



## Article

# Branch and Bound Algorithm Based on Prediction Error of Metamodel for Computational Electromagnetics

Reda El Bechari <sup>1,2,\*</sup> , Stéphane Brisset <sup>1</sup> , Stéphane Clénet <sup>1</sup> , Frédéric Guyomarch <sup>1</sup>  
and Jean Claude Mipo <sup>2</sup>

<sup>1</sup> University Lille, Arts et Metiers Institute of Technology, Centrale Lille, Junia, ULR 2697-L2EP, F-59000 Lille, France; stephane.brisset@centralelille.fr (S.B.); stephane.clenet@ensam.eu (S.C.); frederic.guyomarch@univ-lille.fr (F.G.)

<sup>2</sup> Valeo Powertrain Systems, F-94000 Créteil, France; jean-claude.mipo@valeo.com

\* Correspondence: reda.el-bechari@centralelille.fr

Received: 16 November 2020; Accepted: 17 December 2020; Published: 21 December 2020



**Abstract:** Metamodels proved to be a very efficient strategy for optimizing expensive black-box models, e.g., Finite Element simulation for electromagnetic devices. It enables the reduction of the computational burden for optimization purposes. However, the conventional approach of using metamodels presents limitations such as the cost of metamodel fitting and infill criteria problem-solving. This paper proposes a new algorithm that combines metamodels with a branch and bound (B&B) strategy. However, the efficiency of the B&B algorithm relies on the estimation of the bounds; therefore, we investigated the prediction error given by metamodels to predict the bounds. This combination leads to high fidelity global solutions. We propose a comparison protocol to assess the approach's performances with respect to those of other algorithms of different categories. Then, two electromagnetic optimization benchmarks are treated. This paper gives practical insights into algorithms that can be used when optimizing electromagnetic devices.

**Keywords:** electromagnetic; finite element method; metamodel; optimization methods; TEAM Workshop problems

## 1. Introduction

The increasing constraints on the design of electromagnetic devices require numerical tools that can model the electromagnetic fields with high precision on local quantities. The Finite Element (FE) method is the most used method to satisfy such a requirement. However, this method may be costly in terms of computational time due to nonlinear behavior, 3D geometry, and time dependency. Thus, its usage for optimization, i.e., iterative process, should be made with caution since only a limited number of evaluations of the simulation tool is possible. It also suffers from numerical noise due to the remeshing when dealing with geometric parameters. Thus, these issues have been extensively studied in the literature using the metamodel approach and the stochastic algorithms [1,2], e.g., genetic algorithms (GA), etc. On the other hand, optimization using gradient-based algorithms is less studied due to the remeshing error that perturbs the gradient's computation using the finite difference technique [3].

Metamodels are widely used in the context of optimization of electromagnetic devices [4,5]. They are constructed based on a small sample from the expensive model, i.e., FEM simulation. The optimization is performed on the metamodel rather than the expensive model. This metamodel could be improved using infill criteria that enable the sampling of new promising designs from the

full model. The infill criterion has a strong influence on how efficiently and accurately the algorithm locates the optimum.

In this paper, we developed a new approach to the usage of metamodels in optimization because the conventional approach of fitting one metamodel and enriching it has some drawbacks:

1. The infill criteria that focus more on the local search tend to sample points close to each other, which decays the conditioning of the correlation matrix and, consequently, the construction of the metamodel [6,7].
2. The number of samples exponentially increases as the optimization problem's dimensionality does (the curse of dimension). Thus, the size of the correlation matrix increases and consequently the time needed to fit the metamodel. This time can even exceed the evaluation time of the FE model.
3. The infill criteria optimization problem is highly multimodal, and their modality is highly correlated to the sample size of the expensive model evaluations; when the number of samples increases, the number of local optima increases too. Therefore, it becomes highly difficult to solve these optimization problems [8].
4. When dealing with constrained optimization, the infill criteria tend to sample inside the feasible region but not on the feasibility boundaries, which affects the solution accuracy [9].

