On the Impact of Covariance Functions in Multi-Objective Bayesian Optimization for Engineering Design

Pramudita Satria Palar* and Lavi Rizki Zuhal†
Institut Teknologi Bandung, Jl. Ganesha No. 10, West Java, Indonesia

Tinkle Chugh‡
University of Exeter, Exeter, United Kingdom
Palacky University in Olomouc, Czech Republic

Alma Rahat§
Swansea University, Swansea, SA1 8EN United Kingdom.

Multi-objective Bayesian optimization (BO) is a highly useful class of methods that can effectively solve computationally expensive engineering design optimization problems with multiple objectives. However, the impact of covariance function, which is an important part of multi-objective BO, is rarely studied in the context of engineering optimization. We aim to shed light on this issue by performing numerical experiments on engineering design optimization problems, primarily low-fidelity problems so that we are able to statistically evaluate the performance of BO methods with various covariance functions. In this paper, we performed the study using a set of subsonic airfoil optimization cases as benchmark problems. Expected hypervolume improvement was used as the acquisition function to enrich the experimental design. Results show that the choice of the covariance function give a notable impact on the performance of multi-objective BO. In this regard, Kriging models with Matérn-3/2 is the most robust method in terms of the diversity and convergence to the Pareto front that can handle problems with various complexities.

I. Introduction

Kriging is one of the most widely used surrogate models that is frequently employed to aid the process of engineering design optimization. The main advantage of Kriging is that it provides both the prediction and the mean-squared error that can be used for adaptive sampling strategies. This is very useful because it allows efficient optimization and refinement of the surrogate model as compared to non-probabilistic surrogates that provide only the prediction structure (e.g. polynomial regression and support vector regression [1]). Due to this advantage, Kriging is a popular choice for Bayesian optimization (BO) [2, 3] which has found a renewed interest in aerospace engineering [4–6]. The use of Kriging in aerospace engineering itself can be traced to some earlier works by Simpson et al. [7, 8]. BO is an attractive surrogate-based optimization technique to deal with computationally expensive optimization problems. The main components of BO are the surrogate model and the acquisition function. An acquisition function determines the sample to be evaluated next. Among popular choices of acquisition functions for single-objective BO are expected improvement (EI) [9], knowledge gradient [10], and predictive entropy search [11]. Besides, it is also possible to create a portfolio of acquisition functions [12]. Of interest here is the multi-objective BO that is highly useful to discover the trade-off between competing objectives in engineering design optimization. In contrast to single-objective BO, the goal of multi-objective BO is to find a set of non-dominated solutions. There are several variants of acquisition functions for multi-objective BO which include expected hypervolume improvement (EHVI) [13], EI based on penalty boundary intersection [14], Euclidean expected improvement [15], and predictive entropy search for multi-objective optimization [16]. It is also possible to use the EI of the scalarized form of multiple objectives as in ParEGO [17].

---

*Lecturer, Faculty of Mechanical and Aerospace Engineering, Institut Teknologi Bandung.

†Associate Professor, Faculty of Mechanical and Aerospace Engineering, Institut Teknologi Bandung.

‡Postdoctoral Research Fellow, Department of Computer of Science, University of Exeter. Also affiliated with Faculty of Science, Palacky University in Olomouc, Czech Republic

§Lecturer in Big Data/Data Science, Swansea University.

Multi-objective BO is particularly useful if designers wish to perform design exploration since it reveals important insight that is useful for actual design processes. Some studies compare these different multi-objective BO variants in various artificial or engineering design problems [6, 18].

One aspect that is rarely addressed is the impact of covariance functions on the performance of the Kriging model, either for pure prediction or as a surrogate for Bayesian optimization. We need to emphasize that it is necessary to perform such an investigation with non-synthetic data sets so that the investigation is more meaningful. The choice of covariance function might affect the performance of the Kriging model in prediction and optimization; thus, it is important to properly select the covariance function. Arguably, Gaussian is the most widely used covariance function in engineering design optimization. However, Stein suggested that Matérn covariance functions are more capable of handling various complexities in real-world data sets [19]. We strongly believe that it is necessary to carry on empirical experiments to gain insight into the performance of these covariance functions in non-synthetic data sets/problems. This is because real-world problems have some characteristics that are difficult to be fully replicated by artificial problems (e.g., small noises due to truncation and meshing errors).

Previous research on single-objective BO shows that Matérn covariance functions are more capable than Gaussian for handling real-world single-objective optimization problems [20]. However, it is still necessary to perform similar investigations in the context of multi-objective BO. This is because the goal in multi-objective optimization is different, that is, the optimizer should be able to find a diverse set of solutions that are also close to the Pareto front. Decision-makers can then pick one or several solutions from the Pareto front or use the knowledge for improving the design or decision-making processes. However, from the viewpoint of pure multi-objective optimization, the main concern is how to find good representatives of the Pareto front (although it is possible for decision-makers and the optimization process to interact in the form of preference or other means). In this paper, we aim to gain important insight into the choice of covariance function in multi-objective optimization for handling real-world problems. It is hoped that the insight gained from this study will be highly useful for researchers and practitioners, especially in the field of engineering design optimization. In this paper, we investigate the impact of covariance functions by using a multi-objective optimization of a fixed lift subsonic airfoil as a benchmark. It is worth noting that we can set different covariance functions for different objective functions. However, we typically do not have prior knowledge regarding which covariance function is the most suitable. Therefore, we use the same covariance function for the different objective functions. As for the acquisition function, we deploy EHVI due to its good performance on various engineering problems [6].

This paper is structured as follows: Section 2 explains the methodology used in this paper, which include Kriging surrogate model, covariance function, and multi-objective BO; Section 3 details the computational experiments and the results; the conclusions and the future works are given in Section 4.

II. Methodology

A. Kriging surrogate model

The task of surrogate modeling is to approximate a black-box function $f(\boldsymbol{x})$, where $\boldsymbol{x} = \{x_1, x_2, \ldots, x_m\}^T$ and m is the dimensionality, with an analytical model $\hat{f}(\boldsymbol{x})$. In this paper, the surrogate model that we use is the ordinary Kriging (OK) which uses a zeroth order polynomial for the trend function, that is

$$\hat{y}(\boldsymbol{x}) = \beta + Z(\boldsymbol{x}), \tag{1}$$

where the first term on the right-hand side of the equation is the Kriging trend (or mean) function and the second term is the stochastic process of Kriging (i.e., a Gaussian process). A Kriging model is created by using a finite set of experimental design $X = \{\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots, \boldsymbol{x}^{(n)}\}^T$, where n is the sample size. For multi-objective optimization with K objectives, we can build K surrogate models, in which one surrogate approximates one objective. Notice that methods such as ParEGO use only one surrogate model that is constructed from scalarized objective functions [17].

Covariance function is a fundamental part of the Kriging model. This is because it encapsulates how much the predictive function response is permitted to change given a change in the input space. Thus, we can control the nature of the predictive relationship between the input and the function space by changing the precise convariance function used in Kriging. It is, therefore, likely that for different problems as the nature of the true function landscape change, a covariance function that best approximate the true function will yield best overall performance for multi-objective BO. The main aim in this paper is to identify whether there is a covariance function that performs well in general.

