Delft University of Technology

High-Dimensional Bayesian Optimisation with Large-Scale Constraints - An Application to Aeroelastic Tailoring

Maathuis, H.F.; De Breuker, R.; Castro, Saullo G.P.

Citation (APA)
Maathuis, H. F., De Breuker, R., & Castro, S. G. P. (2024). High-Dimensional Bayesian Optimisation with Large-Scale Constraints - An Application to Aeroelastic Tailoring. In *Proceedings of the AIAA SCITECH 2024 Forum* Article AIAA 2024-2012 (AIAA SciTech Forum and Exposition, 2024). American Institute of Aeronautics and Astronautics Inc. (AIAA). https://doi.org/10.2514/6.2024-2012

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

High-Dimensional Bayesian Optimisation with Large-Scale Constraints - An Application to Aeroelastic Tailoring

Hauke Maathuis*, Roeland De Breuker [†] and Saullo G.P. Castro [‡]
Delft University of Technology, Delft, The Netherlands

Design optimisation potentially leads to lightweight aircraft structures with lower environmental impact. Due to the high number of design variables and constraints, these problems are ordinarily solved using gradient-based optimisation methods, leading to a local solution in the design space while the global space is neglected. Bayesian Optimisation is a promising path towards sample-efficient, global optimisation based on probabilistic surrogate models. While for problems with a low number of design variables, Bayesian Optimisation methods have demonstrated their strength, the scalability to high-dimensional problems while incorporating large-scale constraints is still lacking. Especially in aeroelastic tailoring where directional stiffness properties are embodied into the structural design of aircraft, to control aeroelastic deformations and to increase the aerodynamic and structural performance, the safe operation of the system needs to be ensured by involving constraints resulting from different analysis disciplines. Hence, a global design space search becomes even more challenging. The present study attempts to tackle the problem by using high-dimensional Bayesian Optimisation in combination with a dimensionality reduction approach to solve the optimisation problem occurring in aeroelastic tailoring, presenting a novel approach for high-dimensional problems with large-scale constraints. Experiments on well-known benchmark cases with black-box constraints show that the proposed approach can incorporate large-scale constraints.

I. Introduction

Humanity is driving towards a greener future, particularly within sustainable aviation with an ever-growing demand for more efficient and environmentally friendly aircraft. Improving high-performance aircraft is a crucial step towards a sustainable and widely accepted aviation sector. Achieving this involves optimising structural designs to reduce energy consumption. Therefore, developing new methods and technologies to mitigate environmental impact becomes inevitable.

Aeroelastic tailoring is a promising technique for weight reduction in high aspect ratio wings to enhance their performance. Therein, directional stiffness properties are embodied into the structural design of aircraft to control aeroelastic deformations to increase the aerodynamic and structural performance by optimising, for instance, ply thicknesses and angles (Shirk et al. 1986). This process involves multiple disciplines, known as multi-disciplinary design optimisation (MDO), ensuring the system's feasibility and safe operation.

Evaluating these complex aeroelastic models' computational expense and time-consuming nature demand efficient optimisation algorithms. Gradient-based methods are often utilised in this context to find a local optimum with a reasonable amount of model evaluations. To allow the use of gradient-based optimisation in the aeroelastic tailoring of composite wings, a convenient route makes use of the so-called lamination parameters to describe the laminated composite parts, allowing a condensed description of the membrane, bending, and coupled stiffness terms using continuous variables (Dillinger et al. 2013). With these continuous variables, it is possible to compute gradients in high-dimensional optimisation problems. At the same time, large-scale constraints arise from considering various structural analysis disciplines, such as buckling and other failure criteria. However, multiple challenges emerge from this gradient-based approach. Firstly, the computation of gradients is not always easy, e.g., if the model's source code is not available, thus relying on approaches like Finite Differences, leading to prohibitive computational costs, which motivates the use of gradient-free methods. Secondly, the response surface for feasible designs in aeroelastic tailoring is

---

*Ph.D. Candidate, Aerospace Structures and Materials, Faculty of Aerospace Engineering, Delft University of Technology, h.f.maathuis@tudelft.nl
[†]Associate Professor, Aerospace Structures and Materials, Faculty of Aerospace Engineering, Delft University of Technology, r.debreuker@tudelft.nl, Associate Fellow AIAA
[‡]Associate Professor, Aerospace Structures and Materials, Faculty of Aerospace Engineering, Delft University of Technology, s.g.p.castro@tudelft.nl

known to be multi-modal, trapping gradient-based methods in local optima and neglecting the global design space, potentially hindering the discovery of better designs. Hence, there is a desire to develop methods exploring the global design space, optimising structures to achieve lighter aircraft configurations.

The optimisation problem at hand can be formulated as follows:

$$\min_{\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^D} f(\mathbf{x}) \text{ s.t.} \forall i \in \{1, ..., G\}, c_i(\mathbf{x}) \leq 0, \tag{1}$$

where $\mathcal{X} \subset \mathbb{R}^D$ is a D-dimensional space of potential designs, $f(\mathbf{x}) : \mathbf{x} \in \mathcal{X} \to \mathbb{R}$ the objective function and G constraints arising from the multi-disciplinary analyses. Overall, this optimisation problem can also be seen as a multi-output problem where the model maps from a vector of design variables to the objective function and G constraints.

Due to the expensive nature of evaluating the objective function and especially the associated constraints, a sample-efficient algorithm is crucial. Bayesian Optimization (BO) is a powerful method for complex and costly problems, extensively applied across various domains, including aircraft design Saves et al. (2022). In BO, the expensive-to-evaluate functions that represent the objective and constraints, in many problems seen as a black box, are replaced by a computationally cheap surrogate model by using e.g. Gaussian Processes (GP) (Frazier 2018). While many authors have shown that for lower dimensional problems, BO methods perform well, high-dimensional cases pose significant challenges due to the curse of dimensionality (Eriksson and Jankowiak 2021; Priem 2020), resulting from the fact that high dimensional search spaces are difficult to explore exhaustively. However, BO offers a probabilistic approach to efficiently guide through the design space to find promising regions and reduce the computational burden. While these algorithms bring a variety of advantages, their scalability to high-dimensional problems with many constraints, as is often the case in engineering design, remains a great challenge.

The present study focuses on employing high-dimensional BO algorithms for aeroelastic tailoring while considering large-scale constraints arising from the multi-disciplinary analyses, as formulated in Equation 1. First, BO for unconstrained and constrained problems is introduced, and the difficulties in terms of scalability are highlighted. Subsequently, dimensionality reduction in the context of constrained BO is presented before the theory is applied to the aeroelastic tailoring optimisation problem.

The novelty of this paper lies in the formulation of a high-dimensional BO method with a dimensionality reduction approach that lowers the computational burden arising from the incorporation of a large number of constraints. Subsequently, the methodology is applied to the $10D$ Ackley function with two black-box constraints as well as to the $7D$ speed reducer problem with 11 black box constraints before some preliminary results are shown for the use in aeroelastic tailoring.

II. High-Dimensional Constrained Bayesian Optimisation

This section briefly introduces Bayesian Optimization (BO) within the context of high dimensionality and constraints. Gaussian Processes (GPs) are introduced as the preferred surrogate modelling technique. Subsequently, GPs are linked to unconstrained BO, which is then expanded to address the constrained scenario, followed by an outline of the challenges encountered in this work.

A. Gaussian Processes

A Gaussian Process in the context of BO is used to compute a surrogate model which is fast to evaluate and from which the optimum can be obtained cheaply. Recall that $\mathcal{X} \subset \mathbb{R}^D$ is a D-dimensional domain and the corresponding minimisation problem is presented in Equation 1. Starting from a Design of Experiments (DoE), written as $\mathcal{D} = \{\mathbf{x}_i, f(\mathbf{x}_i)\}_{i=1,...,N}$, where $\mathbf{x}_i \in \mathbb{R}^D$ is the i-th of N samples and $f(\mathbf{x}_i) : \mathcal{X} \to \mathbb{R}$ the objective function, mapping from the design space to a scalar value. Typically, GPs are employed within BO to learn a surrogate model $\hat{f}(\mathbf{x}) : \mathcal{X} \to \mathbb{R}$ of the objective function f from this given data set $\mathcal{D}$. Therefore, it is assumed that the objective function f follows a GP, also called a multivariate normal distribution $\mathcal{N}$. By defining the mean $m : \mathcal{X} \to \mathbb{R}$ and covariance $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, the surrogate can be denoted as

$$f(\mathbf{x}) \sim \hat{f}(\mathbf{x}) = \mathcal{GP}\left(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')\right) = \mathcal{N}\left(\mu(\mathbf{x}), \sigma(\mathbf{x})^2\right), \tag{2}$$

also known as the prior. This encodes the a priori belief that the observations are distributed normally. A common covariance function, sometimes referred to as kernel, is, for instance, the so-called squared exponential kernel $k(\mathbf{x}, \mathbf{x}')$

defined as

$$k(\mathbf{x}, \mathbf{x}') = s^2 \exp\left(-\frac{1}{2} \sum_{i=1}^{D} \left(\frac{x_i - x_i'}{l_i}\right)^2\right), \tag{3}$$

encoding the similarity between two chosen points $\mathbf{x}$ and $\mathbf{x}'$ (Rasmussen and Williams 2006). The parameter l_i for $i = 1, ..., D$ is called the length scale and measures the distance for being correlated along x_i. Together with s^2, the parameters form a set of so-called hyperparameters $\boldsymbol{\theta} = \{l_1, ..., l_D, s^2\}$ (in total $D + 1$ parameters) which need to be determined to train the model with respect to the target function. The kernel matrix is defined as $\mathbf{K} = \left[k(\mathbf{x}_i, \mathbf{x}_j)\right]_{i,j=1,...,N} \in \mathbb{R}^{N \times N}$. The kernel needs to be defined such that $\mathbf{K}$ is symmetric positive definite to ensure its invertibility. The positive definite symmetry is guaranteed if and only if the used kernel is a positive definite function, as detailed in Schoenberg (1938).

