its magnitude. Thus, optimising with the PI generally emphasises exploitation, as points close to the current optimum have higher probabilities of further improving the solution, even by a small amount.

Expected Improvement (EI): The expected improvement quantity can be computed by taking the expected value of the improvement of the currently best-observed value over the posterior [50],

$$\alpha^{\text{EI}}(\theta) = \mathbb{E}[\max\{0, f(\theta) - f^*\}] \quad (10)$$

Unlike the PI acquisition, the expected improvement returns values that are dependent on the magnitude of the improvement and, therefore, is generally more explorative than the PI. Nevertheless, it is accepted that the EI generally also favours exploitation over exploration of the domain. To solve the problem of vanishing gradients for a large portion of the design space, the Log Expected Improvement (LogEI) has been recently proposed [51]. In practice, LogEI retains the same properties as the EI but is evaluated in a more numerically stable fashion in the logarithmic space, which has been shown to drastically improve the performance of high-dimensional problems. In this work, this variant is preferred over the standard EI.

Upper Confidence Bound (UCB): This family of acquisition functions is derived from the concept of confidence bounds in statistical analysis [52]. Frequently written for Gaussian posteriors, the upper confidence bound can be computed as

$$\alpha^{\text{UCB}}(\theta) = \mu(\theta) + \beta\sigma(\theta), \quad (11)$$

where $\mu(\theta)$ and $\sigma(\theta)$ are the posterior mean and standard deviation, and β is a user-defined hyperparameter. Higher values favour exploration, whilst smaller ones promote exploitation.

All these functions can be evaluated analytically for Gaussian posteriors, which makes them particularly attractive for the Bayesian optimisation framework. Examples of other relevant acquisition functions include the lookahead Knowledge Gradient (KG) [53,54], as well as the entropy search (ES) [55] and the related predictive entropy search (PES) [56] and max-value entropy search (MES) [57], which are not further explored in this work for the sake of brevity.

The choice of the next candidate point to sample is made by maximising the acquisition function over the design space. Since $\alpha(\theta)$ can be cheaply evaluated from the surrogate model, this step can be performed efficiently. Furthermore, since both the acquisition

function and the GP are differentiable, the optimisation can be carried out with gradient-based methods. Notwithstanding, the acquisition function is generally non-convex and the design space may contain multiple local optima. A robust approach is to use a multi-restart gradient-based method to efficiently explore the design space and find the global optimum of the acquisition function. In this work, the built-in optimiser of the BoTorch package is used to maximise the acquisition function, which adopts a multi-restart L-BFGS-B algorithm.

2.3. Composite Bayesian optimisation

The objective function described in Eq. (4) is a scalar quantity that depends on the output of a numerical simulation. It has been established as a composite function of the reduction function \mathcal{L} and the numerical response of the solver with the time and data points $t(\theta)$ and $Y(\theta)$. In this context, the reduction function \mathcal{L} is generally known through an analytical expression. Thus, this function is generally cheap to evaluate and its derivatives can be computed. The non-differentiability of the objective function and its time-consuming evaluation arises from the inner part of the composition, that is, the numerical simulation. The Bayesian optimisation framework can be drastically improved by exploiting the composite nature of the objective function [15,43], as described throughout this section. The main requirements for establishing the composite Bayesian optimisation framework are:

1. Adopt a transformation \mathcal{T} and its inverse \mathcal{T}^{-1} of the numerical simulation data $t(\theta)$ and $Y(\theta)$ into a fixed-size vector $p(\theta)$ defined on a *latent response space*;
2. Build a fixed-size vector-valued multi-output Gaussian process as the surrogate model for the *outputs* of the numerical simulations $p(\theta)$ on the latent response space;
3. Leverage (quasi-)Monte Carlo methods to sample the posterior and evaluate the acquisition functions;
4. Establish the evaluation of the optimisation objective in terms of the latent response space samples $p_q(\theta)$, and, critically allowing the *gradients* of this transformation to be evaluated;
5. Optionally, employ dimensionality reduction schemes, such as Principal Component Analysis (PCA), to transform the latent response space $p(\theta)$ into the *reduced latent space* $\hat{p}(\theta)$.

In conjunction with Fig. 3, these steps constitute a high-level overview of the composite Bayesian optimisation framework. Each of these ingredients is described in detail in Sections 2.3.1–2.3.5. As shown throughout the work, when embedded into the curve optimisation problem for material design, the composite Bayesian optimisation framework has numerous advantages, namely:

- The entire response is observed by the optimiser instead of a single scalar objective value — this drastically reduces the loss of information when computing the objective function;
- A surrogate model is built for the responses of the numerical simulation with multi-output Gaussian processes — leading to an increased predicting ability of the surrogate model and allowing learning of the response space;
- The reduction function does not need to be learned by the surrogate model — this information is embedded in the evaluation of the posterior of the objective function.

These advantages lead to an increase in optimisation performance, particularly in reducing the number of required function evaluations. However, it should be remarked that the composite strategy is accompanied by an increase in the computational cost of the optimiser. In the limit case, where the function call is expected to be the computational bottleneck, the increased optimisation performance justifies the additional computational cost.

2.3.1. Transformation of the numerical simulation data

Due to the nature of the Bayesian optimisation framework, one is required to build a surrogate model for relevant quantities of the numerical simulation. In the classical Bayesian optimisation framework, the surrogate model is built for the objective function $J(\theta)$, which is a scalar quantity. However, to leverage the composition aspect of the objective function in the composite Bayesian optimisation framework, we wish to build a surrogate model for the *responses* of the numerical simulations. From a broad perspective, this requires constructing a dataset of the numerical simulation outputs $t(\theta)$ and $Y(\theta)$. However, from a practical point of view, it should not be assumed that the dimensionality of the discrete data points of $t(\theta)$ and $Y(\theta)$ is fixed — it may vary from one simulation to another.³ Therefore, consider a transformation \mathcal{T} that maps the numerical simulation data into a fixed-size vector $p(\theta)$,

$$p(\theta) = \mathcal{T}(t(\theta), Y(\theta)), \quad (12)$$

where the responses $t(\theta)$ and $Y(\theta)$ are assumed to be of variable size and $p(\theta)$ is a fixed-size vector that represents the output of the numerical simulation. It is assumed that the vector $p(\theta)$ is defined in the so-called *latent response space*. Furthermore, the transformation \mathcal{T} must be invertible; that is, there exists an inverse transformation \mathcal{T}^{-1} such that

$$[t(\theta), Y(\theta)] = \mathcal{T}^{-1}(p(\theta)). \quad (13)$$

Note that this last equivalence is not a strict equality, as the transformation \mathcal{T} may introduce some information loss. However, considering an adequately designed transformation, it is assumed that this information loss is negligible in this work.

³ Several factors may contribute to this observation. For instance, the time grid may be sensitive to the parameters, or a sub-stepping strategy may introduce variability in the number of discrete time points.

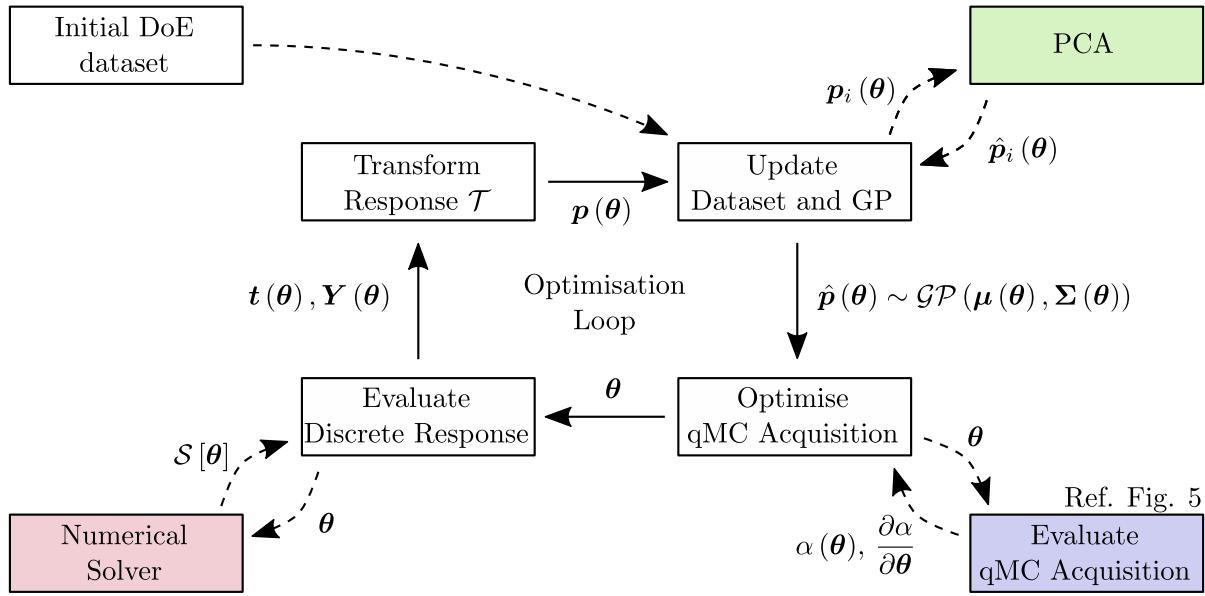


Fig. 3. High-level overview of the composite Bayesian optimisation framework.

Naturally, the choice of the transformation is problem-dependent. To showcase this point, we illustrate two options for the transformation in the context of two different response optimisation scenarios, depicted in [Fig. 4](#) and described here:

Curve Fitting - pointwise errors: In this scenario, we aim to compare the output of a numerical simulation with a set of reference data points to find the set of parameters that better fit the reference data. In this case, since the reference grid is fixed, the transformation \mathcal{T} can be defined as the computation of the pointwise errors $e(\theta)$ between the simulation and the reference data, as described in detail in [Cardoso Coelho et al. \[15\]](#). In this context, the objective function $J(\theta)$ is generally defined as the sum of the squared pointwise errors, that is,

$$J(\theta) = \frac{1}{N} \sum_{i=1}^N e_i^2(\theta), \quad (14)$$

where N is the number of data points in the reference grid, so that the minimisation of this function leads to the best fit of the simulation to the reference data. [Fig. 4\(a\)](#) illustrates this scenario. With this setting, the latent response $p(\theta)$ has the format

$$p(\theta) = [e_1(\theta), e_2(\theta), \dots, e_N(\theta)]. \quad (15)$$

Response Design - endpoint interpolation: For general responses, the shape of the output of the numerical simulation may vary significantly from one simulation to another. A simple strategy to address this issue is to interpolate the output of the numerical simulation at a fixed set of N evenly spaced points between the initial and final time steps t_{\min} and t_{\max} , respectively. Therefore, assuming $y_i(\theta)$ as the i th interpolated value, the latent response $p(\theta)$ can be defined as the concatenation of the interpolated values with the endpoints of the time grid, that is,

$$p(\theta) = [y_1(\theta), y_2(\theta), \dots, y_N(\theta), t_{\min}, t_{\max}], \quad (16)$$

thus defining the latent response space with $N + 2$ fixed dimensions. Naturally, this requires setting an appropriate value for N so that the interpolation captures the relevant features of the output of the numerical simulation. [Fig. 4\(b\)](#) schematically depicts this transformation. This strategy can then be easily reversed to obtain an approximation of the original response of the numerical simulation.

This strategy allows the construction of a fixed-size vector for a single response curve. Notwithstanding, this can be easily extended to multiple responses or full-field measurements by concatenating the latent response vectors into a single vector. In piglot, this is implemented in the so-called `ConcatUtility`, which, given the sizes of each vector $p(\theta)$, handles the concatenation and de-concatenation of the latent response vectors.

Remark 1. The transformation of the numerical simulation data into a fixed-size vector is a critical step in the composite Bayesian optimisation framework. The choice of the transformation is problem-dependent and should be carefully designed to capture the

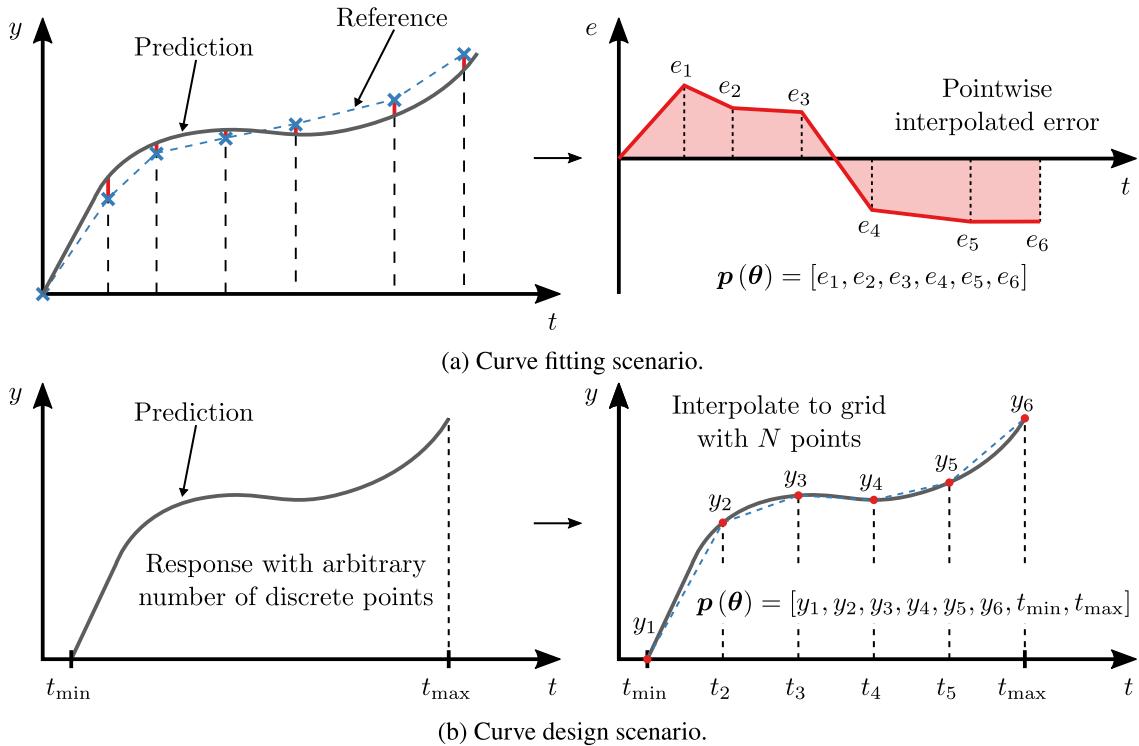


Fig. 4. Illustration of the proposed transformations for two different problems.

relevant features of the output of the numerical simulation. While the transformation for the curve fitting scenario is straightforward, the transformation for the response design scenario requires a careful choice of the interpolation strategy. The transformation adopted in this work has been shown to be effective for the classes of problems at hand, but it may pose limitations for responses with severe differences in length or with varying length scales. While this topic is not further explored in this work due to brevity, it remains an open research question. In addition, the number of interpolation points emerges as a hyperparameter that, at first glance, needs to be carefully chosen to balance the computational cost of the optimisation and the accuracy of the surrogate model. However, as demonstrated in Section 2.3.5, a PCA strategy completely mitigates the dependency on the computational cost. Thus, the number N can be set to a high value without any significant increase in the computational cost.

2.3.2. Multi-output surrogate model

To model the response within of the latent response space, we employ a multi-output Gaussian process. Within this scope, since our surrogate model must account for the multiple outputs of the numerical simulation, there are two main strategies to build the multi-output Gaussian process:

1. **Single Multi-task Gaussian process:** In this approach, the vector-valued output of the numerical simulation is directly modelled using a GP, so that the vector $p(\theta)$ is assumed to follow a multivariate normal distribution,

$$p(\theta) \sim \mathcal{N}(\mu(\theta), \Sigma(\theta)), \quad (17)$$

where $\mu(\theta)$ is the vector-valued mean function and $\Sigma(\theta)$ is the covariance matrix. Since the outputs follow a multivariate normal distribution, the correlation between the outputs is naturally accounted for. However, the computational cost of this approach is significantly higher than the alternative.

2. **Multiple Single-task Independent Gaussian processes:** In this approach, a separate Gaussian process is built for each output of the numerical simulation. As a consequence, the individual outputs are modelled independently, and are assumed to follow a normal distribution with known mean and variance functions,

$$p_i(\theta) \sim \mathcal{N}(\mu_i(\theta), \sigma_i^2(\theta)), \quad (18)$$

where $\mu_i(\theta)$ and $\sigma_i^2(\theta)$ are the scalar-valued mean and variance functions of the i th output. Note that this strategy does not consider the potential correlations between the outputs. However, for noiseless observations, it can be shown that the independent models are equivalent to the multi-task Gaussian process — a property known as *autokrigability* [58], which is particularly useful in the context of this work. The main advantage of this approach is that the computational cost of both training and inference is significantly lower than the multi-task Gaussian process, and can be easily batched for parallel evaluation in the GPU.

In this work, multiple single-task independent Gaussian processes are adopted as the surrogate model for the responses of the numerical simulation. While the single multi-task Gaussian process is a more general approach, its increased computational cost makes this strategy prohibitive for the high-dimensional problems considered in this work. As a final note, it should be remarked that, while the autokrigability property ensures that the surrogate models for the latent response space are equivalent, there is a notable difference in the evaluation of the posterior for the objective function and, particularly, a dependency on the number of points in the latent response space. This topic is described in more detail in Section 2.3.3.

2.3.3. Posterior of the objective function

Within the context of Bayesian optimisation, it is essential to estimate the posterior distribution of the objective function. In classical Bayesian optimisation, the surrogate model is built for the objective function $J(\theta)$. Thus, the posterior distribution is known and analytically computable — it is a Gaussian distribution. On the other hand, in the composite Bayesian optimisation framework, the posterior for the objective function is not directly available — while $p(\theta)$ is assumed to be Gaussian, the distribution of $J(\theta)$ cannot be directly inferred.

In the composite setting, the standard approach for estimating the posterior distribution of the objective function is via Monte Carlo sampling. To this end, for a given point θ_q , a set of Q samples $p_q(\theta_q)$ is drawn from the posterior distribution of the latent response space, that is,

$$p_q(\theta_q) \sim \mathcal{N}(\mu(\theta_q), \Sigma(\theta_q)). \quad (19)$$

Then, the samples need to be transformed back to the original response space, that is,

$$[t_q(\theta_q), Y_q(\theta_q)] = \mathcal{T}^{-1}(p_q(\theta_q)), \quad (20)$$

where $t_q(\theta_q)$ and $Y_q(\theta_q)$ are the transformed samples of the time grid and the output of the numerical simulation. Finally, the objective function $J(\theta_q)$ is evaluated for each of these samples through Eq. (4),

$$J_q(\theta_q) = \mathcal{L}(\theta_q, t_q(\theta_q), Y_q(\theta_q)), \quad (21)$$

where the samples $J_q(\theta_q)$ are assumed to be drawn from the posterior distribution of the objective function and can be used for its estimation. While these samples should be sufficient for fitting a known probability distribution, in practice, as shown in Section 2.3.4, the acquisition functions can be directly evaluated from these samples efficiently.

As a final note, as highlighted in Section 2.3.4, it is particularly useful to compute the gradients of the objective function with respect to the input parameters to allow an efficient optimisation of the acquisition functions. In the context of the composite Bayesian optimisation framework, the gradients of the objective function can be computed by the chain rule, that is,

$$\frac{\partial J_q}{\partial \theta_q} = \frac{\partial \mathcal{L}}{\partial \theta_q} + \frac{\partial \mathcal{L}}{\partial t_q} \frac{\partial t_q}{\partial \theta_q} + \frac{\partial \mathcal{L}}{\partial Y_q} \frac{\partial Y_q}{\partial \theta_q}. \quad (22)$$

The gradients of the reduction function \mathcal{L} are assumed to be known, and the gradients of the response quantities $t_q(\theta_q)$ and $Y_q(\theta_q)$ can be computed by the inverse transformation \mathcal{T}^{-1} , that is,

$$\left[\frac{\partial t_q}{\partial \theta_q}, \frac{\partial Y_q}{\partial \theta_q} \right] = \frac{\partial \mathcal{T}^{-1}}{\partial p_q} \frac{\partial p_q}{\partial \theta_q}. \quad (23)$$

Finally, the gradient of the latent response space $p_q(\theta_q)$ with respect to the input parameters θ_q is assumed to be computable via the surrogate model.⁴ Thus, both the gradients of the reduction function \mathcal{L} and of the inverse transformation \mathcal{T}^{-1} are required to compute the gradients of the objective function.⁵ In `piglot`, these gradients are not explicitly computed; instead, the functions \mathcal{L} and \mathcal{T}^{-1} are written in terms of PyTorch operations, which allows the automatic differentiation via backpropagation. This approach allows increased flexibility in the definition of the reduction function and the transformation. As a final note, these functions are designed to allow their evaluation for large batches of points, so that the computation can be efficiently parallelised.

Before proceeding, it is worth highlighting that when adopting the multiple single-task independent Gaussian processes as the surrogate model for the responses of the numerical simulation, the components of the latent response space are assumed to be independent. Due to autokrigability, this assumption is equivalent to the multi-task Gaussian process in the noise-free scenario, so that the model can learn the same information as the multi-task model. However, a critical difference arises when *sampling* from the posterior distribution of the latent response space: the samples of $p(\theta)$ are independent and do not reflect possible correlations between them. This has the hidden consequence that the posterior of the objective function $J(\theta)$ becomes sensitive to the number of points in the latent response space; from the central limit theorem, as the number of points increases, the posterior standard deviation should decrease proportionally to $1/\sqrt{N}$, while the mean should remain unaffected. Therefore, even if the reduction function \mathcal{L} is independent of the number of points, the posterior of the objective function and, consequently, the Bayesian optimisation process are dependent on the number of points in the latent response space. To mitigate this issue, two approaches can be considered:

⁴ Note that these are the derivatives of the surrogate model, not of the numerical response. For Gaussian processes, these gradients are often easily computable. The gradient of the original numerical response does not need to be known.

⁵ Note that, with this approach, the gradients of the forward transformation \mathcal{T} do not need to be computed.

1. **Multivariate normal sampler:** Instead of using independent samplers for each component of $p(\theta)$, a multivariate normal sampler can be used to sample the latent response space. To this end, the covariance matrix of the latent response space must be estimated from the training data. Broadly speaking, this is similar to using Eq. (19) for sampling, with $\mu(\theta_q)$ and $\Sigma(\theta_q)$ estimated from the training data and Eq. (18). Naturally, there is an increase in the computational cost when using this strategy.
2. **Dimensionality reduction to weakly correlated features:** A more efficient approach is to reduce the dimensionality of the latent response space to a set of weakly correlated features. This can be achieved by employing dimensionality reduction techniques, such as Principal Component Analysis (PCA), to transform the latent response space $p(\theta)$ into a reduced latent space $\hat{p}(\theta)$ where the components are not strongly correlated and can be assumed to be independent. This strategy is described in more detail in Section 2.3.5.