Typical covariance functions used in engineering design take two samples, say $x^{(i)}$ and $x^{(j)}$, and the lengthscale θ as inputs. To be more specific, we typically use $h = |x^{(i)} - x^{(j)}|$, i.e., the Euclidean distance between $x^{(i)}$ and $x^{(j)}$ as the

input. The first covariance function that we use is Gaussian, reads as

$$k(h, \theta) = \sigma^2 \exp\left(-\frac{|h|^2}{2\theta^2}\right). \tag{2}$$

Notice that we use the Kriging/signal variance equals to one (i.e., $\sigma^2 = 1$) in the formulation of the covariance function. The main criticism of the Gaussian covariance function is that its smoothness assumption might be unrealistic for approximating real-world processes [19]. To that end, other popular alternatives than Gaussian is the Matérn covariance function reads as

$$k(h, \theta, \nu) = \frac{1}{2^{\nu-1}\Gamma(\nu)}\left(2\sqrt{\nu}\frac{|h|}{\theta}\right)\mathcal{K}_\nu\left(2\sqrt{\nu}\frac{|h|}{\theta}\right). \tag{3}$$

where $\nu \geq 1/2$ is the shape parameter, Γ is the Gamma function, and $\mathcal{K}_\nu$ is the modified Bessel function of the second kind. Note that the choice of ν defines the form of the Matérn covariance function. In various applications, the most popular choice of ν is the half integer (i.e., $\nu = p + 1/2$, where p is a non-negative integer). To be more precise, Matérn covariance functions with $\nu = 3/2$ and $\nu = 5/2$ are the most widely used; they are, respectively, described as

$$k(h, \theta, \nu = 3/2) = \left(1 + \frac{\sqrt{3}|h|}{\theta}\right)\exp\left(-\frac{\sqrt{3}|h|}{\theta}\right), \tag{4}$$

and

$$k(h, \theta, \nu = 5/2) = \left(1 + \frac{\sqrt{5}|h|}{\theta} + \frac{5h^2}{3\theta^2}\right)\exp\left(-\frac{\sqrt{5}|h|}{\theta}\right). \tag{5}$$

The three covariance functions and also their effects on the Kriging model in approximating the one-dimensional Forrester function are depicted in Fig. 1. As can be seen from Fig. 1a, the similarity between $x^{(i)}$ and $x^{(j)}$ decreases as the distance increases; it is also clear here that the three covariance functions differ in how they model the covariance between different points in the input space. This, in turn, affects the Kriging prediction and the mean-squared error as can be seen in Figs. 1b and 1c, respectively. For this example, Gaussian is the best accurate covariance function in terms of accuracy followed by Matérn-5/2 and Matérn-3/2 (see Fig. 1d).

For higher-dimensional approximations, a multi-dimensional covariance function can be simply constructed by the product of one-dimensional covariance functions. In this case, we have a vector of lengthscales in which its size depends on the dimensionality m, i.e., $\boldsymbol{\theta} = \{\theta_1, \theta_2, \ldots, \theta_m\}^T$. Covariance function enters the Kriging formulation via the covariance matrix $\boldsymbol{R}$ and covariance vector $\boldsymbol{r}$. Here, $\boldsymbol{R}$ contains the covariance information between every sample in the experimental design, while $\boldsymbol{r}$ indicates the covariance between an arbitrary point and the experimental design. The size of $\boldsymbol{R}$ is $n \times n$ with the (i, j) input is $R_{i,j} = k(\boldsymbol{\theta}; \boldsymbol{x}^i, \boldsymbol{x}^j)$. The i input for $\boldsymbol{r}$ is $k(\boldsymbol{\theta}; \boldsymbol{x}, \boldsymbol{x}^i)$, where $\boldsymbol{x}$ is an arbitrary sample point in this context.

The prediction and the mean-squared error at any arbitrary point are, respectively, computed by

$$\hat{y}(\boldsymbol{x}) = \beta + \boldsymbol{r}(\boldsymbol{x})^T \boldsymbol{R}^{-1}(y - \mathbf{1}\beta) \tag{6}$$

and

$$\hat{s}^2(\boldsymbol{x}) = \sigma^2(1 - (\boldsymbol{r}(\boldsymbol{x})^T \boldsymbol{R}^{-1}\boldsymbol{r}(\boldsymbol{x})) + (\mathbf{1}^T \boldsymbol{R}^{-1}\boldsymbol{r}(\boldsymbol{x}) - 1(\boldsymbol{x}))^T(\mathbf{1}^T \boldsymbol{R}^{-1}\mathbf{1})^{-1}(\mathbf{1}^T \boldsymbol{R}^{-1}\boldsymbol{r}(\boldsymbol{x}) - 1(\boldsymbol{x})), \tag{7}$$

where $\mathbf{1}$ is the vector of ones, σ^2 is the Kriging variance, and β is computed by a generalized least-squares procedure, reads as.

$$\hat{\beta} = (\mathbf{1}^T \boldsymbol{R}^{-1}\mathbf{1})\mathbf{1}^T \boldsymbol{R}^{-1}\boldsymbol{y}. \tag{8}$$

In this paper, the Kriging hyperparameters are found by optimizing the log-likelihood function, reads as

$$\ln(\mathcal{L}) = -\frac{n}{2}\ln 2\pi - \frac{1}{2}\ln|\boldsymbol{R}| - \frac{1}{2}(\boldsymbol{y} - \mathbf{1}\beta)^T \boldsymbol{R}^{-1}(\boldsymbol{y} - \mathbf{1}\beta). \tag{9}$$

More details of Kriging construction can be found elsewhere, such as in Jones et al. [2]