Considering a new query point $\mathbf{x}_* \in \mathcal{X}$, the stochastic process in Equation 2 can be used to predict the new query point using Bayes' rule through the posterior distribution, defined as

$$f_* \sim \mathcal{N}\left(\mu(\mathbf{x}_*), k(\mathbf{x}_*, \mathbf{x}_*)\right) \tag{4}$$

The posterior mean $\hat{\mu}(\bullet)$ and covariance function $\hat{\sigma}(\bullet)$ are computed with

$$\hat{\mu}(\mathbf{x}_*) = \mathbf{k}(\mathbf{x}_*, \mathbf{x})\mathbf{K}(\mathbf{x}, \mathbf{x})^{-1}\mathbf{f}, \tag{5}$$

$$\hat{\sigma}(\mathbf{x}_*) = k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}(\mathbf{x}_*, \mathbf{x})\mathbf{K}(\mathbf{x}, \mathbf{x})^{-1}\mathbf{k}(\mathbf{x}, \mathbf{x}_*), \tag{6}$$

where $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N]$ is the collection of samples and $\mathbf{f} = [f_1, f_2, ..., f_N]$ of computed objective values in $\mathcal{D}$. The values of the hyperparameters $\boldsymbol{\theta}$ are determined by maximising the marginal likelihood. More detailed information can be found in Rasmussen and Williams (2006).

B. Unconstrained Bayesian Optimisation

Up to this point, the GP has been computed based on the initial samples contained in $\mathcal{D}$. BO now aims to iteratively increase the accuracy of the surrogate model, enriching the DoE while exploring the design space. Thus, leveraging the acquired data, the endeavour is to pinpoint regions where optimal values are anticipated. For a thorough derivation, unconstrained BO is considered first. The problem at hand can be written as

$$\min_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}). \tag{7}$$

An acquisition function $\alpha : \mathcal{X} \to \mathbb{R}$ is used to guide the optimisation through the design space while trading off exploration and exploitation based on the posterior mean and variance defined in Equation 5. The former describes the exploration of the whole design space, whereas the latter tries converging to an optimum based on the data observed. However, a multitude of acquisition functions exist. A popular choice is the so-called Expected Improvement (EI) (Mockus, J. et al. 1978), denoted as

$$\alpha_{\text{EI}}(\mathbf{x}) = \begin{cases} 0 & \text{if } \hat{\sigma}(\mathbf{x}) = 0 \\ (f_{\min} - \hat{\mu}(\mathbf{x})) \Phi\left(\frac{f_{\min} - \hat{\mu}(\mathbf{x})}{\hat{\sigma}(\mathbf{x})}\right) + \hat{\sigma}(\mathbf{x})\phi\left(\frac{f_{\min} - \hat{\mu}(\mathbf{x})}{\hat{\sigma}(\mathbf{x})}\right) & \text{else} \end{cases} \tag{8}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the cummulative distribution and probability density function, whereas $f_{\min}$ represents the observed minimum. By maximising Equation 8 over the design space $\mathcal{X}$, the new query point $\mathbf{x}_*$ can be found (Frazier 2018)

$$\mathbf{x}_* \in \operatorname*{argmax}_{\mathbf{x} \in \mathcal{X}} \alpha_{\text{EI}}(\mathbf{x}). \tag{9}$$

C. Constrained Bayesian Optimisation

Most engineering design problems involve constraints, which can be integrated into the previously introduced BO method. There are plenty of algorithms to do so, e.g. (Gardner, Kusner, et al. 2014; Michael A Gelbart et al. 2014; Hernández-Lobato, Michael A. Gelbart, et al. 2016). Assuming that the output of a model evaluation at design point $\mathbf{x}_i$ is not only the objective function $f(\mathbf{x}_i)$ but also contains a mapping from the design space to a collection of G constraints $\mathbf{c}(\mathbf{x}_i) : \mathcal{X} \to \mathbb{R}^G$, the DoE for this case can be written as $\mathcal{D} = \{\mathbf{x}_i, f(\mathbf{x}_i), \mathbf{c}(\mathbf{x}_i)\}_{i=1,..,N}$. The new design point found

in Equation 9 needs to lie in the feasible space $\mathcal{X}_f$, written as $\mathbf{x}_* \in \mathcal{X}_f \subset \mathcal{X}$ where $\mathcal{X}_f := \{\mathbf{x} \in \mathcal{X} \text{ s.t. } \hat{c}_{1:G}(\mathbf{x}) \leq 0\}$. Among others, Gardner, Kusner, et al. (2014) propose to model each constraint $c_j(\mathbf{x})$, $j = 1, ..., G$ by an independent surrogate model, the same way as it is done for the objective function

$$c_i(\mathbf{x}) \sim \hat{c}_i(\mathbf{x}) = \mathcal{GP}\left(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')\right) = \mathcal{N}\left(\mu(\mathbf{x}), \sigma(\mathbf{x})^2\right), \tag{10}$$

leading to $G + 1$ GP models in total, enabling the extension of Equation 8 for constrained problems, also referred to as Expected Feasible Improvement (EFI), written as

$$\alpha_{\text{EFI}} = \alpha_{\text{EI}} \prod_{i=1}^{G} \Pr\left(\hat{c}_i(\mathbf{x} \leq 0)\right). \tag{11}$$

Accordingly, within the acquisition strategy, the sub-problem

$$\mathbf{x}_* \in \underset{\mathbf{x} \in \mathcal{X}_f \subset \mathcal{X}}{\arg\max} \; \alpha_{\text{EFI}}(\mathbf{x}) \tag{12}$$

has to be solved. This subsection solely aims to introduce the crucial aspects of constrained BO briefly and shall stress the fact that each constraint needs to be modelled via a separate GP model.

Of course, a multitude of alternatives to incorporate constraints exists. Among these approaches, for instance, is the use of Thompson Sampling (TS) (Hernández-Lobato, Requeima, et al. 2017) in the constrained setting, as proposed by Eriksson, Pearce, et al. (2019) and employed in the course of this work.

D. High-Dimensional Problems

As presented, BO algorithms consist of two components, namely the GPs to model the surrogate based on Bayesian statistics (Rasmussen and Williams 2006) and the acquisition function to determine where to query the next point to converge towards the minimiser of the objective function. While these algorithms have been proven very efficient for lower-dimensional problems (Binois and Wycoff 2022), the scaling to higher dimensions implies some difficulties:

- The curse of dimensionality, which essentially states that with increasing number of dimensions, the size of the design space increases exponentially, resulting in an intractable search through the whole design space
- As the dimensions increase, so does the number of tunable hyperparameters $\boldsymbol{\theta} \in \mathbb{R}^{D+1}$, leading to a computationally costly learning of the GP
- Higher-dimensional problems usually require more samples N to construct the surrogate model accurately. Since the covariance matrix is $\mathbf{K} \in \mathbb{R}^{N \times N}$, the inversion of this matrix becomes more costly with a complexity of $O(N^3)$
- Not enough information can be collected, leading to the fact that in the D-dimensional hyperspace, most of the samples have a big distance to each other, not allowing for an efficient correlation.
- Acquisition optimisation suffers from large uncertainty in a high-dimensional setting and thus requires more surrogate model evaluation (Binois and Wycoff 2022)

Different approaches have been used to mitigate the problem of high dimensionality when no or only a few constraints are involved. In Wang et al. (2016), the authors use random projection methods to project the high-dimensional inputs to a lower dimensional subspace, ending up by constructing the GP model directly on the lower dimensional space, drastically reducing the number of hyperparameters. Raponi et al. (2020) and Antonov et al. (2022) use (kernel) Principal Component Analysis on the input space to identify a reduced set of dimensions based on the evaluated samples. Afterwards, the surrogate model is trained in this reduced dimensional space. Eriksson and Jankowiak (2021) use a hierarchical Bayesian model, assuming that some design variables are more important than others. Employing a sparse axis-aligned prior on the length scale will remove dimensions unless the accumulated data tells otherwise. However, a high computational overhead is shown in (Santoni et al. 2023). Similarly, decomposing methods exist where the original space is decomposed using additive methods (Kandasamy et al. 2016; Ziomek and Bou-Ammar 2023). Within the Trust-Region Bayesian Optimisation (TuRBO) algorithm, (Eriksson, Pearce, et al. 2019) split up the design space in multiple trust regions. These trust regions are defined as hyper-rectangles of size $L \in \mathbb{R}$. The size is then determined

with the help of the length scale l_i, already defined in Equation 3, and a base length scale L as

$$L_i = \frac{l_i L}{\left(\prod_{j=1}^{D} l_j\right)^{1/D}}. \tag{13}$$

In this approach, an independent GP model is constructed within each trust region and batched Thompson Sampling (TS) Thompson (1933) is used as the acquisition function.

These methods are all considering unconstrained problem, although Eriksson and Jankowiak (2021) apply the algorithm for the constrained MOPTA08 (Anjos, M. 2008) problem by using penalty methods. Subsequently, in Eriksson and Poloczek (2021), the TuRBO algorithm is extended to take into account multiple constraints. Here, each constraint is modelled by an independent GP process and batched TS is extended for constrained problems.