2.3.4. Evaluating acquisition functions

By definition, an acquisition function is a utility function that quantifies the value of sampling a given point in the design space. This requires the knowledge of the posterior probability distribution of the objective function $J(\theta)$ to optimise, which has been described in Section 2.3.3. In the composite setting, it is not practical to estimate the posterior distribution of the objective function directly; alternatively the acquisition functions can be evaluated from the samples of the latent response space. This can be easily accomplished by establishing the acquisition functions an expectation over some quantity computed from the samples of the objective function and evaluating via Monte Carlo, that is,

$$\alpha(\theta) = \mathbb{E}[\alpha(\xi) | \xi \sim J(\theta)] \approx \frac{1}{Q} \sum_{q=1}^Q \alpha(J_q(\theta)), \quad (24)$$

where Q is the number of samples drawn for the Monte Carlo estimation. From the family of acquisitions in Section 2.2.2, only the EI is derived from an expectation that can be trivially computed with Monte Carlo. For the PI and UCB, as shown in Wilson et al. [59], Cardoso Coelho et al. [15], they can be written as expectations according to the following transformations:

$$\alpha^{\text{PI}}(\theta) = \mathbb{E}[H(\max\{0, f(\theta) - f^*\})], \quad (25)$$

$$\alpha^{\text{UCB}}(\theta) = \mathbb{E}\left[\mu(\theta) + \beta \sqrt{\frac{\pi}{2}} |\hat{y}(\theta) - \mu(\theta)|\right], \quad (26)$$

where H is the Heaviside step function and $\hat{y}(\theta)$ is a sample from the posterior distribution of the objective function. In addition, the gradients of these acquisitions can be computed using the reparameterisation trick, as described by Wilson et al. [59]. Thus, with this procedure, the acquisitions can be efficiently evaluated using the samples from the posterior obtained with the procedure in Section 2.3.3, and the objective posterior does not need to be explicitly evaluated.

In practice, the evaluation of the acquisition function and its derivatives requires a series of steps, as illustrated in Fig. 5. For a given parameter θ , the samples of the latent response space $p_q(\theta)$ are drawn from their posterior distribution at θ . Then, the samples are transformed back to the original response space $t_q(\theta)$ and $Y_q(\theta)$ by the inverse transformation \mathcal{T}^{-1} . At this stage, these quantities can be interpreted as several Monte Carlo samples of possible numerical responses for the given parameters θ . Finally, the objective function $J_q(\theta)$ is evaluated for each response sample, and the acquisition function is computed from the expectation in Eq. (24). As all of these operations are implemented using PyTorch tensors, both (i) the gradients are automatically computed via backpropagation⁶ and (ii) the computations can be efficiently parallelised for multiples batches of points θ . These aspects strongly motivate using multi-start gradient-based optimisation methods to maximise the acquisition functions,⁷ as described in Section 2.2.2.

Regarding the sampling of the posterior, two additional improvements are employed to improve the efficiency of the method. Firstly, the samples of the latent response space are drawn using a quasi-Monte Carlo (q-MC) method based on Sobol sequences, which helps to reduce the variance of the Monte Carlo estimation. Then, as described by Balandat et al. [46], fixed base samples are used for the sampling, which increases the smoothness of q-MC evaluated acquisitions. This allows for increased efficiency in the optimisation of these functions with gradient-based methods, at the cost of a small bias in their function values. In this work, the BoTorch package is used to evaluate the acquisition functions, which automatically handles the q-MC sampling and the fixed base samples.

2.3.5. Dimensionality reduction with PCA

In the context of data-driven strategies, it is very common to carry an intermediate step of dimensionality reduction to reduce the number of features in the dataset. One of the most popular techniques for linear dimensionality reduction is Principal Component Analysis (PCA). At its core, the PCA algorithm aims to project the original data into a new space defined by the principal components (PC), which are the set of orthogonal directions that maximise the explained variance of the original data. The main steps for this strategy are depicted in Fig. 6 and can be summarised as follows:

⁶ The backpropagation for the posterior sampling stage requires the reparameterisation trick [60].

⁷ Parallelism can be exploited by both the initial candidate generation stage and the concurrent running of gradient-based optimisation on multiple points.

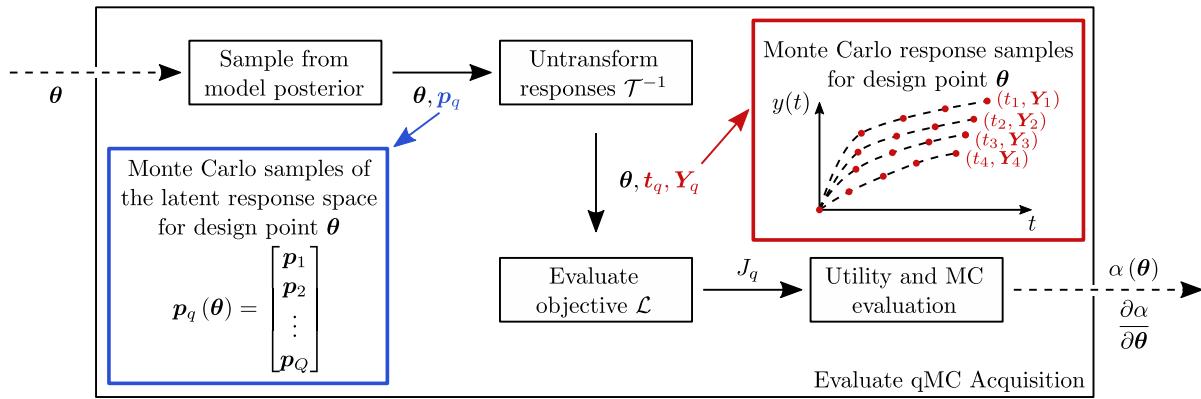


Fig. 5. Illustration of the acquisition function evaluation in the composite Bayesian optimisation framework.

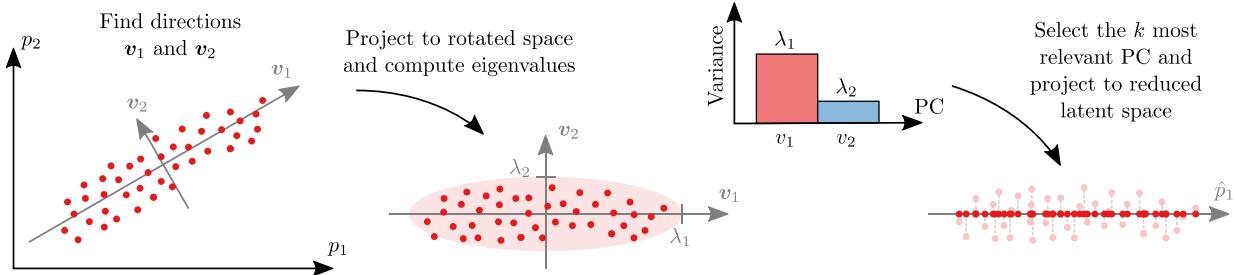


Fig. 6. Illustration of the PCA transformation in the composite Bayesian optimisation framework.

- 1. Standardise the data and get covariance matrix:** It is well established that PCA should only be used with zero-mean data and, furthermore, that the data should be scaled to have unit variance. Therefore, the data is first standardised to ensure these properties are met. Then, the covariance matrix Σ of the standardised data is computed to infer the correlation between the different features.
- 2. Find the principal components:** Given the covariance matrix Σ , the eigenvalues λ_i and eigenvectors v_i are computed. The principal components are the eigenvectors of the covariance matrix, and the eigenvalues represent the amount of variance explained by each component. Using the eigenvalues, the principal components are then ordered by decreasing variance. The eigenvectors are the orthogonal set of directions that maximise the variance of the data.
- 3. Project to the lower dimensionality space:** Finally, with the principal components ordered, we can choose the first k components that, cumulated, explain a certain portion of the total variance of the data. Having k selected, the $N \times k$ transformation matrix W is built with the first k eigenvectors,

$$W = [v_1, v_2, \dots, v_k]. \quad (27)$$

With this matrix, any point in the latent response space $p(\theta)$ can be projected to the reduced latent space $\hat{p}(\theta)$ by the matrix multiplication,

$$\hat{p}(\theta) = p(\theta) W, \quad (28)$$

together with the reverse transformation,

$$p(\theta) = \hat{p}(\theta) W^T. \quad (29)$$

In piglot, PCA is employed to reduce the dimensionality of the latent response space $p(\theta)$ to a reduced latent space $\hat{p}(\theta)$. The transformation is recomputed at every iteration of the Bayesian optimisation, and the number of components k can change from one iteration to another. As highlighted in Cardoso Coelho et al. [15], the computational cost of the composite Bayesian optimisation is highly sensitive to the number of points in the latent response space, so it is critical to carry out this dimensionality reduction. Furthermore, one advantage of this procedure is that, in the worst-case scenario, all the principal components are selected and the original latent response space is preserved. In addition, the PCA transforms the space into a set of uncorrelated features that can be assumed to be independent, thus mitigating some of the issues raised in Section 2.3.3 for the single-task models. Finally, since the reduced latent space is a linear transformation of the original latent response space, the gradients of the objective function can be easily computed in the reduced latent space and transformed back to the original latent response space. Recalling the acquisition function evaluation flowchart in Fig. 5, the PCA transformation is carried out between the sampling of the posterior of the objective

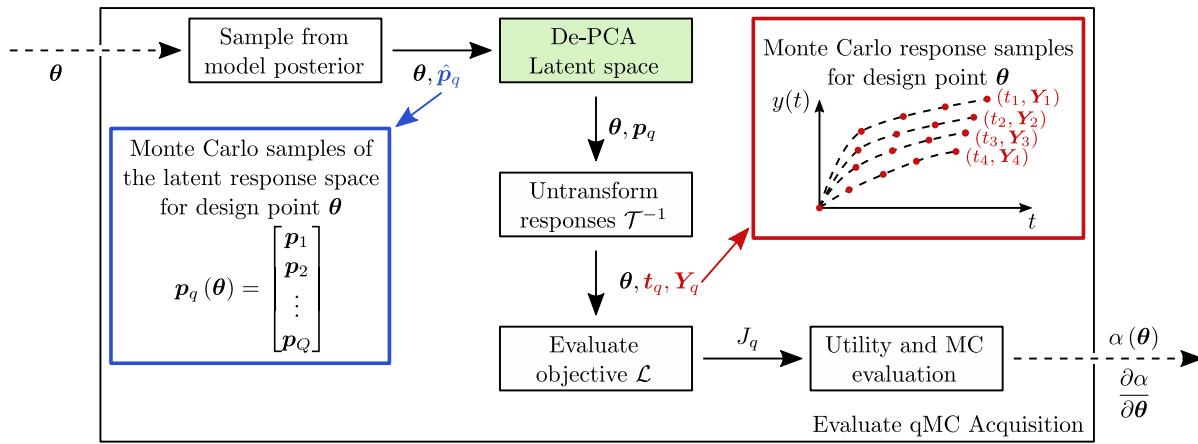


Fig. 7. Illustration of the acquisition function evaluation in the composite Bayesian optimisation framework with PCA.

function and the inverse transformation of the response, leading to the diagram depicted in Fig. 7. Therefore, this component can be easily introduced in the procedure without significant changes to the overall workflow. The impact of this PCA transformation is discussed in Section 3.2, along with a parametric study concerning the unexplained PCA variance threshold.

2.4. Multi-objective Bayesian optimisation

Up to this point, the Bayesian optimisation framework has been described for the single-objective case. However, in many engineering problems, the optimisation is characterised by multiple objectives that need to be simultaneously optimised, i.e., solving a multi-objective optimisation problem. In this scope, considering a set of m objectives, the optimisation problem can be defined as

$$\max_{\theta \in \mathcal{B}} [J_1(\theta), J_2(\theta), \dots, J_m(\theta)], \quad (30)$$

where $J_i(\theta)$ is the i th objective function. Note that, for notational convenience, the problem has been written as the maximisation of all objectives, but it can instead involve the minimisation of some objectives.⁸ In the multi-objective scenario, there is generally no single solution θ^* that is optimal for all objectives; instead, in these problems, we seek to find the so-called *Pareto front* of the problem. The Pareto front is defined in the objective space and, broadly speaking, denotes the regions where the improvement of one objective leads to the worsening of another. A solution θ_1 *dominates* another solution θ_2 if it improves at least one objective without worsening any other. Therefore, a solution that lies in the Pareto front is referred to as Pareto optimal or non-dominated, while others are known as dominated.

Bayesian optimisation can be used to solve the multi-objective optimisation problem. In this context, it can be shown that finding the Pareto front of the problem is equivalent to maximising its *hypervolume indicator* [61,62], which is the m -dimensional volume between the region dominated by the Pareto front and a reference point [63]. Naturally, this allows the problem to be cast into a single-objective case, with hypervolume as the objective function. This approach allows the construction of the so-called Expected Hypervolume Improvement (EHVI) acquisition function, which is the multi-objective counterpart of the Expected Improvement (EI) acquisition function.⁹ There are known challenges in the evaluation of the hypervolume and its derivatives, particularly in high-dimensional problems, which require a partitioning of the objective space into hyper-rectangles. In this work, we adopt the EHVI acquisition function described in Daulton et al. [66], which allows the efficient and exact computation of the hypervolume and its gradients and the subsequent acquisition evaluation through Monte Carlo sampling. For noisy objective observations, the Noisy Expected Hypervolume Improvement (NEHVI) acquisition function can also be explored [67]. A schematic representation of the relevant concepts in multi-objective Bayesian optimisation is depicted in Fig. 8.

3. Hyperparameter estimation

Throughout Sections 2.2 and 2.3, the Bayesian optimisation strategy and its composite counterpart have been introduced and, crucially, two hyperparameters arose from the discussion: (i) the number of initial DoE points and (ii) the unexplained PCA variance threshold. Note that, while the budget for function evaluations needs to be defined by the user, this is considered a design choice rather than a hyperparameter. In this section, we discuss the estimation of the two hyperparameters mentioned above, which are crucial for the performance of the optimisation strategy.

⁸ Alternatively, to mimic minimisation, we can maximise the negated objective.

⁹ Other variants using the Probability of Improvement [64] and the Knowledge Gradient [65] acquisitions for hypervolume optimisation have also been explored in the literature, which, for brevity, have not been considered in this work.

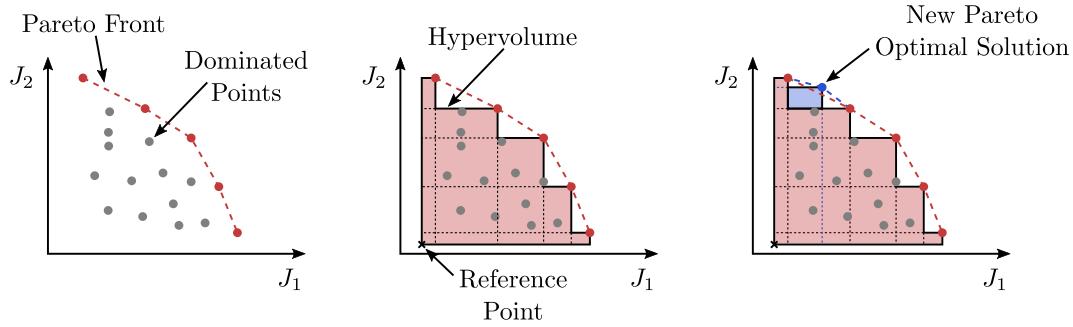


Fig. 8. Multi-objective optimisation problem: Pareto front, hypervolume and influence of new Pareto optimal solutions.

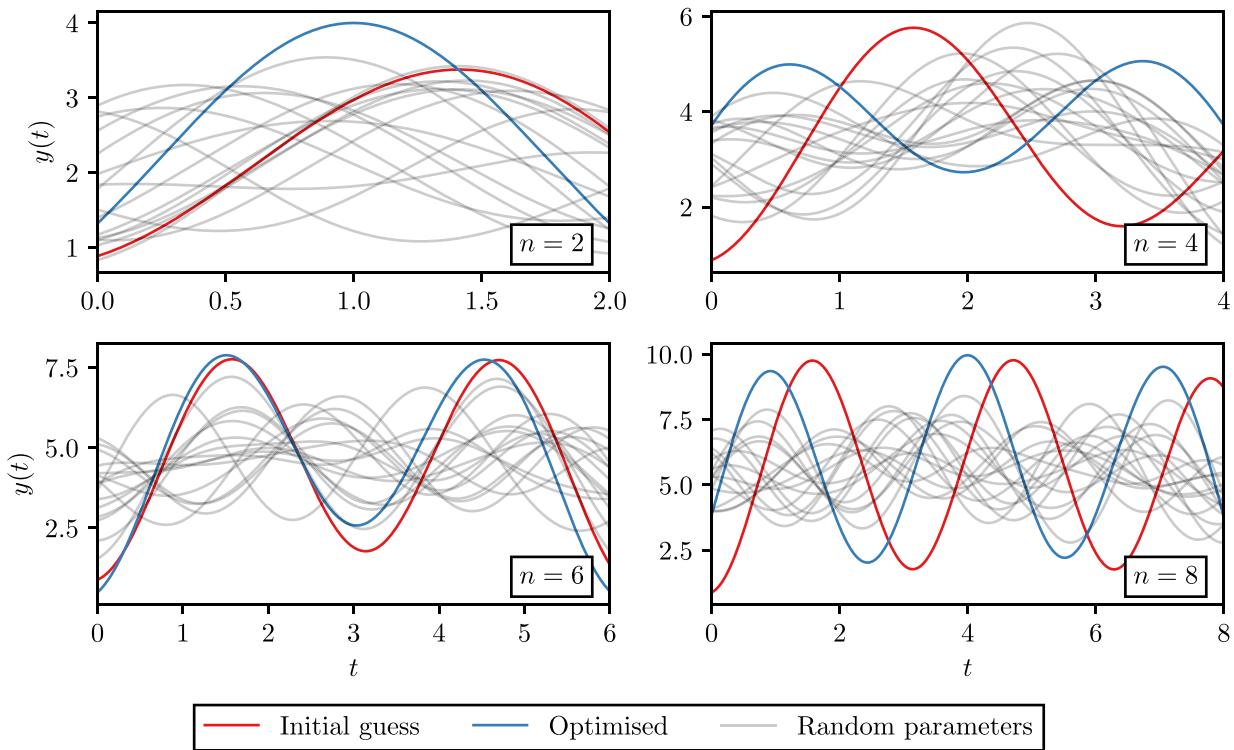


Fig. 9. Samples of analytical response from the objective function in Eq. (31) for different values of \$n\$. The objective value refers to the integral of this response in the represented domain. The initial guess refers to the midpoint of every optimisation dimension, whilst the optimised response has been selected from one of the realisations. In addition, some random samples are shown for comparison.

A numerical example has been adopted for the parametric studies of these hyperparameters. To allow for a varying number of dimensions in the design space, the following objective is considered:

$$\theta^* = \arg \max_{\theta \in \mathcal{B}} \left[\int_0^n f_n(\theta, t) dt \right], \quad \text{with } f_n(\theta, t) = \sum_{i=1}^n \cos^2(t - \pi\theta_i) + \sum_{i=1}^n \exp \left[-(t - \theta_{i+n})^2 \right], \quad (31)$$

where the bounds for the parameters are generically expressed as

$$\begin{cases} \theta_i \in [i-1, i], & \forall i \in \{1, 2, \dots, n\} \\ \theta_{i+n} \in [i-1, i], & \forall i \in \{1, 2, \dots, n\}. \end{cases} \quad (32)$$

Therefore, with this approach, depending on the parameter \$n\$, the design space is defined by the number of dimensions \$2n\$. Some examples of responses generated for different values of \$\theta\$ are shown in Fig. 9 for different values of \$n\$. All the optimisations were performed using composite Bayesian optimisation strategy, the LogEI acquisition function and a budget of 128 function evaluations.

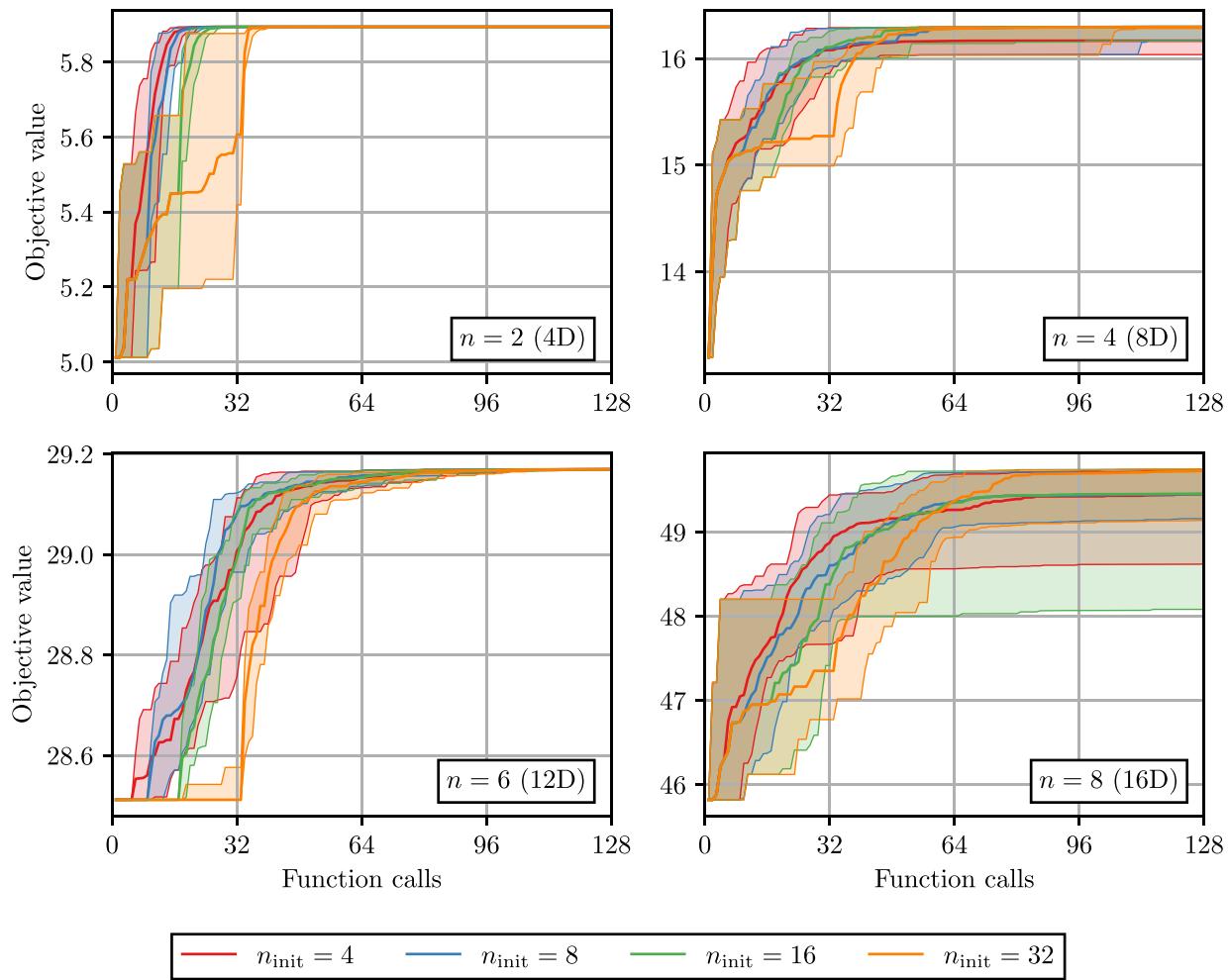


Fig. 10. Median, maximum and minimum best observed objective value for different numbers of initial DoE points for the four different problems.

