



OPEN

Low-cost quasi-global optimization of expensive electromagnetic simulation models by inverse surrogates and response features

Slawomir Koziel^{1,2} & Anna Pietrenko-Dabrowska^{1,2}✉

Conceptual design of contemporary high-frequency structures is typically followed by a careful tuning of their parameters, predominantly the geometry ones. The process aims at improving the relevant performance figures, and may be quite expensive. The reason is that conventional design methods, e.g., based on analytical or equivalent network models, often only yield rough initial designs. This is especially the case for miniaturized components featuring considerable electromagnetic (EM) cross couplings, or antenna systems with non-negligible radiator coupling (e.g., MIMO, closely-spaced arrays). For reliability reasons, parametric optimization is carried out using EM simulation tools, which is a time-consuming task. In many cases, designer needs to resort to a global search, especially when handling several objectives and constraints is necessary, or the high-frequency structure under design is overly complex. Combination of both aforementioned factors makes it no longer possible to rely on engineering insight, even to detect a promising region of the design space. Unfortunately, nature-inspired algorithms, commonly employed for solving these tasks typically exhibit significant computational expenditures. This paper proposes a simple yet efficient method for globalized search using a response feature approach and inverse regression surrogates. Owing to less nonlinear dependence of the feature point coordinates on the system variables (as compared to the original responses, e.g., *S*-parameter frequency characteristics), our methodology permits a rapid identification of the most appropriate regions of the parametric space, and further design tuning by means of local routines. At the same time, the overall optimization cost is comparable to the cost of local procedures. The proposed approach is validated using several high-frequency structures (a dual-band antenna, a microstrip coupler, an impedance matching transformer) optimized under different design scenarios. Global search capability and computational efficiency are demonstrated through comprehensive comparisons with multiple-start local search, as well as particle swarm optimizer, a representative nature-inspired algorithm.

List of symbols

A	Regression matrix
a_{red}	Multiplication factor for sampling region reduction
β	Penalty coefficients of the objective function U
$c_1 = c_2 = 2.05$	Control parameters of the PSO optimization algorithm
$\mathbf{d}^{(i)}$	Size vector of the TR algorithm
$E(\mathbf{p})$	Merit function
$E^{(j)}$	Merit function for particular feature vector $\mathbf{p}(\mathbf{x}_r^{(j)})$; $E^{(j)} = E(\mathbf{p}(\mathbf{x}_r^{(j)}))$
E_{reduce}	Threshold value for $E(\mathbf{p}(\mathbf{x}_{best}))$ to reduce the sampling region (infill sampling)
EM	Electromagnetic
f	Frequency
F	Objective space
$\mathbf{F} = [F_1 \dots F_N]^T$	Target objective vector

¹Engineering Optimization and Modeling Center, Reykjavik University, 102 Reykjavik, Iceland. ²Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, 80-233 Gdansk, Poland. ✉email: anna.dabrowska@pg.edu.pl

F^*	Target objective vector specific to the particular design task
F_T	Space of auxiliary objective vectors
F_T^*	Auxiliary objective vector
F_T^*	Particular auxiliary vector corresponding to the target objective vector F^*
$f_0, f_{0,1}, \dots, f_{0,N}$	Circuit operating frequencies
FD	Finite differentiation
$G^{(i)}$	Linear model locally approximating $R(\mathbf{x})$ at $\mathbf{x}^{(i)}$
J_R	Sensitivity matrix of the system responses
K	Power split
k_{\max}	Maximum number of iterations, $k_{\max} = 50$ (PSO optimization algorithm)
l	Lower bound on design variables
$l_k, k = 1, \dots, N_T$	Level coordinates of the response features
$L(\mathbf{s}(\mathbf{p}))$	Inverse regression surrogate
LHS	Latin hypercube sampling
n	Number of design variables
N_1	Number of initial random samples $\mathbf{x}_r^{(j)}$
N_2	Number of best samples selected to construct the inverse surrogate
N_3	Number of infill samples added to the sample pool per iteration
N_{\max_global}	Maximum number of EM evaluations for the global search stage
N_{\max_local}	Maximum number of EM evaluations for the local search state
N_s	Swarm size, $N_s = 10$ (PSO optimization algorithm)
$\mathbf{p}(\mathbf{x}) = [p_1(\mathbf{x}) \dots p_L(\mathbf{x})]^T$	Vector of response features at \mathbf{x}
PSO	Particle swarm optimization
$R(\mathbf{x})$	Response of the circuit EM-simulation model
$s_k(\mathbf{p}), k = 1, \dots, N_T$	Scalar functions mapping the features into F_T
$\mathbf{s}(\mathbf{p}) = [s_1(\mathbf{p}) \dots s_{N_T}(\mathbf{p})]^T$	Vector of scalar functions $s_k(\mathbf{p})$
$S_{kl}(\mathbf{x}_j f)$	Circuit S-parameters, k and l denote circuit ports
TR	Trust region
u	Upper bound on design variables
$U(\mathbf{x}, \mathbf{F})$	Objective function
$\mathbf{x}^*(\mathbf{F}) = \mathbf{x}^*$	Optimal solution
U_{local}	Threshold value for $U(\mathbf{x}_{best}, \mathbf{F}^*)$ to enter the local search stage
W	Weight matrix
\mathbf{x}	Vector of designable (typically, geometry) parameters
\mathbf{x}^*	Optimal solution
$\mathbf{x}^{(i)}, i = 0, 1, \dots$	Approximations of the solution \mathbf{x}^* (TR algorithm)
\mathbf{x}_{best}	Best design found so far in the optimization process
$\mathbf{x}_r^{(j)} = [x_{r,1}^{(j)} \dots x_{r,n}^{(j)}]^T, j = 1, \dots, N_1$	Set of random points in X
\mathbf{x}_T^*	Design rendered by inverse model used as a starting point for local optimization
X	Design space
χ	Control parameter of PSO optimization algorithm, $\chi = 0.73$

Due to a growing complexity of contemporary high-frequency structures (antennas, microwave components, integrated photonic devices, etc.), carrying out a complete design cycle at the level of simplified representations such as analytical/semi-empirical formulas or equivalent networks is no longer feasible. Certain phenomena that exhibit non-negligible effects on the system performance, e.g. electromagnetic (EM) cross-coupling¹, the presence of connectors², housing³, mutual coupling of radiators⁴, can only be accounted for by means of a full-wave EM analysis. Appropriate capturing of these effects, along with the necessity of fulfilling stringent performance requirements (which includes realization of various functionalities such as tunability⁵, multi-band operation⁶, circular polarization⁷, but also maintaining small physical size of the devices^{8–10}), leads to problems with obtaining reasonable initial designs or even locating the regions of the parameter space that contain such designs.

In this situation, parameter tuning becomes imperative. Nowadays, it is often performed using rigorous numerical optimization methods, as they are suitable for handling multiple objectives and constraints, as well as simultaneous adjustment of many parameters. Yet, several practical problems arise. EM-driven design, otherwise dictated by reliability, entails considerable computational costs, even in the case of local optimization. Further, the lack of good initial designs but also multimodality of the functional landscapes pertinent to many problems (e.g., pattern synthesis in antenna arrays¹¹) call for global search, the cost of which may be considerably higher^{12,13}.

The most widely used global optimization methods are by far population-based metaheuristics^{14–20}. Well-known techniques of this sort include genetic¹⁴ and evolutionary algorithms¹⁵, particle swarm optimizers (PSO)¹⁶, differential evolution¹⁷, as well as numerous variations such as firefly algorithm¹⁸, harmony search¹⁹, grey wolf optimization²⁰, and many others (e.g.,^{21–26}). The underlying search mechanism is the exchange of information between a set of individuals (or agents) realized using appropriate exploratory (recombination) and exploitative (mutation) operators. This generally allows for identifying and exploiting the promising regions of the parameter space, in particular, escaping from the local optima. Metaheuristics are easy to implement and handle, but their global search capabilities come at a price of significant computational expenses, which may reach many thousands of objective function evaluations for a single algorithm run^{20,23}. This is acceptable if the system evaluation cost

is not of primary concern (e.g., for pattern synthesis of antenna arrays using analytical array factor models²⁰), but turns prohibitive when directly optimizing full-wave EM models.

Population-based global search may still be a viable option if the EM simulation cost is reasonably low (e.g., a few seconds per analysis), or the available computational resources and licensing permit parallelization. Otherwise, more sophisticated approaches have to be employed, mostly involving surrogate modelling^{27–29}. A representative example constitutes efficient global optimization (EGO) methods³⁰, where a fast surrogate model (typically, kriging interpolation³¹ or Gaussian Process Regression³²) is constructed along with conducting the search process itself. The infill samples are added based on various criteria that may aim at the parameter space exploration (i.e., improving the surrogate predictive power) or exploitation, i.e., finding the global optimum³³. Other, yet related, options include application of machine learning techniques^{34,35}, often in conjunction with sequential sampling methods³⁶ as well as utilization of auxiliary surrogates to perform design space pre-screening^{37,38}.

A construction of a globally accurate surrogate to replace expensive EM simulations altogether is a conceptually attractive alternative. In practice, due to the curse of dimensionality, it is limited to low-dimensional cases within restricted ranges of the system parameters^{39,40}. This difficulty has been recently mitigated by the performance-driven modelling methods^{41–43}, where the surrogate model domain is meticulously allocated to contain the designs being optimum with respect to the selected performance figures. This enables rendering reliable models over wide ranges of parameters and operating conditions^{41,42}. Similar ideas were proposed in the context of multi-objective design, where initial identification of the Pareto front by means of single-objective optimization runs may permit one-shot surrogate construction at practically acceptable computational costs⁴⁴. Such a model is subsequently used to yield initial approximations of the Pareto sets, further refined using local methods⁴⁴. For this, but also other applications, additional speedup can be achieved by employing variable-fidelity EM simulations^{45,46}.

The response feature approach, originally fostered for local optimization⁴⁷, and later for surrogate modeling⁴⁸, offers supplementary advantages. The optimization (or modeling) process is reformulated to be carried out at the level of appropriately defined characteristic points of the system outputs (e.g., frequency/level allocation of multi-band antenna resonances or local maxima of the in-band return loss response of a microwave filter), which is in opposition to handling the original responses, typically, frequency characteristics. The dependence of the feature point coordinates on the geometry parameters is normally much less nonlinear than for the standard outputs, which leads to flattening of the functional landscape to be handled. In the case of local optimization, it results in faster convergence⁴⁷, whereas in the case of surrogate modeling it brings a considerable reduction of the number of training data samples required to render a reliable model⁴⁸. This paper capitalizes on the response feature technology to realize globalized optimization of electromagnetic computational models. The major component is an inverse surrogate constructed at the level of feature points extracted from a set of random observables. The predictions generated by the surrogate allow for rapid identification of the promising regions of the parameter space and yielding a reasonable initial design, which only needs to be tuned in a local sense. At the same time, the computational complexity of the proposed procedure is comparable to that of local procedures. Our methodology is validated using several high-frequency structures (a dual-band antenna, a microstrip coupler, an impedance matching transformer), optimized under different scenarios. The global search capability and computational efficiency are demonstrated through comprehensive comparisons with multiple-start local search, as well as particle swarm optimizer (PSO), a representative nature-inspired population-based procedure.