As shown, to scale BO to high-dimensional problems, strong assumptions have to be made to mitigate the aforementioned obstacles. While the mentioned authors show promising results, an approach of taking into account large-scale constraints, as in the aeroelastic tailoring case where $G > 10^3$, is still lack. In this work we choose to use the constrained TuRBO algorithm (Scalable Constrained Bayesian Optimisation, SCBO) for high-dimensional BO. In the following, an extension to this method is presented to mitigate the problem of large-scale constraints.

III. Large-Scale Constrained BO via Reduced-dimensional Output Spaces

Recall the optimisation problem formulated in Equation 1. By using constrained BO methods, as shown earlier, each of the G constraints needs to be modelled with an independent GP model, denoted as $\hat{c}_i(\mathbf{x})$. This work follows the idea of Higdon et al. (2008) to construct the surrogates on a lower dimensional output subspace. This subspace may be found by using dimensionality reduction methods such as Principal Component Analysis (PCA) (Jolliffe and Cadima 2016) on the training data in $\mathcal{D}$. An extended version of PCA is the kernel PCA (kPCA), presented by Schölkopf et al. (1998). While performing the DoE, apart from the samples $\mathbf{x}_i$ and the corresponding objective function f_i also the constraint values $\mathbf{c} : \mathcal{X} \rightarrow \mathbb{R}^G$ are contained in $\mathcal{D}$, enabling the construction of a matrix

$$\mathbf{C}(\mathbf{x}) = \begin{bmatrix} \mathbf{c}(\mathbf{x}_1)^T \\ \mathbf{c}(\mathbf{x}_2)^T \\ \vdots \\ \mathbf{c}(\mathbf{x}_N)^T \end{bmatrix} = \begin{bmatrix} c_1(\mathbf{x}_1) & c_2(\mathbf{x}_1) & ... & c_G(\mathbf{x}_1) \\ c_1(\mathbf{x}_2) & c_2(\mathbf{x}_2) & ... & c_G(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ c_1(\mathbf{x}_N) & c_2(\mathbf{x}_N) & ... & c_G(\mathbf{x}_N) \end{bmatrix} \in \mathbb{R}^{N \times G} \tag{14}$$

with N samples and G constraints. The superscript T denotes the transpose.

A. Principle Component Analysis (PCA)

Within PCA, a linear combination with maximum variance is sought, such that

$$\mathbf{C}\mathbf{v} = \lambda\mathbf{v} \tag{15}$$

where $\mathbf{v}$ is a vector of constants. These linear combinations are called the principle components of the data contained in $\mathbf{C}$ and can be linked with the Singular Value Decomposition (SVD) (Jolliffe 2002), leading to

$$\mathbf{C} = \mathbf{\Psi}\mathbf{\Sigma}\mathbf{\Phi}^T. \tag{16}$$

The matrix $\mathbf{\Psi} = [\Psi_1, ..., \Psi_r] \in \mathbb{R}^{N \times r}$ has orthonormal columns which are called the left singular eigenvectors, $\mathbf{\Sigma} = diag(\sigma_1, ..., \sigma_r) \in \mathbb{R}^{r \times r}$ is a diagonal matrix, containing the eigenvalues and $\mathbf{\Phi} = [\phi_1, ..., \phi_r] \in \mathbb{R}^{G \times r}$ contains the corresponding right singular eigenvectors. Here, it is assumed that the SVD automatically sorts the eigenvalues and eigenvectors in descending order. By investigating the eigenvalues in $\mathbf{\Sigma}$, and choosing the ones with the g highest values, the truncated decomposition is obtained, consisting of the reduced basis containing g orthogonal basis vectors in $\mathbf{\Psi}_g \in \mathbb{R}^{G \times g}$ with $g \ll G$. The new basis vectors can subsequently be used as a projection $\mathbf{\Psi}_g : \mathbb{R}^G \rightarrow \mathbb{R}^g$ to project the matrix $\mathbf{C}$ onto the reduced subspace $\tilde{\mathbf{C}} \in \mathbb{R}^{N \times g}$, written as

$$\tilde{\mathbf{C}} = \mathbf{\Psi}_g^T\mathbf{C} \tag{17}$$

and for each new vector of constraint values $\mathbf{c}_* \in \mathbb{R}^G$

$$\tilde{\mathbf{c}}_* = \mathbf{\Psi}_g^T \mathbf{c}_* \tag{18}$$

Summarising, the G constraints $\mathbf{c}(\mathbf{x})$ can be represented on a reduced subspace through the mapping $\mathbf{\Psi}_g$ while the eigenvalues σ_i give an indication about the loss of information, potentially drastically lowering the number of constraints that need to be modelled. A graphical interpretation is depicted in Figure 1a. For a more thorough derivation of this method, the reader is referred to Jolliffe and Cadima (2016).

B. Kernel Principle Component Analysis (kPCA)

While PCA can be seen as a linear dimensionality reduction technique, in Schölkopf et al. (1998) the authors present an extension, called kernel PCA, using a nonlinear projection step to depict nonlinearities in the data. Similarly to the PCA algorithm, the starting point are the (centred) samples $\mathbf{c}_i(\mathbf{x}_i) \in \mathbb{R}^G \; \forall i \in \{1, ..., N\}$.

Let $\mathcal{F}$ be a dot product space (in the following, also called feature space) of arbitrary large dimensionality. A nonlinear map $\boldsymbol{\phi}(\mathbf{x})$ is defined as $\boldsymbol{\phi} : \mathbb{R}^G \to \mathcal{F}$. This map is used to construct a covariance matrix C defined as

$$C = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\phi}(\mathbf{c}(\mathbf{x}_i)) \boldsymbol{\phi}(\mathbf{c}(\mathbf{x}_i))^T. \tag{19}$$

The corresponding eigenvalues and eigenvectors in $\mathcal{F}$ are computed by solving

$$C\mathbf{v} = \lambda \mathbf{v}. \tag{20}$$

As stated earlier, since the function $\boldsymbol{\phi}$ maps possibly to a very high-dimensional space $\mathcal{F}$, solving the eigenvalue problem therein may be costly. A workaround is used to avoid computations in $\mathcal{F}$. Therefore, similar to the formulation of the GP models in Section II.A, a kernel $k : \mathbb{R}^G \times \mathbb{R}^G \to \mathbb{R}$ is defined as

$$k(\mathbf{c}(\mathbf{x}_i), \mathbf{c}(\mathbf{x}_j)) = \boldsymbol{\phi}(\mathbf{c}(\mathbf{x}_i))^T \boldsymbol{\phi}(\mathbf{c}(\mathbf{x}_j)) \tag{21}$$

and the corresponding kernel matrix $\mathbf{K}_{ij} \in \mathbb{R}^{N \times N}$ as

$$\mathbf{K}_{ij} := \left(\boldsymbol{\phi}(\mathbf{c}(\mathbf{x}_j)), \boldsymbol{\phi}(\mathbf{c}(\mathbf{x}_j)) \right). \tag{22}$$

By solving the eigenvalue problem for non-zero eigenvalues

$$\mathbf{K}\alpha = \lambda \alpha \tag{23}$$

the eigenvalues $\lambda_1 \leq ... \leq \lambda_N$ and eigenvectors $\boldsymbol{\alpha}^1, ..., \boldsymbol{\alpha}^N$ are obtained. This part can be seen as the linear PCA, as presented before, although in the space $\mathcal{F}$. To map a test point $\mathbf{c}_*(\mathbf{x})$ from the feature space $\mathcal{F}$ to the q-th principle component $\mathbf{v}^q$ of Equation 20, the following relationship is evaluated

$$\left((\mathbf{v}^q)^T \boldsymbol{\phi}(\mathbf{c}_*(\mathbf{x})) \right) = \sum_{i=1}^{N} \alpha_i^q (\boldsymbol{\phi}(\mathbf{c}(\mathbf{x}_i)^T \boldsymbol{\phi}(\mathbf{c}_*(\mathbf{x}))) \equiv \tilde{\mathbf{c}}_*(\mathbf{x}_*). \tag{24}$$

A graphical interpretation can be found in Figure 1b. The kernel function in Equation 21 can also be replaced by another a priori chosen kernel function. Examples of kernels and a more detailed derivation of kPCA can be found in the cited literature.

C. Dimensionality Reduction for Large-Scale Constraints

When large-scale constraints are involved, the computational time scales drastically since for each constraint one GP model has to be constructed and trained. Thus, describing the constraints on a lower dimensional subspace allows to significantly lower the computational burden. This idea is based on the work of Higdon et al. (2008), who project the simulation output onto a lower dimensional subspace where the GP models are constructed. Other works extended this method then by employing, among others, kPCA as well as manifold learning techniques to account for nonlinearities (Wei Xing et al. 2015; W.W. Xing et al. 2016). However, the aforementioned authors try to approximate

PDE model simulations with high-dimensional outputs, whereas, to the best of the authors' knowledge, the combination of dimensionality reduction techniques for use in high-dimensional BO with large-scale constraints is novel.

The methods herein presented are capable of extracting the earlier introduced, most important principle components of available data, reducing the required amount of GP models to g instead of G, with $\mathbf{v}_j$ as the j-th orthogonal basis vector. After projecting the data onto the lower dimensional subspace by using either PCA as in equations 17 or 18 respectively, or kPCA in equation 24, independent GPs are constructed on the reduced-dimensional output space, formulated as

$$\tilde{\mathbf{c}}_i \sim \hat{\tilde{\mathbf{c}}}_i = \mathcal{GP}\left(m_i(\mathbf{x}), k_i(\mathbf{x}, \mathbf{x}')\right) \forall i \in \{1, ..., g\}. \tag{25}$$

A graphical interpretation is depicted in Figure 1.