Both $\hat{y}(\boldsymbol{x})$ and $\hat{s}^2(\boldsymbol{x})$ are important pieces of information for BO, either single- or multi-objective. In this regard, exploitation is enabled by selecting solutions with better mean predictions, while selecting solutions with high uncertainty is equivalent to exploration. Mostly in BO, we aim to strike a balance between exploration and exploitation to locate a promising solution to evaluate expensively in order to sequentially better approximate the global optimum (in single-objective) or the Pareto front (in multi-objective) with as few evaluations of the expensive function(s) as possible.
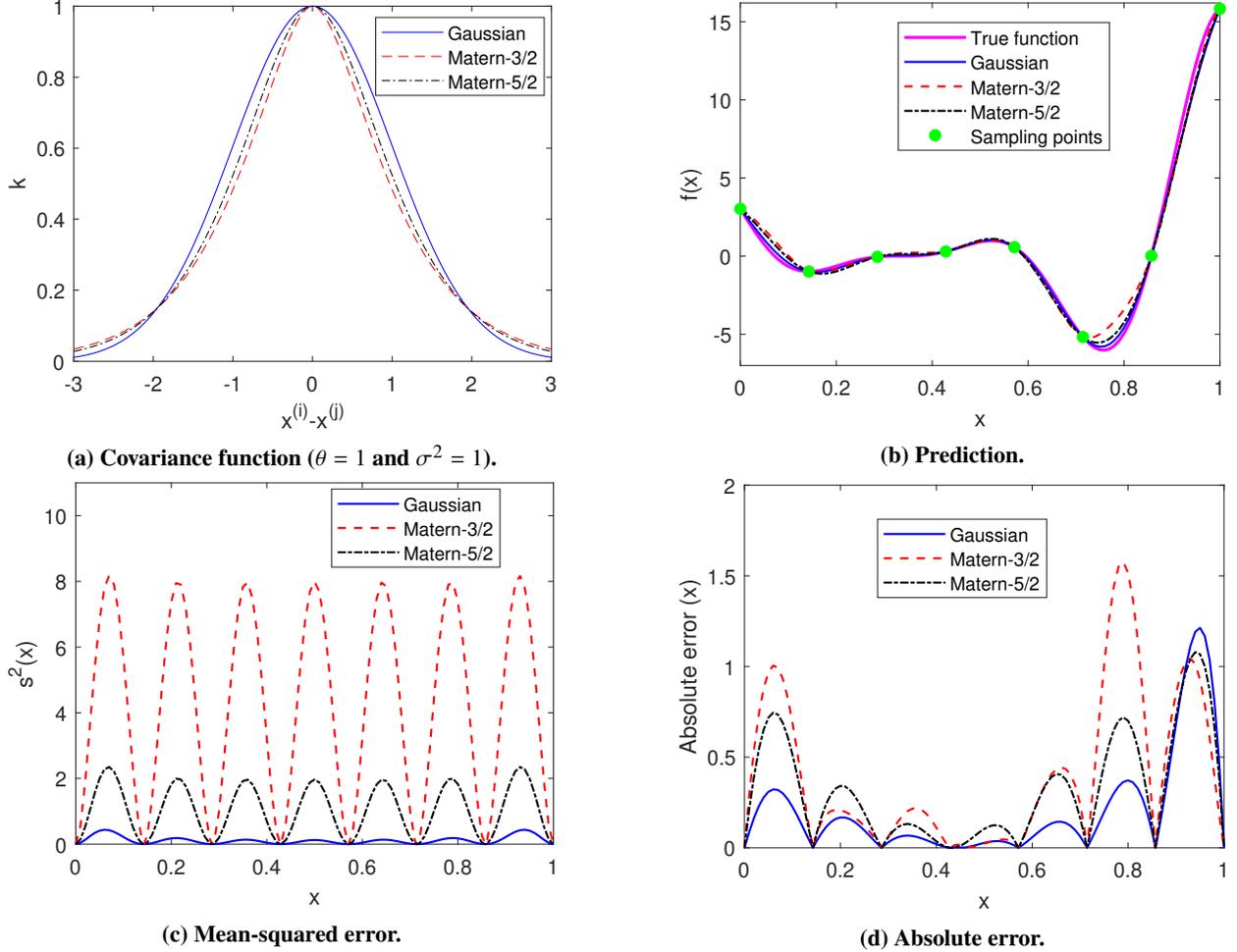
(a) **Covariance function** ($\theta = 1$ and $\sigma^2 = 1$).

(b) **Prediction.**

(c) **Mean-squared error.**

(d) **Absolute error.**

Fig. 1 **Depictions of covariance functions and the effect of different covariance functions on approximating a one-dimensional Forrester function with 8 sampling points. As can be seen in this figure, the choice of covariance function affects the quality of the Kriging model.**

B. Multi-objective Bayesian Optimization

In this paper, we use the EHVI [13] as the acquisition function due to its good performance in many engineering problems [6, 18]. Multi-objective BO with EHVI pursues an improvement in the hypervolume value, which directly corresponds to the quality of solutions in the objective space. Here, a hypervolume is defined as the region that is dominated by the current non-dominated set P and the reference point $f_{HV} = \{f_{r,1}, f_{r,2}, \ldots, f_{r,K}\}$: we denote this function as $S(P)$. When we have a new candidate solution y^{cand}, we can compute the hypervolume improvement defined as $I(y^{cand}, P) = S(P \cup \{y^{cand}\}) - S(P)$. By using the prediction and uncertainty structure of multiple Kriging models, we can compute the expected hypervolume improvement that is computed as $\text{EHVI}(x) = \int_{y \in V_{nd}} I(y, P) \cdot \text{PDF}_x(y) dy$, where PDF_x is the posterior predictive distribution, which is a multi-variate Gaussian (c.f. Fig 2), and V_{nd} is the area of integration (defined from $-\infty$ to the reference point). More details of EHVI can be found in Emmerich et al. [13]. Fig 2 illustrates the hypervolume metric, hypervolume improvement, and also EHVI a in two-dimensional objective space (both objective functions are minimized). In this paper, we use $f_{HV} = [2, 2]$ to calculate the EHVI after the objective space is normalized to $[0, 1]^2$ using the minimum and maximum of the all solutions evaluated so far.

We have a fond interest in investigating the impact of covariance function on the performance of multi-objective BO. Previous studies on multi-objective BO mostly focus on the impact of acquisition functions and not the Kriging model itself. Nevertheless, several studies aim to speed-up multi-objective BO by improving the accuracy of the Kriging model [21]. A proper choice of covariance function can accelerate the process of optimization and also produce high-quality solutions that are useful for tasks such as multi-objective design exploration. Such a study is important so
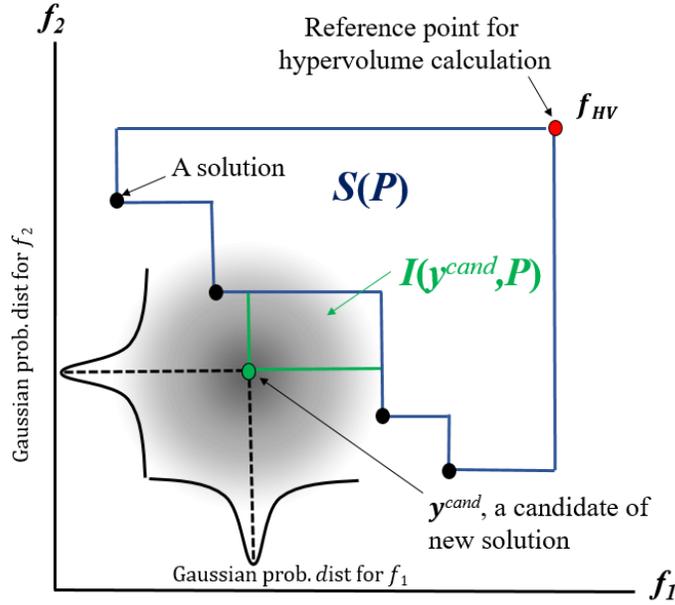
Fig. 2 An illustration of hypervolume, hypervolume improvement, and EHVI for a two-objective problem.

that we can better enhance and integrate our knowledge on covariance function and also acquisition function for efficient multi-objective BO. Therefore, in this paper, our primary aim is to identify the best practice of utilizing multi-objective BO in terms of the choice of the covariance function. EHVI is used as the acquisition function since it has been shown that it well balances the exploration and exploitation of multi-objective design space in a set of synthetic and aerodynamic optimization problems [6]. In practical terms, our goal was to seek which covariance function is the most robust to be implemented.