This paper proposes a novel approach that considers the FEM simulation as a black-box and constructs cheap metamodels to reduce the computational burden while refining in regions with a high error or low objective value. Then we compare it to different approaches that directly use the expensive model; a derivative-based algorithm that uses the derivatives computed using the adjoint variable method and derivative-free ones such as GA and DIRECT [10,11]. The comparison mainly focuses on the implementation and the computational cost to get efficient results. Furthermore, we define metrics and a comparison protocol to ease this task.

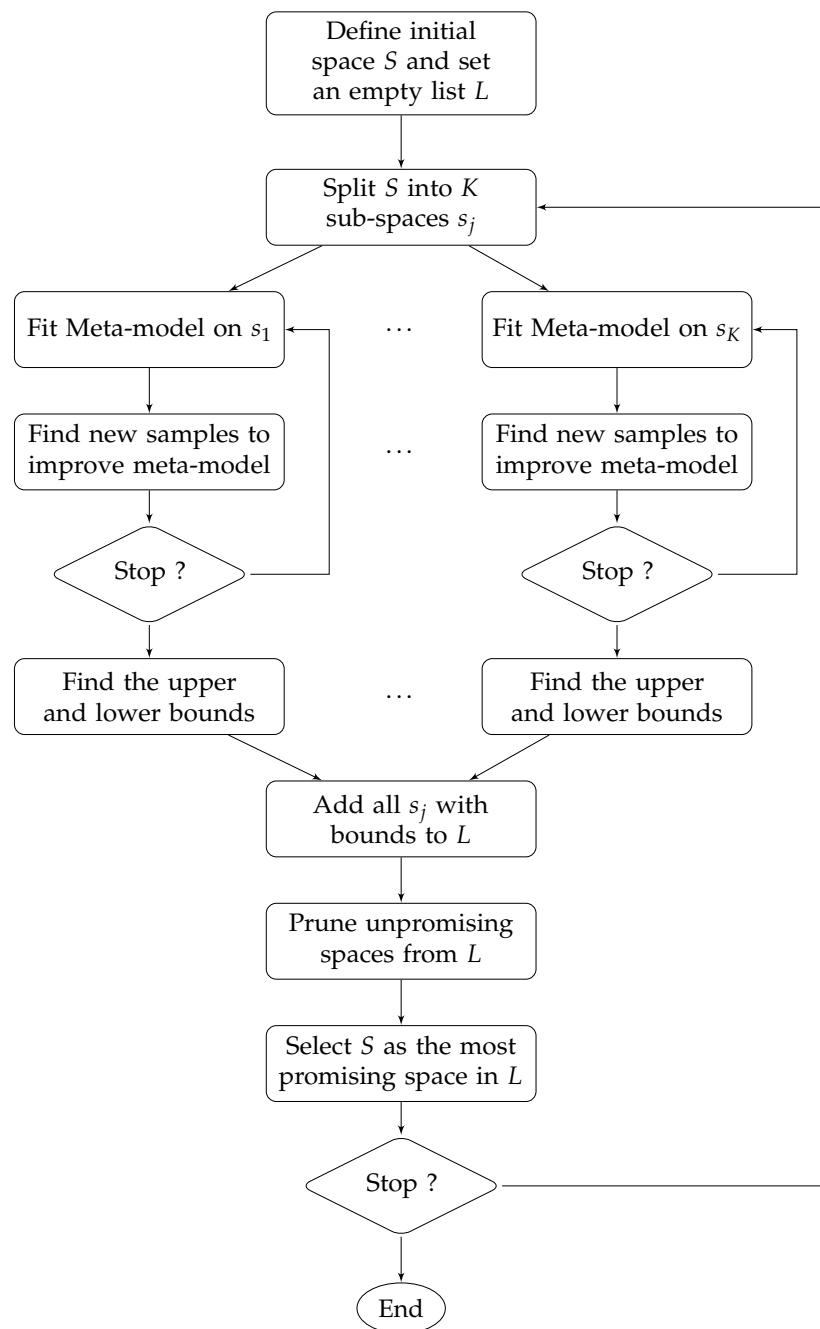
The comparison results highlight the algorithms suitable for the optimization of electromagnetic devices and the pros and cons of each of them.