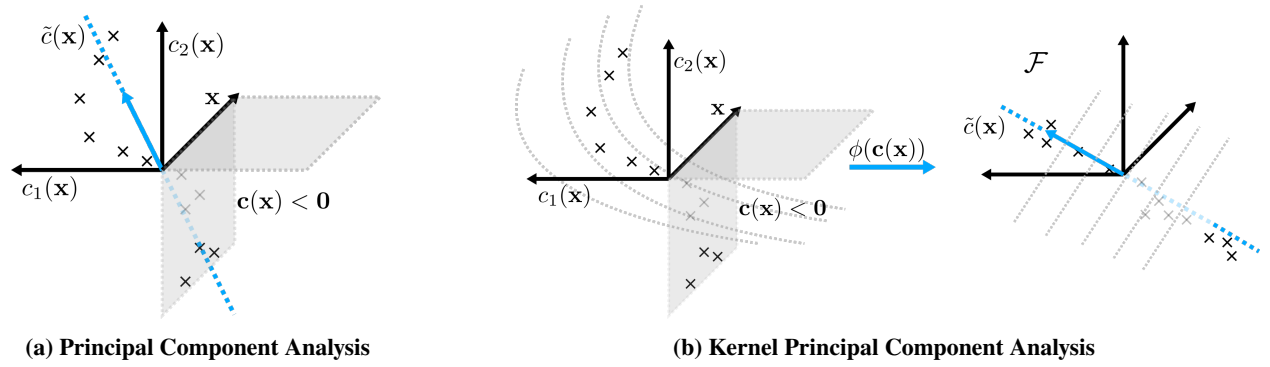**(a) Principal Component Analysis**   **(b) Kernel Principal Component Analysis**

Fig. 1   Graphical interpretation of dimensionality reduction for constraints. On the left, PCA as a linear method is depicted, finding the lower dimensional subspace (blue arrow). On the right, the nonlinear extension, kPCA, is shown, first using a nonlinear kernel to map into the infinite dimensional space $\mathcal{F}$ and subsequently perform the standard PCA. The figure is inspired by Schölkopf et al. (1998).

IV. Aeroelastic Tailoring: A Multi-Disciplinary Design Optimisation Problem

The aeroelastic tailoring model used in this work is based on the preliminary design framework called *Proteus* developed by the Delft University of Technology, Aerospace Structures and Materials, initiated with the work of De Breuker (2011). In general, the framework transforms a three-dimensional wingbox made of laminate panels into a three-dimensional beam model, depicted in Figure 2. A panel in this setting can be, for instance, the upper and lower skin cover or the front and rear spars. The design regions can be discretised by means of various panels along the chord-wise and span-wise directions, ultimately defining the number of design variables.
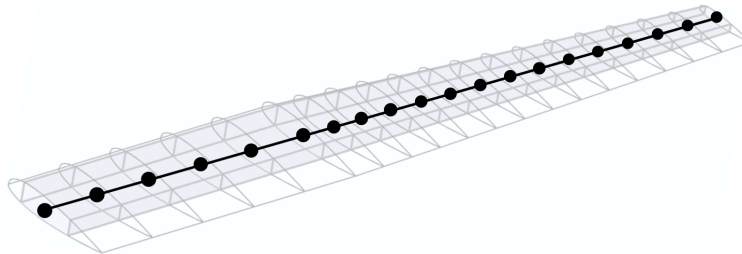

Fig. 2   Beam representation of the wing structure

The design variables consist of the lamination parameters and the thickness of each panel, respectively, denoted with the superscripts "lam" and "t". Note that, in the present study the lamination parameters are denoted by $\mathbf{x}_i^{lam}$, whereas

in the composite community they are often referred to as $\xi_{1,2,3,4}^{A,D}$.

$$\mathbf{x} = \left\{ \mathbf{x}_1^{lam}, x_1^t, ..., \mathbf{x}_{n_P}^{lam}, x_{n_P}^t \right\}. \tag{26}$$

For the sake of illustration, the single panels and the locally optimised thickness and stiffness are depicted in Figure 3 where a chord-wise panel discretisation of 2 is chosen. Recall the already introduced optimisation problem from
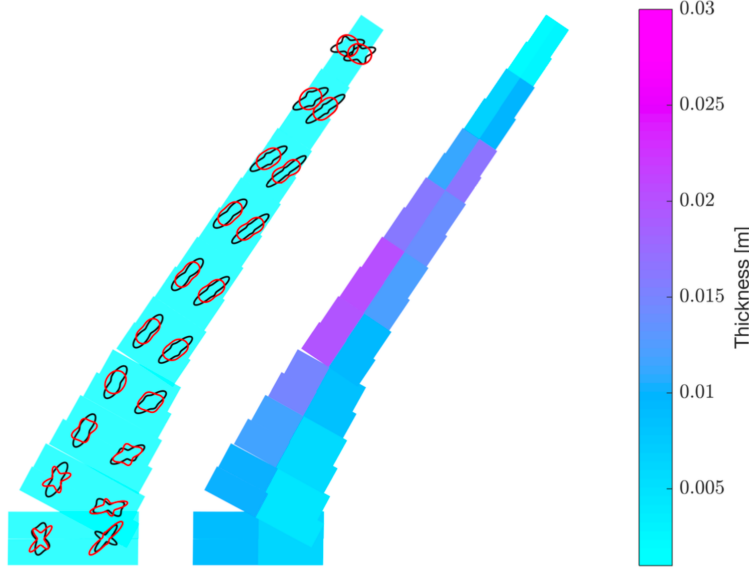

Fig. 3 Stiffness and thickness distributions (In-plane stiffness: black, out-of-plane stiffness: red), adopted from Rajpal (2021)

Equation 1. As mentioned earlier, to ensure the safe operation and feasibility of the system, constraints are computed based on different analysis methods discussed in the following. For the sake of illustration, these constraints are compactly derived here.

The lamination parameter feasibility constraints $\mathbf{c}_{lpf}$ ensure that the laminates satisfy certain interdependencies and are analytic equations, derived in (Raju et al. 2014), ending up in 6 inequality constraints per panel

$$\mathbf{c}_{lpf} = \left[\mathbf{g}_1^T(\mathbf{x})\ \mathbf{g}_2^T(\mathbf{x})\ \mathbf{g}_3^T(\mathbf{x})\ \mathbf{g}_4^T(\mathbf{x})\ \mathbf{g}_5^T(\mathbf{x})\ \mathbf{g}_6^T(\mathbf{x}) \right]^T \leq \mathbf{0}. \tag{27}$$

Note that each of the six constraints is evaluated for all n_P panels, thus $\mathbf{g}_i : \mathcal{X} \to \mathbb{R}^{n_P}$ and $\mathbf{c}_{lpf} : \mathcal{X} \to \mathbb{R}^{6n_P}$. The lamination parameters can be used with the classical laminate theory to construct the following relationship

$$\begin{bmatrix} N \\ M \end{bmatrix} = \begin{bmatrix} \mathbf{A}(\mathbf{x}) & \mathbf{B}(\mathbf{x}) \\ \mathbf{B}(\mathbf{x}) & \mathbf{D}(\mathbf{x}) \end{bmatrix} \begin{bmatrix} \epsilon^0 \\ \kappa \end{bmatrix} \tag{28}$$

This relationship encodes the dependency of the design variables $\mathbf{x}$ with the stiffness of the system. A cross-section modeller (Ferede and Abdalla 2014), based on a variational approach, is used to obtain the element Timoshenko cross-sectional stiffness matrix $\mathbf{C} \in \mathbb{R}^{6 \times 6}$ by relating the strains ϵ to the applied forces and moment σ or F_i, M_i respectively, as in

$$\sigma = \mathbf{C}\epsilon \rightarrow \begin{bmatrix} F_1 & F_2 & F_3 & M_1 & M_2 & M_3 \end{bmatrix}^T = \mathbf{C} \begin{bmatrix} \epsilon_{11} & \epsilon_{12} & \epsilon_{13} & \kappa_1 & \kappa_2 & \kappa_3 \end{bmatrix}^T \tag{29}$$

with κ_1 as the twist and κ_2, κ_3 the bending curvatures. Therewith, the properties of a $2D$ cross section are mapped onto the corresponding beam element node, leading to a 6×6 element Timoshenko beam stiffness matrix. The corresponding

8

beam strain energy in the continuous form can be derived like

$$\mathcal{U} = \frac{l_0}{2} \int_0^1 \boldsymbol{\epsilon}^T \mathbf{C} \boldsymbol{\epsilon} \, d\xi. \tag{30}$$

By discretisation and introduction of the element degrees of freedom $\mathbf{p}$, the linear constitutive stiffness matrix of the beam element, at this point neglecting geometric and material nonlinearities, can be computed by

$$\mathbf{K}_{ij} = \frac{\partial^2 \mathcal{U}}{\partial \mathbf{p}_i \partial \mathbf{p}_j} \qquad\qquad \boldsymbol{\epsilon} = \mathbf{B}\mathbf{p}. \tag{31}$$

where the matrix $\mathbf{B}$ interpolates the nodal quantities to strains $\boldsymbol{\epsilon}$. Please note that in this paragraph $\mathbf{K}$ denotes the structural stiffness and not the covariance matrix. Thus, each beam element has 12 Degrees of Freedom (DoF), 6 DoF per node. The beam model is depicted in Figure 2.