III. Computational Experiments

1. Problem definition

In this paper, we performed numerical experiments on a set of multi-objective optimization problems of a subsonic airfoil. Since it is not trivial to compare the impact of covariance functions if we use different covariance functions for a different objective, we use the same covariance function for the first and second objectives to simplify the analysis. The covariance functions that we investigated are Gaussian, Matérn-3/2, and Matérn-5/2. Notice that it is possible to combine different covariance functions as in the composite kernel learning method [22]; however, we did not consider composite covariance functions since we aim to understand the performance of the standard covariance functions first.

We used a low-fidelity XFOIL solver [23] to evaluate the objective functions of the fixed lift subsonic airfoil problem. The use of a low-fidelity solver allows us to collect several independent runs of multi-objective BO, which is important for statistical analysis. In our study, we use three different sets of objective functions that we explained in detail in the problem definition. For all problems, we employed the PARametric SECtion (PARSEC) [24] to create airfoil shapes by using the NLF(1)-0416 airfoil as the baseline. The PARSEC parameters of NLF(1)-0416, which were found by parameter fitting, were then perturbed by 10% from the original parameters to act as the design variables. The definition of PARSEC variables and also the bounds used for the optimization are shown in Table 1. Fig. 3 shows the baseline design and the two extreme airfoils which are located in the boundaries of the design space. All problems have the same flight conditions, i.e., $C_{l_{tgt}} = 0.5$, $M = 0.3$, and $Re = 4 \times 10^6$. Note that the $C_{l_{tgt}}$ is automatically satisfied by implicitly tuning the angle of attack. In this paper, we set the number of initial sampling points to 45 and we enriched the experimental design with 60 new samples in a sequential manner.

The objectives for the first problem are the maximization of L/D and minimization of $|C_m|$ without any geometric constraint. On the other hand, the second problem shares the same objectives with those of the first problem but with
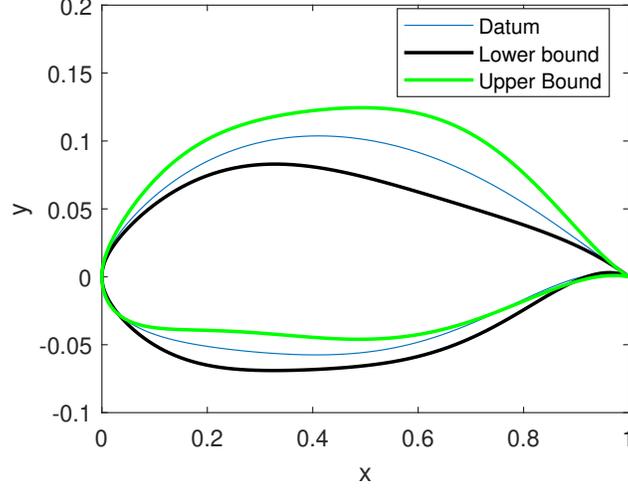
Fig. 3   The baseline airfoil and the extreme airfoils at the upper and the lower bound of the design space.

Table 1   The upper and lower bounds for fixed-lift subsonic airfoil optimization problems

Variable	Definition	l_b	u_b
r_{le}	radius of leading edge	0.0108	0.0162
X_{up}	upper crest position in horizontal coordinates	0.3288	0.4932
Z_{up}	upper crest position in vertical coordinates	0.0830	0.1245
$Z_{XX_{up}}$	upper crest curvature	-0.8700	-0.5800
X_{lo}	lower crest position in horizontal coordinates	0.3254	0.4881
Z_{lo}	lower crest position in vertical coordinates	-0.0690	-0.0460
$Z_{XX_{lo}}$	lower crest curvature	0.3086	0.4629
α_{te}	trailing edge direction	-0.2286	0.1120
β_{te}	trailing edge wedge angle	-0.1524	0.1680

an additional geometric constraint, i.e., the maximum thickness was set to $t/c = 0.12$. The geometric constraint is automatically satisfied by proper scaling of the $y-$coordinate of the airfoil. In contrast to the deterministic optimization nature of the first and the second problem, the third problem features an objective that is formulated as a robust optimization problem. To be more precise, one objective for the third problem is the standard deviation of C_l, i.e, $\sigma(C_l)$ subject to uncertainty in the angle of attack (the other objective is still the maximization of L/D and the geometric constraint from the second problem also applies). The α that yields the target C_l is perturbed by a normally distributed random value with the standard deviation value of 0.25 (the mean of the random input is α). The $\sigma(C_l)$ is then calculated by a simple Monte Carlo simulation with 200 samples. Such a robust optimization problem simulates the case when there is a disturbance in the angle of attack that perturbs the nominal C_L value. This set of the fixed-lift subsonic airfoil optimization problems is summarized in Table 2.

Table 2   The set of fixed-lift subsonic airfoil problems considered in this paper. The flight conditions for all problems were set to $C_{l_{tgt}} = 0.5$, $M = 0.3$, and $Re = 4 \times 10^6$.

Problem no.	Objective 1	Objective 2	Condition
1	Maximize L/D	Minimize $|C_m|$	No constraint applied
2	Maximize L/D	Minimize $|C_m|$	Equality constraint $t/c = 0.12$
3	Minimize $\sigma(C_L)$	Maximize L/D	The random variable is α with distribution $\mathcal{N}(\alpha, 0.25)$

Fig. 4 shows all solutions evaluated for the three problems which clearly shows the trade-off between objectives.