3.1. Number of initial DoE points

For the selection of the number of initial DoE points, a parametric study has been conducted with the number of initial DoE points varying from 4 to 32. In addition, according to the definition in Eq. (31), the number of dimensions in the design space has been varied from 4 to 16 by choosing n between 2 and 8. This combination accounts for a total of 16 different problems and, for each, 20 realisations have been performed. In addition, PCA has been used with a threshold of 10^{-6} for the unexplained variance and 32 points are used for the discretisation of the numerical response.

The convergence of the median, maximum and minimum best observed objective values are shown in Fig. 10 for the four different problems and the four different numbers of initial DoE points. It can be observed that the median objective value generally converges for all of the problems. However, as the number of dimensions increases, the convergence is slower, and the number of initial DoE points has a more significant impact on the convergence. By visual inspection, the best median objective arises when n_{init} is increased, at the cost of a slower initial convergence due to additional random points. To further investigate this, the quantitative results are shown in Table 1 for the median best-observed loss, its minimum and maximum values, and the cumulative regret of the optimisation. The cumulative regret is calculated as the integral along the function calls axis of the difference between the best-observed objective value for the problem and its current value, and we seek to minimise this quantity [15]. The results further reinforce the visual inspection, showing that the best median objective value is generally obtained with a larger number of initial DoE points, but the convergence is slower — as evidenced by the higher cumulative regret. Notwithstanding, these effects are more pronounced as the number of dimensions increases and, for lower dimensional problems, the solution seems to be less sensitive to the number of initial DoE points.

With these results, it is possible to conclude that the number of initial DoE points has a significant impact on the optimisation performance. Particularly, it plays a critical role in the so-called exploration-exploitation tradeoff of Bayesian optimisation, particularly for high-dimensional problems. The results suggest that the number of initial DoE points should be increased as the

Table 1

Median best observed objective value for different numbers of initial DoE points. The minimum and maximum values are also shown, as well as the cumulative regret of the optimisation. The best values are highlighted in bold.

| Problem | n_{init} | Best objective value | | | Regret |
|----------------|-------------------|----------------------|----------------|----------------|----------------|
| | | Median | Minimum | Maximum | |
| 4D $n = 2$ | 4 | 5.8943 | 5.8943 | 5.8943 | 6.053 |
| | 8 | 5.8943 | 5.8943 | 5.8943 | 7.346 |
| | 16 | 5.8943 | 5.8943 | 5.8943 | 10.475 |
| | 32 | 5.8943 | 5.8943 | 5.8943 | 16.165 |
| 8D $n = 4$ | 4 | 16.1666 | 16.0363 | 16.2970 | 37.857 |
| | 8 | 16.2969 | 16.1661 | 16.2970 | 28.441 |
| | 16 | 16.2968 | 16.1666 | 16.2970 | 31.187 |
| | 32 | 16.2958 | 16.2816 | 16.2968 | 43.796 |
| 12D $n = 6$ | 4 | 29.1695 | 29.1691 | 29.1698 | 16.166 |
| | 8 | 29.1695 | 29.1688 | 29.1699 | 15.065 |
| | 16 | 29.1694 | 29.1688 | 29.1698 | 18.106 |
| | 32 | 29.1693 | 29.1687 | 29.1698 | 25.699 |
| 16D $n = 8$ | 4 | 49.4488 | 48.6238 | 49.7319 | 106.442 |
| | 8 | 49.4506 | 49.1587 | 49.7419 | 115.186 |
| | 16 | 49.4538 | 48.0814 | 49.7458 | 124.064 |
| | 32 | 49.7263 | 49.1386 | 49.7476 | 130.726 |

number of dimensions increases, but this should be balanced so as not to slow down the convergence of the optimisation. With this in mind, the following heuristic based on the number of optimisation dimensions n_{dim} is adopted for this work:

$$n_{\text{init}} = \begin{cases} 8, & \text{if } n_{\text{dim}} \leq 4, \\ 2n_{\text{dim}}, & \text{otherwise,} \end{cases} \quad (33)$$

that is, the number of initial points is the largest, between 8 and twice the number of dimensions in the design space. As evidenced in the results in Table 1, this heuristic provides a good balance between the exploration and exploitation of the design space, leading to a good convergence of the optimisation. Even so, the user should be aware that this heuristic is problem-dependent and, for specific cases, a different number of initial DoE points might be more appropriate. As a final note, since it is typical to have some domain knowledge of a good initial guess for the design parameters, we also consider the addition of this initial guess to the initial DoE. Thus, in practice, a dataset of $n_{\text{init}} + 1$ is used in this work for the upcoming examples.

3.2. Unexplained PCA variance threshold

To showcase the impact of the PCA transformation in the optimisation procedure, an experiment is carried out by selecting $n = 4$ in Eq. (31). As a result, the input design space comprises 8 dimensions and is characterised by the optimisation of the integral of a response $f(\theta, t)$ over the interval $[0, 4]$, and aims to mimic the optimisation of a response curve. To use the composite Bayesian optimisation framework, the response $f(\theta, t)$ must be transformed into a fixed-size vector according to the procedure described in Section 2.3.1. Particularly, the transformation depicted in Fig. 4(b) is adopted; therefore, the number of points for the interpolation of the response needs to be defined. In this experiment, the number of points varies from 8 to 128 on a logarithmic scale, and the integration is carried out numerically using a trapezoidal rule. Furthermore, the optimisation is carried out with and without the PCA transformation, and a value of 10^{-6} is selected as the threshold for the unexplained variance. Using 20 realisations, the median values for the best-observed objective value and the computational time after 128 function evaluations are presented in Fig. 11, along with their respective 95 % confidence intervals. The optimisation quality improves with PCA — the best objective value is higher and with a smaller dispersion, particularly for increasing numbers of points. This observation is justified by the remark in Section 2.3.3 that the posterior variance of the objective function is inversely proportional to the number of points in the latent response space and, therefore, the search becomes overly exploitative and may converge to a local optimum. Using PCA for this example, the number of principal components (PC) is usually 8 and seems insensitive to the number of interpolation points, which significantly emphasises exploration when compared to the non-PCA scenario. Finally, there is a notable reduction in the computational time when using PCA, which, in combination with the previous topics, largely justifies the use of PCA in the composite Bayesian optimisation framework.

As a concluding remark, while the PCA transformation is an optional component of the composite Bayesian optimisation framework, it has been shown to be crucial to improve the optimisation quality and efficiency. For this reason, it is applied to the remaining experiments in this work. The choice of the variance threshold is discussed in Section 3.2. By efficiently using this technique, the following advantages are achieved:

- The dimensionality of the latent response space is reduced, which drastically reduces the computational cost of the method;
- The reduced latent space is composed of uncorrelated features, which allows the assumption of independence between the components of the latent response space and, therefore, allows using multiple independent single-task Gaussian processes. This has been observed to improve the exploration-exploitation balance of the optimisation and the convergence of the method;

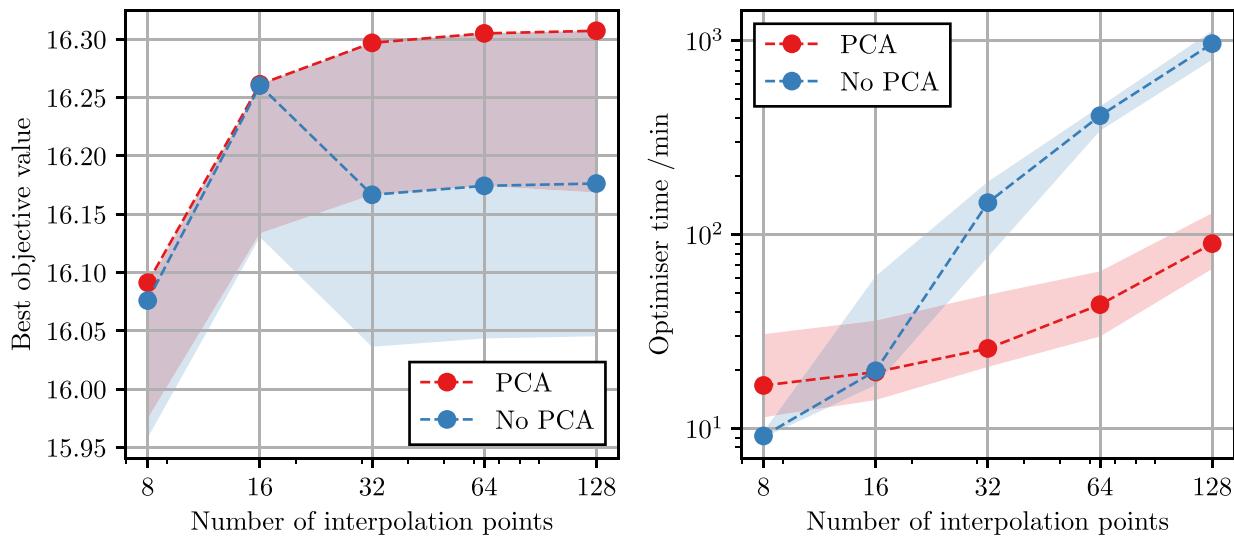


Fig. 11. Median best objective value and average computational time for the optimisation of the test problem with different numbers of points for interpolating the response, with and without PCA.

Table 2

The mean number of principal components (PC), median time and median best objective for different values of unexplained variance were determined using PCA with 32 points for interpolation. For each quantity, 95 % confidence intervals are shown. The case without PCA is also presented for comparison.

| Unexplained variance | Number PC Mean \pm 95 % CI | Time /min Median | Best objective | |
|----------------------|-----------------------------------|---------------------|----------------|---------------|
| | | | 95 % CI | Median |
| 10^{-2} | 3.67 ± 0.95 | 13.3 | [11.9, 14.4] | 16.186 |
| 10^{-3} | 4.98 ± 0.70 | 14.4 | [13.2, 18.6] | 16.293 |
| 10^{-4} | 5.98 ± 0.28 | 24.5 | [16.3, 42.9] | 16.297 |
| 10^{-5} | 7.05 ± 0.91 | 20.9 | [15.7, 38.5] | 16.297 |
| 10^{-6} | 7.97 ± 0.35 | 26.0 | [19.7, 45.6] | 16.297 |
| 10^{-7} | 8.38 ± 0.94 | 27.2 | [19.6, 47.2] | 16.296 |
| No PCA | - | 146.5 | [76.8, 187.7] | 16.167 |

- In the curve design scenario, the hyperparameter N can be set to a high value without a significant increase in the computational cost — the PCA automatically selects the number of appropriate components in the reduced latent space.

For completeness, the same problem is optimised with the PCA transformation, but now for a varying threshold for the unexplained variance. Choosing the number of points for the interpolation as 32, the optimisation is carried out with logarithmic-spaced thresholds from 10^{-2} to 10^{-7} , whose results for 20 realisations are presented in Table 2. As expected, the number of principal components and computational time are correlated with the threshold for the unexplained variance — the smaller the threshold, the more components are required to explain the variance of the data and, consequently, the higher the computational cost. Furthermore, larger thresholds lead to a worsening of the optimisation quality due to an increasing information loss during the PCA transformation. In this example, the threshold of 10^{-6} is selected as a conservative compromise between optimisation quality and computational cost, and is used for the remaining experiments in this work.

4. Examples

Throughout this section, a comparative analysis of the proposed composite Bayesian optimisation strategy is conducted for distinct material and structural design problems. This strategy is compared with three other baseline optimisers in the literature, namely:

Classical Bayesian optimisation: the standard Bayesian optimisation strategy for the black-box function $J(\theta)$. For this task, we use a BoTorch implementation of the method, similar to the composite setting, which shares both the same number of initial design points and acquisition functions.

Random sampling: a naive random sampling strategy based on Sobol sequences for exploration of the design space. All function evaluations are conducted with random points, and the best-observed solution is tracked throughout the optimisation procedure.

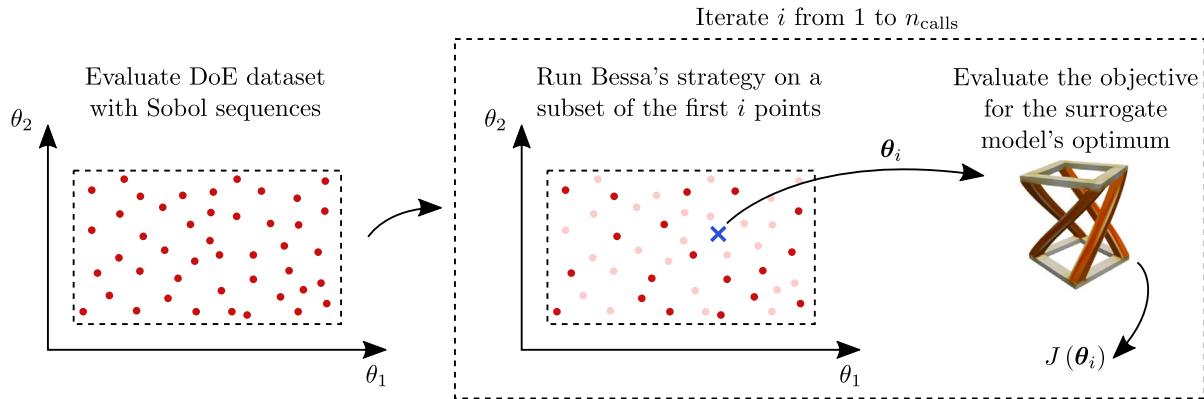


Fig. 12. Schematic representation of the modifications to the Bessa surrogate strategy to ease the comparison with the remaining methods.

Bessa surrogate: the design strategy proposed by Bessa and coworkers [24], which uses a surrogate model to guide the design problem. This framework is aimed at the class of problems where the objective function may not be known beforehand, since the optimisation is only carried out at a later stage without requiring re-evaluating the objective function. We include this approach in the comparative study to highlight the increased efficiency in optimisation when the objective function is known and we guide the sampling process towards finding the optimum. This strategy requires:

1. Generation of an initial dataset using DoE strategies, such as Sobol sequences. This step requires evaluating the objective function for each DoE point and it is assumed that the entire budget for function evaluations has been allocated for this task.
2. Construction of a surrogate model for the observations of the objective function. Various options are available for this task, such as Gaussian processes, neural networks, or support vector machines. Nevertheless, in this work, we use a Gaussian process model for the surrogate due to its straightforward training and availability using the BoTorch package.
3. Optimisation of the surrogate model to find the optimal solution. As before, several methods are available for this step, and depending on the chosen model, even gradient-based methods can be considered. For integration with our Bayesian optimisation framework, we use the same gradient-based multi-start L-BFGS-B optimiser to find the optimal posterior mean of the Gaussian process surrogate. Note that this optimisation is generally cheap, so less sample-efficient methods such as particle swarm and genetic algorithms can also be considered.

Additionally, since we intend to assess the convergence quality of this method, a variant of this framework is considered to ease the comparison with the remaining methods. Firstly, to lessen the computational cost of the examples, we use the dataset generated by the random sampling optimiser as the DoE for the Bessa surrogate.¹⁰ Secondly, since the optimal point is obtained from optimising the surrogate model and, therefore, can be considered as a low-fidelity sample, we re-evaluate the optimal point using the true objective function to ensure a fair comparison with the remaining methods. Finally, to be able to compare convergence rates, one is required to have the history for the evaluated objectives $J(\theta)$. Since this approach assumes that the optimisation only occurs after the DoE dataset is built, that is, after the budget of function evaluations is exhausted, we simulate stopping the generation of the DoE dataset and optimising the Bessa surrogate. In other words, at the i th iteration, a surrogate model is built using only the first i DoE points. Then, the optimal point is found using the surrogate model, and the true objective function is evaluated. This procedure is repeated until the budget for function evaluations is exhausted. A schematic representation of the modifications to the Bessa surrogate strategy is illustrated in Fig. 12. This allows extracting the convergence history $J(\theta_i)$ of the Bessa surrogate strategy and comparing it with the remaining methods.

Since all methods share randomness in the initial design points, each design problem is repeated 20 times for each method to ensure the statistical significance of the results. Similarly to the work of Cardoso Coelho et al. [15], we focus on the median of the various metrics from the optimisation procedure, as well as their 95 % confidence intervals, to assess the performance of the different design strategies. The confidence interval for the median, for a sample size of $n = 20$, is between the observations $nq \pm z\sqrt{nq(1-q)}$, in the ordered sample data, being q the quantile of interest ($q = 0.5$ for the median) and z the z-critical value, which for a 95 % confidence interval is $z = 1.96$ [68].

Finally, to ensure an adequate coverage of the class of problems of interest, we consider four distinct design problems. These problems encompass: (i) the optimisation of an analytical response to maximise the integral of a function (see Section 4.1), (ii) the

¹⁰ Note that this option does not degrade the performance of this method in any way; it simply avoids regenerating a DoE dataset from scratch for the two optimisers.

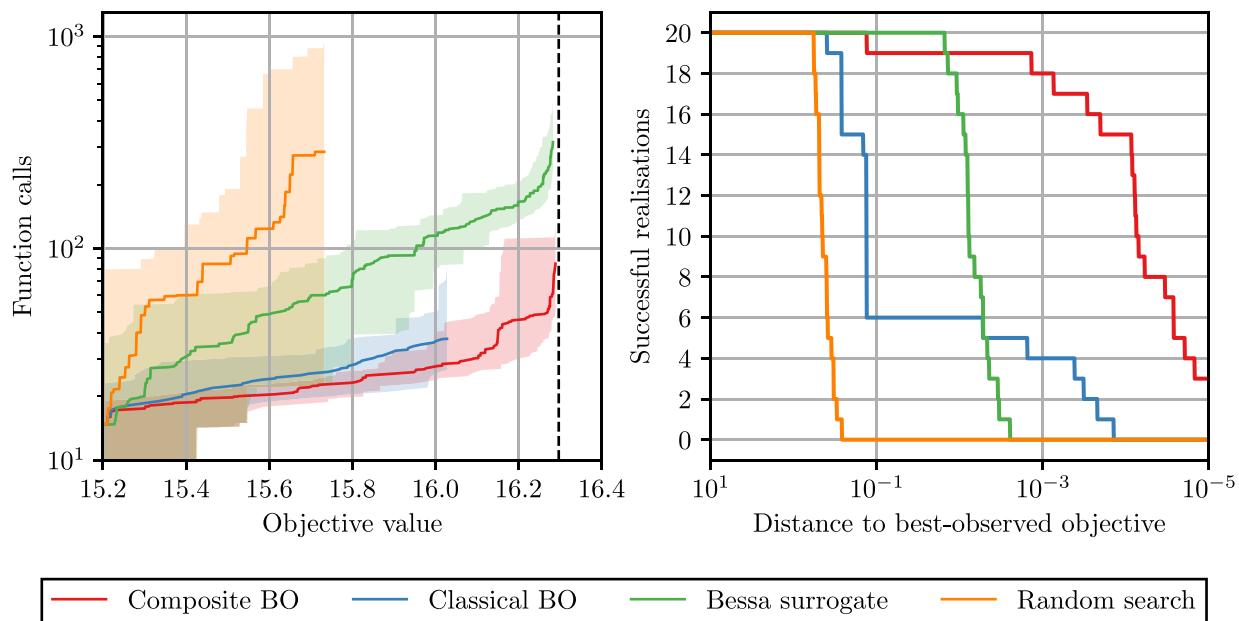


Fig. 13. Median number of function evaluations to reach a given objective value and associated success rate for the analytical design problem. On the left-hand side, the best-observed objective value for all the optimisers in the 20 realisations is also illustrated. The right-hand side of the figure illustrates how many realisations of the optimisation problem reached a given objective value, measured by the distance to the best-observed objective of the 20 realisations.

identification of the hardening parameters of a 3D inelastic particle-matrix RVE (see Section 4.2), (iii) the identification of structural parameters of a compression-torsion coupling metamaterial (see Section 4.3) and (iv) the identification of 16 structural parameters of a bending beam (see Section 4.4). All these examples have been optimised with *piglot* [44]. In addition, unless otherwise stated, the numerical simulations are conducted with the in-house finite element code *Links* (Large Strain Implicit Nonlinear Analysis of Solids Linking Scales), which allows for solving both macroscopic and microscopic single-scale problems, as well as coupled multi-scale problems at large strains.

4.1. Design of an analytical response

To showcase the performance of the proposed composite Bayesian optimisation strategy, we start by recalling the optimisation of the analytical response described in Section 3, characterised by the objective function given in Eq. (31) with $n = 4$. This eight-dimensional optimisation problem focuses on the maximisation of the integral of the function. For each optimiser, 20 realisations have been conducted to extract statistical information about the performance of the different methods. In this comparative example, we focus on finding the number of function evaluations required for the optimiser to reach a given target objective value, as well as the associated success rate. Therefore, an arbitrarily large budget of function evaluations is considered for each optimiser, which has also been adjusted to avoid a high computational cost — BO approaches are limited to 128 function evaluations, Bessa's surrogate strategy to 512 and the random sampling to 1024.