Assuming a force vector $\mathbf{f}$, the static solution $\mathbf{p}$ is calculated from

$$\mathbf{K}(\mathbf{p})\mathbf{p} = \mathbf{f}. \tag{32}$$

In this framework, geometrical nonlinearities are introduced by using the co-rotational framework of (Battini and Pacoste 2002), decomposing large displacements/rotations into rigid body displacements and small elastic deformations, ultimately leading to a dependence of the stiffness matrix $\mathbf{K}$ on the displacements $\mathbf{p}$. After formulating the nonlinear structure, the aerodynamic forces and moments are computed via the unsteady vortex lattice method (UVLM) and mapped onto the structure, resulting in an overall nonlinear aeroelastic system. Due to this nonlinearity, no guarantee exists of finding an equilibrium point right away, motivating the need for an iterative solution to obtain the nonlinear static response. Starting with

$$\mathbf{f}_s(\mathbf{p}) = \mathbf{f}_{ext}(\mathbf{p}), \tag{33}$$

where the subscript s denotes the structural force, a predictor-corrector Newton-Raphson solver is used to solve the nonlinear system given by

$$\left(\frac{\partial \mathbf{f}}{\partial \mathbf{p}} - \frac{\partial \lambda \mathbf{f}_{ext}}{\partial \mathbf{p}} \right) \delta \mathbf{p} = \mathbf{f} - \mathbf{f}_{ext} = \mathbf{R}. \tag{34}$$

The solution can then be used to compute the corresponding stresses that are used to calculate the Tsai-Wu failure criterion $\mathbf{w}(\boldsymbol{\sigma})$, used to assess the strength of the structure. To reduce the number of constraints, only the 8 most critical Tsai-Wu strain factors per panel are considered (Werter 2017), leading to

$$\mathbf{c}_{tw} = \mathbf{w}_{crit}(\boldsymbol{\sigma}) \le 0. \tag{35}$$

The buckling analysis assumes that no global buckling can occur due to sufficient strength of stiffeners and ribs. By additionally computing the geometric stiffness matrix $\mathbf{K}_g$, the buckling factor λ_b can be found by solving the following eigenvalue problem

$$\left(\mathbf{K} + \lambda_b \mathbf{K}_g \right) \mathbf{a} = \mathbf{0}. \tag{36}$$

To further reduce the number of constraints, only the eight most critical buckling eigenvalues per panel are formulated as a constraint, leading to

$$\mathbf{c}_b = -\lambda_{b,crit} + 1 \le 0. \tag{37}$$

To compute the aeroelastic stability of the system, the equilibrium between the internal forces and moments $\mathbf{f}$ and all the external forces and moments must be regarded. The external forces are split up into applied aerodynamic loads $\mathbf{f}_a$ and remaining external forces due to e.g. gravity or thrust $\mathbf{f}_e$, given by

$$\mathbf{f}_s - \mathbf{f}_{ext} = \mathbf{f}_s - \mathbf{f}_a - \mathbf{f}_e = \mathbf{0}. \tag{38}$$

By linearising Equation 38, the corresponding stiffness matrices $\mathbf{K}_a$, $\mathbf{K}_e$ and $\mathbf{K}_s$ can be obtained, and the stability of this static aeroelastic equilibrium is governed by

$$(\lambda_s \mathbf{K}_a + \mathbf{K}_e - \mathbf{K}_s) \Delta \mathbf{p} = \mathbf{0}. \tag{39}$$

To ensure the static aeroelastic stability of the system, thus preventing divergence, the eigenvalues need to lie in $\lambda_s \ge 1$. The dynamic aeroelastic analysis is carried out by linearising the system around the static aerodynamic equilibrium and

by using the state space formulation for both the aerodynamic and the structural part. It should be mentioned that in the present discussion, many steps are left out for the sake of compactness. More details can be found in Werter (2017) and De Breuker (2011). As a result, the well-known continuous-time state-space equation is obtained, which can be written as

$$\dot{\mathbf{s}} = \mathbf{As} + \mathbf{B}\alpha_{air} \tag{40}$$

with α_{air} being the perturbation angle of attack of the induced free stream flow. For dynamic aeroelastic stability, used to prevent flutter, the eigenvalue problem on the state matrix $\mathbf{A}$ has to be solved once again, written as

$$\left(\mathbf{A} - \mathbf{I}\lambda_f\right)\boldsymbol{\varphi} = \mathbf{0}. \tag{41}$$

Anew, only the ten most critical eigenvalues $\lambda_{f,crit}$ are considered (Werter 2017), leading to

$$\mathbf{c}_{ds} = \mathfrak{R}(\lambda_{f,crit}) \leq 0. \tag{42}$$

Furthermore, two more types of constraints are formulated. The aileron effectiveness is constrained as follows

$$c_{ae} = \eta_{eff} - \eta_{eff,min} \leq 0, \tag{43}$$

to ensure safe manoeuvrability of the aircraft as well as the angle of attack α is constrained by using an upper and lower bound, written as

$$\begin{aligned} \mathbf{c}_{AoA,lb} &= -\alpha - \alpha_{lb} \leq 0, \\ \mathbf{c}_{AoA,ub} &= \alpha - \alpha_{ub} \leq 0 \end{aligned} \tag{44}$$

adding two more constraints per aerodynamic cross-section.

Finally, the constraints can be concatenated to form together with the objective function $f(\mathbf{x})$ the outputs of the model, as introduced in Section III, written as $\mathbf{c}(\mathbf{x}) = \{\mathbf{c}_{lpf}, \mathbf{c}_{tw}, \mathbf{c}_b, \mathbf{c}_{ds}, c_{ae}, \mathbf{c}_{AoA}\}^T$. As depicted in Table 1, all categories of constraints besides the lamination parameter feasibility need to be taken into account per load case; thus, with an increasing number of loading conditions, the number of constraints quickly increases to a magnitude of $10^3 - 10^5$.

This section aims to expose the origin of the constraints. Nevertheless, it is not always easy or even possible to obtain gradients of those, which is why gradient-free methods such as the Bayesian optimisation, herein proposed, can be very useful.

Table 1   Aeroelastic Tailoring constrained optimisation problem

Type	Parameter	Symbol	Equation	/Loadcase	
Objective	Minimise Wing Mass	f			
Design Variables (D)	Lamination Parameter	$\mathbf{x}_i^{lam}$	(26)		
	Laminate Thickness	$\mathbf{x}_i^t$			
Constraints (G)	Laminate Feasibility	$\mathbf{c}_{lpf}$	(27)	No	Analytic
	Static Strength	$\mathbf{c}_{tw}$	(35)	Yes	Analysis
	Buckling	$\mathbf{c}_b$	(37)	Yes	Analysis
	Aeroelastic Stability	$\mathbf{c}_{ds}$	(42)	Yes	Analysis
	Aileron Effectiveness	c_{ae}	(43)	Yes	Analysis
	Local Angle of Attack	$\mathbf{c}_{AoA}$	(44)	Yes	Analysis

V. Application

In this section the presented methodology is applied to two well-known benchmark cases before preliminary results for the aeroealastic tailoring optimisation problem are shown. For the sake of comparison, we follow the same approach as Eriksson and Poloczek (2021) and Hernández-Lobato, Michael A. Gelbart, et al. (2016). Any feasible solution is preferred over an infeasible one. That is why the maximum value of all found feasible solutions is taken as the default value for all infeasible solutions and noted as a dotted red line. Moreover, all computations are performed on an Apple M1 Pro chip while using the frameworks BoTorch (Balandat et al. 2020) and GPyTorch (Gardner, Pleiss, et al. 2018).

A. Academic Example: 10D Ackley Function with 2 Black-Box Constraints

The in Section III presented methodology is employed on the well-known Ackley function. This problem has a dimensionality of $D = 10$. Additionally, 2 black-box constraints are considered. The optimisation is performed within the domain $[-5, 10]^{10}$, and can be written as

$$f(\mathbf{x}) = -20 \exp\left(-0.2\sqrt{\frac{1}{d}\sum_d^{i=1} x_i^2}\right) - \exp\left(-0.2\frac{1}{d}\sum_d^{i=1} \cos(2\pi x_i)\right) \tag{45}$$

$$c_1(\mathbf{x}) = \sum_{i=1}^{10} x_i \leq 0 \tag{46}$$

$$c_2(\mathbf{x}) = ||\mathbf{x}||_2 - 5 \leq 0 \tag{47}$$

As mentioned earlier, the constrained TuRBO algorithm, SCBO, is employed. The same hyperparameters are used as presented in Eriksson and Poloczek (2021), a batch size $q = 4$ and $N = 10$ initial samples in $\mathcal{D}$. The two constraints are projected onto a lower dimensional subspace (G=2, g=1) using PCA/kPCA (in the following called as SCBO-PCA/SCBO-kPCA). Within SCBO-kPCA the exponential kernel is chosen. In addition, as proposed by Eriksson and Poloczek (2021), a bilog transformation is employed on the constraints to emphasise the region around zero, which is crucial for whether a design is feasible or not.
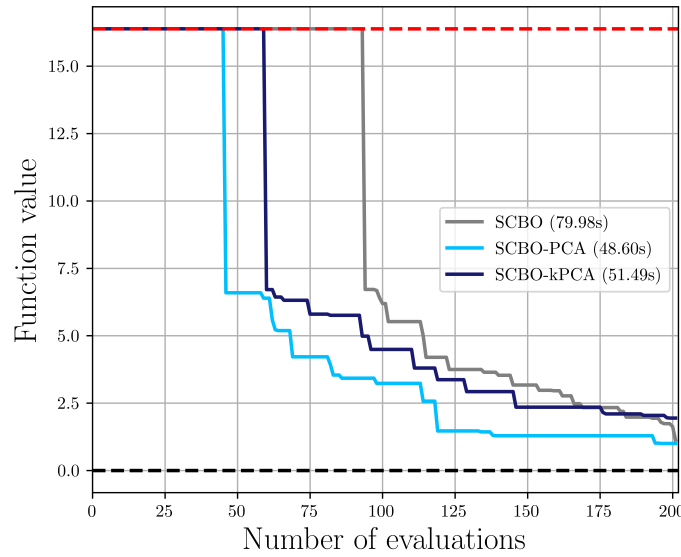

Fig. 4   A comparison of the optimisation of the $10D$ **Ackley function for SCBO and SCBO combined with PCA and kPCA. Only the objective values for feasible designs are plotted. All methods find a feasible optimum.**

As it can be seen in Figure 4, even though SCBO-PCA and SCBO-kPCA construct the GP solely on the lower dimensional subspace, the methods still perform as well as or even outperform the standard SCBO while saving approximately 40% computational time. The performance of SCBO compared to other state-of-the-art optimisation algorithms can be found in Eriksson and Poloczek (2021).