That is, a high L/D can only be achieved by sacrificing $|C_m|$. Similarly, a high L/D is typically achieved by sacrificing robustness in the $\sigma(C_L)$. Notice that shown in Fig. 4 are the estimated Pareto front which we obtained from multiple runs of multi-objective BO using EHVI and various covariance functions. The shape of the Pareto front for all problems are mostly convex in shape without any feasible discontinuity. This shape of the Pareto front is typical for a number of engineering problems [6]. By including geometric constraint, which is implicitly handled, it can be seen that there is a clear change in the shape of the objective space of the second problem as compared to the first problem. One distinct characteristic of the third problem is that there is an almost flat region near the high L/D part of the estimated Pareto front. It was quite difficult to find a wide range of solutions that lie on the Pareto front of the third problem; this is an indication that the third problem is the most difficult of the three problems considered in this paper. The primary cause of this difficulty is the noisy $\sigma(C_L)$ response (due to the use of Monte Carlo simulation) and the response is also likely to be non-linear. For all problems, it is clear that the task is to find solutions that are very close to the Pareto front and also diverse in terms of the coverage of the Pareto front.
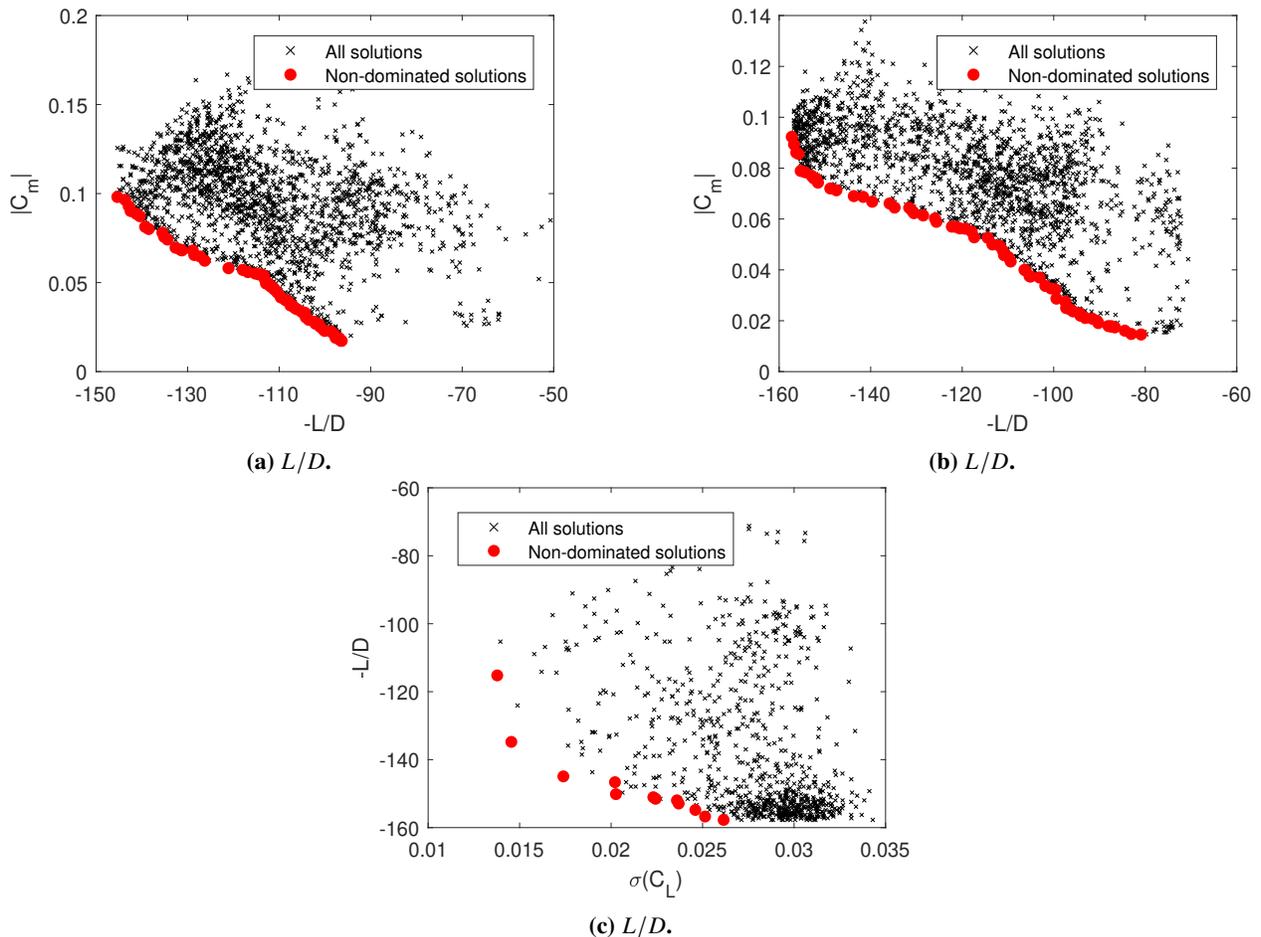

(a) L/D.

(b) L/D.

(c) L/D.

Fig. 4   Depictions of all solutions evaluated from all independent optimization runs. Notice that we depict the negative L/D so that the problem becomes a minimization problem.

2. Performance metrics

In this paper, we use the hypervolume (HV) and the inverted generational distance (IGD) to assess the performance of the optimizer. The HV metric is defined as a hypervolume which is bounded by the reference point and the set of non-dominated solutions P. Since the set with the highest hypervolume is the Pareto front, a high value of hypervolume is desirable as it indicates that we are closer to the true Pareto front. On the other hand, the IGD metric measures the distance between non-dominated solutions to the true Pareto front of the problem. If P and $P*$ are, respectively, the

current non-dominated solutions (estimated Pareto front) and the true Pareto front, the IGD can be calculated as

$$IGD(P*, P) = \frac{\sum_{v \in P*} d(v, P)}{|P*|} \tag{10}$$

where $d(v, P)$ is the minimum Euclidean distance between v (a point in the Pareto front) and points in P. Both the IGD and HV metric measure the diversity and also the proximity of a set of solutions to the true Pareto front. Here, a good multi-objective optimization method yields a high HV and a low IGD value, which indicates a good approximation of the Pareto front. The reference point for EHVI was set to $[1.1, 1.1]$ after we normalized the objectives to $[0, 1]^2$ by using the minimum and maximum of the final non-dominated solutions. The reference set $P*$ used for IGD is the final non-dominated solutions as shown in Fig. 4.

For the following analysis of the accuracy of Kriging and also the results of optimization, we used the Mann-Whitney U test with p value of 0.05 to check the statistical significance of the results.

3. Accuracy of the Kriging model

First, we analyze the quality of the Kriging model with 45 sampling points (i.e., the size of the initial samples) and various covariance functions in approximating all objectives. Notice that in total we have five different responses since the second and the third problem shares one similar objective. The first and the second problem have different responses of L/D and $|C_m|$ since the latter applies scaling to the y-coordinate of the airfoil; thus, changed the mapping between the design variables and the objective functions. The objective of this study is primarily to link the accuracy of the Kriging model and the optimization performance. The initial sampling points were generated using latin hypercube sampling (LHS). For the validation samples, an independent set consisting of 1000 samples were used to measure the accuracy of the surrogate model. The results as measured by the R^2 metric are shown in Fig. 5 (averaged from 50 independent runs). Notice that these are the results for the initial Kriging models. Since it is required to build a number of Kriging models during the course of optimization, the results that are shown in Fig. 5 only give us an initial assessment of the surrogate model.

Here, we can see that the choice of covariance function, to an extent, affects the approximation quality of Kriging. Based on the Mann-Whitney U test with $p = 0.05$, the only statistically significant difference that we observed is between Gaussian and Matérn-3/2 for $|C_m|$ (problem 1) and also $\sigma(C_L)$ (problem 3). There is a slight notable difference between the Gaussian and Matérn-3/2 for L/D (problem 1), however, the statistical test reveals that such a difference is not significant. We also observe that approximating L/D is more difficult than $|C_m|$ for both problem 1 and problem 2, as indicated by the lower R^2 of the former. This implies that the response surfaces of L/D is more non-linear compared to those of $|C_m|$. For problem 3, it also turns out that approximating the $\sigma(C_L)$ response is a highly tedious task; the R^2 of $\sigma(C_L)$ is much lower compared to that of L/D and $|C_m|$. Such a low R^2 is primarily due to the combined effect of nonlinearity and the noisy response surface. However, we observe that equipping Kriging with Matérn-3/2 covariance functions is better than Gaussian for $\sigma(C_L)$. Gaussian covariance function assumes that the function is extremely smooth and this is the most plausible reason for its relatively poor performance.