The results for this design problem are compiled in Fig. 13. On the left-hand side, the relationship between an objective value and the median number of function evaluations required to reach this objective value is illustrated. For this example, 95 % confidence intervals are considered for both the dispersion of the function calls and the cutoff point for the objective value.¹¹ The best-observed objective value for all optimisers and realisations is also depicted. These results reinforce the sample efficiency of the Bayesian optimisation strategies and, particularly, how the composite strategy drastically outperforms the remaining methods. For certain target objective values, the composite strategy may require a full order of magnitude fewer function evaluations than the Bessa surrogate method. Finally, on the right-hand side of the figure, the success rate for reaching a given objective value is illustrated. In this figure, the objective value is quantified by the distance to the best-observed objective of the 20 realisations for all optimisers, which allows for a more accurate comparison when approaching the presumably optimal solution. It can be observed that the composite strategy retains a significantly higher success rate as the objective value increases than the remaining methods. As a concluding remark, this example demonstrates the increased sample-efficiency of the composite Bayesian optimisation strategy, which allows for reaching target objective values with fewer function evaluations than the remaining methods.

¹¹ This indicates that the maximum values depicted in this plot are not necessarily the best-observed objective values by a given optimiser; the maximum value in the plot is related to the appropriate quantile for computing the confidence interval.

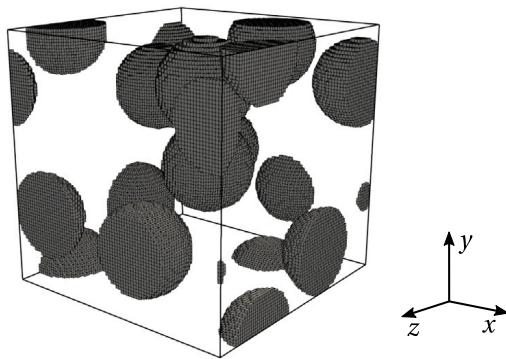


Fig. 14. Geometry and mesh of the matrix-particle Representative Volume Element used to identify two hardening parameters of the particles in order to maximise the toughness, U_T , of the composite.

Table 3

Material parameters of the von Mises elasto-plastic constitutive model using the hardening law expressed in Eq. (34), for both the matrix and the particles of the 3D inelastic particle-matrix RVE.

| Parameter | Matrix | Particles |
|---------------------------|--------|-----------|
| E/GPa | 100 | 500 |
| ν | 0.3 | 0.19 |
| $\sigma_{Y,0}/\text{MPa}$ | 100 | a |
| β_1/MPa | 300 | b |
| β_2 | 0.4 | 0.2 |

4.2. Design of a 3D inelastic particle-matrix RVE

In this example, a three-dimensional particle-matrix RVE is considered to identify two hardening parameters of the particles to maximise the toughness of the composite material, inspired by the work of Bessa et al. [24]. The RVE is assumed to be composed of 10 randomly distributed periodic spherical particles, with a particle volume fraction of 20 %, where a voxel grid of $80 \times 80 \times 80$ is considered, as illustrated in Fig. 14. Furthermore, the RVE is generated with the in-house microstructure generation program for particle-reinforced materials, GMMD, based on an adaptive multi-temperature isokinetic time-driven molecular dynamics scheme [69]. To match the original example in Bessa et al. [24], for this case, all the numerical simulations are conducted with the SCA clustering-based reduced-order model, available as an open-source package, CRATE [70].

Both phases of the composite material are assumed inelastic following the von Mises elasto-plastic model, with a hardening law given by

$$\sigma_Y = \sigma_{Y,0} + \beta_1 (\bar{\varepsilon}^p)^{\beta_2}, \quad (34)$$

where $\bar{\varepsilon}^p$ is the accumulated plastic strain, $\sigma_{Y,0}$ is the initial yield stress and β_1 and β_2 are two hardening parameters. The material properties for each phase are compiled in Table 3, where the parameters a and b are related to the yielding and hardening parameters $\sigma_{Y,0}$ and β_1 of the particles, respectively. Additionally, a simple fracture model is adopted [24]:

- The composite fails when 10 % of the matrix phase reaches a maximum strain component of 0.07;
- The material is considered brittle;
- The particles are considered to fail at a significantly higher strain level, thus the composite failure is uniquely due to the matrix failure.

In this example, the two hardening parameters a and b of a three-dimensional particle-matrix RVE are optimised to maximise the toughness of this composite material, that is,

$$\max_{\theta} J(\theta) = \max_{a,b} J(a, b) = \max_{a,b} U_T(a, b), \quad (35)$$

where the toughness U_T of the composite material is defined as the integral of the stress-strain curve obtained in a uniaxial tension test before fracture. Using a trapezoidal rule for approximating the integral, the toughness is calculated as

$$J(a, b) = \int_0^{\varepsilon_f} \sigma(\varepsilon) d\varepsilon \approx \sum_{i=1}^n \frac{\sigma_i + \sigma_{i-1}}{2} (\varepsilon_i - \varepsilon_{i-1}), \quad (36)$$

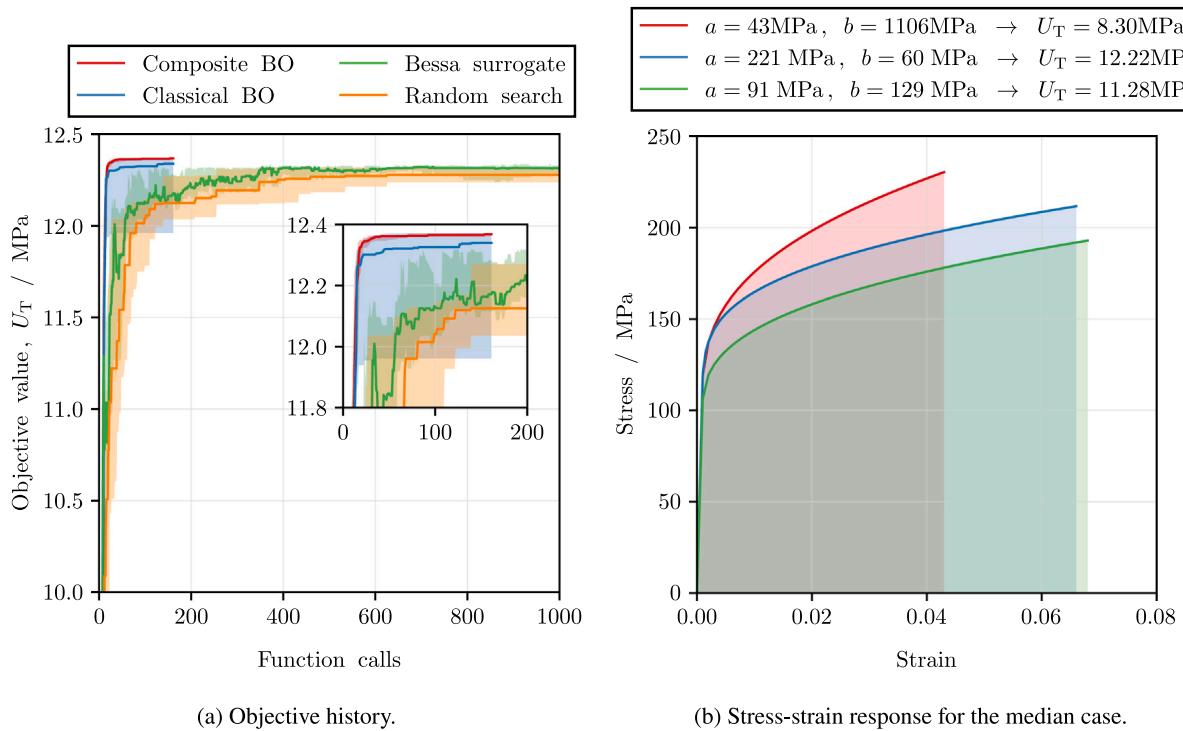


Fig. 15. Results of the design problem of a 3D matrix-particle RVE used to identify two hardening parameters (a and b) in order to maximise the toughness (U_T) of the composite: (a) median toughness evolution for different design strategies and corresponding 95% confidence bounds; (b) stress-strain response, and corresponding toughness, of the RVE for three different sets of hardening parameters.

where ϵ_f is the fracture strain, $\sigma(\epsilon)$ is the stress-strain curve, and n is the number of points in the stress-strain curve.¹²

Considering the stress-strain curve as the output of the response functional S , we can identify the compositional nature of the objective function via

$$J(a, b) = \mathcal{L}(a, b, S[a, b]) = \text{Trapz}(S[a, b]), \quad (37)$$

where the trapezoidal rule is denoted by the operator Trapz, given by,

$$\text{Trapz}(S[a, b]) = \sum_{i=1}^n \frac{\sigma_i + \sigma_{i-1}}{2} (\epsilon_i - \epsilon_{i-1}), \quad (38)$$

whose gradients are computable. By adopting this composition, the composite Bayesian optimisation strategy observes the entire stress-strain curve of the RVE, and the objective is computed using the reduction function \mathcal{L} . This format explicitly highlights the composition of the objective function, where the inner part is the numerical response of the simulation.

The following bounds are considered for the two input design variables:

$$a = [10, 700] \text{ MPa}, \quad b = [50, 2000] \text{ MPa}. \quad (39)$$

In addition, the initial values of $a = 350$ MPa and $b = 1000$ MPa are adopted. It is worth emphasising that the complexity of this design problem is a result of two competing factors, namely the ductility and the strength of the composite material. Both these phenomena strongly affect the overall toughness of the material. In particular, if the particles are characterised by a high yield strength and hardening, the strength of the composite material increases, but the ductility decreases due to a high concentration of strain in the matrix phase. In contrast, if the particles have low yield strength and hardening, the ductility of the material increases at the expense of decreasing its strength. Likewise, it is of utmost interest to find a compromise between the strength and ductility of the particles so that the composite material has maximum toughness.

The results for this design problem are compiled in Fig. 15. The maximum number of function evaluations for each optimiser are defined by — BO approaches are limited to 159 function evaluations, Bessa's surrogate strategy to 1026 and the random sampling to 1025. In Fig. 15(a), the median toughness evolution (U_T) for the different design strategies is illustrated along with the corresponding 95% confidence bounds. In Fig. 15(b), the stress-strain response of the 3D inelastic RVE and its corresponding toughness value are represented for three different sets of parameters.

¹² Note that the dependency of the fracture strain and the stress-strain curve on the parameters a and b has been dropped, for notational convenience.

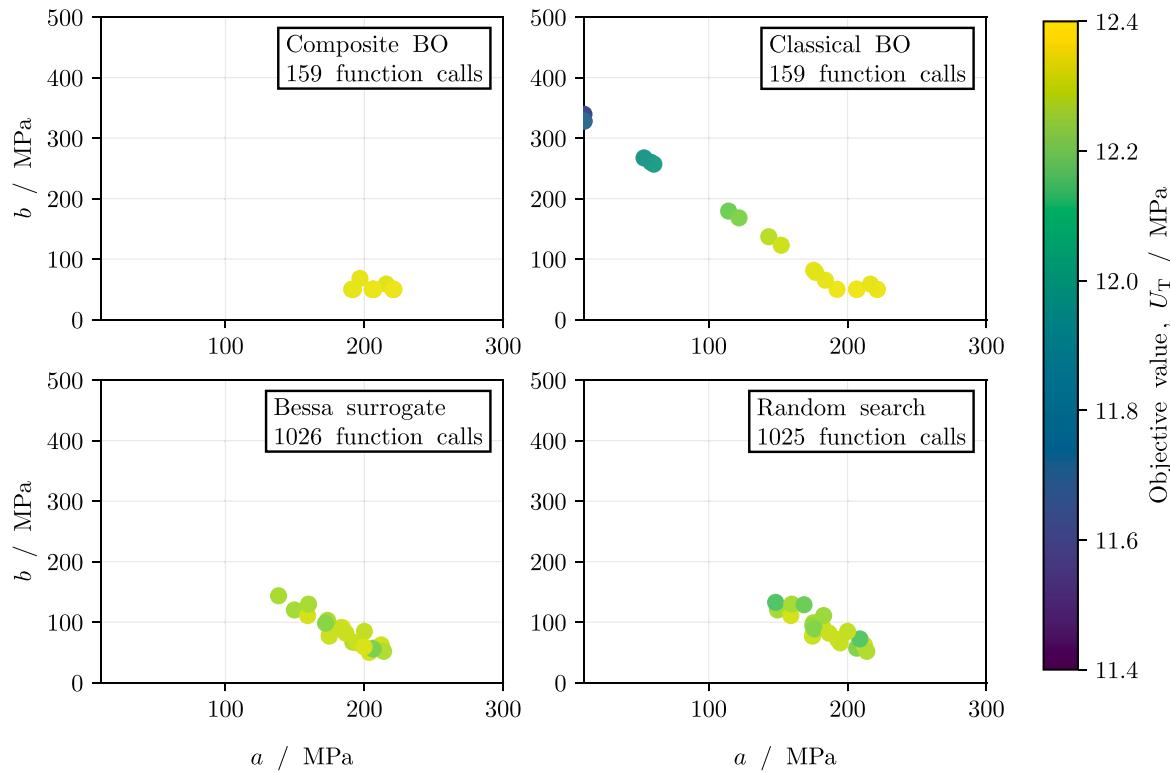


Fig. 16. Distribution of the optimal hardening parameters obtained by the different design strategies in 20 runs of the optimisation problem. Recall that the bounds for the parameters are: $a \in [10, 700]$ MPa and $b \in [50, 2000]$ MPa.

From these results, it is observed that, as expected, the increase of both hardening parameters promotes the increase of strength of the material at the expense of decreasing its ductility. Moreover, it is observed that, from the four design strategies, the composite Bayesian optimisation strategy consistently finds the best solution for this design problem, pinpointing the hardening parameters that provide higher toughness of the material. In addition to its excellent accuracy, this strategy demonstrates low variance between different runs. In contrast, the Bessa surrogate design strategy is not as accurate as the Bayesian optimisation techniques, independently of the number of points in the initial dataset. Recall that while the BO strategies leverage the information of the objective function to guide the design problem, the Bessa surrogate employs a larger initial dataset (not guided by the design objective), which is then cheaply optimised.

In Fig. 16, the distribution of the optimal hardening parameters obtained by the different design strategies in 20 runs of the optimisation procedure is illustrated. From these results, the consistency of the composite Bayesian optimisation is again demonstrated, with all optimal solutions falling within the same region of the design input space. Moreover, it is observed that the classical Bayesian optimisation demonstrates some variation between different runs, and that due to the elevated number of function evaluations, both the Bessa surrogate strategy and the Random Sampling strategy provide similar optimal distributions. Notably, almost all optimal solutions are contained in the lower bound of the b hardening parameter. Recall that the bounds for the parameters are: $a = [10, 700]$ MPa and $b = [50, 2000]$ MPa, which suggests that the lower bound of the b parameter should be reduced. However, in the present contribution, the bounds of the parameters are kept equal to the ones adopted by Bessa et al. [24].

4.3. Design of structural parameters of a compression–torsion coupling metamaterial

This example focuses on the geometric optimisation of a three-dimensional RVE of a chiral compression–torsion coupling (CTC) metamaterial with inclined rods, as described in the work of [71]. Due to these inclined rods, the RVE exhibits macroscopic compression–torsion coupling behaviour — compressive loadings induce torsion relative to the compression axis. The unit cell is composed of two square loops with four inclined rods connecting their vertices, as depicted in Fig. 17. The rods are inclined at an angle θ relative to the x–y plane, and the side length of the square loops is denoted as l_μ . The thickness t is specified by the parameter β , which is a scalar coefficient computed as the ratio of the thickness, t , to the side length of the square loop, l_μ ,

$$\beta = \frac{t}{l_\mu}. \quad (40)$$

While the square loops feature a quadrangular cross-section with side t , the thickness of the tie-rods is dependent on the angle θ and is given by $t \sin \theta$. Additionally, two different materials are considered for each of these components, where the square loops are

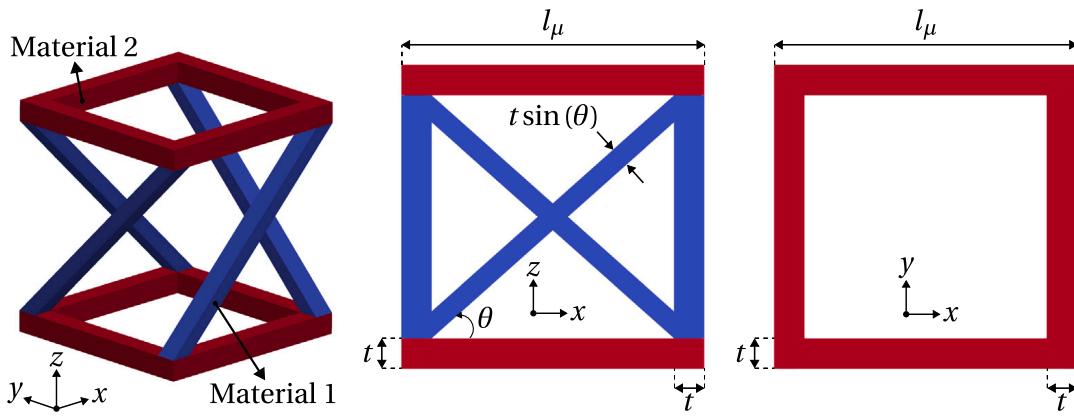


Fig. 17. Geometry of the unit cell of the metamaterial.

Table 4
Material properties of the inclined rods.

| Young's modulus, E_1/GPa | Volumic mass, $\rho_1/\text{kg m}^{-3}$ | Poisson's ratio, ν_1 |
|--------------------------------------|--|-----------------------------|
| 210 | 7850 | 0.3 |

assumed to be rigid and the inclined rods are deformable beams [72,73]. For all of these examples, it is assumed that the bottom square loop is fixed, whilst the top square loop has a prescribed displacement in the z-direction but is free to move and rotate in the other directions.

For this problem, two geometric parameters are considered as design variables: the angle of the inclined rods, θ , and the thickness coefficient, β . The generation of the finite element meshes for this problem has been automated for the given parameters θ and β , where 20-noded hexahedral elements with reduced integration are used. The cross sections of the square loops and the inclined rods are discretised regularly with 3 elements per side (n_{el}). Along their length, the number of elements is not fixed; rather, it is determined by the expression $\lfloor n_{\text{el}} (l_\mu - 2t) / t \rfloor + 2n_{\text{el}}$. This setting has been identified as sufficient for the convergence of the results. Thus, three examples are analysed: (i) the minimisation of the mass and maximisation of the infinitesimal stiffness, (ii) the maximisation of the specific absorbed energy, and (iii) the minimisation of the mass and maximisation of the absorbed energy. The first example is performed using infinitesimal strains, while the second and third examples are performed using a finite strain formulation.

4.3.1. Minimisation of the mass and maximisation of the infinitesimal stiffness

The first example is a multi-objective optimisation problem that aims to minimise the mass and maximise the infinitesimal stiffness of the CTC metamaterial. The Pareto front solution from the optimisation methods is compared with the analytical solution derived in [Appendix](#). The objective functions are defined as:

$$\max_{[\theta, \beta] \in \mathcal{B}} [J_1(\theta, \beta), J_2(\theta, \beta)] = \max_{[\theta, \beta] \in \mathcal{B}} [-m(\theta, \beta), k_z(\theta, \beta)]. \quad (41)$$

where the mass m is calculated analytically using Eq. (82), and the infinitesimal stiffness k_z is calculated numerically from simulations. We start by prescribing a small compressive displacement of $\delta_z = \epsilon H$ at the top square loop, where H is the height of the unit cell

$$H = l_\mu [(1 - \beta) \tan(\theta) + 2\beta], \quad (42)$$

and $\epsilon = 0.01\%$. Extracting the reaction force F , then the infinitesimal stiffness can be computed via the slope of the force-displacement curve with

$$k_z(\theta, \beta) = \frac{F_z(\theta, \beta)}{\delta_z(\theta, \beta)}. \quad (43)$$

This prescription is specified using a single load increment in the numerical simulations, such that the output response is a two-point curve with the points $(0, 0)$ and (δ_z, F_z) . The material properties of the inclined rods are presented in [Table 4](#).

For compatibility with the composite Bayesian optimisation strategy, the objective function is defined as

$$[J_1(\theta, \beta), J_2(\theta, \beta)] = \left[-m(\theta, \beta), \frac{F_z(\theta, \beta)}{\delta_z(\theta, \beta)} \right] \quad (44)$$

$$= \left[-m(\theta, \beta), \frac{\text{Max}(S[\theta, \beta])}{\delta_z(\theta, \beta)} \right] \quad (45)$$

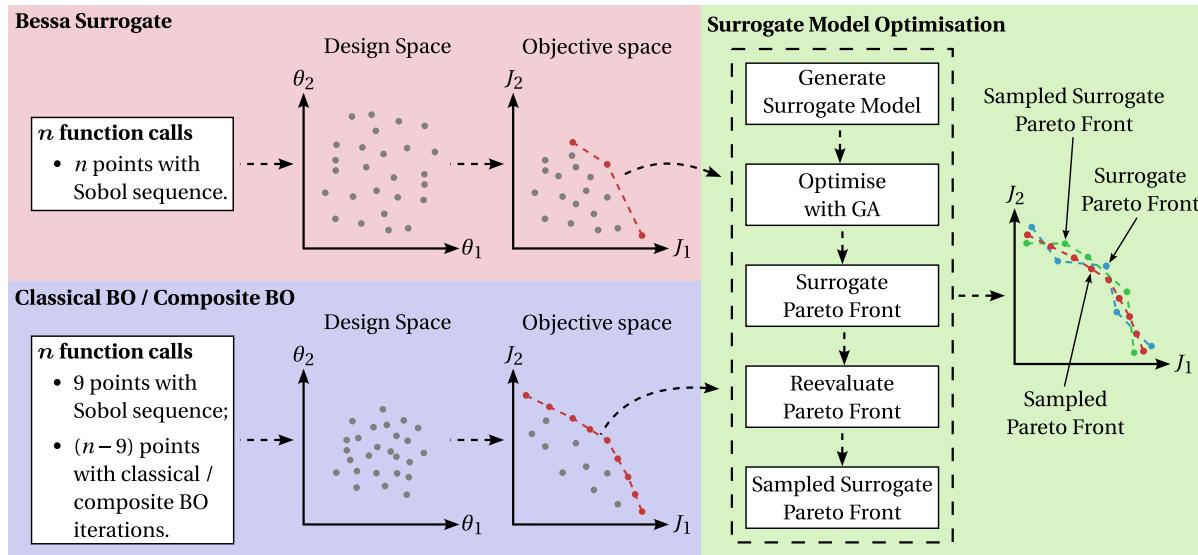


Fig. 18. Schematic representation of the design strategies.

$$= [-m(\theta, \beta), \mathcal{L}(\theta, \beta, S[\theta, \beta])] \quad (46)$$

where \mathcal{L} is the outer composition function of the objective $J_2(\theta, \beta)$ of the multi-objective problem. For this case, we compute the vertical force F_z by extracting the maximum point of the force–displacement curve from the response functional $S[\theta, \beta]$. With this approach, both the mass m and the prescribed displacement δ_z are computed directly from the analytical outer composition and do not need to be learned by the surrogate model. The bounds of the input design variables, the angle θ and the thickness coefficient β , are

$$\theta = [30, 80]^\circ, \quad \beta = [0.05, 0.15], \quad (47)$$

respectively, and the initial values are set to the midpoints $\theta = 55^\circ$ and $\beta = 0.1$.