Methods. The proposed optimization approach has been developed in accordance with the methodology itemized below:

1. Conceptual development of the algorithmic framework for globalized optimization of high-frequency structures;
2. Computer implementation of the algorithm and debugging;
3. Preparing a library of test problems including their computational (EM-simulation) models;
4. Setting up control parameters of the optimization framework;
5. Implementation of benchmark algorithms;
6. Conducting numerical experiments;
7. Comparative analysis of the results;
8. Extraction of the final conclusions and findings.

The proposed optimization framework has been implemented in MATLAB. Whereas the EM models of all the high-frequency components used in this work as verification case studies are evaluated using time-domain solver of CST Microwave Studio. All the simulations have been performed on Intel Xeon 2.1 GHz dual-core CPU with 128 GB RAM.

Globalized search using inverse regression surrogates and response features

This section introduces a globalized search procedure proposed in this work. We begin by discussing its two fundamental components: a response feature approach employed to flatten the functional landscape of the objective function pertinent to the design problem at hand, as well as the inverse regression model utilized to facilitate identification of the promising regions of the parameter space. Both mechanisms are critical to enable global search capability at low computational cost, comparable to that of local optimizers as demonstrated in “[Demonstration case studies](#)” section. Another important ingredient of the procedure is a local optimization routine outlined in “[Design refinement by local search](#)” section, which permits carrying out fine-tuning of the system parameters. The operation of the entire optimization framework is discussed in “[Optimization framework](#)” section.

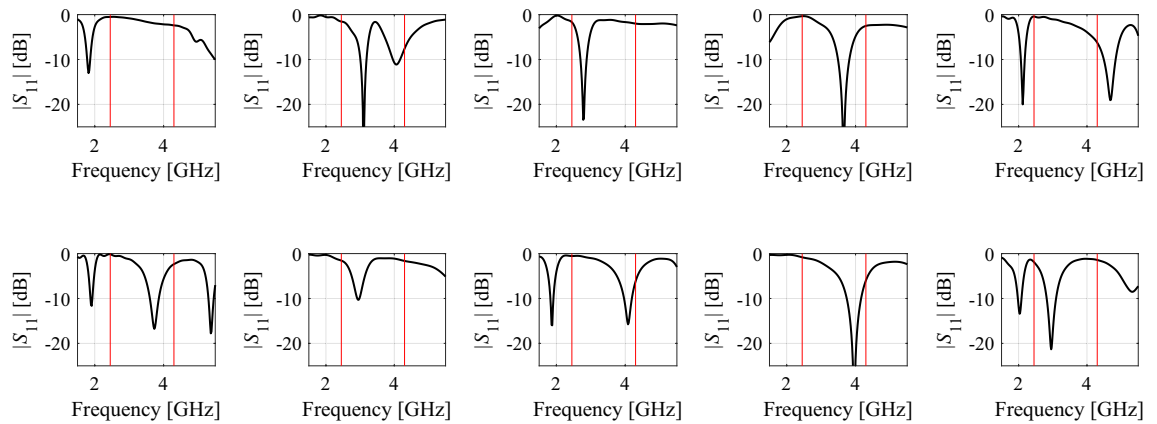


Figure 1. Exemplary reflection responses of a dual-band antenna along with the target operating frequencies marked using vertical lines (here, 2.45 GHz and 4.3 GHz). Launching the local search aimed at, e.g., reduction of the antenna reflection in the vicinity of the target frequencies (and formulated in a minimax sense), from the majority of the presented designs would likely not bring satisfactory results as a direct consequence of a deficient initial allocation of the reflection characteristic minima.

Response features for global search. The need for global search emerges for many high-frequency structures, either due to the lack of a reasonable initial design, or a multimodality (the presence of multiple local optima) of the functional landscape pertinent to the cost function quantifying the design utility. The primary challenge of global optimization is the necessity of carrying of the search process within the entire parameter space, which is often highly dimensional. Furthermore, the responses of electromagnetic computational models exhibit significant variability, e.g., considerable shifts of operating frequencies across the design space. For example, conducting the search process using nature-inspired algorithm, otherwise suitable for such purposes, entails considerable computational expenses, especially when executed at the level of full-wave simulation models.

Figure 1 shows example reflection responses of a dual band antenna at various points of the parameter space. It is clear that using these particular designs as starting points for local optimization may lead to a failure of the optimization process, e.g., because the initial allocation of antenna resonances is incompatible with the target ones, or the resonances are not clearly distinguished. On the other hand, following the underlying physical properties of passive components, the parameters of high-frequency structures corresponding to designs that are optimum with respect to particular performance figures are largely correlated^{49–51}. For example, dimension scaling of antenna or microwave components with respect to the operating frequency (or frequencies in the case of multi-band structures), substrate parameters, power split ratio (in the case of couplers), etc., requires synchronized adjustment of the design variables. In other words, the optimum designs are normally allocated along low-dimensional manifolds (surfaces) of the dimensionality equal to the number of operating conditions or performance figures considered for a given design task⁵². This observation has led to the development of a paradigm of constrained modeling, in which the modeling process is restricted to this part of the design space, where designs of high-quality reside⁵². As the number of operating conditions is normally considerably lower than the number of parameters of the system at hand, constrained modelling effectively reduces the modelling task dimensionality⁵². These manifolds are usually regular as indicated in Fig. 2, using the example of a miniaturized rat-race coupler⁵³.

Capturing the correlations between the parameter values corresponding to optimum designs is a key to accelerate the global optimization process according to the methodology proposed in this paper. Typically, to assess a particular design in relation to the optimum design manifold, it is sufficient to consider only particular portions of the system outputs, here, referred to as response features⁴⁷. The feature points are to be extracted from the complete characteristics, and they determine the relations between these and the design objectives. For example, if a goal is to optimize a dual-band antenna so that its resonances are allocated at given target frequencies and, at the same time, reflection levels therein are to be minimized, appropriate response features may simply be the frequency and level coordinates of the resonances. If, however, we aim at maximization of a fractional bandwidth around the target frequencies, then the frequencies corresponding to -10 dB reflection level would be more suitable. For other structures, the choice of the feature points may be more involved (see, e.g.,^{54,55}).

It should be emphasized that a rigorous mathematical definition of response features has not been formulated. This is because they are very much problem dependent, as the examples of the previous paragraph indicate. Thus, there exist no definition, which would be sufficiently general to cover all possible cases that may emerge when solving practical design tasks. This is a limitation of the response feature technology. A generalized definition of the response features has been given in⁵⁶ for antenna input characteristics. Despite covering merely one type of antenna response, the definition of⁵⁶ is intricate, which confirms that providing a rigorous mathematical definition of response feature for all types of characteristics of antenna and microwave structures that may be of interest in high-frequency design is far from trivial. Therefore, as explained above, the response features need to be defined in relation to a specific design task.

Qualitatively, given a number of random designs (observables) and their response features extracted, one may attempt to use this information to at least roughly approximate the optimum design manifold. The model

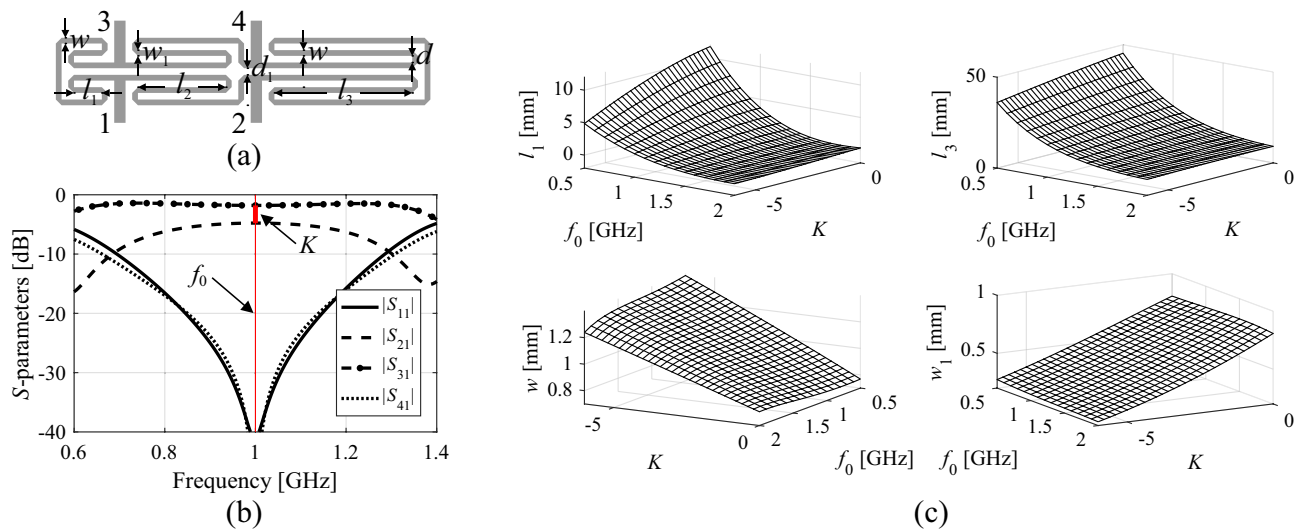


Figure 2. Miniaturized rat-race coupler⁵³: (a) circuit geometry, (b) typical S-parameter response, (c) two-dimensional surfaces corresponding to (approximately) optimum designs obtained for various target operating frequencies and power split ratios, here, shown for the selected geometry parameters.

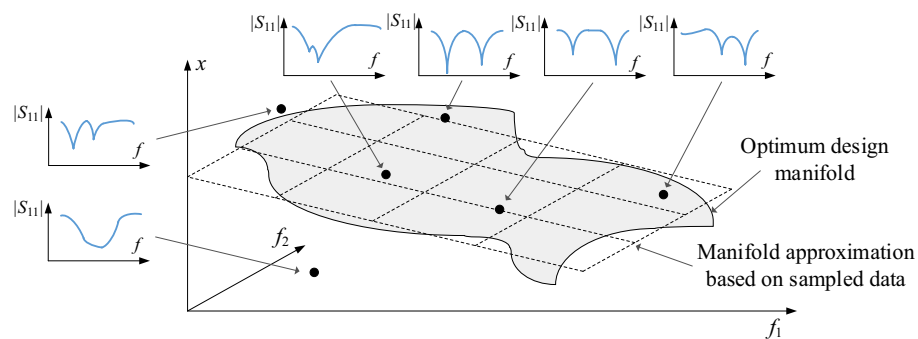


Figure 3. Conceptual illustration of the use of response features for identification of the promising regions of the parameter space. Shown are: optimum design manifold (here, demonstrated for a dual-band antenna optimized with respect to two operating frequencies f_1 and f_2), the vertical axis being one of antenna geometry parameters; black circles corresponding to random observables with their response features (here, antenna resonance allocation) extracted. Note that the observables are not optimized, therefore, they are generally away from the optimum design manifold; however, they can still be used to produce a rough approximation of the manifold by associating the response features with the objective space of f_1 and f_2 (cf. “[Identification of promising regions by inverse regression models](#)” section).

obtained this way can be employed as a predictor rendering a point in a reasonable proximity of the optimum design, corresponding to the target values of the performance figures relevant to the design problem at hand. A graphical illustration of this concept has been shown in Fig. 3. Its quantification will be discussed at length in “[Identification of promising regions by inverse regression models](#)” section.