B. Academic Example: 7D Speed Reducer Problem with 11 Black-Box Constraints

Next, the methodology is applied to the $7D$ speed reducer problem from Lemonge et al. (2010), including 11 black-box constraints. The results can be found in Figure 5, whereas Figure 5a shows the decay of the feasible objective values. Figure 5b, by contrast, depicts the eigenvalues of the constraint matrix $\mathbf{C} \subset \mathcal{D}$. In this example, where $G = 11$,

$g = 4$ principal components are chosen. Again, these are the same hyperparameters as presented in Subsection V.A, meaning a batch size $q = 4$ and $N = 10$ initial samples.
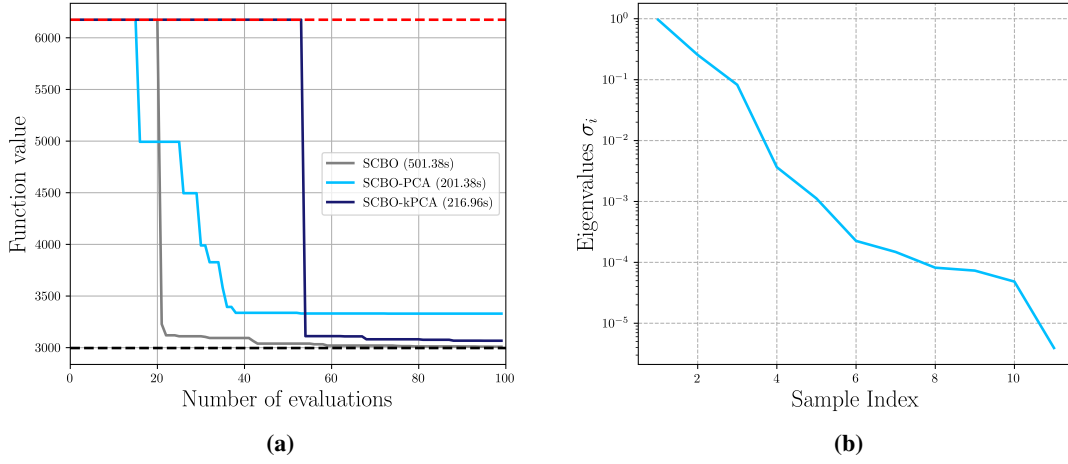

(a) **(b)**

Fig. 5 $7D$ **Speed reducer problem with** 11 **black-box constraints from Lemonge et al. (2010). In (a), SCBO, SCBO-PCA and SCBO-kPCA are compared. In (b), the eigenvalues of the matrix C are plotted**
.

All methods find a feasible design, whereas SCBO-kPCA performs better than SCBO-PCA. Nevertheless, both methods are significantly faster than the standard SCBO. The better performance of SCBO-kPCA might stem from the ability to capture a nonlinear lower subspace and hence offer a better approximation.

However, the lower dimensional subspace is constructed based on the constraint values in $\mathcal{D}$. Assuming that the global optimum lies on the boundary of the feasible space $\mathcal{X}_f$, the success of the method highly depends on how accurately the lower dimensional subspace captures the original space. Hypothesising that the data in $\mathcal{D}$ with $N = 10$ was not sufficient, in Figure 6 the number of initial samples is doubled. It can be seen that by increasing the DoE, even better objectives can be found, presuming that due to the additional data, a more accurate subspace and, thus, an optimum closer to the optimum of SCBO can be found. This leads to the conclusion that a sensitivity analysis of the number N might be very important for the method's success. Furthermore, SCBO-PCA and SCBO-kPCA needed more evaluations to find the first feasible design.

C. Aeroelastic Tailoring: A Multi-Disciplinary Design Optimisation Problem

This work aims to adapt the proposed BO method for the use in aeroelastic tailoring, posing a high-dimensional problem with large-scale constraints, as explained in Section IV. Assuming that $10^3 < G < 10^5$ high-dimensional GPs are not feasible from a computational standpoint, the presented methodology shall speed up the process by modelling the constraints on a lower-dimensional subspace. The number of design regions has been decreased, ending up at $D = 108$, as well as limiting the number of loadcases to one or two, respectively.

By using the approach presented in Section III, this work aims to numerically reduce the number of constraints and construct the surrogate models via GPs directly on the lower-dimensional subspace, as demonstrated in Subsections V.A and V.B. In Table 1, the sources of constraints are shown. The multitude of outputs arises from the inclusion of multiple loadcases. The premise of this approach lies in the consistency of the physics governing the constraints across loadcases, where eventually only the load changes. This stresses the potential for compressing this information due to the unchanged underlying physics for varying loadcases. The aforementioned aeroelastic tailoring model is used to compute the DoE $\mathcal{D}$ with $N = 416$ samples. Sampling was performed via Latin Hypercube Sampling (LHS). Subsequently, PCA is applied on the matrix $\mathbf{C}$ to investigate its eigenvalues. Figure 7a shows the decay of these computed eigenvalues. If the same error is used as in Subsection V.B, eigenvalues up to approx $\sigma_i \approx 10^{-2}$, thus $g \approx 29$ principal components might be enough to construct a lower dimensional subspace of sufficient accuracy. In addition, the projection error can be computed. Therefore, some unseen data $\mathbf{C}_*$, meaning data that has not been used to compute the principal components, is mapped onto the lower dimensional subspace $\tilde{\mathbf{C}}_* = \mathbf{\Psi}_g^T \mathbf{C}_*$. Since PCA is a linear mapping, the inverse
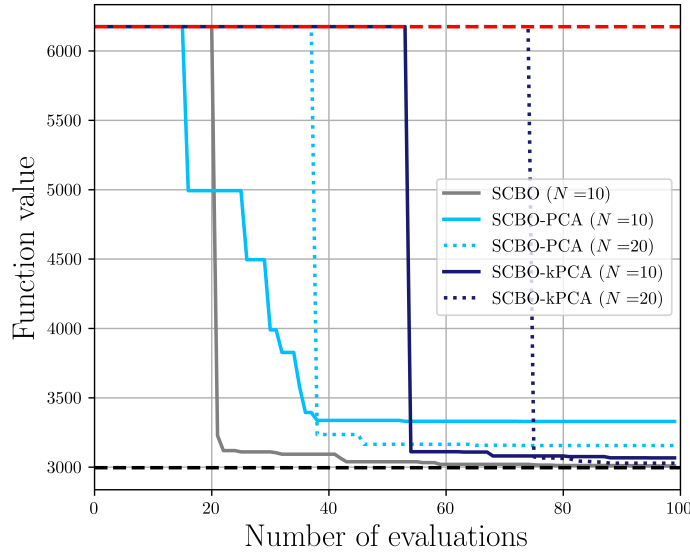
12

Fig. 6 Investigation of number of samples N in C for the $7D$ speed reducer problem with 11 black-box constraints.

mapping can be simply computed by $\hat{\mathbf{C}}_* = \tilde{\mathbf{C}}_* \boldsymbol{\Psi}$. The approximation error can then be computed by

$$\epsilon = \frac{\parallel \mathbf{C}_* - \hat{\mathbf{C}}_* \parallel_F^2}{\parallel \mathbf{C}_* \parallel_F^2}. \tag{48}$$

In Figure 7b, the trend reveals that including more components leads to a reduced error, even for unseen data. Furthermore, to investigate how the construction of the lower-dimensional subspace behaves with sample size variation, the error ϵ is shown for $N = 40$, $N = 416$ and $2N$ samples. It can be seen that the error is approximately the same for the latter two cases. As anticipated, an insufficient initial sample size N results in limited information availability during the subspace construction, consequently leading to a larger error. Moreover, the conclusion drawn is that even with $N = 416$ samples, sufficient data is available to attain a reasonable subspace. Further, increasing the number of samples in the DoE does not contribute to higher accuracy.

As previously noted, the high number of constraints stems from the incorporation of multiple loadcases. Consequently, it becomes intriguing to explore how the eigenvalues vary when the number of loadcases is altered. Recall that the eigenvalues denote the importance of their corresponding eigenvector, which serves as a measure of where to truncate the projection matrix. In Figure 8, it can be observed that, even though the number of constraints in the original space has doubled, from $G = 893$ to $G = 1786$, if the eigenvalues $\sigma_i > 10^{-2}$ are used, no more principal components have to be taken into account. For $\sigma_i > 10^{-3}$, however, only 27 more components are needed to maintain the same error. Beyond that, the threshold of the eigenvalues is commonly set based on experience, thus can be seen as a hyper-parameter of the method.

These promising preliminary results motivate the use of the introduced SCBO algorithm (Eriksson and Poloczek 2021) in combination with a reduced-basis approach to lower the number of constraints in a Bayesian Optimisation to allow a global search of the design space in this high-dimensional problem with large-scale constraints.