To compare the methods, a total of 64 function evaluations are considered. For the Bayesian optimisation strategies, both the classical and composite variants start with an initial dataset of 9 points (initial guess and 8 Sobol sequence points). In the Bessa surrogate strategy, 64 points are randomly sampled from the design space with Sobol sequences. A schematic representation of the design strategies for multi-objective optimisation is presented in Fig. 18. In this context, we introduce and compare three different Pareto fronts:

- Sampled Pareto Front — corresponds to the Pareto front obtained by selecting the non-dominated points in the dataset generated by each optimiser.
- Surrogate Pareto Front — using a surrogate model for the dataset generated by the optimisers, a multi-objective optimisation is performed on this model to obtain an approximate Pareto front for the problem. This is particularly useful since the surrogate model can be cheaply evaluated. Note, however, that these points are guesses from the surrogate model and may not correspond to the true Pareto front.
- Sampled Surrogate Pareto Front — given the Surrogate Pareto front, we re-evaluate the points using the finite element model to obtain a more accurate representation of the Pareto front and assess the quality of the surrogate model.

This approach enhances the fairness of comparison between methods. For instance, directly comparing Sampled Pareto fronts may not be adequate between BO and Bessa surrogate strategies, as the latter has no knowledge of the problem during the sampling stage. Additionally, comparisons with the Surrogate Pareto front may introduce the additional error of the surrogate model into the analysis, since the evaluations are typically low-fidelity approximations of the actual problem. Finally, the Sampled Surrogate Pareto front is expected to provide a more accurate representation of the Pareto front, as it is obtained by re-evaluating the points using the finite element model. However, in practical applications, one typically compares the Sampled Pareto front of the BO strategies with the Surrogate Pareto front of the Bessa surrogate strategy, as these represent the outputs of the optimisation methods.

The Surrogate Pareto front is computed using the well-known NSGA-II algorithm, as implemented in pymoo [74]. This genetic algorithm is also used to determine the Pareto front for the analytical solution. For the genetic algorithm, 40 generations are executed, starting with an initial population of 40 individuals and producing 20 offspring in subsequent generations. This setup ensures convergence to the Pareto front with a total of 820 function evaluations. Both the Surrogate Pareto front and the Sampled Surrogate Pareto front contain 40 points, one for each generation. Although a larger number of points could be selected due to the low computational cost of the genetic algorithm, reevaluating the Surrogate Pareto front is computationally expensive as it requires a finite element simulation for each point. It is important to note that the surrogate model is constructed solely to facilitate a fair comparison of strategies and is not in the optimisation process.

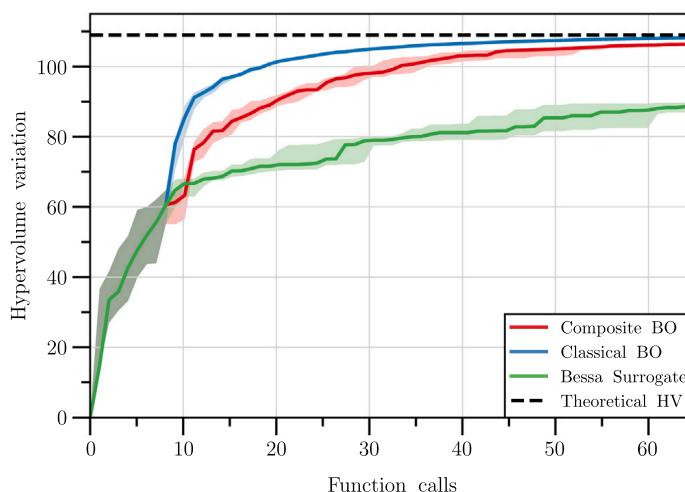


Fig. 19. Median of the Sampled Pareto front's HV variation with an increasing number of function calls for the multi-objective problem of the 3D CTC metamaterial under infinitesimal strains for the minimisation of the mass (m) and the maximisation of the stiffness (k_z) with respect to the geometrical parameters θ and β .

Table 5

Median hypervolume, generational distance and the ratio of non-dominated to dominated solutions for the different design strategies in the multi-objective optimisation problem of the 3D CTC metamaterial under infinitesimal strains.

| Design strategy | Sampled PF | | | Surrogate PF | | Sampled Surrogate PF | | |
|-----------------|-----------------|---------------|---------------|-----------------|---------------|----------------------|---------------|------|
| | HV | GD | ND/D | HV | GD | HV | GD | ND/D |
| Bessa Surrogate | 88.6007 | 19.5990 | 1.1317 | 106.3100 | 3.0359 | 106.6140 | 2.9167 | 40 |
| Classical BO | 108.2127 | 1.8191 | 8.2857 | 106.5828 | 2.9020 | 106.5780 | 2.9443 | 40 |
| Composite BO | 106.4507 | 3.7332 | 3.3333 | 106.6367 | 3.0461 | 106.5761 | 3.0328 | 40 |

Regarding evaluation metrics, the design strategies are assessed using hypervolume (HV), Generational Distance (GD) [75], and the ratio between non-dominated and dominated solutions. The HV metric is a widely recognised performance indicator in multi-objective optimisation, measuring the volume of the dominated space covered by the Pareto front. Higher HV values indicate better-quality solutions. This metric does not require knowledge of the optimal Pareto front but does need a reference point, defined as the worst possible point for each objective (see Fig. 8). In this example, the reference point for the hypervolume metric is set to $[m_{\max}, k_{z\min}]$. Considering the previous bounds, the maximum mass is approximately 0.576 mg, and the minimum stiffness is 0 N mm $^{-1}$. Therefore, the chosen reference point is [0.58, 0]. On the other hand, the GD metric requires a reference Pareto front for calculation and evaluates the convergence of the obtained Pareto front to the true Pareto front. It calculates the Euclidean distance from each point on the reference Pareto front (derived from the analytical solution) to the nearest point on the Pareto front produced by the design strategy. Smaller GD values suggest a better approximation of the true Pareto front. However, it is important to note that the analytical model is an approximation of the real structure and will inherently have some degree of error and, thus, this metric should be interpreted with caution. Lastly, the ratio between non-dominated and dominated solutions is a straightforward metric that assesses the effectiveness of the design strategy in identifying non-dominated solutions. A higher ratio should indicate a more efficient sampling of the design strategy.

The median hypervolume convergence for the Sampled Pareto front is shown in Fig. 19. The Pareto fronts produced by both the classical and composite Bayesian optimisation strategies are significantly superior to those generated by the Bessa surrogate strategy. Additionally, both Bayesian strategies exhibit much less dispersion, with the classical Bayesian strategy achieving a slightly higher hypervolume value in this instance. Furthermore, Fig. 20 illustrates the Pareto fronts for the various strategies, corresponding to the evaluation where the hypervolume of the Sampled Pareto front is highest. All strategies successfully recovered the analytical solution with the Sampled Surrogate Pareto front. However, it is noteworthy that the Bayesian strategies fully recovered the analytical solution on the initial computation of the Pareto front (Sampled Pareto front). This presents a significant advantage, as the number of function calls is considerably reduced since there is no need to construct and subsequently reevaluate a Surrogate Pareto front. Table 5 presents the median hypervolume, generational distance, and the ratio of non-dominated to dominated solutions for all Pareto fronts across the different strategies. The results show that for the Sampled Pareto front, both Bayesian strategies outperform the Bessa surrogate strategy in terms of hypervolume and generational distance. They also have a higher ratio of non-dominated to dominated solutions, demonstrating greater efficiency in identifying non-dominated solutions. For the Surrogate and Sampled Surrogate Pareto fronts, the results are quite similar across all strategies. Due to the simplicity of the problem, the composite Bayesian optimisation strategy does not particularly stand out in this example. However, as will be shown in the following examples, the composite Bayesian optimisation strategy is expected to be more advantageous in more complex and nonlinear problems.

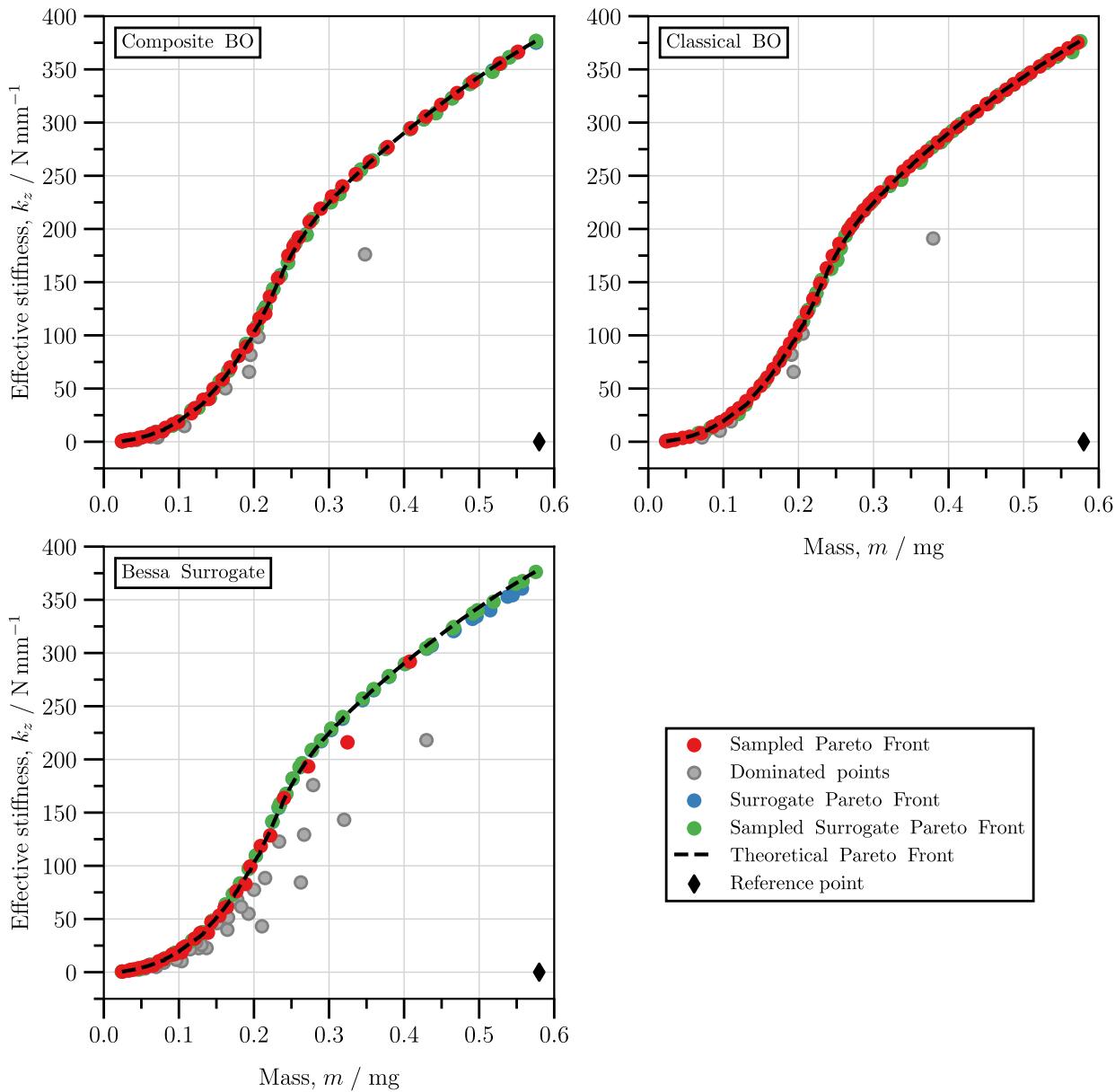


Fig. 20. Results of the Pareto fronts for the multi-objective problem of the 3D CTC metamaterial under infinitesimal strains for the minimisation of the mass (m) and the maximisation of the stiffness (k_z).

4.3.2. Maximisation of the ratio of the absorbed energy to the mass

This example focuses on the maximisation of the ratio of the absorbed energy to the mass of the CTC metamaterial, that is, the specific absorbed energy, under finite strains. In addition to the geometry design, we also aim to select the material of the inclined rods; particularly whether a fragile high-strength steel or a ductile low-strength steel should be used. For this task, data from several steels were used to fit a curve of the elongation at break versus the tensile strength, as shown in Fig. 21, which is then used to predict the maximum plastic strain of the material. For this case, the assumptions are:

- The unit cell is under a large deformation range, so the deformation caused by the axial force of the polygon loop cannot be ignored, and the analytical solution is no longer valid.
- An elasto-plastic isotropic von Mises model with perfect plasticity is assumed. The mechanical properties of the material are the same as in the previous case (see Table 4).
- The unit cell is subjected to uniaxial load. A compressive displacement of $\delta_z = \varepsilon_z H$ is prescribed in the z-direction with H being the height of the unit cell (Eq. (42)). In this case, because the unit cell undergoes large deformations, it is crucial to validate all simulations to prevent excessive prescribed displacement that could cause penetration. To address this, a single-objective optimisation with contact detection was conducted to determine the minimum strain the unit cell can endure without

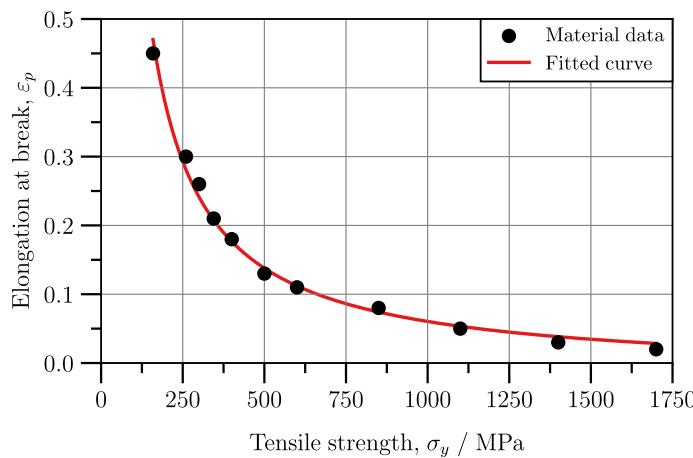


Fig. 21. Elongation at break vs Tensile strength for different steels.

penetration, given the bounds of the input design variables. The minimum strain identified was $\varepsilon_z = 9.29\%$, thus a $\varepsilon_z = 8\%$ was selected.

- The bottom square loop is constrained in all degrees of freedom and the top square loop is free to move and rotate.
- A straightforward failure criterion is used to predict material failure. The material is considered to have failed when a Gauss point in the FEM mesh reaches the maximum accumulated plastic strain component. The maximum plastic strain is calculated accordingly

$$\varepsilon_p(\sigma_y) = \frac{a}{\sigma_y} + b, \quad (48)$$

where the parameters a and b were calculated by fitting a curve of several data points of yield stress and the corresponding elongation at failure for different steels, as can be seen in Fig. 21, and σ_y is the yield stress from the design input variables. Thus, the parameters are $a = 77.4737$ and $b = -0.0171$.

The input design variables are the angle θ , the thickness coefficient β , as in the previous case, and the yield strength of the material σ_y . The bounds for the design variables are defined as:

$$\theta = [30, 80]^\circ, \quad \beta = [0.05, 0.15], \quad \sigma_y = [159, 1700] \text{ MPa}. \quad (49)$$

The initial values are set to $\theta = 55^\circ$, $\beta = 0.1$ and $\sigma_y = 300$ MPa. The objective function for this example is defined as:

$$\max_{[\theta, \beta, \sigma_y] \in \mathcal{B}} J(\theta, \beta, \sigma_y) = \max_{[\theta, \beta, \sigma_y] \in \mathcal{B}} \frac{E_{\text{abs}}(\theta, \beta, \sigma_y)}{m(\theta, \beta)}, \quad (50)$$

where the mass m is calculated analytically using Eq. (82) and the absorbed energy E_{abs} is calculated numerically from simulations. The absorbed energy is calculated by prescribing a compressive displacement of δ_z at the top square loop and measuring the area under the force-displacement curve by numerically integrating the response. The absorbed energy is then computed as

$$E_{\text{abs}}(\theta, \beta, \sigma_y) = \int_0^{\delta_{z_{\text{max}}}} F_z(\theta, \beta, \sigma_y) d\delta_z. \quad (51)$$

This prescription is specified using 50 load increments in the numerical simulations. For compatibility with the composite Bayesian optimisation strategy, the objective function is defined as

$$J(\theta, \beta, \sigma_y) = \frac{\int_0^{\delta_{z_{\text{max}}}} F_z(\theta, \beta, \sigma_y) d\delta_z}{m(\theta, \beta)} = \frac{\text{Trapz}(S[\theta, \beta, \sigma_y])}{m(\theta, \beta)} = \mathcal{L}(\theta, \beta, \sigma_y, S[\theta, \beta, \sigma_y]), \quad (52)$$

where \mathcal{L} is the outer composition function of the objective function of the single-objective problem using a trapezoidal integration scheme for the response. This approach introduces the computation of the mass inside \mathcal{L} , and thus, it does not need to be learned by the surrogate model.

To compare the strategies in this example, a total of 128 function evaluations are considered. For the Bayesian optimisation strategies, default settings are used, with an initial dataset of 9 points (one initial guess and 8 Sobol points) for both the classical and composite variants. The number of PCA points for the composite strategy varied between 6 and 12, depending on the complexity of the numerical response. In the random sampling strategy, 130 points are randomly sampled from the design space. Additionally, the Bessa surrogate strategy employs 128 function evaluations, using the same dataset as the random sampling strategy to build the surrogate model.

The median objective evolution is depicted in Fig. 22. Both the classical and composite Bayesian optimisation strategies outperform the random sampling and Bessa surrogate strategies. Furthermore, the composite Bayesian optimisation strategy is more efficient than its classical counterpart. Notably, after approximately 50 function evaluations, the composite Bayesian optimisation

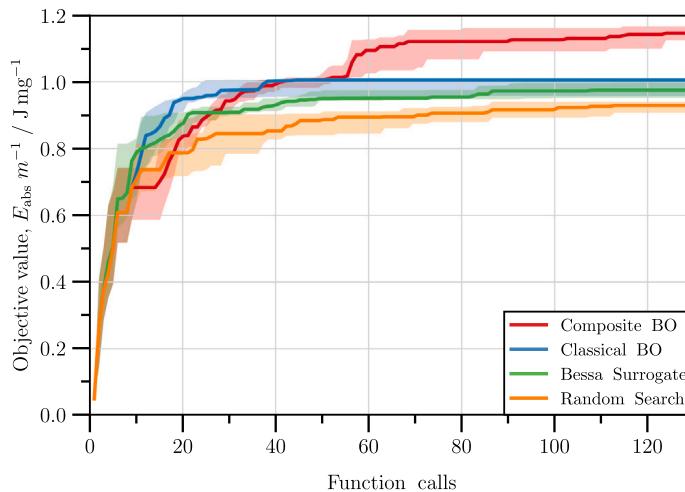


Fig. 22. Objective evolution for the maximisation of the ratio of the absorbed energy to the mass.

Table 6

The median of the best objective and parameter values obtained by the different design strategies.

| Design strategy | $E_{\text{abs}} \text{ } m^{-1} / \text{J mg}^{-1}$ | $\theta / {}^\circ$ | β | σ_y / MPa |
|-----------------|---|---------------------|---------------|-------------------------|
| Random | 0.9292 | 72.3101 | 0.1399 | 1523.5030 |
| Bessa Surrogate | 0.9750 | 71.2628 | 0.1499 | 1550.9063 |
| Classical BO | 1.0052 | 72.3009 | 0.1500 | 1678.5970 |
| Composite BO | 1.1533 | 74.6237 | 0.1470 | 1699.9126 |

strategy consistently surpasses the classical Bayesian optimisation strategy. **Table 6** presents the median best objective and parameter values obtained by the different design strategies. The results indicate that the composite Bayesian optimisation strategy outperforms the other strategies in terms of the objective function, demonstrating its effectiveness in solving this problem. The median best objective value is 1.1533 J mg^{-1} , which surpasses those obtained by the other strategies. The parameter values corresponding to this best objective value are an angle of 74.6237° , a thickness coefficient of 0.147, and a yield strength of 1699.9126 MPa. In addition, the force–displacement response for the parameters and design strategies presented in **Table 6** is shown in **Fig. 23**. The geometry of the optimal solution (Composite BO) is also depicted with its respective accumulated plastic strain after deformation. The differences between the responses clearly demonstrates the effectiveness of the Bayesian optimisation strategies, particularly the composite one.