Identification of promising regions by inverse regression models. As indicated in “[Response features for global search](#)” section, the response features can be effectively employed to identify the promising regions of the parameter space, i.e., the vicinity of the optimum design manifold. This section provides quantification of this concept by means of inverse regression surrogates. We start by introducing the necessary notation. Let $\mathbf{R}(\mathbf{x})$ denote the response (e.g., S-parameters vs. frequency) of the EM-simulation model of the structure at hand, where \mathbf{x} is a vector of designable variables.

The task is to find

$$\mathbf{x}^* = \mathbf{x}^*(\mathbf{F}) \in \arg \min_{\mathbf{x} \in X} U(\mathbf{x}, \mathbf{F}) \quad (1)$$

where $U(\mathbf{x}, \mathbf{F})$ is a scalar merit function to be minimized, $\mathbf{F} = [F_1 \dots F_N]^T$ is a target vector of design objectives (hereafter it will be referred to as the target objective vector), whereas X denotes the parameter space usually



delimited by lower and upper bounds on design variables. Two examples follow. If the goal is to minimize the reflection coefficient at the intended operating frequencies F_k of a multi-band antenna, the merit function could be defined in a minimax sense as

$$U(\mathbf{x}, \mathbf{F}) = \max \{ |S_{11}(\mathbf{x}, F_1)|, \dots, |S_{11}(\mathbf{x}, F_N)| \} \quad (2)$$

If the goal is to design a coupler for a particular power split ratio $K = |S_{21}(\mathbf{x}, \mathbf{F})| - |S_{31}(\mathbf{x}, \mathbf{F})|$ at the operating frequency f_0 , then one may define $F_1 = f_0, F_2 = K$ and

$$U(\mathbf{x}, \mathbf{F}) = \max \{ |S_{11}(\mathbf{x}, F_1)|, |S_{41}(\mathbf{x}, F_1)| \} + \beta (F_2 - (|S_{21}(\mathbf{x}, F_1)| - |S_{31}(\mathbf{x}, F_1)|))^2 \quad (3)$$

Here, the primary objective is the minimization of the reflection and isolation at the operating frequency. The required power split ($F_2 = K$) is enforced by adding a regularization factor (second term in (3)) with β being a penalty coefficient.

The optimum design manifold $O(F)$ is the set of points of the form $\mathbf{x}^*(F)$ for all $F \in F$, where F is the objective space pertinent to the design task (e.g., determined by the lower and upper bounds for the objectives F_k). We also denote as \mathbf{F}^* a particular target objective vector being a target for the specific design task considered for the structure at hand.

Let $\mathbf{p}(\mathbf{x}) = [p_1(\mathbf{x}) \dots p_L(\mathbf{x})]^T$ be the vector of response features extracted from the simulated characteristics of the structure under design at \mathbf{x} . Further, let $s_k(\mathbf{p}(\mathbf{x}))$, $k = 1, \dots, N_T$, be the scalar functions mapping the features into the space F_T of auxiliary objective vectors $F_T \in F_T$. Generally, F_T coincides with the objective space F (in particular, $N_T = N$); however, in some instances, they may be different (cf. Section "Demonstration case studies") section, for example, when the assessment of the design utility (for the purpose of yielding the initial design for further refinement) requires a larger number of scalar quantifiers than those contained in the target objective vectors F .

The response features, aggregated in the vector \mathbf{p} , are defined for specific circuit characteristics based on their shape and by taking into account the design objectives (operating parameters). The reason for employing the function $s_k(\mathbf{p}(\mathbf{x}))$ is that the response features and the design objectives may be in some situations identical, yet, this is not always the case. For example, if the design objective is to allocate the antenna resonance at a required frequency f_0 , then the design objective F and the response feature f_0 are identical, i.e., we have $F = f_0$. Still, in many cases, the relationship between the response features and the design objectives is not so straightforward. Consider the microwave coupler, which is to operate at a target center frequency f_0 . For this structure, the response features are defined as the minima of $S_{11}(\mathbf{x}, \mathbf{F})$ and $S_{41}(\mathbf{x}, \mathbf{F})$ characteristics, and they need to be controlled by the optimization process. We have $\mathbf{p} = [f_{0,S11}, f_{0,S41}]^T$, where $f_{0,S11}$ and $f_{0,S41}$ are the frequencies corresponding to minima of the reflection and isolation characteristics. However, the relevant design objective in this case is the circuit operating frequency f_0 , i.e., $F = f_0$. Therefore, a function $s(\mathbf{p})$ allows us to map the space of the response features into the objective space if necessary. In this case it might be defined as $s_1(\mathbf{p}(\mathbf{x})) = (f_{0,S11} + f_{0,S41})/2$, i.e., the operating frequency of the circuit is estimated as the average of the frequencies of the said minima.

The functions s_k are defined by the designer and determine the approximate allocation of the design \mathbf{x} in terms of its performance figure values in relation to $\mathbf{x}^*(F)$. We also denote $\mathbf{s}(\mathbf{p}) = [s_1(\mathbf{p}) \dots s_{N_T}(\mathbf{p})]^T$. For simplicity of notation, we omit the dependence of the vector of features \mathbf{p} on the parameter vector \mathbf{x} . Further, let $F_T = [F_{T,1} \dots F_{T,N_T}]^T$ be a target (auxiliary) objective vector, also user-defined, determining the design quality in terms of response features. The corresponding merit function $E(\mathbf{p})$ quantifying the design quality in the above sense is $E(\mathbf{p}) = \|\mathbf{s}(\mathbf{p}) - F_T\|^2$. We will denote by F_T^* the particular auxiliary vector corresponding to the target objective vector F^* (pertinent to the specific design task).

Going back to the example of a multi-band antenna considered before and the merit function (2), the appropriate choice of the response features would be $\mathbf{p} = [f_{0,1} \dots f_{0,N}, l_1 \dots l_N]^T$, where $f_{0,k}$ and l_k are the frequency and the $|S_{11}(\mathbf{x}, \mathbf{F})|$ -value at the k th antenna resonance. In this case, one may define $s_k(\mathbf{p}) = f_{0,k}$ if the assessment of the allocation of a particular design is to be based on the distance between the actual and the target operating frequencies. Then, the vector F_T coincides with the target objective vector F (in particular, $N_T = N$), and $E(\mathbf{p}) = \|[f_{0,1} \dots f_{0,N}]^T - F\|^2$.

In the case of the coupler (cf. (3)), the response feature vector may be defined as $\mathbf{p} = [f_{0,S11}, f_{0,S41}, l_{S11}, l_{S21}, l_{S31}, l_{S41}]^T$, where $f_{0,S11}$ and $f_{0,S41}$ are the frequencies corresponding to minima of the reflection and isolation characteristics with l_{S11} and l_{S41} being the corresponding levels, whereas l_{S21} and l_{S31} are the levels of the transmission characteristics at the frequency $(f_{0,S11} + f_{0,S41})/2$. Here, the functions s_k would be defined as $s_1(\mathbf{p}) = (f_{0,S11} + f_{0,S41})/2$, and $s_2(\mathbf{p}) = l_{S21} - l_{S31}$. Again, the vector F_T coincides with F and $E(\mathbf{p}) = \|[f_{0,S11}, f_{0,S41}, l_{S21}, l_{S31}]^T - F\|^2$. As mentioned before, although in general, $N_T = N$ and $F_T = F$, this may not be the case for specific situations (cf. Section "Demonstration case studies").

The aforementioned concepts will be employed to identify the promising regions of the parameter space X , and delimited using the lower and upper bounds \mathbf{l} and \mathbf{u} , respectively. Let $\mathbf{x}_r^{(j)} = [x_{r,1}^{(j)} \dots x_{r,n}^{(j)}]^T$, $j = 1, \dots, N_1$, be the set of random points allocated (preferably uniformly) in X . With each point, we associate the corresponding feature vector $\mathbf{p}(\mathbf{x}_r^{(j)})$ and $E^{(j)} = E(\mathbf{p}(\mathbf{x}_r^{(j)}))$. Here, the dependence of the response feature on the particular random observable vector $\mathbf{x}_r^{(j)}$ has been explicitly given to indicate that the value of the merit function $E^{(j)}$ corresponds to the particular vector $\mathbf{x}_r^{(j)}$, in the following, for simplicity of notation, this dependence will be omitted and we will simply use $E^{(j)}$. If some of the pre-defined feature points do not exist (e.g., one or more of the multi-band antenna resonances is not clearly distinguished) the corresponding sample is removed from the set. Without loss of generality, we assume that the observables $\mathbf{x}_r^{(j)}$ are ordered so that $E^{(1)} \leq E^{(2)} \leq \dots \leq E^{(N_1)}$. In particular, the first sample point represents the design that is the closest to the auxiliary objective vector F_T in the sense described in the previous paragraphs.