VI. Conclusion and Future Work

Aeroelastic tailoring can be seen as a high-dimensional multi-disciplinary design optimisation problem with large-scale constraints. Since the global design space search in this use case is not a trivial task, this work uses BO to do so. However, the application of constrained BO is not straightforward due to the poor scalability in terms of the number of constraints. This work introduces a novel approach where a large number of constraints is mapped onto a lower dimensional subspace where the surrogate models are constructed.
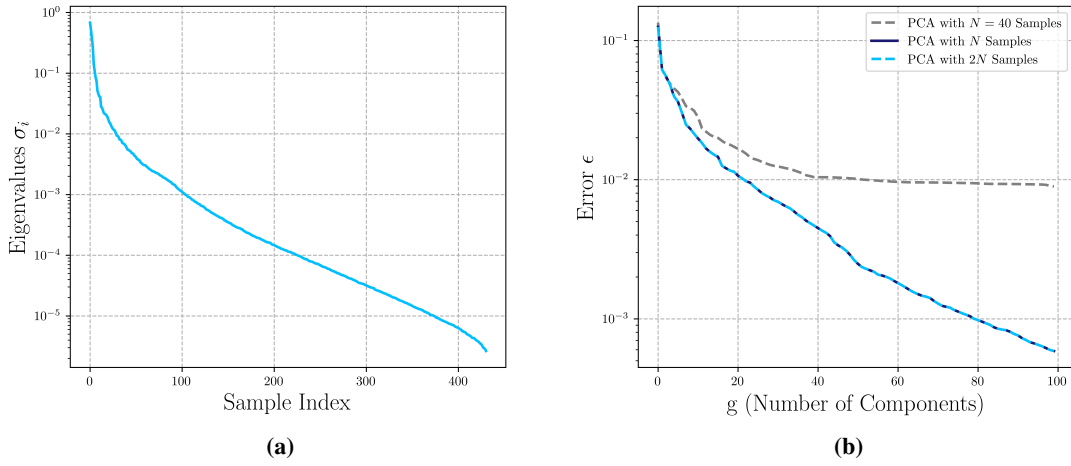
(a)   **(b)**

Fig. 7   Investigating the constraints in $\mathcal{D}$. (a) shows the decay of the eigenvalues for $N = 416$. (b) computes the error depending on how many principal components are taken into account (Equation 48)
.

The herein-presented numerical findings clearly indicate the applicability of this approach. As it can be seen in Section V, SCBO with kPCA performs similarly to SCBO while being computationally more efficient. It should be noted that this computational saving can become even more significant when working in a high-dimensional setting where the training of each GP becomes crucial. Thus, by drastically reducing the number of needed GPs, major computational savings can be obtained.

Furthermore, this work entails preliminary investigations for the use of this methodology in aeroelastic tailoring, likewise showing promising results. Until now, PCA has been used solely to perform the presented investigations. However, as shown, kPCA can be seen as a nonlinear extension of PCA, which is why even better performance within the optimisation of the aeroelastic tailoring problem is expected due to the nonlinear nature of the constraints. Thus, follow-up studies will investigate these aspects and will aim to incorporate thousands of constraints into the optimisation process.

Even though this work has been performed within the realm of aeroelastic tailoring, it is important to stress the generality of the herein-proposed method. As indicated by the numerical investigations, this approach can easily be applied to all sorts of problems where large-scale constraints are involved.

To critically reflect the methodology adopted in this work, the following statements can be made. Some authors proposed so-called Multi-Output GP (MTGP) (Maddox et al. 2021) models, which essentially model all the outputs in parallel while additionally taking into account their correlations. The computational burden is excessive, especially for high-dimensional problems. This is why the presented methodology has been chosen over MTGP. Another point which might be addressed in the future is the use of Bayesian Neural Networks (BNN) instead of GPs for surrogate modelling. As presented in Section II, the complexity of one GP depends on the number of samples N. Especially in high-dimensional problems, the number of samples might be very high, thus increasing the computational cost. As the authors in Snoek et al. (2015) show, BNN does not scale with the number of samples N but with the dimension D, thus staying constant over the whole optimisation process. This may lead to an improved efficiency. Additionally, there is a notable computational expense during hyperparameter tuning of the surrogate in the high-dimensional case. To mitigate this challenge, employing methods such as REMBO (Wang et al. 2016) and ALEBO (Letham et al. 2020) or (k)PCA-BO (Raponi et al. 2020; Antonov et al. 2022) presents an avenue for further reducing the computational cost. These methods operate under the assumption that certain dimensions are more significant than others, consequently reducing the number of tunable hyperparameters.
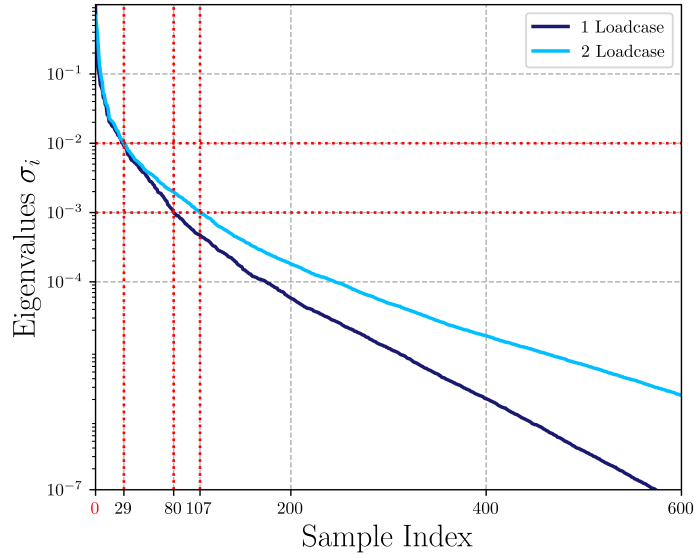
14

Fig. 8   Performing PCA on the matrix C for one and two loadcases.

References

Anjos, M. (2008). *The MOPTA 2008 Benchmark*. URL: `http://www.miguelanjos.com/jones-benchmark`.

Antonov, K., E. Raponi, H. Wang, and C. Doerr (2022). "High Dimensional Bayesian Optimization with Kernel Principal Component Analysis". en. In: DOI: `arXiv:2204.13753v2`. (Visited on 11/15/2023).

Balandat, Maximilian, Brian Karrer, Daniel R. Jiang, Samuel Daulton, Benjamin Letham, Andrew Gordon Wilson, and Eytan Bakshy (2020). "BoTorch: A Framework for Efficient Monte-Carlo Bayesian Optimization". In: *Advances in Neural Information Processing Systems 33*. URL: `http://arxiv.org/abs/1910.06403`.

Battini, Jean-Marc and Costin Pacoste (Feb. 2002). "Co-rotational beam elements with warping effects in instability problems". en. In: *Computer Methods in Applied Mechanics and Engineering* 191.17-18, pp. 1755–1789. ISSN: 00457825. DOI: `10.1016/S0045-7825(01)00352-8`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S0045782501003528` (visited on 05/15/2023).

Binois, Mickaël and Nathan Wycoff (June 2022). "A Survey on High-dimensional Gaussian Process Modeling with Application to Bayesian Optimization". en. In: *ACM Transactions on Evolutionary Learning and Optimization* 2.2, pp. 1–26. ISSN: 2688-299X, 2688-3007. DOI: `10.1145/3545611`. URL: `https://dl.acm.org/doi/10.1145/3545611` (visited on 10/27/2023).

De Breuker, Roeland (2011). "Energy-based aeroelastic analysis and optimisation of morphing wings". eng. OCLC: 840445505. PhD thesis. S.l.: s.n.

Dillinger, J. K. S., T. Klimmek, M. M. Abdalla, and Z. Gürdal (July 2013). "Stiffness Optimization of Composite Wings with Aeroelastic Constraints". en. In: *Journal of Aircraft* 50.4, pp. 1159–1168. ISSN: 0021-8669, 1533-3868. DOI: `10.2514/1.C032084`. URL: `https://arc.aiaa.org/doi/10.2514/1.C032084` (visited on 05/15/2023).

Eriksson, David and Martin Jankowiak (June 2021). *High-Dimensional Bayesian Optimization with Sparse Axis-Aligned Subspaces*. en. arXiv:2103.00349 [cs, stat]. URL: `http://arxiv.org/abs/2103.00349` (visited on 05/12/2023).

Eriksson, David, Michael Pearce, Jacob Gardner, Ryan D Turner, and Matthias Poloczek (2019). "Scalable Global Optimization via Local Bayesian Optimization". en. In.

Eriksson, David and Matthias Poloczek (Feb. 2021). *Scalable Constrained Bayesian Optimization*. en. arXiv:2002.08526 [cs, stat]. URL: `http://arxiv.org/abs/2002.08526` (visited on 05/12/2023).

Ferede, Etana and Mostafa Abdalla (Jan. 2014). "Cross-sectional modelling of thin-walled composite beams". en. In: *55th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*. National Harbor, Maryland: American Institute of Aeronautics and Astronautics. ISBN: 978-1-62410-314-8. DOI: `10.2514/6.2014-0163`. URL: `https://arc.aiaa.org/doi/10.2514/6.2014-0163` (visited on 05/12/2023).

Frazier, Peter I. (July 2018). *A Tutorial on Bayesian Optimization*. en. arXiv:1807.02811 [cs, math, stat]. URL: `http://arxiv.org/abs/1807.02811` (visited on 05/12/2023).

Gardner, Jacob R, Matt J Kusner, and Gardner Jake (2014). "Bayesian Optimization with Inequality Constraints". en. In.

Gardner, Jacob R, Geoff Pleiss, David Bindel, Kilian Q Weinberger, and Andrew Gordon Wilson (2018). "GPyTorch: Blackbox Matrix-Matrix Gaussian Process Inference with GPU Acceleration". In: *Advances in Neural Information Processing Systems*.