4.3.3. Minimisation of the mass and maximisation of the absorbed energy

This third example is a multi-objective optimisation problem aimed at minimising the mass and maximising the absorbed energy of the CTC metamaterial. The assumptions for this example are consistent with those in Section 4.3.2, under finite strains, rendering the analytical solution inapplicable. Furthermore, the material itself is considered a design variable in this example. The objective functions are defined as follows:

$$\max_{[\theta, \beta, \sigma_y] \in \mathcal{B}} [J_1(\theta, \beta), J_2(\theta, \beta, \sigma_y)] = \max_{[\theta, \beta, \sigma_y] \in \mathcal{B}} [-m(\theta, \beta), E_{\text{abs}}(\theta, \beta, \sigma_y)]. \quad (53)$$

where the mass m is calculated analytically using Eq. (82) and the absorbed energy E_{abs} is calculated numerically from simulations, accordingly with the previous example (Section 4.3.2). For compatibility with the composite Bayesian optimisation strategy, the objective function is defined as

$$[J_1(\theta, \beta), J_2(\theta, \beta, \sigma_y)] = \left[-m(\theta, \beta), \int_0^{\delta z_f} F_z(\theta, \beta, \sigma_y) d\delta_z \right] \quad (54)$$

$$= \{-m(\theta, \beta), \text{Trapz}(S[\theta, \beta, \sigma_y])\} \quad (55)$$

$$= \{-m(\theta, \beta), \mathcal{L}(\theta, \beta, \sigma_y, S[\theta, \beta, \sigma_y])\}, \quad (56)$$

where \mathcal{L} is the outer composition function of the objective $J_2(\theta, \beta, \sigma_y)$ of the multi-objective problem using a Trapz function for the response. As before, the mass is computed directly from the analytical solution and does not need to be learned by the surrogate model. The bounds of the input design variables, the angle θ , the thickness coefficient β and the yield stress σ_y , are respectively

$$\theta = [30, 80] {}^\circ, \quad \beta = [0.05, 0.15], \quad \sigma_y = [159, 1700] \text{ MPa}. \quad (57)$$

The initial values are set to $\theta = 55^\circ$, $\beta = 0.1$ and $\sigma_y = 300 \text{ MPa}$. The reference point for the hypervolume metric in this scenario is selected as $[m_{\text{max}}, E_{\text{abs min}}]$. Taking into account the previous bounds, the maximum mass is approximately 0.576 mg and the minimum

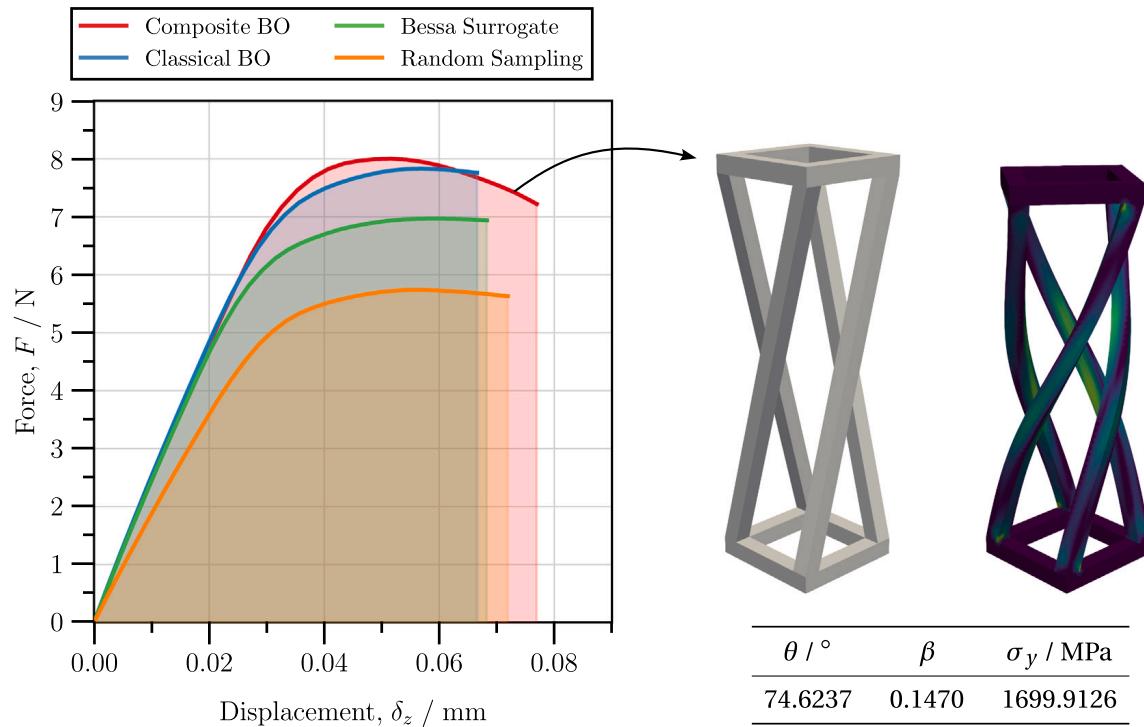


Fig. 23. The median of the force–displacement response for the maximisation of the ratio of the absorbed energy to the mass and the structure geometry of the Composite BO solution. The geometry of the optimal solution is depicted with its respective accumulated plastic strain after deformation.

Table 7

Median hypervolume and the ratio of non-dominated to dominated solutions for the different design strategies in the multi-objective optimisation problem of the 3D CTC metamaterial under finite strains.

| Design strategy | Sampled PF | | Surrogate PF | | Sampled Surrogate PF | |
|-----------------|---------------|---------------|---------------|---------------|----------------------|--------|
| | HV | ND/D | HV | ND/D | HV | ND/D |
| Bessa Surrogate | 0.1154 | 0.1944 | 0.1330 | 0.1356 | 8.2250 | |
| Classical BO | 0.1524 | 0.7552 | 0.1519 | 0.1493 | | 4.8571 |
| Composite BO | 0.1565 | 0.5829 | 0.1554 | 0.1539 | | 3.3278 |

energy is zero. Consequently, [0.58, 0] is chosen as a reference point. To compare the strategies in this example, a total of 128 function evaluations are considered, and the Pareto front comparison strategy in Section 4.3.1 is adopted.

The median hypervolume convergence is displayed in Fig. 24. The Pareto fronts generated by both the classical and composite Bayesian optimisation strategies are significantly better than those sampled by the Bessa Surrogate strategy. Furthermore, the composite Bayesian strategy exhibits significantly less dispersion, particularly after approximately 60 function calls, where it becomes almost negligible. It also achieves the highest hypervolume among all strategies, indicating superior performance. Naturally, for all approaches, the hypervolume monotonically increases with the number of function calls. Fig. 25 illustrates the various Pareto fronts for the different optimisation strategies. Particularly, this highlights the difference between the Sampled, Surrogate and Sampled Surrogate Pareto fronts. This example also illustrates how the low-fidelity Surrogate Pareto front is generally over-optimistic when compared to the Sampled Pareto front, which further motivates using high-fidelity evaluations (as those obtained directly from BO strategies) to build the MO solution. Table 7 presents the median hypervolume and the ratio of non-dominated to dominated solutions for all Pareto fronts across the different strategies. The results show that the composite Bayesian optimisation strategy outperforms the others in terms of hypervolume for all Pareto fronts. Despite their lower hypervolume indicator, the classical Bayesian strategy has a higher ratio of non-dominated to dominated solutions in the Sampled Pareto front, while the Bessa Surrogate strategy shows a higher ratio in the Surrogate Sampled Pareto front. Finally, in Fig. 26, the optimal Sampled Pareto front for the three design strategies is displayed. Additionally, the geometries of three distinct solutions for the Composite BO strategy are illustrated, emphasising the significant variations among the potential optimal design solutions found on the Pareto front.

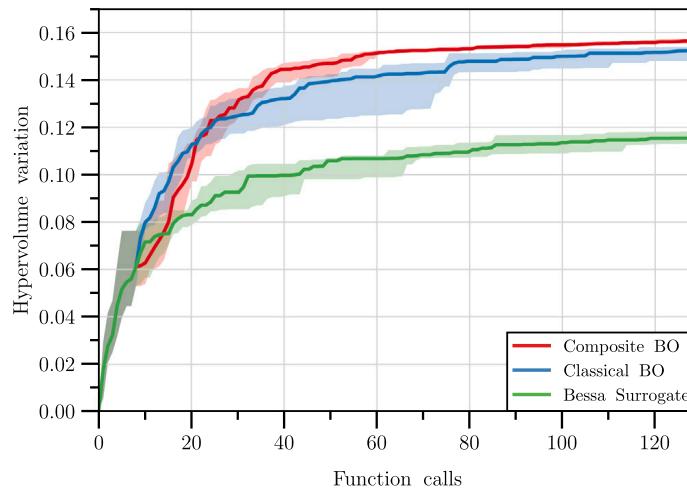


Fig. 24. The median of HV variation with an increasing number of function calls for the multi-objective problem of the 3D CTC metamaterial under finite strains for the minimisation of the mass (m) and the maximisation of the absorbed energy (E_{abs}) with the variation of the geometrical parameters θ , β and the tensile strength σ_y .

4.4. Layered metallic foam beam under bending

In this example, the optimisation of the cross-sectional material of a beam under bending is considered. The beam is composed of 16 layers of metallic foam, with each layer having a different relative foam density and, hence, different mechanical properties. In this case, the geometry consists of a beam with a length of 4 m and a height of 1 m under plane strain conditions. The leftmost end of the beam is fixed, and a moment M is applied at the rightmost end. A regular mesh with 32×16 quadratic 8-node quadrilateral elements is used to discretise the beam; thus we consider a row of elements for each layer of the beam. A schematic representation of the beam is shown in Fig. 27.

In this example, a simplistic approach has been adopted to model the behaviour of the metallic foam. Considering the review from Smith et al. [76], we adopt a foam obtained from a powder metallurgical process, with a relative density ranging from 0.04 to 0.65. Then, according to the Gibson and Ashby model [77], the elastic modulus and the yield stress of the foam can be estimated from a set of empirical equations relying on specific parameters. In this work, we consider the following equations to estimate the elastic modulus and the yield stress of the foam:

$$E = E_s (0.5\xi^2 + 0.3\xi), \quad (58)$$

$$\sigma_{y,c} = 0.5\sigma_{y,s} (0.5\xi^{2/3} + 0.3\xi), \quad (59)$$

$$\sigma_{y,t} = 1.2\sigma_{y,c}, \quad (60)$$

where E_s and $\sigma_{y,s}$ are the elastic modulus and the yield stress of the solid material, $\sigma_{y,t}$ and $\sigma_{y,c}$ denote the tensile and compressive yield stress of the foam and ξ is the relative foam density,

$$\xi = \frac{\rho}{\rho_s}. \quad (61)$$

For the properties of the solid material, $E_s = 210$ GPa, $\nu = 0.33$ and $\sigma_{y,s} = 200$ MPa have been selected. Since different tensile and compressive yield stresses are considered, a Drucker–Prager plasticity model is adopted to model the foam behaviour, whose yield criterion is given by,

$$\tau + p \tan \phi = c, \quad (62)$$

where τ is the shear stress, p is the hydrostatic pressure, ϕ is the internal friction angle and c is the cohesion. According to Eq. (60), and considering a scalar $\alpha = \sigma_{y,t}/\sigma_{y,c} = 1.2$, it can be shown that the internal friction angle is constant and given by

$$\tan \phi = -\frac{3\alpha - 1}{2\alpha + 1} \Rightarrow \phi = -7.765^\circ, \quad (63)$$

and the cohesion is given by

$$c = \sigma_{y,c} \left(1 + \frac{\alpha - 1}{\alpha + 1} \right). \quad (64)$$

Finally, for simplicity, the associative Drucker–Prager variant is used, and perfect plasticity is assumed. Therefore, the foam constitutive response is modelled as an elastic–plastic material whose elastic and plastic properties depend on the relative foam density. Note that more sophisticated models can be adopted to model the foam behaviour, such as the ones in Miller

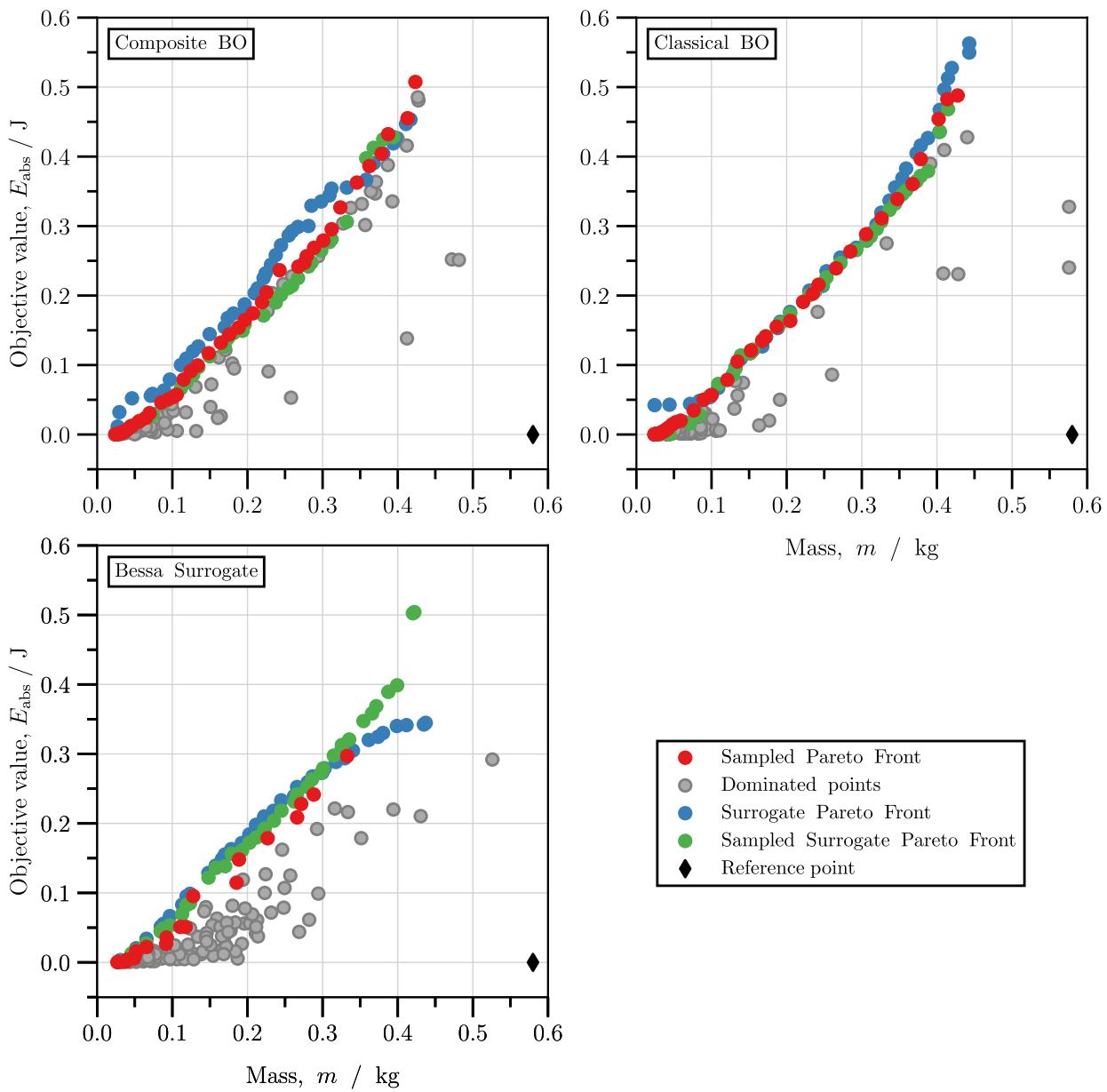


Fig. 25. Results of the Pareto fronts for the multi-objective problem of the 3D CTC metamaterial under finite strains for the minimisation of the mass (m) and the maximisation of the absorbed energy (E_{abs}) with the variation of the geometrical parameters θ , β and the tensile strength σ_y .

[78], Deshpande and Fleck [79], Reyes et al. [80], but for the sake of simplicity, the Gibson and Ashby model is adopted in this example.

The optimisation procedure focuses on finding the optimal distribution of the relative foam density of each layer according to a given objective, which is defined on a per-example basis. Thus, we define the 16-dimensional design space for the relative foam density, ξ_i , using the theoretical bounds expected for this foam type,

$$\xi_i \in [0.04, 0.65], \quad i = 1, \dots, 16. \quad (65)$$

For this problem setup, two examples are analysed: (i) the maximisation of the infinitesimal specific beam bending stiffness, and (ii) the maximisation of the specific absorbed energy for a prescribed rotation of 5° at the rightmost edge.

4.4.1. Maximisation of the infinitesimal specific beam bending stiffness

The first example aims to maximise the infinitesimal specific beam bending stiffness, defined as the ratio between the infinitesimal bending stiffness K and the density of the beam,

$$\max_{\xi} J(\xi) = \max_{\xi} \frac{K(\xi)}{\rho(\xi)}, \quad (66)$$

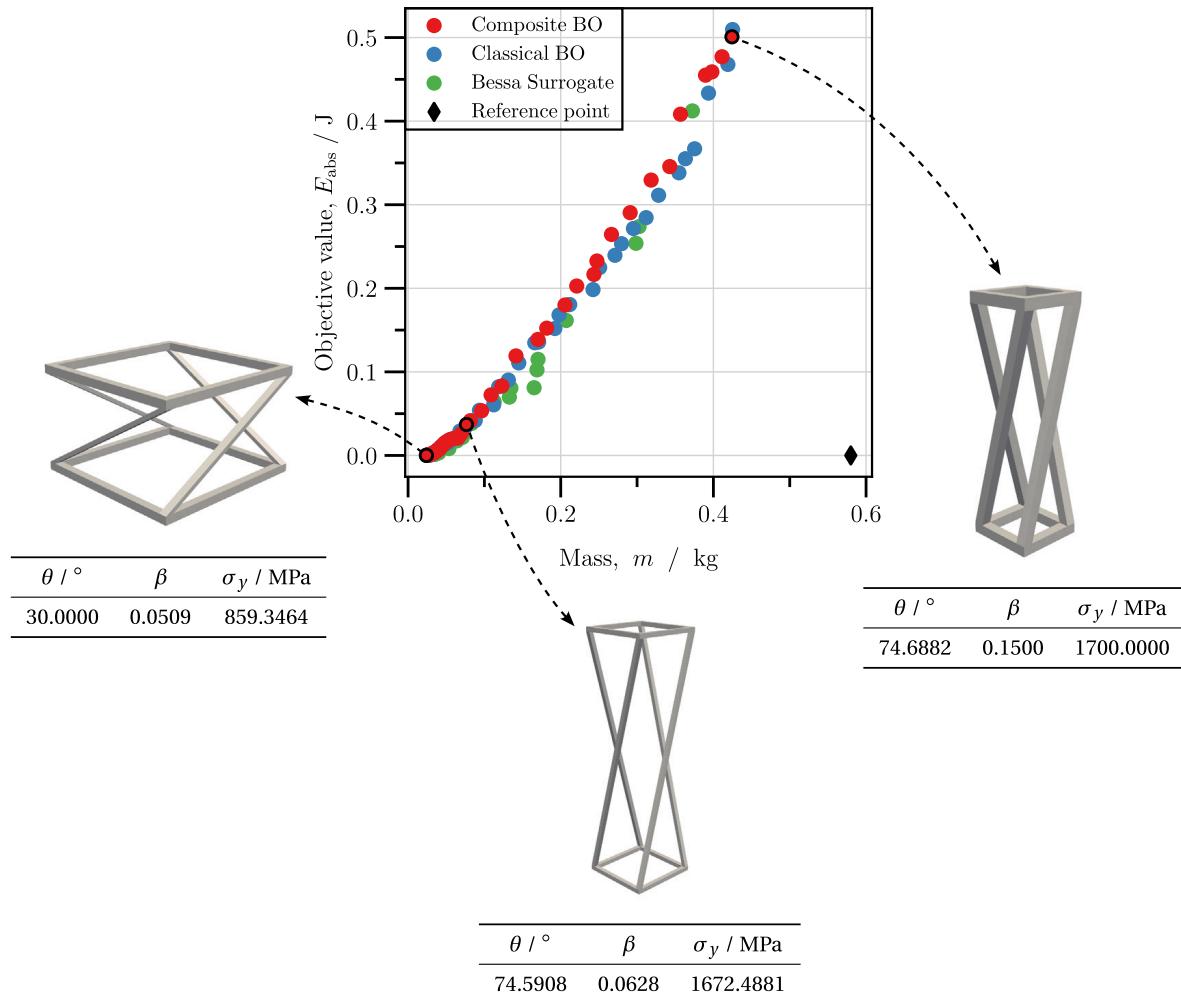


Fig. 26. The best Pareto fronts for the three design strategies in the multi-objective optimisation problem of the 3D CTC metamaterial under finite strains for the minimisation of the mass (m) and the maximisation of the absorbed energy (E_{abs}). Three geometries of the Composite BO strategy are also presented.

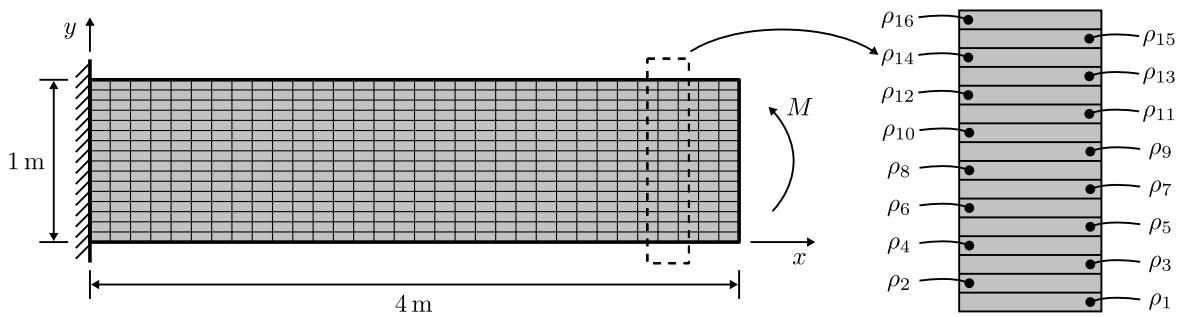


Fig. 27. Schematic representation of the beam with 16 layers of metallic foam under bending. Each layer is assigned a different relative foam density.

where the density can be evaluated analytically from the mean relative density of the beam,

$$\rho(\xi) = \frac{\rho_s}{16} \sum_{i=1}^{16} \xi_i. \quad (67)$$

We compute the infinitesimal bending stiffness via simulations by prescribing a small rotation of $\phi = 0.01^\circ$ at the rightmost edge of the beam, measuring the reaction moment M and then compute the slope of the moment–rotation curve with

$$K(\xi) = \frac{M(\xi)}{\phi}. \quad (68)$$

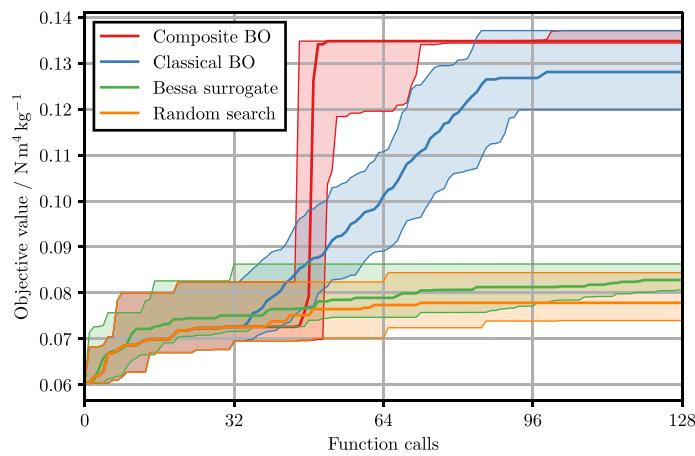


Fig. 28. Median objective evolution for the maximisation of the infinitesimal specific beam bending stiffness.