In the next step, we use $N_2 \leq N_1$ of the best sample points to establish an inverse regression surrogate $L(\mathbf{s}(\mathbf{p}))$ defined over F_T with the values in X , which will be used to yield an approximation of the design that is as close

as possible to the target one, i.e., the design that minimizes $E(\mathbf{p})$. Here, the surrogate is assumed to be a linear function of \mathbf{s} due to the fact that N_1 is generally small, and practical allocation of the observables may be far from uniform. Under these circumstances, linear approximation ensures better extrapolation capabilities than interpolative models (e.g., kriging³¹). Thus, we have

$$\mathbf{L}(\mathbf{s}(\mathbf{p})) = \begin{bmatrix} l_{0,1} + \sum_{i=1}^{N_T} l_{i,1} s_k(\mathbf{p}) \\ \vdots \\ l_{0,n} + \sum_{i=1}^{N_T} l_{i,n} s_k(\mathbf{p}) \end{bmatrix} \quad (4)$$

The model coefficients are obtained by solving the regression problems

$$\mathbf{L}(\mathbf{s}(\mathbf{x}_r^{(j)})) = \mathbf{x}_r^{(j)}, \quad j = 1, \dots, N_2 \quad (5)$$

where the dependence of the regression surrogate $\mathbf{L}(\mathbf{s}(\mathbf{x}_r^{(j)}))$ on the specific random observable vector $\mathbf{x}_r^{(j)}$ has been explicitly shown. In other words, the affine plane corresponding to the surrogate model output provides the best possible approximation of the observable set in the least square sense. In practice, we intend to put more emphasis on the high-quality samples, i.e., those characterized by small values of $E^{(j)}$. Toward this end, let us define the weights $w_j = 1/E^{(j)}$ and the weight matrix $\mathbf{W} = \text{diag}(w_1, \dots, w_{N_2})$. The original regression problem (5) is then replaced by the weighted one of the form

$$\mathbf{A}^T \mathbf{W} \mathbf{A} \mathbf{L} = \mathbf{A}^T \mathbf{W} \mathbf{X} \quad (6)$$

where

$$\mathbf{L} = \begin{bmatrix} l_{0,1} & l_{1,1} & \cdots & l_{N_T,1} \\ \vdots & \vdots & \ddots & \vdots \\ l_{0,n} & l_{1,n} & \cdots & l_{N_T,n} \end{bmatrix}^T \quad (7)$$

$$\mathbf{A} = \begin{bmatrix} 1 & s_1(\mathbf{x}_r^{(1)}) & \cdots & s_{N_T}(\mathbf{x}_r^{(1)}) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & s_1(\mathbf{x}_r^{(N_2)}) & \cdots & s_{N_T}(\mathbf{x}_r^{(N_2)}) \end{bmatrix} \quad (8)$$

and

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_r^{(1)T} \\ \vdots \\ \mathbf{x}_r^{(N_2)T} \end{bmatrix} \quad (9)$$

The least-square solution to (6) is given as

$$\mathbf{L}_L = (\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W} \mathbf{X} \quad (10)$$

The inverse surrogate can be used to produce the design \mathbf{x}_r^* that provides the best approximation of the target objective vector \mathbf{F}_T^* as follows

$$\mathbf{x}_r^* = \mathbf{L} \left(\begin{bmatrix} 1 & (\mathbf{F}_T^*)^T \end{bmatrix}^T \right) = \mathbf{L}_L^T \begin{bmatrix} 1 & (\mathbf{F}_T^*)^T \end{bmatrix}^T \quad (11)$$

Provided this design is of sufficient quality, the vector \mathbf{x}_r^* may be used as a starting point for local optimization. Otherwise, it might be further refined by supplementing the existing observable set with the infill samples, rebuilding the inverse model, and making another prediction according to (11). “[Optimization framework](#)” section provides the details of the optimization framework that implements the aforementioned iterative procedure.

The fundamental advantage of constructing the feature-based inverse surrogate is only somewhat nonlinear dependence between the feature point coordinates and the geometry parameters of the structure under design (cf. Fig. 2). Owing to this, even a relatively small number of observables $\mathbf{x}_r^{(j)}$ may be sufficient to identify a promising region of the parameter space, i.e., containing the optimum design $\mathbf{x}^*(\mathbf{F}^*)$.

Design refinement by local search. Local optimization is performed using the trust-region (TR) gradient search algorithm with numerical derivatives⁵⁷. The algorithm produces a sequence $\mathbf{x}^{(i)}$, $i=0, 1, \dots$, of approximations to \mathbf{x}^* as

$$\mathbf{x}^{(i+1)} = \arg \min_{\mathbf{x}; -d^{(i)} \leq \mathbf{x} - \mathbf{x}^{(i)} \leq d^{(i)}} U(\mathbf{G}^{(i)}(\mathbf{x}), \mathbf{F}^*) \quad (12)$$

The model $\mathbf{G}^{(i)}(\mathbf{x})$ locally approximates $\mathbf{R}(\mathbf{x})$ at $\mathbf{x}^{(i)}$; here, we employ a first-order Taylor expansion given as

1. Set iteration index $i = 1$;
2. Assign N_1 random observables $\mathbf{x}_r^{(j)}$, $j = 1, \dots, N_1$ (here, using LHS);
3. For all samples, calculate $\mathbf{p}(\mathbf{x}_r^{(j)})$ and $E^{(j)}$; samples for which the prescribed feature points do not exist are removed from the pool;
4. Use N_2 best samples to construct the inverse surrogate $\mathbf{L}(\mathbf{s}(\mathbf{p}(\mathbf{x})))$ (cf. (4));
5. Obtain the design \mathbf{x}_T^* providing the best approximation of \mathbf{F}_T^* (cf. (11));
6. Assign \mathbf{x}_{best} as $\mathbf{x}_{best} = \arg \min \{ \mathbf{x} \in \{ \mathbf{x}_T^*, \mathbf{x}_r^{(1)}, \dots, \mathbf{x}_r^{(N_1)} \} : U(\mathbf{x}, \mathbf{F}^*) \}$;
7. **If** $U(\mathbf{x}_{best}, \mathbf{F}^*) < U_{local}$ OR $N_1 + i + (i-1)N_3 > N_{max_global}$ **then** go to 11; **else** go to 8;
8. **If** $E(\mathbf{p}(\mathbf{x}_{best})) < E_{reduce}$ **then** reduce the parameter space as $\mathbf{l} \leftarrow a\mathbf{l} + (1-a)\mathbf{x}_{best}$, $\mathbf{u} \leftarrow a\mathbf{u} + (1-a)\mathbf{x}_{best}$, with $a = 1/a_{red}$; **else** go to 8;
9. Assign N_3 infill random observables within the interval $[\mathbf{l}, \mathbf{u}]$ (here, using LHS) and construct the updated sample set $\{\mathbf{x}_r^{(j)}\}$ by concatenating the infill points and the samples already existing in $[\mathbf{l}, \mathbf{u}]$;
10. Set $i = i + 1$ and go to 3;
11. Perform local search (12) starting from $\mathbf{x}^{(0)} = \mathbf{x}_{best}$ (cf. Section 2.3); set $i = 0$; termination condition based on convergence in argument $\|\mathbf{x}^{(i+1)} - \mathbf{x}^{(i)}\| < \varepsilon$ (user-defined) and maximum computational budget N_{max_local} .

Figure 4. The workflow of the proposed procedure for quasi-global optimization of high-frequency structures.

$$\mathbf{G}^{(i)}(\mathbf{x}) = \mathbf{R}(\mathbf{x}^{(i)}) + \mathbf{J}_R(\mathbf{x}^{(i)}) \cdot (\mathbf{x} - \mathbf{x}^{(i)}) \quad (13)$$

In (13), the Jacobian $\mathbf{J}_R(\mathbf{x})$ estimation requires finite differentiation. The trust region is delimited by $\mathbf{x}^{(i)} - \mathbf{d}^{(i)} \leq \mathbf{x} \leq \mathbf{x}^{(i)} + \mathbf{d}^{(i)}$ (i.e., its lower and upper bounds), with the inequalities interpreted component-wise. The size $\mathbf{d}^{(i)}$ of the search region is established with conformance to the standard TR algorithm setup⁵⁷. The adopted definition of the trust region permits handling variables featuring considerably different ranges, as $\mathbf{d}^{(0)}$ (i.e., the size of the initial TR region) becomes proportional to the parameter space size $\mathbf{u} - \mathbf{l}$, thus eliminating the necessity of scaling the antenna parameters.

Upon each successful iteration, the TR algorithm (12) adjusts the local model (i.e., re-evaluates the Jacobian), hence, its cost equals $n + 1$ EM simulation of the component under design. The local search process may be accelerated using the expedited versions of the TR algorithm involving sparse Jacobian updates, as described in⁵⁸ or⁵⁹. Here, after the first iteration, we simply replace finite differentiation by a rank-one Broyden formula⁶⁰, which is justified by the availability of good-quality initial design found by the global search stage.

Optimization framework. This section describes the entire optimization framework involving the components discussed in “Response features for global search” through “Design refinement by local search” sections. The control parameters of the algorithm (some of which have been already mentioned before) include:

- N_1 —the number of initial random samples $\mathbf{x}_r^{(j)}$, here, using LHS⁶¹ (cf. “Identification of promising regions by inverse regression models” section);
- N_2 —the number of best samples selected to construct the inverse surrogate;
- N_3 —the number of infill samples added to the sample pool per iteration;
- N_{max_global} —maximum number of EM evaluations for the global search stage;
- N_{max_local} —maximum number of EM evaluations for the local search stage;
- \mathbf{x}_{best} —best design found so far in the process;
- U_{local} —threshold value for $U(\mathbf{x}_{best}, \mathbf{F}^*)$ to enter the local search stage;
- E_{reduce} —threshold value for $E(\mathbf{p}(\mathbf{x}_{best}))$ to reduce the sampling region (infill sampling);
- a_{red} —multiplication factor for sampling region reduction.

The operation of the algorithm is shown in Fig. 4. In Fig. 4, Steps 3 through 10 pertain to the global search step, whereas Step 11 corresponds to the local refinement. Similarly as in “Identification of promising regions by inverse regression models” section, \mathbf{F}^* and \mathbf{F}_T^* denote the target objective and auxiliary objective vectors; \mathbf{l} and \mathbf{u} are lower and upper bounds determining the parameter space.

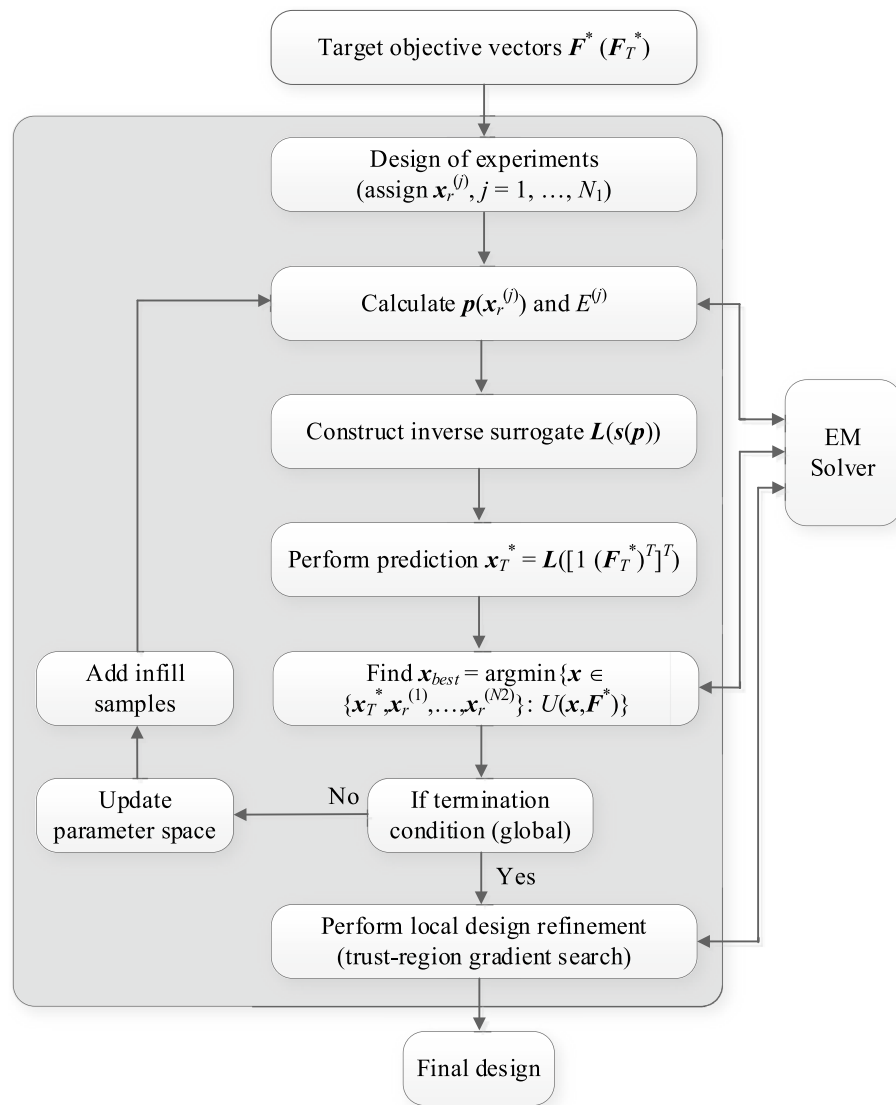


Figure 5. Flow diagram of the proposed globalized optimization framework involving response features and inverse surrogates.