Gelbart, Michael A, Jasper Snoek, and Ryan P Adams (2014). "Bayesian Optimization with Unknown Constraints". en. In.

Hernández-Lobato, José Miguel, Michael A. Gelbart, Ryan P. Adams, Matthew W. Hoffman, and Zoubin Ghahramani (Sept. 2016). *A General Framework for Constrained Bayesian Optimization using Information-based Search*. en. arXiv:1511.09422 [stat]. URL: `http://arxiv.org/abs/1511.09422` (visited on 05/12/2023).

Hernández-Lobato, José Miguel, James Requeima, Edward O. Pyzer-Knapp, and Alán Aspuru-Guzik (2017). "Parallel and Distributed Thompson Sampling for Large-scale Accelerated Exploration of Chemical Space". In: Publisher: arXiv Version Number: 1. DOI: `10.48550/ARXIV.1706.01825`. URL: `https://arxiv.org/abs/1706.01825` (visited on 10/27/2023).

Higdon, Dave, James Gattiker, Brian Williams, and Maria Rightley (June 2008). "Computer Model Calibration Using High-Dimensional Output". en. In: *Journal of the American Statistical Association* 103.482, pp. 570–583. ISSN: 0162-1459, 1537-274X. DOI: `10.1198/016214507000000888`. URL: `https://www.tandfonline.com/doi/full/10.1198/016214507000000888` (visited on 05/12/2023).

Jolliffe, Ian T. (2002). *Principal component analysis*. eng. 2nd ed. Springer series in statistics. New York: Springer. ISBN: 978-0-387-95442-4.

Jolliffe, Ian T. and Jorge Cadima (Apr. 2016). "Principal component analysis: a review and recent developments". en. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2065, p. 20150202. ISSN: 1364-503X, 1471-2962. DOI: `10.1098/rsta.2015.0202`. URL: `https://royalsocietypublishing.org/doi/10.1098/rsta.2015.0202` (visited on 05/12/2023).

Kandasamy, Kirthevasan, Jeff Schneider, and Barnabas Poczos (May 2016). *High Dimensional Bayesian Optimisation and Bandits via Additive Models*. arXiv:1503.01673 [cs, stat]. URL: `http://arxiv.org/abs/1503.01673` (visited on 05/15/2023).

Lemonge, A.C.C., H.J.C. Barbosa, C.C.H. Borges, and F.B.S. Silve (2010). "Constrained Optimization Problems in Mechanical Engineering Design Using a Real-Coded Steady-State Genetic Algorithm". en. In: *Mecánica Computacional Vol XXIX*, pp. 9287–9303.

Letham, Benjamin, Roberto Calandra, Akshara Rai, and Eytan Bakshy (Oct. 2020). *Re-Examining Linear Embeddings for High-Dimensional Bayesian Optimization*. en. arXiv:2001.11659 [cs, stat]. URL: `http://arxiv.org/abs/2001.11659` (visited on 05/12/2023).

Maddox, Wesley J., Maximilian Balandat, Andrew Gordon Wilson, and Eytan Bakshy (Oct. 2021). *Bayesian Optimization with High-Dimensional Outputs*. en. arXiv:2106.12997 [cs, stat]. URL: `http://arxiv.org/abs/2106.12997` (visited on 05/12/2023).

Mockus, J., Tiesis, V., and Zilinskas, A. (1978). "The Application of Bayesian Methods for Seeking the Extremum". In: pp. 117–129.

Priem, Remy (2020). "Optimisation bayésienne sous contraintes et en grande dimension appliquée à la conception avion avant projet". fr. PhD thesis.

Rajpal, D. (2021). "Dynamic aeroelastic optimization of composite wings including fatigue considerations". PhD thesis. Delft University of Technology. DOI: `10.4233/UUID:FC33C568-B2F2-48C0-96CC-48221C69C2BB`. URL: `http://resolver.tudelft.nl/uuid:fc33c568-b2f2-48c0-96cc-48221c69c2bb` (visited on 05/23/2023).

Raju, Gangadharan, Zhangming Wu, and Paul Weaver (Jan. 2014). "On Further Developments of Feasible Region of Lamination Parameters for Symmetric Composite Laminates". en. In: *55th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*. National Harbor, Maryland: American Institute of Aeronautics and Astronautics. ISBN: 978-1-62410-314-8. DOI: `10.2514/6.2014-1374`. URL: `https://arc.aiaa.org/doi/10.2514/6.2014-1374` (visited on 11/01/2023).

Raponi, E., H. Wang, M. Bujny, S. Boria, and C. Doerr (2020). "High Dimensional Bayesian Optimization Assisted by Principal Component Analysis". en. In: DOI: `arXiv:2007.00925v1`. (Visited on 11/15/2023).

Rasmussen, Carl Edward and Christopher K. I. Williams (2006). *Gaussian processes for machine learning*. en. Adaptive computation and machine learning. OCLC: ocm61285753. Cambridge, Mass: MIT Press. ISBN: 978-0-262-18253-9.

Santoni, M.L., E. Raponi, R. De Leone, and C. Doerr (2023). "Comparison of High-Dimensional Bayesian Optimization Algorithms on BBOB". en. In: DOI: `arXiv:2303.00890v2`. (Visited on 11/15/2023).

Saves, Paul, Nathalie Bartoli, Youssef Diouane, Thierry Lefebvre, Joseph Morlier, Christophe David, Eric Nguyen Van, and Sébastien Defoort (Jan. 2022). "Multidisciplinary design optimization with mixed categorical variables for aircraft design". en. In: *AIAA SCITECH 2022 Forum*. San Diego, CA & Virtual: American Institute of Aeronautics and Astronautics. ISBN: 978-1-62410-631-6. DOI: `10.2514/6.2022-0082`. URL: `https://arc.aiaa.org/doi/10.2514/6.2022-0082` (visited on 05/15/2023).

Schoenberg, I. J. (1938). "Metric spaces and positive definite functions". en. In: *Transactions of the American Mathematical Society* 44.3, pp. 522–536. ISSN: 0002-9947, 1088-6850. DOI: `10.1090/S0002-9947-1938-1501980-0`. URL: `https://www.ams.org/tran/1938-044-03/S0002-9947-1938-1501980-0/` (visited on 11/15/2023).

Schölkopf, Bernhard, Alexander Smola, and Klaus-Robert Müller (July 1998). "Nonlinear Component Analysis as a Kernel Eigenvalue Problem". en. In: *Neural Computation* 10.5, pp. 1299–1319. ISSN: 0899-7667, 1530-888X. DOI: `10.1162/089976698300017467`. URL: `https://direct.mit.edu/neco/article/10/5/1299-1319/6193` (visited on 05/12/2023).

Shirk, Michael H., Terrence J. Hertz, and Terrence A. Weisshaar (Jan. 1986). "Aeroelastic tailoring - Theory, practice, and promise". en. In: *Journal of Aircraft* 23.1, pp. 6–18. ISSN: 0021-8669, 1533-3868. DOI: `10.2514/3.45260`. URL: `https://arc.aiaa.org/doi/10.2514/3.45260` (visited on 11/02/2023).

Snoek, Jasper, Oren Rippel, Kevin Swersky, Ryan Kiros, Nadathur Satish, Narayanan Sundaram, Md Mostofa Ali Patwary, Prabhat, and Ryan P. Adams (July 2015). *Scalable Bayesian Optimization Using Deep Neural Networks*. en. arXiv:1502.05700 [stat]. URL: `http://arxiv.org/abs/1502.05700` (visited on 05/12/2023).

Thompson, William R. (Dec. 1933). "On the Likelihood that One Unknown Probability Exceeds Another in View of the Evidence of Two Samples". In: *Biometrika* 25.3/4, p. 285. ISSN: 00063444. DOI: `10.2307/2332286`. URL: `https://www.jstor.org/stable/2332286?origin=crossref` (visited on 05/13/2023).

Wang, Ziyu, Frank Hutter, Masrour Zoghi, David Matheson, and Nando de Freitas (Jan. 2016). *Bayesian Optimization in a Billion Dimensions via Random Embeddings*. en. arXiv:1301.1942 [cs, stat]. URL: `http://arxiv.org/abs/1301.1942` (visited on 05/12/2023).

Werter, N.P.M. (2017). "Aeroelastic Modelling and Design of Aeroelastically Tailored and Morphing Wings". en. PhD thesis. Delft University of Technology. DOI: `10.4233/UUID:74925F40-1EFC-469F-88EE-E871C720047E`. URL: `http://resolver.tudelft.nl/uuid:74925f40-1efc-469f-88ee-e871c720047e` (visited on 05/12/2023).

Xing, W.W., V. Triantafyllidis, A.A. Shah, P.B. Nair, and N. Zabaras (Dec. 2016). "Manifold learning for the emulation of spatial fields from computational models". en. In: *Journal of Computational Physics* 326, pp. 666–690. ISSN: 00219991. DOI: `10.1016/j.jcp.2016.07.040`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S0021999116303722` (visited on 10/27/2023).

Xing, Wei, Akeel A. Shah, and Prasanth B. Nair (Feb. 2015). "Reduced dimensional Gaussian process emulators of parametrized partial differential equations based on Isomap". en. In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 471.2174, p. 20140697. ISSN: 1364-5021, 1471-2946. DOI: `10.1098/rspa.2014.0697`. URL: `https://royalsocietypublishing.org/doi/10.1098/rspa.2014.0697` (visited on 10/27/2023).

Ziomek, Juliusz and Haitham Bou-Ammar (Jan. 2023). *Are Random Decompositions all we need in High Dimensional Bayesian Optimisation?* arXiv:2301.12844 [cs, stat]. URL: `http://arxiv.org/abs/2301.12844` (visited on 05/15/2023).