## 2. Metamodel Approach

The proposed approach is called Branch and Bound based on Metamodel (B2M2). It is based on the principle of the branch and bound algorithms and consists of a systematic enumeration of candidate regions by splitting the search space into smaller spaces. The algorithm explores each smaller space then estimates the upper and lower bounds for each one of them. A subspace is discarded if it cannot produce a better solution than the best found so far by the algorithm.

### 2.1. B2M2 Algorithm

We propose an algorithm that builds many metamodels, each on a specified region of the design space; each one of them is relatively easy to fit and is used to estimate the upper and lower bounds. Then, iteratively, it excludes the regions that are not promising. By adopting this algorithm, the drawbacks mentioned earlier are tackled, namely, the numerical ones related to the correlation matrix. The flowchart of the algorithm is shown in Figure 1.



**Figure 1.** Flowchart of B2M2.

### 2.1.1. Initialization

In the beginning, the initial space  $S$  is defined, and an empty list  $L$  is set.  $L$  will store the potentially optimal regions. Additionally, a maximum depth level  $l$  is defined to limit the exploration of the algorithm. The depth level can be seen as a metric of how many times the initial design space  $S$  can be subdivided.

### 2.1.2. Branching

The branching stands for subdividing the initial design space  $S$ . The most intuitive rule is splitting each dimension of  $S$  by two. This will lead to  $K = 2^{n_v}$  subspaces  $s_j$  ( $n_v$  is the number of variables).

Unfortunately, the number of subspaces increases exponentially as the number of variables grows. This is not desired, and a more practical rule is proposed.

For metamodels, the purpose behind branching is to improve upon its accuracy by reducing the size of the design space. Thus, the space  $S$  is split, such as the produced subspaces are of better accuracy. Therefore, the metamodel fit to  $S$  is used to identify the dimensions of high inaccuracy.

The stochastic term in the Kriging predictor enables the modeling of the error not captured by the regression term. The correlation function characterizes this term. Commonly, the exponential function is used to express the correlation between two points  $\mathbf{p}^k$  and  $\mathbf{p}^l$ , and it is given by

$$\text{Corr}[F(\mathbf{p}^k), F(\mathbf{p}^l)] = \exp\left(-\sum_{i=1}^{n_v} \theta_i |p_i^k - p_i^l|^\alpha\right) \quad (1)$$

where  $F(\mathbf{p}^k)$  is the evaluation of the expensive model at  $\mathbf{p}^k$ ,  $n_v$  is the number of variables,  $\theta_i$  and  $\alpha$  are the parameters of the correlation.

The value of  $\alpha$  characterizes the smoothness of the function, while the values of  $\theta_i$  enable the modeling of the changes over distance. Large values of  $\theta_i$  serve to model rapid changes even over small distances. Hence, the values of  $\theta_i$  are indicators of how the function varies.

If the value of  $\theta_i$ , corresponding to a variable  $p_i$ , is the largest. Then, the change in the  $i$ th-dimension is more important than other dimensions (when the variables are normalized). Consequently, splitting the space  $S$  by dividing its dimension  $i$  will lead to metamodels on the subspaces that are better than the one on  $S$ .