4. Results of optimization

The mean convergence of HV and IGD for all problems (averaged from five independent runs) are depicted in Figs. 6 and 7, respectively. As can be seen from the results, the choice of covariance function influences the performance of multi-objective BO. However, statistical tests are necessary to check the significance of these differences. For the first problem, we observe that the Kriging models with Gaussian and Matérn-3/2 covariance functions yielded good performance in terms of both IGD and HV, which indicate that both methods are capable of finding solutions that are diverse and close to the Pareto front of the first problem. Nevertheless, only Matérn-3/2 that yielded statistically significant better performance compared to Matérn-5/2 in both HV and IGD (Gaussian is only significantly better when measured by the IGD metric). We specifically highlight the good convergence of the multi-objective BO with the Matérn-3/2 covariance function; here, Kriging with Matérn-3/2 performed well from the viewpoint of both accuracy and optimization. Although Kriging with Matérn-5/2 performed relatively well in terms of accuracy, it turns out that it did not yield a good optimization performance relative to the others for problem 1. Conversely, Kriging with Matérn-5/2 significantly outperformed Matérn-3/2 on the second problem in terms of IGD; while there is no statistically significant difference in terms of HV. On the other hand, multi-objective BO with Gaussian and Matérn-3/2 performed roughly similar to each other for the second problem, as confirmed by the statistical hypothesis test. Notice that it is difficult to predict such a trend if we solely observe the prediction accuracy (see Fig. 5).
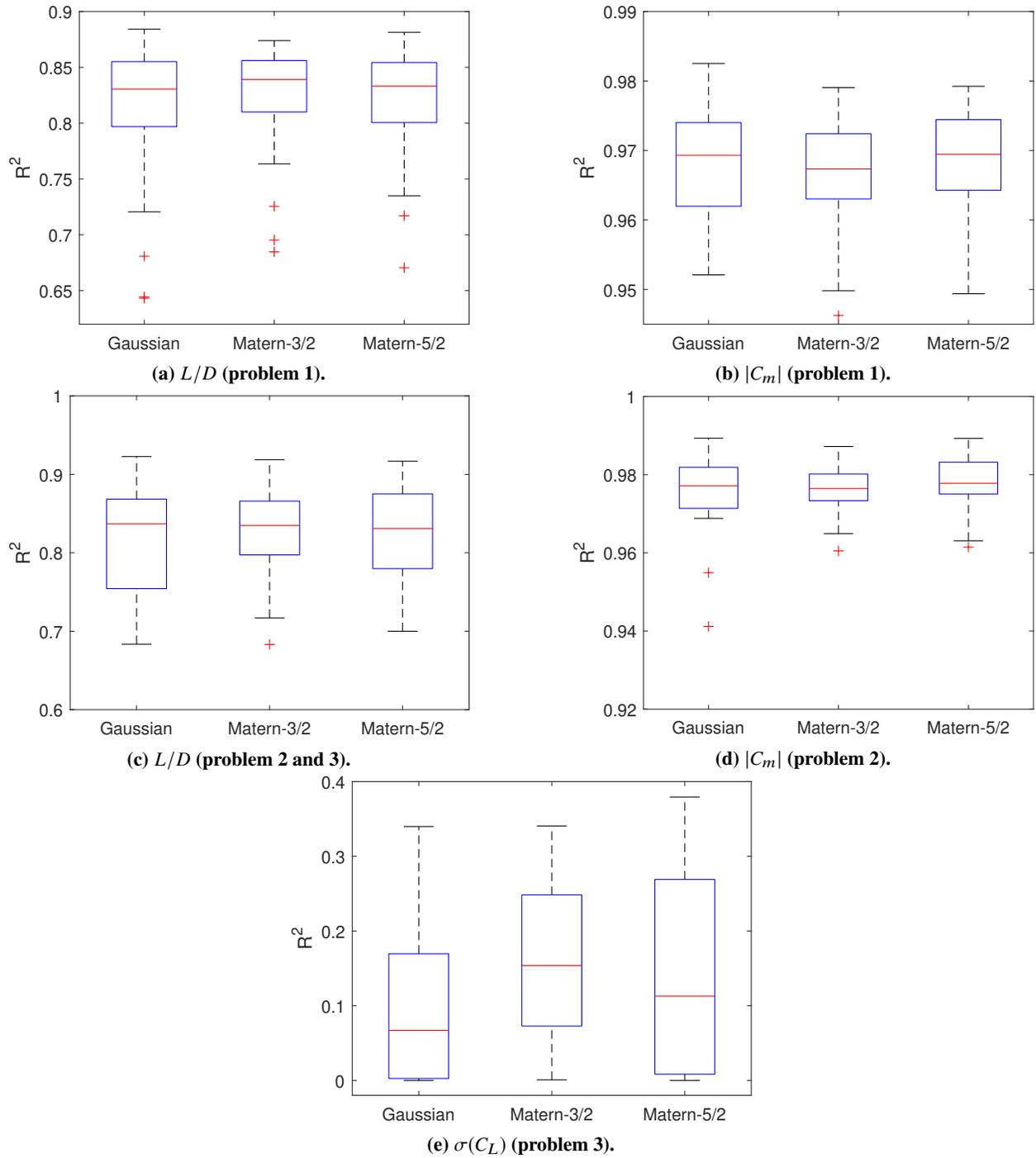
(a) L/D **(problem 1).**