This prescription is specified using a single load increment in the finite element analysis, so the output response of the program is a two-point curve with the points $(0, 0)$ and (ϕ, M) .

For compatibility with the composite Bayesian optimisation setting, the objective function is cast to

$$J(\xi) = \frac{16}{\rho_s \sum_{i=1}^{16} \xi_i} \frac{M(\xi)}{\phi} = \frac{16}{\rho_s \sum_{i=1}^{16} \xi_i} \frac{\text{Max}(S[\xi])}{\phi} = \mathcal{L}(\xi, S[\xi]), \quad (69)$$

where \mathcal{L} is the outer composition of the objective function and we find the maximum value of the moment-angle response to compute the stiffness. This approach introduces the computation of the density directly inside \mathcal{L} and, therefore, it does not need to be learned by the surrogate model. Note that the numerical response for this case is a two-point curve of the elastic response of the beam under bending.

To compare the strategies in this example, a total of 128 function evaluations are considered. For the Bayesian optimisation strategies, the default settings are used, with an initial dataset of 33 points (initial guess and 32 Sobol points) for both the classical and composite variants. In the random sampling strategy, 128 points are randomly sampled from the design space. Finally, the Bessa surrogate strategy is used with 128 function evaluations, where the surrogate model is built using the same dataset as the random sampling strategy.

The median objective evolution is depicted in Fig. 28. Both the classical and composite Bayesian optimisation strategies drastically outperform both the random sampling and the Bessa surrogate strategies. In addition, the composite Bayesian optimisation strategy is slightly more efficient than the classical Bayesian optimisation strategy — it showcases a higher median value. Particularly, after 128 function evaluations, the dispersion in the objective value is almost insignificant for the composite Bayesian optimisation strategy, while the classical Bayesian optimisation strategy still presents a considerable dispersion. Furthermore, the cross-section associated with the median best-observed objective value for each optimiser is shown in Fig. 29. Both the classical and composite Bayesian optimisation strategies find a similar cross-section, with a higher density in the outer layers and a lower density in the inner layers — a result that is consistent with the expected stresses in the cross-section under bending. On the other hand, both the random sampling and the Bessa surrogate strategies find a cross-section with more randomly distributed densities and, particularly, with increased asymmetry.

4.4.2. Maximisation of the specific absorbed energy

In the second example, we aim to optimise a quantity computed from the response of the beam under finite strains bending with plastic activity. Therefore, the objective function is defined as the maximisation of the specific absorbed energy, which is computed as the ratio between the absorbed energy E_{abs} and the density of the beam,

$$\max_{\xi} J(\xi) = \max_{\xi} \frac{E_{\text{abs}}(\xi)}{\rho(\xi)}, \quad (70)$$

where the absorbed energy is computed as the integral of the moment-rotation curve up to a given maximum prescribed rotation $\phi_{\text{max}} = 5^\circ$, that is,

$$E_{\text{abs}}(\xi) = \int_0^{\phi_{\text{max}}} M(\xi, \phi) \, d\phi. \quad (71)$$

The moment-rotation response is computed with the finite element solver, using 50 load increments up to the prescribed rotation. Then, the integration is carried out numerically using a trapezoidal rule. As in the previous example, we can exploit the composition to embed the density computation inside the objective function,

$$J(\xi) = \frac{16}{\rho_s \sum_{i=1}^{16} \xi_i} E_{\text{abs}}(\xi) = \frac{16}{\rho_s \sum_{i=1}^{16} \xi_i} \text{Trapz}(S[\xi]) = \mathcal{L}(\xi, S[\xi]), \quad (72)$$

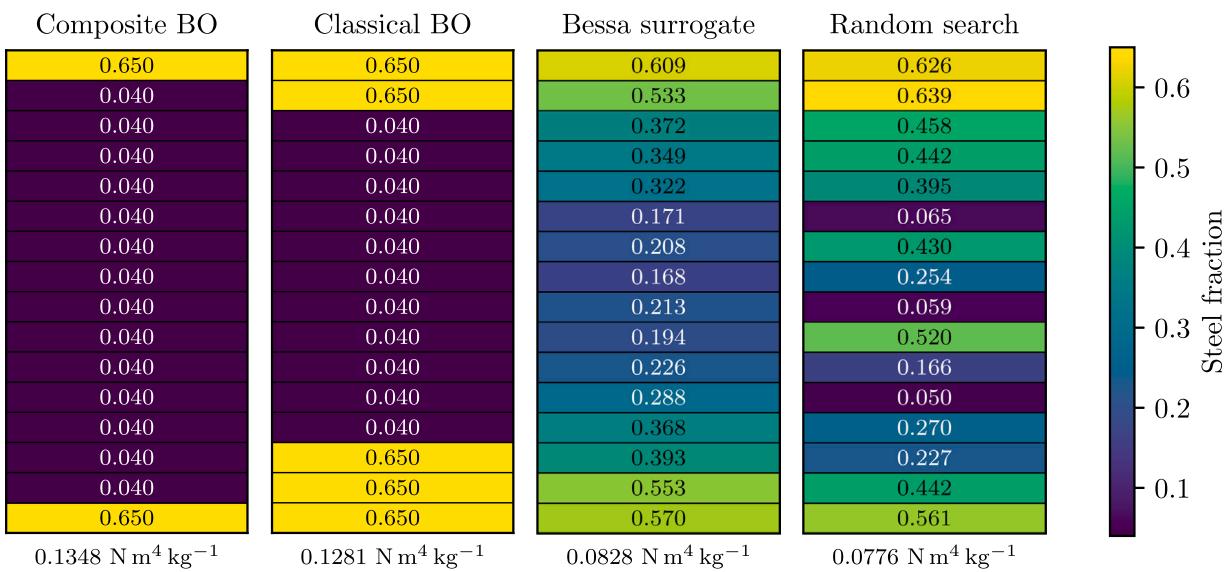


Fig. 29. Median best-observed cross-section for the maximisation of the infinitesimal specific beam bending stiffness.

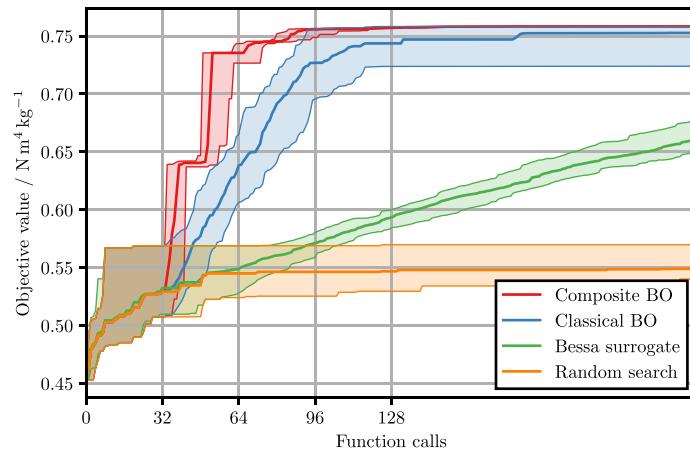


Fig. 30. Median objective evolution for the maximisation of the specific absorbed energy.

where \mathcal{L} is the outer composition of the objective function and Trapz denotes the trapezoidal integration rule to compute the absorbed energy. In contrast with the previous case, the numerical response is an elasto-plastic curve with 50 points. For comparison purposes, all optimisers are evaluated with a budget of 256 function calls, using 20 realisations.

The median objective evolution and its 95 % confidence interval are depicted in Fig. 30. Similarly to the previous case, the classical and composite Bayesian optimisation strategies drastically outperform the random sampling and the Bessa surrogate strategies. Furthermore, the composite Bayesian optimisation converges faster than the classical Bayesian optimisation and shows a significantly smaller dispersion in the objective value. These results further confirm the efficiency of the composite Bayesian optimisation strategy in this example. In addition, the cross-section associated with the median best-observed objective value for each optimiser is shown in Fig. 31. Both BO strategies show similar cross-sections, with a higher density in the outer layers and a lower density in the inner layers, as well as a given asymmetry associated with the pressure-sensitive behaviour of the foam. On the other hand, Bessa's strategy effectively finds a cross-section where the density is localised in the outer layers, but fails to refine it further to achieve a higher objective value. Finally, the random sampling strategy finds a cross-section with a more random distribution of densities, which performs poorly for this example.

5. Conclusions

In the present contribution, a novel framework for material and structural design based on a tailored composite Bayesian Optimisation strategy is proposed. Unlike traditional optimisation methods, this composite variant leverages the inherent structure of the objective function. This approach enables the optimiser to (i) observe the complete response of the numerical simulation,

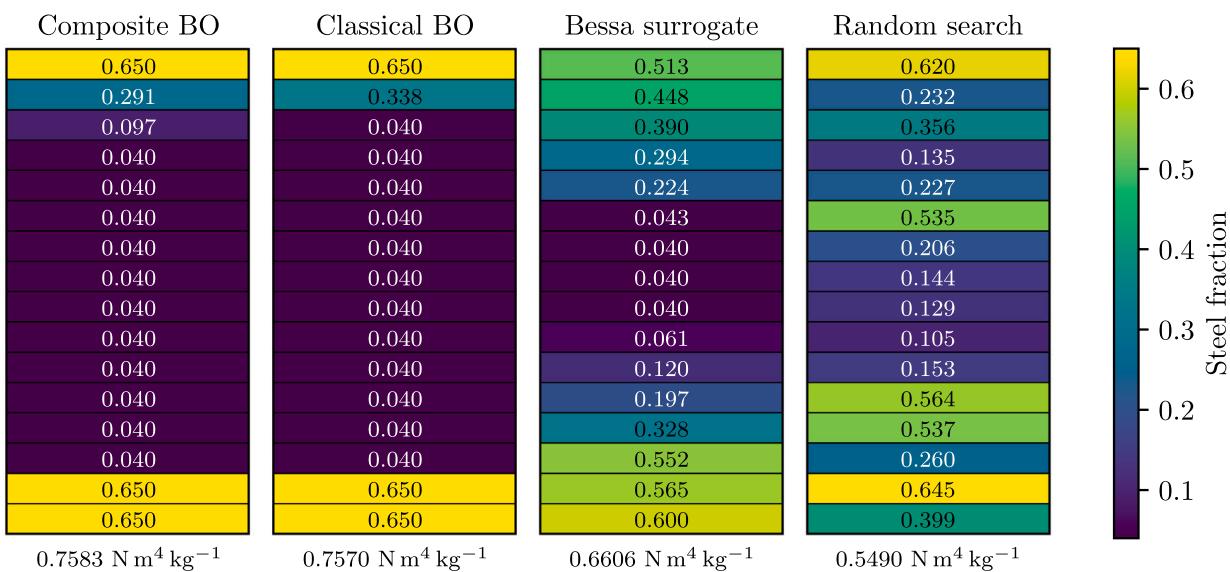


Fig. 31. Median best-observed cross-section for the maximisation of the specific absorbed energy.

(ii) enhance the predictive accuracy of the surrogate model, and (iii) seamlessly integrate the objective function's structure into the optimisation framework.

The formulation of the composite problem incorporates numerous elements essential for optimising generic responses, which we analyse in detail. Key aspects include the reparameterisation of the response into a latent response space, the selection of the surrogate model, and the use of gradient-enabled Monte Carlo acquisition functions. Additionally, we explore the application of Principal Component Analysis (PCA) for dimensionality reduction. While optional, PCA not only reduces the computational burden of the optimisation but also improves the efficiency due to its weakly correlated feature set.

The composite Bayesian optimisation strategy involves defining two hyperparameters: the number of initial Design of Experiments (DoE) points and the PCA variance threshold. The former is particularly vital for balancing the exploration-exploitation tradeoff in Bayesian optimisation strategies. We address the selection of these values through a parametric study, providing heuristics for their definition. Our findings indicate that the number of initial DoE points should increase with the number of dimensions but must be balanced to avoid slowing down optimisation convergence. Conversely, a small, constant PCA variance threshold is sufficient to achieve a favourable tradeoff between computational time and accuracy. While optimal hyperparameters are problem-specific, our recommendations perform well across the problems analysed in this work.

The efficacy of the proposed design framework is evaluated across various linear and nonlinear material and structural design problems, encompassing single-objective and multi-objective scenarios, both with and without observation noise. We compare our approach against three baseline optimisation strategies. Our analyses demonstrate that the composite strategy achieves faster convergence, reduced variability between runs, and superior optima compared to other methods. These benefits are particularly pronounced with increasing objective function complexity, such as in cases involving geometric or material non-linearity. The enhanced sampling efficiency of our method reduces the number of required experiments to achieve a specified objective value.

Overall, the results exhibit significant improvements in performance and quality, especially in nonlinear settings, underscoring the potential of our framework to advance design methodologies in material and structural engineering.

CRediT authorship contribution statement

R.P. Cardoso Coelho: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **A. Francisca Carvalho Alves:** Validation, Software, Methodology, Investigation, Formal analysis. **T.M. Nogueira Pires:** Validation, Methodology, Investigation, Formal analysis. **F.M. Andrade Pires:** Writing – review & editing, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

R. P. Cardoso Coelho and A. Francisca Carvalho Alves gratefully acknowledge the support provided by Fundação para a Ciência e a Tecnologia (FCT), Portugal through the scholarships with references 2020.07159.BD and 2020.07279.BD, respectively. This research has also been supported by Instituto de Ciência e Inovação em Engenharia Mecânica e Engenharia Industrial (INEGI).

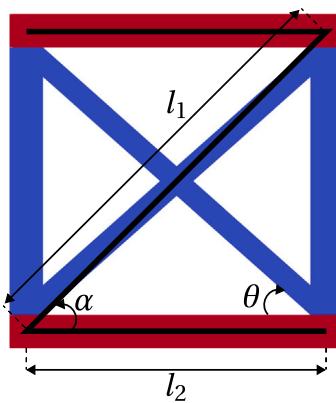


Fig. 32. Approximation model of the 3D unit cell.

Appendix. Infinitesimal strain solution for the compression-torsion coupled metamaterial

To validate the numerical example in Section 4.3.1, an analytical model has been derived to predict the response of the unit cell under infinitesimal strains and isotropic linear elastic material behaviour. The analytical solution was primarily built upon the work established by [81,82]. However, some modifications have been introduced to accommodate slight differences in the unit cell geometry and to improve the quality of the solution, since the previous works had some limitations. The developed analytical solution shows a close agreement with numerical results and is based on the Euler–Bernoulli beam theory. In this model, the inclined rods are represented as beams with rectangular cross-sections, while the top and bottom square loops are treated as rigid bodies¹³ with quadrangular cross-sections.

For this model, a Z-shaped structure is considered as the unit cell, as shown in Fig. 32. The two critical lengths l_1 and l_2 are introduced to define the geometry of the unit cell, which are closely related to the length of the inclined rods and the square loops. With this approach, note that the angle of the inclined rods is defined as α instead of θ , such that the length of the square loop is $l_\mu - t$ which guarantees a height (distance between the top and bottom square loops) equal to the distance of the two neutral lines of the bottom and top square loops cross-section. According to these assumptions, the necessary equations to define the geometry of the unit cell are given by

$$\alpha = \arctan\left(\frac{\beta + (1 - \beta)\tan(\theta)}{1 - \beta}\right), \quad (73)$$

$$l_2 = l_\mu - t = l_\mu(1 - \beta), \quad (74)$$

$$l_1 = \frac{l_2}{\cos(\alpha)}. \quad (75)$$

The degrees of freedom for this model are defined as δ_z , δ_c and φ_z , which represent the relative displacement along the z-direction, the increment of the circumference (expansion) of the top square loop and the twisting angle of the unit cell that is defined as the rotation angle in the x-y plane after deformation, respectively. The degrees of freedom are represented in Fig. 33. Thus, the compression–torsion coupling constitutive relation of the unit cell can be defined as

$$\begin{Bmatrix} F_z \\ F_c \\ T_z \end{Bmatrix} = \begin{bmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \end{bmatrix} \begin{Bmatrix} \delta_z \\ \delta_c \\ \varphi_z \end{Bmatrix}. \quad (76)$$

Therefore, under uniaxial loading, an effective compressive stiffness can be devised from the stiffness matrix as

$$k_z = \frac{k_{11}\delta_z + k_{12}\delta_c + k_{13}\varphi_z}{\delta_z} = \frac{k_{11}(k_{22}k_{33} - k_{23}^2) - k_{12}^2k_{33} + 2k_{12}k_{13}k_{23} - k_{13}^2k_{22}}{k_{22}k_{33} - k_{23}^2}, \quad (77)$$

whose analytical expression is given by

$$k_z = \frac{E_1 A_1 \{ (\xi_1 - 1) \cos(\alpha) [\chi l_2 \pi \sin(\alpha) + 360 \cos(\alpha)] + 360 \}}{90 l_1}, \quad (78)$$

where the subscripts 1 and 2 represent the bodies of the inclined rods and the top and bottom square loops, respectively. The parameter χ is defined as the ratio of the twisting angle and the prescribed displacement in $^\circ \text{mm}^{-1}$ and characterises the CTC

¹³ Similar assumptions have been considered by Fu et al. [73] and Qi et al. [72] without significantly affecting the accuracy of the results.

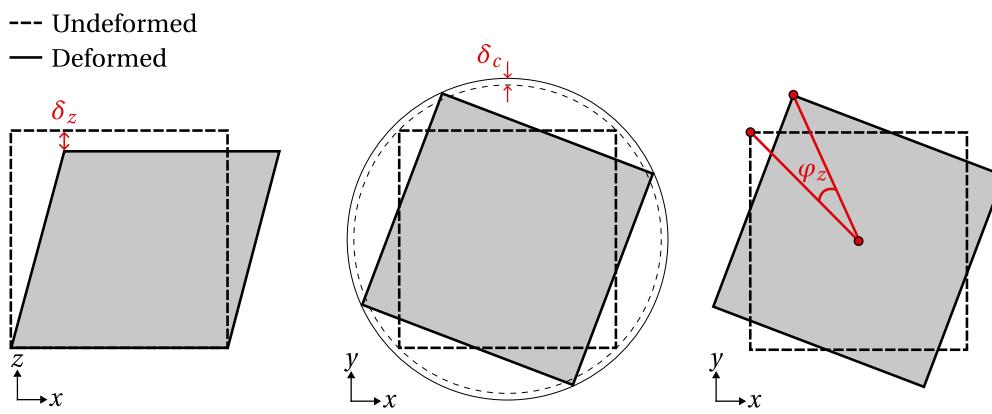


Fig. 33. Degrees of freedom of the unit cell under uniaxial load.

property of the unit cell. The equation for χ is:

$$\chi = \frac{\varphi_z}{\delta_z} = \frac{1080l_2(1 - \xi_1)\sin(\theta)\cos(\theta)}{\pi l_1^2 \{ \cos^2(\theta) [4(3\eta_1 - \xi_1) - 3(1 - \xi_1)\cos^2(\theta)] - 12\eta_1 \}}, \quad (79)$$

and, the ξ_i and η_i are dimensionless parameters defined as:

$$\xi_i = \frac{12I_i}{l_i^2 A_i}, \quad (80)$$

$$\eta_i = \frac{G_i J_i}{E_i A_i K_i l_i^2}, \quad (81)$$

where I_i and J_i are the second moment of area and the polar moment of area of the cross-section, G_i is the shear modulus and K_i is a correction factor considering the inhomogeneous distribution of shear force along the section, which is adopted as 6/5 for rectangular/quadrangular cross-sections. As a final note, the mass of the unit cell can also be computed as

$$m(\theta, \beta) = \frac{4\rho_1\beta^2l_\mu^3(1 - \beta)[2\cos(\theta) + \sin(\theta)]}{\cos(\theta)}. \quad (82)$$

Data availability

I have shared the link to my data.

References

- [1] S. Sridhara, A. Chandrasekhar, K. Suresh, A generalized framework for microstructural optimization using neural networks, Mater. Des. 223 (2022) 111213, <http://dx.doi.org/10.1016/j.matdes.2022.111213>, Retrieved from <https://www.sciencedirect.com/science/article/pii/S0264127522008358>.
- [2] M. Tanaka, H.D. Bui, Inverse Problems in Engineering Mechanics, in: IUTAM Symposium Tokyo 1992, Springer Berlin / Heidelberg, Berlin, Heidelberg, 1993.
- [3] J.M.P. Martins, A. Andrade-Campos, S. Thuillier, Comparison of inverse identification strategies for constitutive mechanical models using full-field measurements, Int. J. Mech. Sci. 145 (2018) 330–345, <http://dx.doi.org/10.1016/j.ijmecsci.2018.07.013>.
- [4] J. Juang, C. Sun, System identification of large flexible structures by using simple continuum models, 1983, <https://repository.exst.jaxa.jp/dspace/handle/a-is/404503>.
- [5] H.T. Banks, Parameter Identification Techniques for Physiological Control Systems, Tech. Rep., Brown Univ Providence Ri Lefschetz Center for Dynamical Systems, 1980.
- [6] H.T. Banks, P. Kareiva, Parameter estimation techniques for transport equations with application to population dispersal and tissue bulk flow models, J. Math. Biol. 17 (3) (1983) 253–273, <http://dx.doi.org/10.1007/BF00276516>.
- [7] D. Lynch, C. Officer, Nonlinear parameter estimation for sediment cores, Chem. Geol. - Chem. Geol. 44 (1984) 203–225, [http://dx.doi.org/10.1016/0009-2541\(84\)90073-1](http://dx.doi.org/10.1016/0009-2541(84)90073-1).
- [8] S. Pal, G. Wije Wathugala, S. Kundu, Calibration of a constitutive model using genetic algorithms, Comput. Geotech. 19 (4) (1996) 325–348, [http://dx.doi.org/10.1016/S0266-352X\(96\)00006-7](http://dx.doi.org/10.1016/S0266-352X(96)00006-7), Retrieved from <https://www.sciencedirect.com/science/article/pii/S0266352X96000067>.
- [9] B. Pichler, R. Lackner, H.A. Mang, Back analysis of model parameters in geotechnical engineering by means of soft computing, Internat. J. Numer. Methods Engrg. 57 (14) (2003) 1943–1978, <http://dx.doi.org/10.1002/nme.740>, arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1002/nme.740>, Retrieved from <https://onlinelibrary.wiley.com/doi/abs/10.1002/nme.740>.
- [10] M. Rokonuzzaman, T. Sakai, Calibration of the parameters for a hardening–softening constitutive model using genetic algorithms, Comput. Geotech. 37 (4) (2010) 573–579, <http://dx.doi.org/10.1016/j.compgeo.2010.02.007>, Retrieved from <https://www.sciencedirect.com/science/article/pii/S0266352X10000339>.
- [11] A. Andrade-Campos, S. Thuillier, P. Pilvin, F. Teixeira-Dias, On the determination of material parameters for internal variable thermoelastic–viscoplastic constitutive models, Int. J. Plast. 23 (8) (2007) 1349–1379, <http://dx.doi.org/10.1016/j.ijplas.2006.09.002>.
- [12] O. Colak, Y. Cakir, Material model parameter estimation with genetic algorithm optimization method and modeling of strain and temperature dependent behavior of epoxy resin with cooperative-VBO model, Mech. Mater. 135 (2019) 57–66, <http://dx.doi.org/10.1016/j.mechmat.2019.04.023>.