Note that in the above algorithm, global search is followed by the local refinement, which concludes the process. In principle, the procedure can be iterated; however, in order to demonstrate the benefits of involving the response features into the process, for all demonstration examples considered in “[Demonstration case studies](#)” section, only one pass is executed. The number $N_1 + i + (i - 1)N_3$ in Step 7 is the overall number of EM simulations executed at this stage of the process, including: (i) the initial sampling, (ii) system evaluation at the surrogate-predicted design x_T^* , and (iii) the infill sampling. If, upon removing samples for which the prescribed feature points do not exist, the remaining number of samples is smaller than N_2 , the surrogate is constructed if possible, otherwise (i.e., if the number of samples is smaller than $N_1 + 1$), the surrogate is not constructed in a given iteration and $x_{best} = \operatorname{argmin} \{x \in \{x_r^{(1)}, \dots, x_r^{(N_2)}\} : U(x, F^*)\}$. A graphical illustration of the optimization process has been shown in Fig. 5 in the form of a flow diagram.

Demonstration case studies

This section demonstrates the operation and performance of the proposed optimization technique using three high-frequency structures: a dual-band dipole antenna, a miniaturized rat-race coupler, and a two-section compact impedance matching transformer. In all cases, the results obtained using the algorithm of “[Globalized search using inverse regression surrogates and response features](#)” section have been compared to a representative nature-inspired method (here, particle swarm optimizer, PSO⁶²), in terms of computational cost and design quality. Furthermore, the verification structures have been optimized using gradient-based search executed

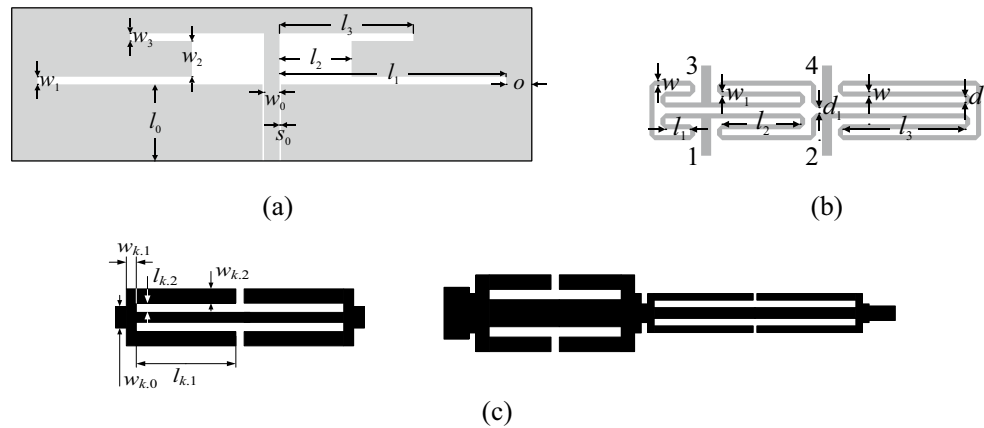


Figure 6. Verification structures: (a) uniplanar dual-band dipole antenna⁶³, (b) miniaturized rat-race coupler (RRC)⁶⁴, (c) compact two-section impedance matching transformer: CMRC cell (left) and complete transformer structure (right)⁶⁵; k refers to a section index.

from random starting points. Less-than-perfect success rate of the latter indicates that globalized optimization is indeed required for the considered design tasks, except for the last structure.

Verification case studies. Figure 6 shows the verification devices utilized to demonstrate and validate the proposed optimization approach. These include:

- A uniplanar dual-band dipole antenna shown in Fig. 6a (Structure I)⁶³;
- A miniaturized rat-race coupler (RRC) shown in Fig. 6b (Structure II)⁶⁴;
- A compact two-section CMRC-based impedance matching transformer shown in Fig. 6c (Structure III)⁶⁵.

The experimental validation of the high-frequency structures utilized as verification case studies has been provided in their respective source papers^{63–65}. Moreover, Structure I and II have been experimentally verified in our previous work, e.g.,^{66,67}. Thus, the experimental validation has not been provided as being immaterial to the scope of the paper. For additional verification, Fig. 7 presents the families of the responses of Structures I through III corresponding to various model fidelities. In our work, the model fidelity is set using the parameter LPW (lines per wavelength) controlling the mesh density in CST Microwave Studio, which is utilized for performing the EM simulations.

Table 1 shows the details concerning all structures, including material parameters of the dielectric substrate, designable parameters, as well as performance specifications. The objective functions are determined in a similar way as delineated in “Identification of promising regions by inverse regression models”. In particular, the objective function for Structure I is defined according to (2) with $F_1 = 2.5$ GHz, and $F_2 = 4.8$ GHz. The merit function for Structure II follows (3) with $F_1 = 1$ GHz and $F_2 = 0$ dB. In the case of Structure III, the objective function takes the form $U(\mathbf{x}, \mathbf{F}) = \max\{F_1 \leq f \leq F_2 : |S_{11}(\mathbf{x}, f)|\}$. It should be noted that the parameter spaces (last row of Table 1) are large: the average upper-to-lower bound ratio is around five, with the maximum values reaching ten for Structures I and III. The verification examples are described by six parameters (Structure I and II), and ten parameters (Structure III). In the future work, for a more comprehensive assessment of the performance of our approach, higher- and lower-dimensional examples, as well as structures characterized by multimodality will be taken into account.

Experimental setup and results. The three verification structures described in “Verification case studies” section have been optimized using the proposed algorithm of “Globalized search using inverse regression surrogates and response features” section, the PSO algorithm, as well as gradient-based search (benchmark methods). In all cases, ten independent runs have been executed, and the results’ statistic have been performed, in particular, the average objective function value, as well as its standard deviation as a measure of repeatability of solutions. In the case of gradient-based optimization, the search process was executed using ten random initial designs.

The setup of all algorithms was as indicated below, where n stands for the dimensionality of the parameter space (the meaning of the control parameters of the algorithm of “Globalized search using inverse regression surrogates and response features” section can be found in “Optimization framework” section):

- Proposed algorithm (Section Globalized search using inverse regression surrogates and response features): $N_1 = N_2 = 2n$, $N_3 = n$, $N_{max_global} = N_{max_local} = 10n$, and $a_{red} = 1.5$. The threshold values U_{local} and E_{reduce} are problem dependent and set to correspond to relatively mild conditions; we set $U_{local} = -5$ dB and $E_{reduce} = 0.5$ (Structure I), $U_{local} = 10$ and $E_{reduce} = 0.3$ (Structure II), $U_{local} = -15$ and $E_{reduce} = 0.2$ (Structure III);

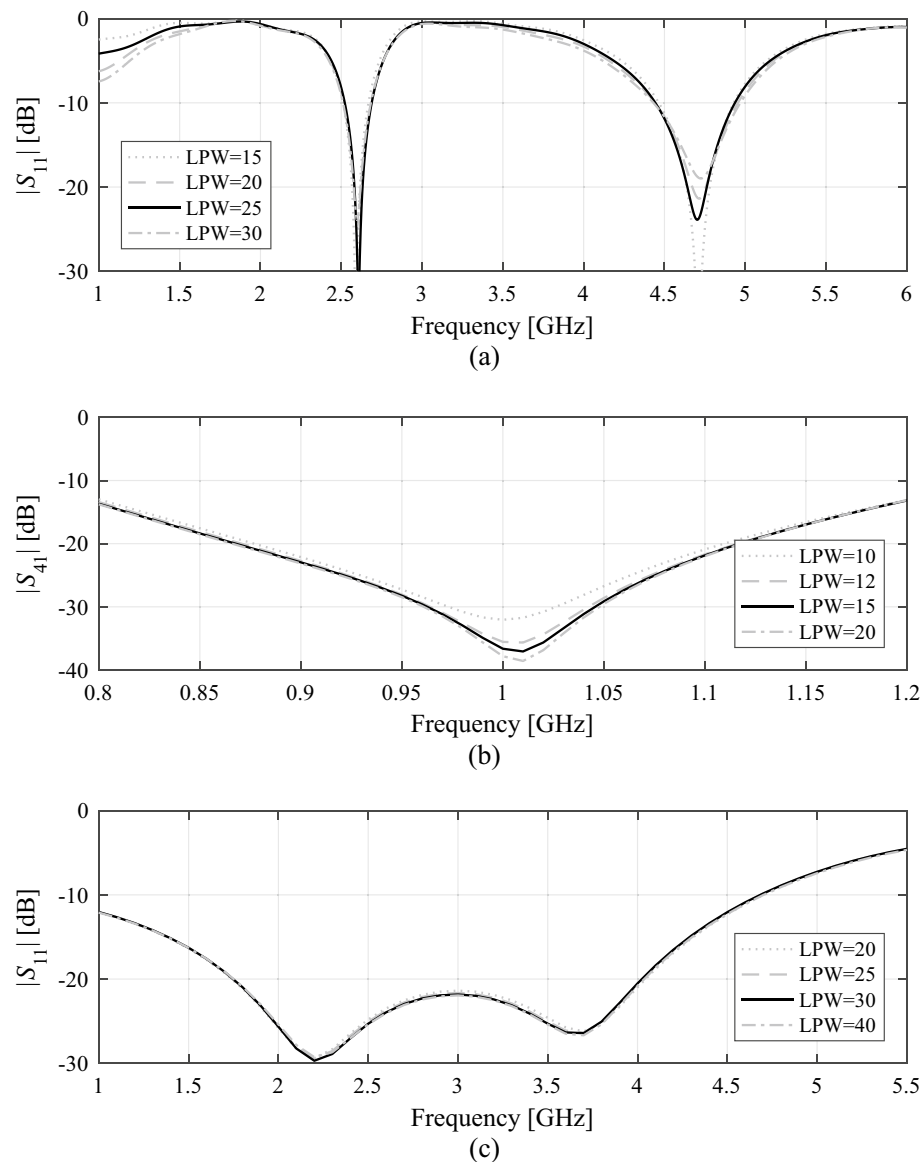


Figure 7. Grid convergence plots for the most representative responses of: (a) Structure I, (b) Structure II, and (c) Structure III for various values of LPW (lines per wavelength) parameter controlling the mesh density in CST Microwave Studio. Black lines pertain to the LPW values used in numerical experiments: 25, 15, and 30 for Structures I through III, respectively.