(b) $|C_m|$ **(problem 1).**

(c) L/D **(problem 2 and 3).**

(d) $|C_m|$ **(problem 2).**

(e) $\sigma(C_L)$ **(problem 3).**

Fig. 5   Accuracy of various Kriging models using 45 sampling points for all objective functions.
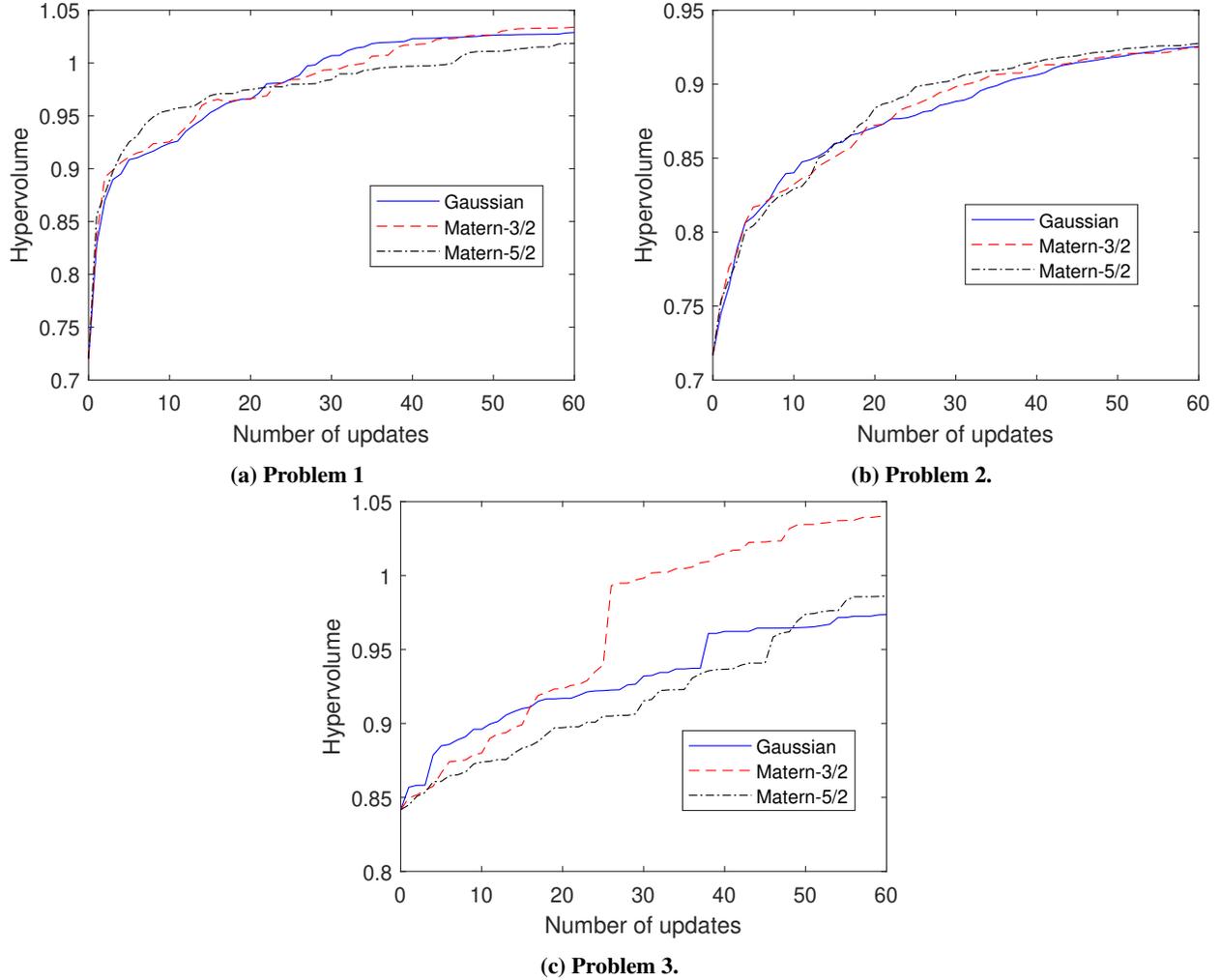
(a) Problem 1

(b) Problem 2.

(c) Problem 3.

Fig. 6 **Convergence of *HV* for the fixed lift subsonic airfoil problems.**

The third problem is the most difficult problem among the three problems considered in this paper. The second objective (i.e., $\sigma(C_L)$) is the main root of this difficulty, primarily due to the noisy response and also the non-linearity. In this problem, we can see that the choice of covariance function, also to a certain extent, affects the performance of multi-objective BO. Due to the limited number of solutions in the estimated Pareto front, we only depict the hypervolume convergence since IGD needs reference solutions that cover the Pareto front well. However, we think that using just HV is enough since the difference between the results in terms of the HV is significant. From the results, we observe that the Matérn-3/2 covariance function produced statistically significant superior results compared to the other two covariance functions. The Matérn-3/2 covariance function features a rough sampling path and this translates into a faster convergence to the Pareto front of the third problem. In terms of accuracy, Kriging with Matérn-3/2 also yielded higher accuracy for the approximation of $\sigma(C_L)$ compared to Gaussian. Since the second and the third problem shares the same objective (i.e., L/D), the better performance of Matérn-3/2 in the third problem can only be attributed to its use on the $\sigma(C_L)$ response. Similar phenomenon was also observed in single-objective Bayesian optimization, in which Matérn-3/2 bested the others when noise corrupts the black-box function [20].

By comparing the initial Kriging accuracy and the optimization performance, we observe that a good accuracy of initial Kriging does not necessarily translate to good performance when it is used in a multi-objective BO. It is also worth noting that the Gaussian covariance function, despite its criticism [19], still performed relatively well on all problems in terms of optimization performance. This implies that Kriging models with Gaussian can point out the directions that improve the quality of solutions in a multi-objective space. Nevertheless, when exploitation of solutions
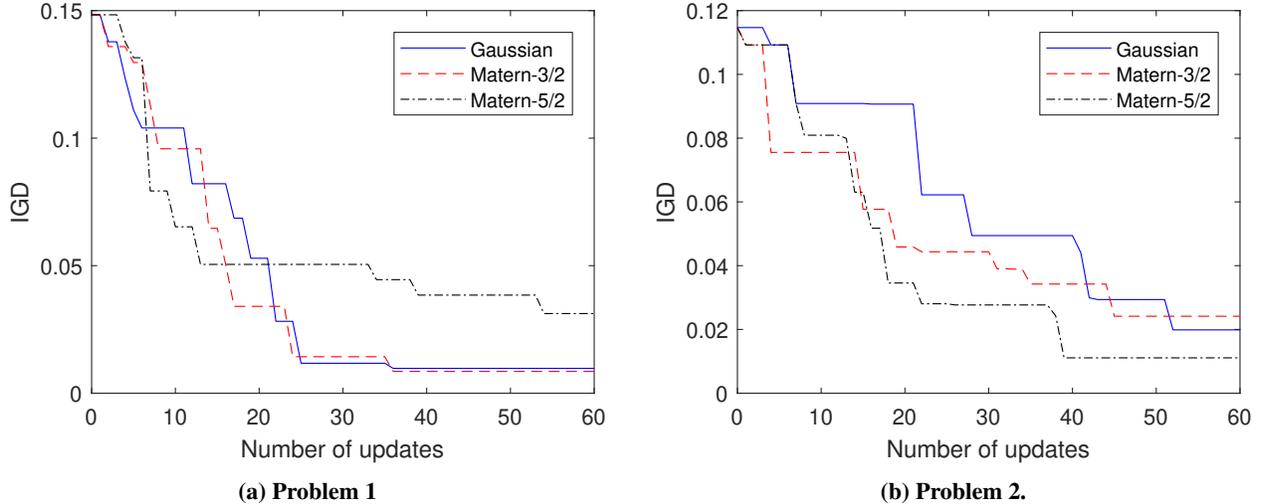
(a) Problem 1

(b) Problem 2.

Fig. 7   Convergence of IGD for the fixed lift subsonic airfoil problems.

is the main aim, Gaussian might not be the best choice due to its lower exploitative power compared to the Matérn covariance function [20]. However, for multi-objective optimization with a limited budget, the main concern is typically to produce solutions that spread out across the Pareto front; therefore, the solutions do not really need to be extremely close to the Pareto front. In contrast to the results in single-objective BO [22], the choice of the covariance function also affects the exploration phase, which is reasonable since exploration is important to find solutions that spread across the Pareto front.

Summarizing the results, we observe that the Matérn-3/2 covariance function provides good performance, meaning that it is capable of finding a good set of solutions on the three problems considered in this paper. Our observation is on the same line with Palar et al. [22], which investigated the same set of covariance functions in single-objective BO, who agreed that Matérn-3/2 covariance function performed well on non-synthetic and noisy problems. For non-algebraic problems (including the three problems that we considered), it is reasonable to assume that the function is corrupted by numerical noise, although it might be very small. However, based on the evidence that we have right now, it is still difficult to take solid conclusions regarding which covariance function yields the best average performance for multi-objective BO. We think that it is necessary to perform tests on other types of problems (e.g. structural optimization problem or problems with higher dimensionality and different levels of non-linearity). Nevertheless, the current results show a good tendency toward the Matérn-3/2 covariance function.