- [13] K. Sedighiani, M. Diehl, K. Traka, F. Roters, J. Sietsma, D. Raabe, An efficient and robust approach to determine material parameters of crystal plasticity constitutive laws from macro-scale stress-strain curves, *Int. J. Plast.* 134 (2020) 102779, <http://dx.doi.org/10.1016/j.ijplas.2020.102779>.
- [14] D.J. Savage, Z. Feng, M. Knezevic, Identification of crystal plasticity model parameters by multi-objective optimization integrating microstructural evolution and mechanical data, *Comput. Methods Appl. Mech. Engrg.* 379 (2021) 113747, <http://dx.doi.org/10.1016/j.cma.2021.113747>.
- [15] R.P. Cardoso Coelho, A.F. Carvalho Alves, F.M. Andrade Pires, Efficient constitutive parameter identification through optimisation-based techniques: A comparative analysis and novel composite Bayesian optimisation strategy, *Comput. Methods Appl. Mech. Engrg.* 427 (2024) 117039, <http://dx.doi.org/10.1016/j.cma.2024.117039>.
- [16] S. Avril, F. Pierron, Y. Pannier, R. Rotinat, Stress reconstruction and constitutive parameter identification in plane-stress elasto-plastic problems using surface measurements of deformation fields, *Exp. Mech.* 48 (4) (2008) 403–419, <http://dx.doi.org/10.1007/s11340-007-9084-2>.
- [17] N. Souto, A. Andrade-Campos, S. Thuillier, Mechanical design of a heterogeneous test for material parameters identification, *Int. J. Mater. Form.* 10 (3) (2017) 353–367, <http://dx.doi.org/10.1007/s12289-016-1284-9>.
- [18] J. Cao, J. Lin, A study on formulation of objective functions for determining material models, *Int. J. Mech. Sci.* 50 (2) (2008) 193–204, <http://dx.doi.org/10.1016/j.ijmecsci.2007.07.003>.
- [19] H. Abdul-Hameed, T. Messager, F. Zaïri, M. Naït-Abdelaziz, Large-strain viscoelastic-viscoplastic constitutive modeling of semi-crystalline polymers and model identification by deterministic/evolutionary approach, *Comput. Mater. Sci.* 90 (2014) 241–252, <http://dx.doi.org/10.1016/j.commatsci.2014.03.043>.
- [20] J. Kuhn, J. Spitz, P. Sonnweber-Ribic, M. Schneider, T. Böhleke, Identifying material parameters in crystal plasticity by Bayesian optimization, *Optim. Eng.* 23 (2021) <http://dx.doi.org/10.1007/s11081-021-09663-7>.
- [21] C. Malherbe, N. Vayatis, Global optimization of Lipschitz functions, 2017, [arXiv:1703.02628](https://arxiv.org/abs/1703.02628) [stat].
- [22] A.F.C. Alves, B.P. Ferreira, F.A. Pires, On the modeling of cavitation and yielding in rubber-toughened amorphous polymers: Fully implicit implementation and optimization-based calibration, *Int. J. Solids Struct.* (2023) 112488, <http://dx.doi.org/10.1016/j.ijsolstr.2023.112488>, Retrieved from <https://www.sciencedirect.com/science/article/pii/S0020768323003852>.
- [23] R.P. Cardoso Coelho, M. Vieira de Carvalho, F.M. Andrade Pires, A multi-scale model combining martensitic transformations with multi-phase crystallographic slip, *Comput. Struct.* 289 (2023) 107174, <http://dx.doi.org/10.1016/j.compstruc.2023.107174>.
- [24] M. Bessa, R. Bostanabad, Z. Liu, A. Hu, D.W. Apley, C. Brinson, W. Chen, W.K. Liu, A framework for data-driven analysis of materials under uncertainty: Countering the curse of dimensionality, *Comput. Methods Appl. Mech. Engrg.* 320 (2017) 633–667, <http://dx.doi.org/10.1016/j.cma.2017.03.037>, Retrieved from <https://www.sciencedirect.com/science/article/pii/S0045782516314803>.
- [25] M. Bessa, S. Pellegrino, Design of ultra-thin shell structures in the stochastic post-buckling range using Bayesian machine learning and optimization, *Int. J. Solids Struct.* 139–140 (2018) 174–188, <http://dx.doi.org/10.1016/j.ijsolstr.2018.01.035>, Retrieved from <https://www.sciencedirect.com/science/article/pii/S0020768318300441>.
- [26] C. Settgast, G. Hüttler, M. Kuna, M. Abendroth, A hybrid approach to simulate the homogenized irreversible elastic–plastic deformations and damage of foams by neural networks, *Int. J. Plast.* 126 (2020) 102624, <http://dx.doi.org/10.1016/j.ijplas.2019.11.003>.
- [27] I. Rocha, P. Kerfriden, F. van der Meer, On-the-fly construction of surrogate constitutive models for concurrent multiscale mechanical analysis through probabilistic machine learning, *J. Comput. Phys.: X* 9 (2021) 100083, <http://dx.doi.org/10.1016/j.jcpex.2020.100083>, Retrieved from <https://www.sciencedirect.com/science/article/pii/S2590055220300354>.
- [28] S. Herath, X. Xiao, F. Cirak, Computational modeling and data-driven homogenization of knitted membranes, *Internat. J. Numer. Methods Engrg.* 123 (3) (2022) 683–704, <http://dx.doi.org/10.1002/nme.6871>.
- [29] A. Couto Carneiro, A.F. Carvalho Alves, R. Cardoso Coelho, J.S. Cardoso, F. Andrade Pires, A simple machine learning-based framework for faster multi-scale simulations of path-independent materials at large strains, *Finite Elem. Anal. Des.* 222 (2023) 103956, <http://dx.doi.org/10.1016/j.finel.2023.103956>, Retrieved from <https://www.sciencedirect.com/science/article/pii/S0168874X23000495>.
- [30] Z. Liu, M. Bessa, W.K. Liu, Self-consistent clustering analysis: An efficient multi-scale scheme for inelastic heterogeneous materials, *Comput. Methods Appl. Mech. Engrg.* 306 (2016) 319–341, <http://dx.doi.org/10.1016/j.cma.2016.04.004>, Retrieved from <https://www.sciencedirect.com/science/article/pii/S0045782516301499>.
- [31] D.W. Abuieida, M. Almasri, R. Ammourah, U. Ravaioli, I.M. Jasiuk, N.A. Sobh, Prediction and optimization of mechanical properties of composites using convolutional neural networks, *Compos. Struct.* 227 (2019) 111264, <http://dx.doi.org/10.1016/j.compstruct.2019.111264>, Retrieved from <https://www.sciencedirect.com/science/article/pii/S0263822319312383>.
- [32] C. Bisagni, L. Lanzi, Post-buckling optimisation of composite stiffened panels using neural networks, *Compos. Struct.* 58 (2) (2002) 237–247, [http://dx.doi.org/10.1016/S0263-8223\(02\)00053-3](http://dx.doi.org/10.1016/S0263-8223(02)00053-3), Retrieved from <https://www.sciencedirect.com/science/article/pii/S0263822302000533>.
- [33] H. Huang, S. Mojumder, D. Suarez, A.A. Amin, M. Fleming, W.K. Liu, Knowledge database creation for design of polymer matrix composite, *Comput. Mater. Sci.* 214 (2022) 111703, <http://dx.doi.org/10.1016/j.commatsci.2022.111703>, Retrieved from <https://www.sciencedirect.com/science/article/pii/S0927025622004268>.
- [34] Y. Zhang, D. Apley, W. Chen, Bayesian optimization for materials design with mixed quantitative and qualitative variables, *Sci. Rep.* 10 (2020) [10.1038/s41598-020-60652-9](https://doi.org/10.1038/s41598-020-60652-9).
- [35] P.I. Frazier, J. Wang, Bayesian optimization for materials design, in: T. Lookman, F.J. Alexander, K. Rajan (Eds.), *Information Science for Materials Discovery and Design*, Springer International Publishing, Cham, 2016, pp. 45–75, http://dx.doi.org/10.1007/978-3-319-23871-5_3.
- [36] K. Wang, A.W. Dowling, Bayesian optimization for chemical products and functional materials, *Curr. Opin. Chem. Eng.* 36 (2022) 100728, <http://dx.doi.org/10.1016/j.coche.2021.100728>, Retrieved from <https://www.sciencedirect.com/science/article/pii/S2211339821000605>.
- [37] A. Tran, J. Sun, J.M. Furlan, K.V. Pagalthivarthy, R.J. Visintainer, Y. Wang, pBO-2GP-3B: A batch parallel known/unknown constrained Bayesian optimization with feasibility classification and its applications in computational fluid dynamics, *Comput. Methods Appl. Mech. Engrg.* 347 (2019) 827–852, <http://dx.doi.org/10.1016/j.cma.2018.12.033>.
- [38] Y. Zhang, D.W. Apley, W. Chen, Bayesian optimization for materials design with mixed quantitative and qualitative variables, *Sci. Rep.* 10 (1) (2020) 4924, <http://dx.doi.org/10.1038/s41598-020-60652-9>.
- [39] A. Tran, J. Tranchida, T. Wildey, A.P. Thompson, Multi-fidelity machine-learning with uncertainty quantification and Bayesian optimization for materials design: Application to ternary random alloys, *J. Chem. Phys.* 153 (7) (2020) 074705, <http://dx.doi.org/10.1063/5.0015672>.
- [40] H.M. Sheikh, T.A. Callan, K.J. Hennessy, P.S. Marcus, Optimization of the shape of a hydrokinetic turbine's draft tube and hub assembly using Design-by-Morphing with Bayesian optimization, *Comput. Methods Appl. Mech. Engrg.* 401 (2022) 115654, <http://dx.doi.org/10.1016/j.cma.2022.115654>.
- [41] Z.Z. Foumani, M. Shishehbor, A. Yousefpour, R. Bostanabad, Multi-Fidelity cost-aware Bayesian optimization, *Comput. Methods Appl. Mech. Engrg.* 407 (2023) 115937, <http://dx.doi.org/10.1016/j.cma.2023.115937>, [arXiv:2211.02732](https://arxiv.org/abs/2211.02732).
- [42] J.M. Winter, R. Abaidi, J.W.J. Kaiser, S. Adami, N.A. Adams, Multi-fidelity Bayesian optimization to solve the inverse Stefan problem, *Comput. Methods Appl. Mech. Engrg.* 410 (2023) 115946, <http://dx.doi.org/10.1016/j.cma.2023.115946>.
- [43] R. Astudillo, P. Frazier, Bayesian optimization of composite functions, in: *Proceedings of the 36th International Conference on Machine Learning, PMLR, 2019*, pp. 354–363.
- [44] R.P. Cardoso Coelho, A.F. Carvalho Alves, T.M. Nogueira Pires, F.M. Andrade Pires, Piglot: An open-source package for derivative-free optimisation of numerical responses, *J. Open Source Softw.* 9 (99) (2024) 6652, <http://dx.doi.org/10.21105/joss.06652>.
- [45] P.I. Frazier, A tutorial on Bayesian optimization, 2018, [arXiv:1807.02811](https://arxiv.org/abs/1807.02811) [cs, math, stat].

- [46] M. Balandat, B. Karrer, D.R. Jiang, S. Daulton, B. Letham, A.G. Wilson, E. Bakshy, BoTorch: A framework for efficient Monte-Carlo Bayesian optimization, 2020, [arXiv:1910.06403](https://arxiv.org/abs/1910.06403) [cs, math, stat].
- [47] C.E. Rasmussen, C.K.I. Williams, in: F. Bach (Ed.), *Gaussian Processes for Machine Learning*, in: *Adaptive Computation and Machine Learning Series*, MIT Press, Cambridge, MA, USA, 2005.
- [48] E. Brochu, V.M. Cora, N. de Freitas, A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and Hierarchical reinforcement learning, 2010, [arXiv:1012.2599](https://arxiv.org/abs/1012.2599) [cs].
- [49] H.J. Kushner, A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise, *J. Basic Eng.* 86 (1) (1964) 97–106, <http://dx.doi.org/10.1115/1.3653121>.
- [50] J. Močkus, On Bayesian methods for seeking the extremum, in: G.I. Marchuk (Ed.), *Optimization Techniques IFIP Technical Conference*: Novosibirsk, July 1–7, 1974, Springer, Berlin, Heidelberg, 1975, pp. 400–404, http://dx.doi.org/10.1007/978-3-662-38527-2_55.
- [51] S. Ament, S. Daulton, D. Eriksson, M. Balandat, E. Bakshy, Unexpected improvements to expected improvement for Bayesian optimization, 2024, [arXiv:2310.20708](https://arxiv.org/abs/2310.20708) [cs, math, stat].
- [52] N. Srinivas, A. Krause, S.M. Kakade, M. Seeger, Gaussian process optimization in the bandit setting: no regret and experimental design, *IEEE Trans. Inform. Theory* 58 (5) (2012) 3250–3265, <http://dx.doi.org/10.1109/TIT.2011.2182033>, [arXiv:0912.3995](https://arxiv.org/abs/0912.3995).
- [53] P.I. Frazier, W.B. Powell, S. Dayanik, A knowledge-gradient policy for sequential information collection, *SIAM J. Control Optim.* 47 (5) (2008) 2410–2439, <http://dx.doi.org/10.1137/070693424>.
- [54] P. Frazier, W. Powell, S. Dayanik, The knowledge-gradient policy for correlated normal beliefs, *INFORMS J. Comput.* 21 (4) (2009) 599–613, <http://dx.doi.org/10.1287/ijoc.1080.0314>.
- [55] P. Hennig, C.J. Schuler, Entropy search for information-efficient global optimization, 2011, [arXiv:1112.1217](https://arxiv.org/abs/1112.1217) [cs, stat].
- [56] J.M. Hernández-Lobato, M.W. Hoffman, Z. Ghahramani, Predictive entropy search for efficient global optimization of black-box functions, 2014, [arXiv:1406.2541](https://arxiv.org/abs/1406.2541) [cs, stat].
- [57] Z. Wang, S. Jegelka, Max-value entropy search for efficient Bayesian optimization, 2018, [arXiv:1703.01968](https://arxiv.org/abs/1703.01968) [cs, math, stat].
- [58] H. Wackernagel, *Multivariate Geostatistics: An Introduction with Applications: 7 Tables*, 3., completely revised ed, Springer, Berlin Heidelberg, 2003.
- [59] J.T. Wilson, R. Moriconi, F. Hutter, M.P. Deisenroth, The reparameterization trick for acquisition functions, 2017, [arXiv:1712.00424](https://arxiv.org/abs/1712.00424) [cs, math, stat].
- [60] D.P. Kingma, M. Welling, Auto-Encoding Variational Bayes, 2022, [arXiv:1312.6114](https://arxiv.org/abs/1312.6114) [cs, stat].
- [61] M. Fleischer, The measure of Pareto optima applications to multi-objective metaheuristics, in: G. Goos, J. Hartmanis, J. van Leeuwen, C.M. Fonseca, P.J. Fleming, E. Zitzler, L. Thiele, K. Deb (Eds.), in: *Evolutionary Multi-Criterion Optimization*, vol. 2632, Springer Berlin Heidelberg, Berlin, Heidelberg, 2003, pp. 519–533, http://dx.doi.org/10.1007/3-540-36970-8_37.
- [62] M. Emmerich, K. Giannakoglou, B. Naujoks, Single- and multiobjective evolutionary optimization assisted by Gaussian random field metamodels, *IEEE Trans. Evol. Comput.* 10 (4) (2006) 421–439, <http://dx.doi.org/10.1109/TEVC.2005.859463>.
- [63] E. Zitzler, L. Thiele, M. Laumanns, C. Fonseca, V. da Fonseca, Performance assessment of multiobjective optimizers: An analysis and review, *IEEE Trans. Evol. Comput.* 7 (2) (2003) 117–132, <http://dx.doi.org/10.1109/TEVC.2003.810758>.
- [64] I. Couckuyt, D. Deschrijver, T. Dhaene, Fast calculation of multiobjective probability of improvement and expected improvement criteria for Pareto optimization, *J. Global Optim.* 60 (3) (2014) 575–594, <http://dx.doi.org/10.1007/s10898-013-0118-2>.
- [65] S. Daulton, M. Balandat, E. Bakshy, Hypervolume knowledge gradient: a lookahead approach for multi-objective Bayesian optimization with partial information, in: *Proceedings of the 40th International Conference on Machine Learning*, PMLR, 2023, pp. 7167–7204.
- [66] S. Daulton, M. Balandat, E. Bakshy, Differentiable expected hypervolume improvement for parallel multi-objective Bayesian optimization, 2020, <http://dx.doi.org/10.48550/ARXIV.2006.05078>, [object Object].
- [67] S. Daulton, M. Balandat, E. Bakshy, Parallel Bayesian optimization of multiple noisy objectives with expected hypervolume improvement, 2021, <http://dx.doi.org/10.48550/ARXIV.2105.08195>, [object Object].
- [68] W.J. Conover, Practical Nonparametric Statistics, third ed., Wiley.com, 1999, Retrieved from <https://www.wiley.com/en-us/Practical+Nonparametric+Statistics%2C+3rd+Edition-p-9780471160687>.
- [69] J.L. Vila-Chá, B.P. Ferreira, F.A. Pires, An adaptive multi-temperature isokinetic method for the RVE generation of particle reinforced heterogeneous materials, part I: Theoretical formulation and computational framework, *Mech. Mater.* 163 (2021) 104069, <http://dx.doi.org/10.1016/j.mechmat.2021.104069>, Retrieved from <https://www.sciencedirect.com/science/article/pii/S016766362100291X>.
- [70] B.P. Ferreira, F.M.A. Pires, M.A. Bessa, CRATE: A Python package to perform fast material simulations, *J. Open Source Softw.* 8 (87) (2023) 5594, <http://dx.doi.org/10.21105/joss.05594>.
- [71] W.F. Dos Santos, I.A. Rodrigues Lopes, F.M. Andrade Pires, S.P. Proença, Exploring novel mechanical metamaterials: Unraveling deformation mode coupling and size effects through second-order computational homogenisation, *Int. J. Solids Struct.* (2024) 112724, <http://dx.doi.org/10.1016/j.ijssolstr.2024.112724>, Retrieved from <https://linkinghub.elsevier.com/retrieve/pii/S0020768324000817>.
- [72] D. Qi, P. Zhang, W. Wu, K. Xin, H. Liao, Y. Li, D. Xiao, R. Xia, Innovative 3D chiral metamaterials under large deformation: Theoretical and experimental analysis, *Int. J. Solids Struct.* 202 (2020) 787–797, <http://dx.doi.org/10.1016/j.ijssolstr.2020.06.047>, Retrieved from <https://linkinghub.elsevier.com/retrieve/pii/S0020768320302699>.
- [73] M. Fu, F. Liu, L. Hu, A novel category of 3D chiral material with negative Poisson's ratio, *Compos. Sci. Technol.* 160 (2018) 111–118, <http://dx.doi.org/10.1016/j.compscitech.2018.03.017>, Retrieved from <https://linkinghub.elsevier.com/retrieve/pii/S0266353817325435>.
- [74] J. Blank, K. Deb, pymoo: Multi-objective optimization in Python, *IEEE Access* 8 (2020) 89497–89509.
- [75] D.A. Van Veldhuizen, *Multiobjective Evolutionary Algorithms: Classifications, Analyses, and New Innovations* (Ph.D. thesis), Air Force Institute of Technology, USA, 1999, AAI9928483.
- [76] B.H. Smith, S. Szyniszewski, J.F. Hajjar, B.W. Schafer, S.R. Arwade, Steel foam for structures: A review of applications, manufacturing and material properties, *J. Constr. Steel Res.* 71 (2012) 1–10, <http://dx.doi.org/10.1016/j.jcsr.2011.10.028>.
- [77] M.F. Ashby, *Metal Foams: A Design Guide*, Butterworth-Heinemann, Boston, 2000.
- [78] R.E. Miller, A continuum plasticity model for the constitutive and indentation behaviour of foamed metals, *Int. J. Mech. Sci.* 42 (4) (2000) 729–754, [http://dx.doi.org/10.1016/S0020-7403\(99\)00021-1](http://dx.doi.org/10.1016/S0020-7403(99)00021-1).
- [79] V.S. Deshpande, N.A. Fleck, Isotropic constitutive models for metallic foams, *J. Mech. Phys. Solids* 48 (6) (2000) 1253–1283, [http://dx.doi.org/10.1016/S0022-5096\(99\)00082-4](http://dx.doi.org/10.1016/S0022-5096(99)00082-4).
- [80] A. Reyes, O.S. Hopperstad, T. Berstad, A.G. Hanssen, M. Langseth, Constitutive modeling of aluminum foam including fracture and statistical variation of density, *Eur. J. Mech. A Solids* 22 (6) (2003) 815–835, <http://dx.doi.org/10.1016/j.euromechsol.2003.08.001>.
- [81] L.-R. Long, M.-H. Fu, L.-L. Hu, Novel metamaterials with thermal-torsion and tensile-torsion coupling effects, *Compos. Struct.* 259 (2021) 113429, <http://dx.doi.org/10.1016/j.compstruct.2020.113429>, Retrieved from <https://linkinghub.elsevier.com/retrieve/pii/S0263822320333584>.
- [82] C. Yang, K. Yang, Y. Tian, M. Fu, L. Hu, Theoretical analysis on the stiffness of compression-torsion coupling metamaterials, *Extreme Mech. Lett.* 46 (2021) 101336, <http://dx.doi.org/10.1016/j.eml.2021.101336>, Retrieved from <https://linkinghub.elsevier.com/retrieve/pii/S2352431621000997>.