A simple example is used to test this property of the Kriging model (2). The studied function is a three-dimensional one but depends explicitly on only two variables, and one of them presents more variation than the other. In this example, we expect the value of  $\theta_2$  to be higher than  $\theta_1$  and  $\theta_3$ . So, a Kriging model is fitted to a Design of Experiment (DoE) of 15 samples (five times the number of variables).

$$f(p_1, p_2, p_3) = p_1 p_2^2, \text{ with } 0 \leq p_i \leq 10, \quad i = 1, 2, 3 \quad (2)$$

Indeed, the metamodel fitting led to  $\theta = (0.072, 0.171, 0.004)$ , which confirms the proposed idea. Ordering the value of  $\theta$  will lead to the variables that have the highest variation. The function does not depend on  $p_3$  then  $\theta_3 = 0.004$  which is the smallest allowed value for  $\theta$ .

More than one variable can be chosen for separation; these variables are picked in the decreasing order of their corresponding  $\theta$  values. Each variable is then subdivided into two parts which will lead to  $2^d$  subspaces, where  $d \leq n_v$  is the number of variables picked.

### 2.1.3. MetaModeling

Once the subspaces are determined using the branching rule (see the previous section), Kriging models are built for each one of them. The initial DoE, such as Hammersley, Latin Hypercube Sampling (LHS), etc., is determined, and the black-box-model is evaluated.

Additional samples are added using infill sampling criteria (ISC) to improve the metamodel approximation. The choice here is crucial for estimating of the bounds and rely on the prediction error. New samples are added—where the prediction error is high—that consequently improve the metamodel globally on the subspace. Additionally, a limited number of points are sampled using the expected improvement (EI) criterion to enhance local exploitation (or EI subjected to the constraints metamodel for constrained problems). This will also improve the global performance of the algorithm.

### 2.1.4. Bounding

Once the metamodel is fit to each subspace, we look for the lower and upper bounds on each of them as the Kriging can give the predictor and its error estimation. The Kriging model is constructed based on the assumption that the prediction is normally distributed with mean  $\hat{f}$  and standard deviation  $\hat{\sigma}_f$  [1].

The aim is to find the interval  $[f, \bar{f}]$  that enclose the solution on a subspace given a specified probability, e.g.,  $k = 3$  corresponds to a probability of 99.73% that the solution is enclosed between the bounds. We exploit this feature for the computation of bounds using the following formulations (3) and (4):

$$\begin{aligned} \underline{f} &= \min_p \hat{f}(p) - k\hat{\sigma}_f(p) \\ \text{s.t. } \hat{g}_i(p) - k\hat{\sigma}_{g_i}(p) &\leq 0, \quad i = 1, \dots, n_c. \end{aligned} \quad (3)$$

$$\begin{aligned} \bar{f} &= \min_p \hat{f}(p) + k\hat{\sigma}_f(p) \\ \text{s.t. } \hat{g}_i(p) + k\hat{\sigma}_{g_i}(p) &\leq 0, \quad i = 1, \dots, n_c. \end{aligned} \quad (4)$$

where  $\hat{f}$  and  $\hat{g}_i$  are the predictors of the objective function  $f$  and the constraints  $g_i$  respectively and  $\hat{\sigma}_f, \hat{\sigma}_{g_i}$  are their error estimations given by the Kriging metamodel. The parameter  $k$  can be seen as the confidence level required.

A smaller value of  $k$  reduces the probability of bounding the solution. It, therefore, increases the risk of eliminating a promising subspace. In contrast, higher values of  $k$  widen the bounding interval  $[f, \bar{f}]$ , which makes the elimination more difficult.

The optimization problems (3) and (4) have to be solved using appropriate optimization algorithms due to the multimodal behavior introduced by the  $k\hat{\sigma}$  terms. We use sequential quadratic programming (SQP) from Matlab for this task improved by a parallel multistart strategy with points uniformly spread on the design space. The number of starting points is much higher than the number of samples of the black-box model. Additionally, the gradient information is computed easily from the metamodel, then it is exploited to increase the efficiency of the SQP algorithm.