- PSO: swarm size $N_s = 10$, maximum number of iterations $k_{\max} = 50$, weight coefficients for velocity updating: $\chi = 0.73$, $c_1 = c_2 = 2.05$, cf.⁶²;
- Gradient-based algorithm: trust-region framework (cf.⁵⁷) with finite-differentiation-based sensitivity updates; termination conditions: convergence in argument $\|\mathbf{x}^{(i+1)} - \mathbf{x}^{(i)}\| < 10^{-3}$ or reduction of the trust region size beyond 10^{-3} .

It should be noted that, despite a formally large number of the control parameters of the algorithm of “Globalized search using inverse regression surrogates and response features” section, the actual setup is simple: the number of random samples and the infill points is related to the parameter space dimensionality, whereas the threshold values are set to correspond to rather mild conditions concerning the design quality and are not critical. This is because the global search stage is only supposed to find the designs that are allocated in the region of attraction of the optimum ones (from the point of view of local search). The numerical values can be easily established by the user familiar with a given design task and the structure at hand.

For the PSO algorithm, the setup is very restrictive, which is because of high computational costs of massive EM analyses of the considered structures, about two, three and six minutes per simulation for Structures I through III, respectively. As a result, the running time of the algorithm is as high as one to two days even for the prescribed swarm size and the number of iterations. Another reason for choosing this PSO setup is to verify

	Case study		
	Structure I	Structure II	Structure III*
Substrate	$\epsilon_r = 3.5$ $h = 0.76$ mm	$\epsilon_r = 3.5$ $h = 0.76$ mm	$\epsilon_r = 3.5$ $h = 0.76$ mm
Design parameters	$\mathbf{x} = [l_1 \ l_2 \ l_3 \ w_1 \ w_2 \ w_3]^T$	$\mathbf{x} = [l_1 \ l_2 \ l_3 \ d \ w \ w_1]^T$	$\mathbf{x} = [l_{1,1} \ l_{1,2} \ w_{1,1} \ w_{1,2} \ w_{1,0} \ l_{2,1} \ w_{2,1} \ w_{2,2} \ w_{2,0}]^T$
Other parameters	$l_0 = 30, w_0 = 3, s_0 = 0.15, \sigma = 5$	$d_1 = d + w - w_1 , d = 1.0, w_0 = 1.7, \text{ and } l_0 = 15$	$w_{in} = 1.7, w_{out} = 0.4$
Design specifications	Minimize reflection coefficient $ S_{11} $ at two operating frequencies, 2.5 GHz and 4.8 GHz	Ensure 3 dB power split, $ S_{31} - S_{21} = 3$ dB, at the operating frequency $f_0 = 1$ GHz, and minimize $\max\{ S_{11} , S_{41} \}$ at f_0	Minimize the maximum in-band reflection coefficient $ S_{11} $ within the range 1.8 GHz to 4.0 GHz
LPW	25	15	30
Simulation time [#]	60 s	160 s	120 s
Parameter space	$\mathbf{l} = [15 \ 5 \ 5 \ 0.2 \ 1.5 \ 0.5]^T$ $\mathbf{u} = [50 \ 15 \ 30 \ 0.6 \ 5.0 \ 5.0]^T$	$\mathbf{l} = [1.0 \ 5.0 \ 10.0 \ 0.2 \ 0.5 \ 0.2]^T$ $\mathbf{u} = [6.0 \ 15.0 \ 25.0 \ 1.2 \ 1.5 \ 1.2]^T$	$\mathbf{l} = [2.0 \ 0.1 \ 0.5 \ 0.1 \ 0.2 \ 2.0 \ 0.1 \ 0.1 \ 0.1 \ 0.2]^T$ $\mathbf{u} = [5.5 \ 0.5 \ 1.1 \ 0.8 \ 1.1 \ 4.5 \ 1.0 \ 0.6 \ 0.6 \ 1.3]^T$

Table 1. Verification structures. *Input and output line widths, w_{in} and w_{out} , set to ensure 50-Ω input and 100-Ω output impedance. #EM simulations have been performed on Intel Xeon 2.1 GHz machine with 64 GB RAM.

Optimization method	Algorithm of Section “Globalized search using inverse regression surrogates and response features” (this work)	PSO		Trust-region gradient-based algorithm
		20 iterations	50 iterations	
Average objective function value (dB)	-20.7	-14.8	-21.9	-12.7
Computational cost [§]	73.3	200	500	63.8
Success rate [#]	10/10	8/10	8/10	6/10

Table 2. Optimization results for Structure I. §The cost calculated as the number of EM simulations of the structure under design, averaged over ten algorithm runs. #Number of algorithms runs with the operating frequency allocated with sufficient accuracy, i.e., to satisfy $U(\mathbf{x}^*, \mathbf{F}) \leq U_{local}$.

Optimization method	Algorithm of Section “Globalized search using inverse regression surrogates and response features” (this work)	PSO		Trust-region gradient-based algorithm
		20 iterations	50 iterations	
Average objective function value (dB)	-36.1	-12.4	-25.8	-16.5
Computational cost [§]	59.6	200	500	75.4
Success rate [#]	10/10	6/10	9/10	5/10

Table 3. Optimization results for Structure II. §The cost calculated as the number of EM simulations of the structure under design, averaged over ten algorithm runs. #Number of algorithms runs with the operating frequency allocated with sufficient accuracy, i.e., to satisfy $U(\mathbf{x}^*, \mathbf{F}) \leq U_{local}$.

the quality of the solutions produced by the algorithm when its computational budget is limited to the levels comparable (although still larger) than for the proposed algorithm. The literature offers several approaches to tuning PSO algorithm to a specific design optimization case and to enhance its computational efficiency^{68–71}. One of the available methods involves scaling the size of the swarm and the number of iterations of the PSO optimizer based on the number of design variables of a design optimization task at hand⁶⁸. Due to the extremely high-cost of executing numerical experiments with the use of EM simulation tools, the analysis of the setup of the parameters of the PSO algorithm has not been conducted. This will be a guide for the future work.

Gradient-based search is used as a benchmark method in order to demonstrate that the considered design tasks are, in fact, multimodal. Therefore, it has been executed ten times from random initial design. For some of the runs, the local optima found are away from satisfactory designs, which means that the operating frequencies are not aligned with the assumed targets. For such runs, the algorithm outcome is marked as unsuccessful. The overall success rate is quantified as the relative number of runs, for which the mentioned alignment has been achieved.

The numerical results have been gathered in Tables 2, 3 and 4. The data therein includes the average objective function value, the computational cost of the optimization process calculated as the number of EM simulations of the structure at hand, as well as the success rate, i.e., the number of runs for which the algorithm found the design at which the operating parameters are sufficiently close to the target. Figures 8, 9 and 10 illustrate the responses of the considered structures obtained during the selected runs of the proposed algorithm.

Optimization method	Algorithm of Section "Globalized search using inverse regression surrogates and response features" (this work)	PSO		Trust-region gradient-based algorithm
		20 iterations	50 iterations	
Average objective function value (dB)	-21.3	-19.9	-20.7	-21.5
Computational cost [§]	63.2	200	500	85.8
Success rate [#]	10/10	10/10	10/10	10/10

Table 4. Optimization results for Structure III. [§]The cost calculated as the number of EM simulations of the structure under design, averaged over ten algorithm runs. [#]Number of algorithms runs with the operating frequency allocated with sufficient accuracy, i.e., to satisfy $U(\mathbf{x}^*, \mathbf{F}) \leq U_{local}$.

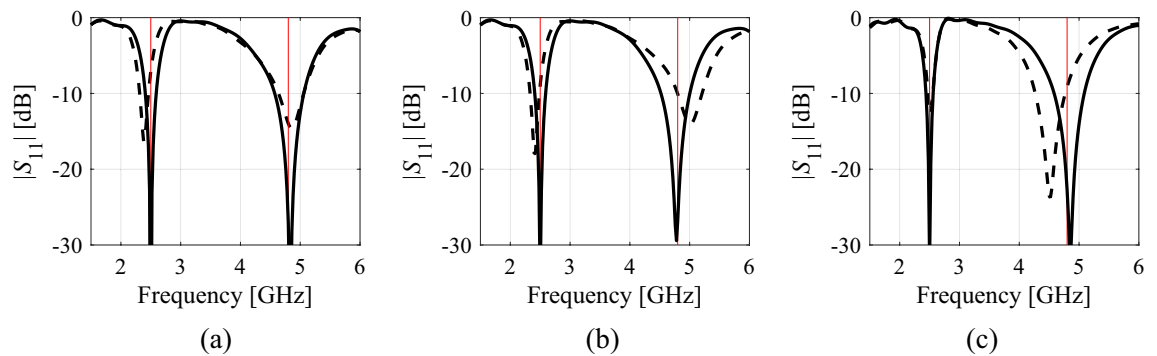


Figure 8. Structure I responses obtained in selected runs of the proposed algorithm: (a) run 1, (b) run 2, (c) run 3: design found after the global search stage (---), final design (—). Vertical lines mark the target operating frequencies, here 2,45 GHz and 5,3 GHz.

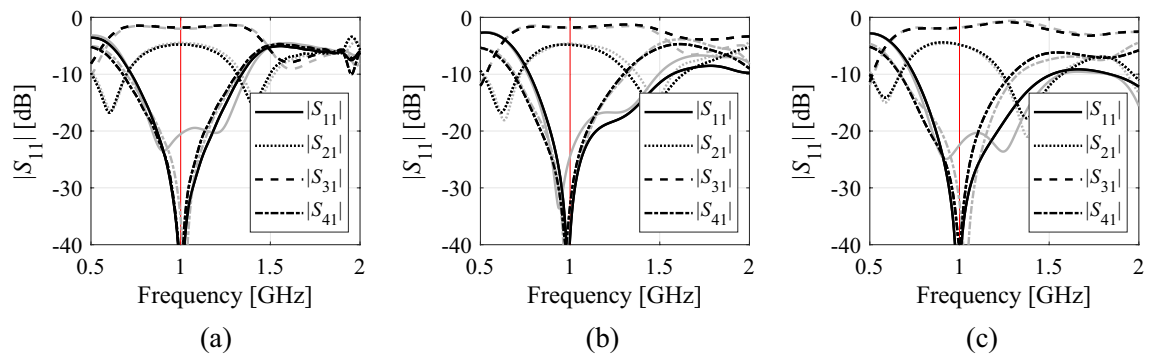


Figure 9. Structure II responses obtained in selected runs of the proposed algorithm: (a) run 1, (b) run 2, (c) run 3: design found after the global search stage (gray), final design (black). The vertical line marks the target operating frequency of 1 GHz.