IV. Conclusions and future works

In this paper, we studied the impact of various covariance functions on the performance of multi-objective BO with Kriging. We used a set of fixed-lift low-fidelity subsonic airfoil optimization for benchmarking Kriging with various covariance functions, namely, Gaussian, Matérn-3/2, and Matérn-5/2. The main goal of this study is to find the best practice of performing multi-objective BO in terms of the covariance function. In this paper, we use the same covariance function for all objectives since, without prior knowledge, it is not trivial to pick different covariance functions for different objectives. We utilized EHVI as the acquisition function to enrich the experimental design sequentially.

The results show that the choice of covariance function affects the performance of multi-objective BO. In general, we observe that Matérn-3/2 is the most robust covariance function that, when combined with Kriging, is capable of handling various challenges (e.g., noisy and nonlinear functions) and discovering good quality solutions. It is also worth noting the accuracy of the Kriging model does not strictly correspond to a better optimization performance, especially when there is a small difference between the accuracy of various surrogate models. The main advantage of the Kriging with Matérn-3/2 covariance function is that it performs well in terms of both accuracy and optimization. Another important observation is that Gaussian performed relatively well, despite its high smoothness assumption which might not be true for the applications considered in this paper.

For future works, it would be interesting to investigate the use of combined covariance functions in both single- and multi-objective BO. However, tests on other types of problems in different fields are still necessary to draw a

more solid conclusion regarding the impact of covariance functions. For now, although Matérn-3/2 is the most robust choice for the three problems considered in this paper, it is not necessarily the best performing covariance function for a specific problem. Incorporating reward-based machine learning techniques (i.e., reinforcement learning) is an interesting research avenue so that the BO algorithm can automatically pick the best performing covariance function that yields the fastest convergence to the Pareto front. This study focuses more on the empirical observations, however, more works still should be done from the theoretical viewpoint to get a better understanding of the impact of covariance functions.

Acknowledgements

References

[1] Drucker, H., Burges, C. J., Kaufman, L., Smola, A. J., and Vapnik, V., "Support vector regression machines," *Advances in neural information processing systems*, 1997, pp. 155–161.

[2] Jones, D. R., Schonlau, M., and Welch, W. J., "Efficient global optimization of expensive black-box functions," *Journal of Global optimization*, Vol. 13, No. 4, 1998, pp. 455–492.

[3] Chugh, T., Rahat, A., Volz, V., and Zaefferer, M., *Towards Better Integration of Surrogate Models and Optimizers*, Springer, 2020, High-Performance Simulation Based Optimization, Studies in Computational Intelligence, Vol. 833, pp. 137–163.

[4] Lam, R., Poloczek, M., Frazier, P., and Willcox, K. E., "Advances in bayesian optimization with applications in aerospace engineering," *2018 AIAA Non-Deterministic Approaches Conference*, 2018, p. 1656.

[5] Hebbal, A., Brevault, L., Balesdent, M., Talbi, E.-G., and Melab, N., "Multi-objective optimization using Deep Gaussian Processes: Application to Aerospace Vehicle Design," *AIAA Scitech 2019 Forum*, 2019, p. 1973.

[6] Zuhal, L. R., Palar, P. S., and Shimoyama, K., "A comparative study of multi-objective expected improvement for aerodynamic design," *Aerospace Science and Technology*, 2019.

[7] Simpson, T., Mistree, F., Korte, J., and Mauery, T., "Comparison of response surface and kriging models for multidisciplinary design optimization," *7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, 1998, p. 4755.

[8] Simpson, T. W., Mauery, T. M., Korte, J. J., and Mistree, F., "Kriging models for global approximation in simulation-based multidisciplinary design optimization," *AIAA journal*, Vol. 39, No. 12, 2001, pp. 2233–2241.

[9] Močkus, J., "On Bayesian methods for seeking the extremum," *Optimization Techniques IFIP Technical Conference*, Springer, 1975, pp. 400–404.

[10] Frazier, P., Powell, W., and Dayanik, S., "The knowledge-gradient policy for correlated normal beliefs," *INFORMS journal on Computing*, Vol. 21, No. 4, 2009, pp. 599–613.

[11] Hernández-Lobato, J. M., Hoffman, M. W., and Ghahramani, Z., "Predictive entropy search for efficient global optimization of black-box functions," *Advances in neural information processing systems*, 2014, pp. 918–926.

[12] Hoffman, M. D., Brochu, E., and de Freitas, N., "Portfolio Allocation for Bayesian Optimization." *UAI*, Citeseer, 2011, pp. 327–336.

[13] Emmerich, M. T., Giannakoglou, K. C., and Naujoks, B., "Single-and multiobjective evolutionary optimization assisted by Gaussian random field metamodels," *IEEE Transactions on Evolutionary Computation*, Vol. 10, No. 4, 2006, pp. 421–439.

[14] Namura, N., Shimoyama, K., and Obayashi, S., "Expected improvement of penalty-based boundary intersection for expensive multiobjective optimization," *IEEE Transactions on Evolutionary Computation*, Vol. 21, No. 6, 2017, pp. 898–913.

[15] Keane, A. J., "Statistical improvement criteria for use in multiobjective design optimization," *AIAA journal*, Vol. 44, No. 4, 2006, pp. 879–891.

[16] Hernández-Lobato, D., Hernandez-Lobato, J., Shah, A., and Adams, R., "Predictive entropy search for multi-objective bayesian optimization," *International Conference on Machine Learning*, 2016, pp. 1492–1501.

[17] Knowles, J., "ParEGO: a hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems," *IEEE Transactions on Evolutionary Computation*, Vol. 10, No. 1, 2006, pp. 50–66.

[18] Shimoyama, K., Sato, K., Jeong, S., and Obayashi, S., "Updating kriging surrogate models based on the hypervolume indicator in multi-objective optimization," *Journal of Mechanical Design*, Vol. 135, No. 9, 2013, p. 094503.

[19] Stein, M. L., *Interpolation of spatial data: some theory for kriging*, Springer Science & Business Media, 2012.

[20] Palar, P. S., and Shimoyama, K., "Efficient global optimization with ensemble and selection of kernel functions for engineering design," *Structural and Multidisciplinary Optimization*, Vol. 59, No. 1, 2019, pp. 93–116.

[21] Palar, P. S., and Shimoyama, K., "On multi-objective efficient global optimization via universal Kriging surrogate model," *2017 IEEE Congress on Evolutionary Computation (CEC)*, IEEE, 2017, pp. 621–628.

[22] Palar, P. S., and Shimoyama, K., "Kriging with Composite Kernel Learning for Surrogate Modeling in Computer Experiments," *AIAA Scitech 2019 Forum*, 2019, p. 2209.

[23] Drela, M., "XFOIL: An analysis and design system for low Reynolds number airfoils," *Low Reynolds number aerodynamics*, Springer, 1989, pp. 1–12.

[24] Sobieczky, H., "Parametric airfoils and wings," *Recent development of aerodynamic design methodologies*, Springer, 1999, pp. 71–87.