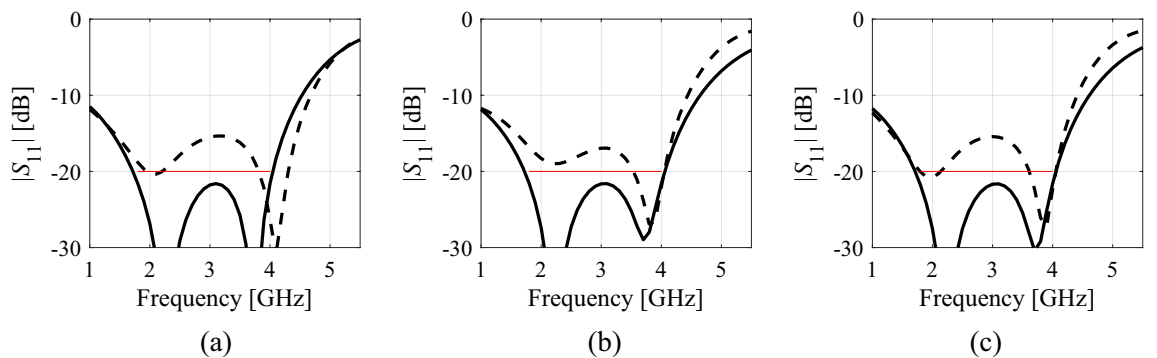


Figure 10. Structure III responses obtained in selected runs of the proposed algorithm: (a) run 1, (b) run 2, (c) run 3: design found after the global search stage (---), final design (—). The horizontal line marks the intended operating bandwidth, here, from 1.8 GHz and 4.0 GHz.

Discussion. The results presented in “[Experimental setup and results](#)” section allow us to draw several conclusions pertaining to the performance of the proposed optimization framework, including both its computational efficiency and reliability. These can be summarized as follows:

- The considered problems, except for Structure III, are indeed multimodal. It has been verified that the outcome of gradient-based optimization depends on the initial design, and the success rate is only sixty percent for Structure I and fifty percent for Structure II;
- Globalized search capability has been corroborated by 100-percent success rate (in terms of identifying the design that satisfies the performance specifications) for all the considered structures;
- The quality of designs rendered by the proposed procedure is competitive over the benchmark. In particular, the average objective function values are comparable or better than for PSO, and better than the gradient-based search (except Structure III, which turns out not to be a multimodal problem).
- The computational cost of our algorithm is comparable to that of gradient-based search, which is remarkable given the problem complexity and added global search capability. As a matter of fact, the average cost is even slightly lower for Structures II and III. The reason for excellent computational efficiency is that the global search stage is capable of yielding good initial designs as indicated in Figs. 8 through 10. Consequently, subsequent local tuning can be concluded in a smaller number of iterations than the local search carried out from a random initial design. The running time is further reduced by utilization of the Broyden update in the later stages of the process, rather than full finite differentiation, FD (recall that the cost of FD-based iterations is at least $n + 1$ EM analyses, with n being the parameter space dimensionality).
- Based on the results obtained for the PSO algorithm, the computational budget required to surpass the design quality obtained by the proposed algorithm can be estimated as being between 500 (for Structure I) and at least 1,000 objective function evaluations (for Structures II and III). This means that the computational cost of the proposed procedure is about an order of magnitude lower than the cost of nature-inspired population-based procedures.
- In order to further investigate the performance of PSO algorithm, for each verification structure, a single optimization run has been executed with the maximum number of iterations set to 200. The obtained objective function values have been equal to -20.4 dB, -33.6 dB and -21.0 dB for Structures I through III, whereas the computational cost of each run has been equal to 2000 EM simulations, which corresponds to the simulation time equal to 33 h, 88 h, and 66 h for Structures I through III, respectively. Note that the obtained objective function values are only better than for PSO executed with 500 function evaluations for Structure II and III. The reasons is poor repeatability of solutions of the PSO algorithm, meaning, that the results of individual runs may readily be worse than the average performance (evaluated over the number of runs).

At this point, one needs to reiterate that the excellent performance of the presented optimization framework comes with some limitations. In particular, the underlying assumption is that the response features are identifiable in the EM-simulated responses of the considered structure (e.g., resonances, bandwidths), at least for the designs that are of decent quality. This is generally the case for structures such as narrow- or multi-band antennas, microwave couplers, power dividers, and so on. Broadband and ultra-broadband components may be trickier, yet the definition and extraction of the response features is normally realized on case-to-case basis. On the other hand, the specific verification devices considered in this work feature distinct characteristics that are representative to a range of high-frequency passive components. This indicates a relative wide spectrum of possible applications of the proposed methodology.

Conclusion

This paper presented a computationally efficient approach to globalized optimization of electromagnetic computational models. Our methodology involves inverse regression models constructed at the level of characteristics points of the system responses. The latter are extracted to estimate the actual operating parameters of the structure at hand, and—upon being embedded into the inverse model—guide the optimization process towards more promising regions of the parameter space. The global search stage is followed by gradient-based parameter tuning accelerated by means of sparse sensitivity updates, carried out using a rank-one Broyden formula. The proposed technique capitalizes on weakly nonlinear relationship between the operating parameters (e.g., the resonant frequencies of an antenna) and geometry variables, as opposed to highly-nonlinear responses when considered in their entirety, i.e., frequency characteristics. Further efficacy improvements are due to a normally small number of operating conditions as compared to the dimensionality of the parameter space, which is explored by the employment of inverse models. The proposed algorithm has been comprehensively validated using three high-frequency structures, a dual-band antenna, a compact microstrip coupler, and a miniaturized impedance matching transformer. The results demonstrate global search capability with a remarkably low computational cost of a few dozens of EM analyses, which is comparable to the cost of gradient-based local optimization. The technique discussed in this paper can be useful for a rapid parameter tuning of high-frequency circuits in situations where globalized search is necessary (e.g., due to the lack of reasonably good initial design), whereas computational budget is severely limited. A limitation of the method is that response features being a foundation of the inverse model, have to be extracted from EM-simulation results, which requires problem-specific routines. The future work will be focused on generalization and automation of this step of the procedure.

Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Received: 16 September 2022; Accepted: 11 November 2022

Published online: 18 November 2022

References

- Ting, H. L., Hsu, S. K. & Wu, T. L. A novel and compact eight-port forward-wave directional coupler with arbitrary coupling level design using four-model control theory. *IEEE Trans. Microw. Theory Technol.* **65**(2), 467–475 (2017).
- Ullah, U. & Koziel, S. A broadband circularly polarized wide-slot antenna with a miniaturized footprint. *IEEE Antennas Wirel. Propag. Lett.* **17**(12), 2454–2458 (2018).
- Yazeen, P. S. M., Vinisha, C. V., Vandana, S., Suprava, M. & Nair, R. U. Electromagnetic performance analysis of graded dielectric inhomogeneous streamlined airborne radome. *IEEE Trans. Antennas Propag.* **65**(5), 2718–2723 (2017).
- Kim, S. & Nam, S. A compact and wideband linear array antenna with low mutual coupling. *IEEE Trans. Antennas Propag.* **67**(8), 5695–5699 (2019).
- Farooqui, M. F. & Kishk, A. 3-D-printed tunable circularly polarized microstrip patch antenna. *IEEE Antennas Wirel. Propag. Lett.* **18**(7), 1429–1432 (2019).
- Qian, J. F., Chen, F. C., Xiang, K. R. & Chu, Q. X. Resonator-loaded multi-band microstrip slot antennas with bidirectional radiation patterns. *IEEE Trans. Antennas Propag.* **67**(10), 6661–6666 (2019).
- Mener, S., Gillard, R. & Roy, L. A dual-band dual-circular-polarization antenna for Ka-band satellite communications. *IEEE Antennas Wirel. Propag. Lett.* **16**, 274–277 (2016).
- Conteras, A. *et al.* Compact fully uniplanar bandstop filter based on slow-wave multimodal CPW resonators. *IEEE Microw. Wirel. Comput. Lett.* **28**(9), 780–782 (2018).
- Tseng, C. H. & Chang, C. L. A rigorous design methodology for compact planar branch-line and rat-race couplers with asymmetrical T-structures. *IEEE Trans. Microw. Theory Technol.* **60**(7), 2085–2092 (2012).
- Wang, J. *et al.* An implantable and conformal antenna for wireless capsule endoscopy. *IEEE Antennas Wirel. Propag. Lett.* **17**(7), 1153–1157 (2018).
- Sommer, A., Farle, O. & Dyczij-Edlinger, R. Certified dual-corrected radiation patterns of phased antenna arrays by offline-online order reduction of finite-element models. *J. Comput. Phys.* **299**, 22–44 (2015).
- Etmiman, A. & Moghaddam, M. Electromagnetic imaging of dielectric objects using a multidirectional-search-based simulated annealing. *IEEE J. Multiscale Multiphys. Comput. Technol.* **3**, 167–175 (2018).
- Zhang, H. & Zeng, F. A Fibonacci branch search (FBS)-based optimization algorithm for enhanced nulling level control adaptive beamforming technique. *IEEE Access* **7**, 160800–160818 (2019).
- Nicholson, G. L. & Lancaster, M. J. Coupling matrix synthesis of cross-coupled microwave filters using a hybrid optimization algorithm. *IET Microw. Antennas Propag.* **3**(6), 950–958 (2009).
- Zhao, W. J., Liu, E. X., Wang, B., Gao, S. P. & Png, C. E. Differential evolutionary optimization of an equivalent dipole model for electromagnetic emission analysis. *IEEE Trans. Electromagn. Comput.* **60**(6), 1635–1639 (2018).
- Darvish, A., Zakeri, B. & Radkani, N. An optimized hybrid convolutional perfectly matched layer for efficient absorption of electromagnetic waves. *J. Comput. Phys.* **356**, 31–45 (2018).
- Goudos, S. K. & Sahalos, J. N. Pareto optimal microwave filter design using multiobjective differential evolution. *IEEE Trans. Antennas Propag.* **58**(1), 132–144 (2010).
- Mahaboob, S., Ajithan, S. K. & Jayaraman, S. Optimal design of shunt active power filter for power quality enhancement using predator-prey based firefly optimization. *Swarm Evol. Comput.* **44**, 522–533 (2019).
- Yang, S. H. & Kiang, J. F. Optimization of sparse linear arrays using harmony search algorithms. *IEEE Trans. Antennas Propag.* **63**(11), 4732–4738 (2015).
- Li, X. & Luk, K. M. The grey wolf optimizer and its applications in electromagnetics. *IEEE Trans. Antennas Propag.* **6**, 66 (2019).
- Ram, G., Mandal, D., Kar, R. & Ghoshal, S. P. Cat swarm optimization as applied to time-modulated concentric circular antenna array: analysis and comparison with other stochastic optimization methods. *IEEE Trans. Antennas Propag.* **63**(9), 4180–4183 (2015).
- Chen, X., Tianfield, H. & Li, K. Self-adaptive differential artificial bee colony algorithm for global optimization problems. *Swarm Evol. Comput.* **45**, 70–91 (2019).
- Rahi, K. H., Singh, H. K. & Ray, T. Feasibility-ratio based sequencing for computationally efficient constrained optimization. *Swarm Evol. Comput.* **62**, 100850 (2021).
- Bayraktar, Z., Komurcu, M., Bossard, J. A. & Werner, D. H. The wind driven optimization technique and its application in electromagnetics. *IEEE Trans. Antennas Propag.* **61**(5), 2745–2757 (2013).
- Al-Azza, A. A., Al-Jodah, A. A. & Harackiewicz, F. J. Spider monkey optimization: a novel technique for antenna optimization. *IEEE Antennas Wirel. Propag. Lett.* **15**, 1016–1019 (2016).
- Darvish, A. & Ebrahimzadeh, A. Improved fruit-fly optimization algorithm and its applications in antenna arrays synthesis. *IEEE Trans. Antennas Propag.* **66**(4), 1756–1766 (2018).
- Peherstorfer, B., Kramer, B. & Willcox, K. Combining multiple surrogate models to accelerate failure probability estimation with expensive high-fidelity models. *J. Comput. Phys.* **341**, 61–75 (2017).
- Couckuyt, I., Declercq, F., Dhaene, T., Rogier, H. & Knockaert, L. Surrogate-based infill optimization applied to electromagnetic problems. *Int. J. RF Microw. Comput. Aided Eng.* **20**(5), 492–501 (2010).
- Dong, J., Qin, W. & Wang, M. Fast multi-objective optimization of multi-parameter antenna structures based on improved BPNN surrogate model. *IEEE Access* **7**, 77692–77701 (2019).
- Jones, D. R., Schonlau, M. & Welch, W. J. Efficient global optimization of expensive black-box functions. *J. Glob. Optim.* **13**(4), 455–492 (1998).
- Queipo, N. V. *et al.* Surrogate-based analysis and optimization. *Prog. Aerosp. Sci.* **41**(1), 1–28 (2005).
- Li, J. & Tartakovsky, A. M. Gaussian process regression and conditional polynomial chaos for parameter estimation. *J. Comput. Phys.* **416**, 109520 (2020).
- Forrester, A. I. J. & Keane, A. J. Recent advances in surrogate-based optimization. *Prog. Aerosp. Sci.* **45**, 50–79 (2009).
- Alzayed, A. M., Mikki, S. M. & Antar, Y. M. M. Nonlinear mutual coupling compensation operator design using a novel electromagnetic machine learning paradigm. *IEEE Antennas Wirel. Propag. Lett.* **18**(5), 861–865 (2019).
- Tak, J., Kantemur, A., Sharma, Y. & Xin, H. A 3-D-printed W-band slotted waveguide array antenna optimized using machine learning. *IEEE Antennas Wirel. Propag. Lett.* **17**(11), 2008–2012 (2018).
- Torun, H. M. & Swaminathan, M. High-dimensional global optimization method for high-frequency electronic design. *IEEE Trans. Microw. Theory Technol.* **67**(6), 2128–2142 (2019).
- Liu, B. *et al.* An efficient method for antenna design optimization based on evolutionary computation and machine learning techniques. *IEEE Trans. Antennas Propag.* **62**(1), 7–18 (2014).
- Liu, B., Koziel, S. & Zhang, Q. A multi-fidelity surrogate-model-assisted evolutionary algorithm for computationally expensive optimization problems. *J. Comput. Sci.* **12**, 28–37 (2016).
- de Villiers, D. I. L., Couckuyt, I., & Dhaene, T. Multi-objective optimization of reflector antennas using kriging and probability of improvement. In *International Symposium on Antennas Propagation* 985–986 (San Diego, USA, 2017).

40. Easum, J. A., Nagar, J., Werner, P. L. & Werner, D. H. Efficient multi-objective antenna optimization with tolerance analysis through the use of surrogate models. *IEEE Trans. Antennas Propag.* **66**(12), 6706–6715 (2018).
41. Koziel, S. Low-cost data-driven surrogate modeling of antenna structures by constrained sampling. *IEEE Antennas Wirel. Propag. Lett.* **16**, 461–464 (2017).
42. Koziel, S. & Sigurdsson, A. T. Triangulation-based constrained surrogate modeling of antennas. *IEEE Trans. Antennas Propag.* **66**(8), 4170–4179 (2018).
43. Koziel, S. & Pietrenko-Dabrowska, A. Performance-based nested surrogate modeling of antenna input characteristics. *IEEE Trans. Antennas Propag.* **67**(5), 2904–2912 (2019).
44. Koziel, S. & Ogurtsov, S. Multi-objective design of antennas using variable-fidelity simulations and surrogate models. *IEEE Trans. Antennas Propag.* **61**(12), 5931–5939 (2013).
45. Koziel, S. & Bekasiewicz, A. *Multi-objective Design of Antennas Using Surrogate Models* (World Scientific, 2016).
46. Koziel, S. & Leifsson, L. *Simulation-Driven Design by Knowledge-Based Response Correction Techniques* (Springer, 2016).
47. Koziel, S. Fast simulation-driven antenna design using response-feature surrogates. *Int. J. RF Microw. CAE* **25**(5), 394–402 (2015).
48. Koziel, S. & Pietrenko-Dabrowska, A. Design-oriented computationally-efficient feature-based surrogate modelling of multi-band antennas with nested kriging. *AEU Int. J. Electron. Commun.* **120**, 66 (2020).
49. Balanis, C. A. *Antenna Theory: Analysis and Design* (Wiley, 2012).
50. Pozar, D. M. *Microwave Engineering* (Wiley, 2011).
51. Volakis, J. L. *Antenna Engineering Handbook* (McGraw-Hill, 2007).
52. Koziel, S. & Pietrenko-Dabrowska, A. *Performance-Driven Surrogate Modeling of High-Frequency Structures* (Springer, 2020).
53. Koziel, S. & Sigurdsson, A. T. Performance-driven modeling of compact couplers in restricted domains. *Int. J. RF Microw. CAE* **28**(6), 66 (2018).
54. Koziel, S. & Bandler, J. W. Reliable microwave modeling by means of variable-fidelity response features. *IEEE Trans. Microw. Theory Technol.* **63**(12), 4247–4254 (2015).
55. Koziel, S. & Bekasiewicz, A. Expedited simulation-driven design optimization of UWB antennas by means of response features. *Int. J. RF Microw. CAE* **27**(6), e21102 (2017).
56. Pietrenko-Dabrowska, A. & Koziel, S. Generalized formulation of response features for reliable optimization of antenna input characteristics. *IEEE Trans. Antennas Propag.* **70**(5), 3733–3748 (2022).
57. Conn, A. R., Gould, N. I. M., & Toint, P. L. *Trust Region Methods* (MPS-SIAM Series on Optimization, 2000).
58. Koziel, S. & Pietrenko-Dabrowska, A. Reduced-cost electromagnetic-driven optimization of antenna structures by means of trust-region gradient-search with sparse Jacobian updates. *IET Microw. Antennas Propag.* **13**(10), 1646–1652 (2019).
59. Koziel, S. & Pietrenko-Dabrowska, A. Efficient gradient-based algorithm with numerical derivatives for expedited optimization of multi-parameter miniaturized impedance matching transformers. *Radioengineering* **28**(3), 572–578 (2019).
60. Broyden, C. G. A class of methods for solving nonlinear simultaneous equations. *Math. Comput.* **19**, 577–593 (1965).
61. Beachkofski, B., & Grandhi, R. Improved distributed hypercube sampling. *American Institute of Aeronautics and Astronautics*, paper AIAA 2002-1274 (2002).
62. Kennedy, J. & Eberhart, R. C. *Swarm Intelligence* (Morgan Kaufmann, 2001).
63. Chen, Y.-C., Chen, S.-Y., & Hsu, P. Dual-band slot dipole antenna fed by a coplanar waveguide. In *Proceedings of the IEEE Antennas Propagation Society International Symposium*, Albuquerque, NM, USA, pp. 3589–3592 (2006).
64. Koziel, S. & Pietrenko-Dabrowska, A. Reduced-cost surrogate modeling of compact microwave components by two-level kriging interpolation. *Eng. Opt.* **52**(6), 960–972 (2019).
65. Koziel, S. & Pietrenko-Dabrowska, A. Reliable data-driven modeling of high-frequency structures by means of nested kriging with enhanced design of experiments. *Eng. Comput.* **36**(7), 2293–2308 (2019).
66. Koziel, S. & Pietrenko-Dabrowska, A. Fast and reliable knowledge-based design closure of antennas by means of iterative prediction-correction scheme. *Eng. Comput.* **38**(10), 3710–3731 (2021).
67. Bekasiewicz, A., & Koziel, S. Novel structure and EM-driven design of miniaturized microstrip rat-race coupler, in *2018 International Applied Computing Electromagnetics Society Symposium (ACES)* 1–2, Denver, CO, USA, 25–29 March, 2018.
68. Serani, A. *et al.* Parameter selection in synchronous and asynchronous deterministic particle swarm optimization for ship hydrodynamics problems. *Appl. Soft Comput.* **49**, 313–334 (2016).
69. Serani, A., Diez, M., Leotardi, C., Peri, D., Fasano, G., Iemma, U., & E. Campana, On the use of synchronous and asynchronous single-objective deterministic particle swarm optimization in ship design problems, in *International Conference Engineering Applied Sciences Optimization*, Kos, Greece (2014).
70. Campana, E. F., Fasano, G. & Pinto, A. Dynamic analysis for the selection of parameters and initial population, in particle swarm optimization. *J. Glob. Optim.* **48**, 347–397 (2010).
71. Diez, M., Serani, A., Leotardi, C., Campana, E. F., Peri, D., Iemma, U., Fasano, G., Giove, S. A proposal of PSO particles' initialization for costly unconstrained optimization problems: ORTHOinit, in: Y. Tan, Y. Shi, & C.A.C. Coello (eds.) *Advances in Swarm Intelligence*, ICSI 2014, Lecture Notes in Computer Science, vol. 8794 (Springer, 2014).

Acknowledgements

The authors would like to thank Dassault Systemes, France, for making CST Microwave Studio available. This work is partially supported by the Icelandic Centre for Research (RANNIS) Grant 206606 and by National Science Centre of Poland Grant 2020/37/B/ST7/01448.

Author contributions

Conceptualization, S.K. and A.P.; methodology, S.K. and A.P.; software, S.K. and A.P.; validation, S.K. and A.P.; formal analysis, S.K.; investigation, S.K. and A.P.; resources, S.K.; data curation, S.K. and A.P.; writing—original draft preparation, S.K. and A.P.; writing—review and editing, S.K.; visualization, S.K. and A.P.; supervision, S.K.; project administration, S.K.; funding acquisition, S.K. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.P.-D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022