The optimization problem (4) can present difficulties in solving since the objective function and constraints are mostly concave. Furthermore, our experience has shown that  $\bar{f}$  is, most of the time, equal to the best-sampled point (the point that has the smallest value of the objective function and satisfies the constraints). Thus, this optimization can be discarded, and  $\bar{f}$  is set to be equal to the sample having the smallest value of the objective function and satisfying the constraints on the subspace. When no feasible sample is available for highly constrained problems,  $\bar{f}$  is set to be equal to a large value.

In most constrained optimization problems, the problem (3) may not have a solution on some subspaces. It means that these subspaces have no potentially feasible solution, and therefore the values of the bounds are set to some high value.

#### 2.1.5. Elimination

The subspaces are added to  $L$  with their lower  $\underline{f}$  and upper  $\bar{f}$  bounds. A comparison between the newly added subspaces to  $L$  and the existing ones is conducted to eliminate those that cannot produce a better solution.

First, we set  $\bar{b}$  as the **smallest upper bound** of all the spaces in  $L$ , and then, we prune the spaces from  $L$  that have a lower bound  $\underline{f}$  that is greater than  $\bar{b}$ .

#### 2.1.6. Selection Priority

The remaining subspaces in  $L$  are all promising ones; thus, one of them has to be chosen for the next iteration to be subdivided. Many strategies could be considered: by preferring depth-first for prioritizing local search or breadth-first for prioritizing exploration of the search space. In our work, we use the best-first; it is based on the lower bound  $\underline{f}$  of the objective function on the subspaces. Indeed, the subspace chosen from  $L$  has the smallest lower bound, and its depth level is smaller than the maximum specified depth level.

This strategy will prioritize subspaces that have the maximum potential for providing the best solution without any criteria on the size of the subspace.

The subspaces that have attained the maximum allowed depth level are stored in a list  $L_s$ .

### 2.1.7. Termination

The algorithm will naturally iterate until the list  $L$  is empty. It is challenging to develop a robust termination criterion. However, the exploration of a subspace can be stopped if the difference between its lower bound and upper bound is less than a small threshold specified by the designer. It remains possible to stop the algorithm by setting a maximum number of evaluations of the black-box model.

The list  $L_s$  contains the remaining subspaces that are at the maximum specified depth level, or their bounds are close. One should investigate them and search for the optimum in one of them without considering the prediction error. Otherwise, the algorithm could be restarted with the list  $L_s$  and a higher maximum depth level if no solution is found.

### 2.2. Test Function

The Rosenbrock function is used as a test problem for this optimization algorithm. As shown in (5), the problem can be extended to many variables. Here we treat the bidimensional problem ( $n_v = 2$ ).

$$\begin{aligned} \min_p f(p) &= \sum_{i=1}^{n_v-1} \left[ (1 - p_i)^2 + 100 (p_{i+1} - p_i^2)^2 \right] \\ \text{s.t. } -2.4 &\leq p_i \leq 2.4, \quad i = 1 \dots n_v \end{aligned} \quad (5)$$

Rosenbrock problem has one global minimum at  $p^0 = (1, \dots, 1)$  and  $f(p^0) = 0$ .

The setup of the algorithm parameters is as follows.

- The maximum depth level is  $l = 4$ .
- The parameter  $k$  in (3) and (4) is equal to 3
- On each subspace, the initial sample is chosen using Hammersley sequence with 4 sample points, the same number of sample points is added using the prediction error as an ISC and two points are added using EI.

Figure 2 shows the evolution of the subspace division. As shown, we compute the lower and upper bounds for each subspace, i.e.,  $\underline{f}$  and  $\bar{f}$ , respectively. The new subspaces are compared with those in the list  $L$  to discard the unpromising ones. The shaded area shows the subspaces discarded from the list  $L$ .

After only four iterations, the algorithm has located the subspace where the optimum lies. The grey region is the only space that was left in the list  $L$ .

To locate the solution in the remaining space, we consider only the predictor of the objective function and the constraints without their prediction errors. Then, we solve the optimization problem defined by these predictors to find the minimum. The solution found is close to the global minimum of the problem.

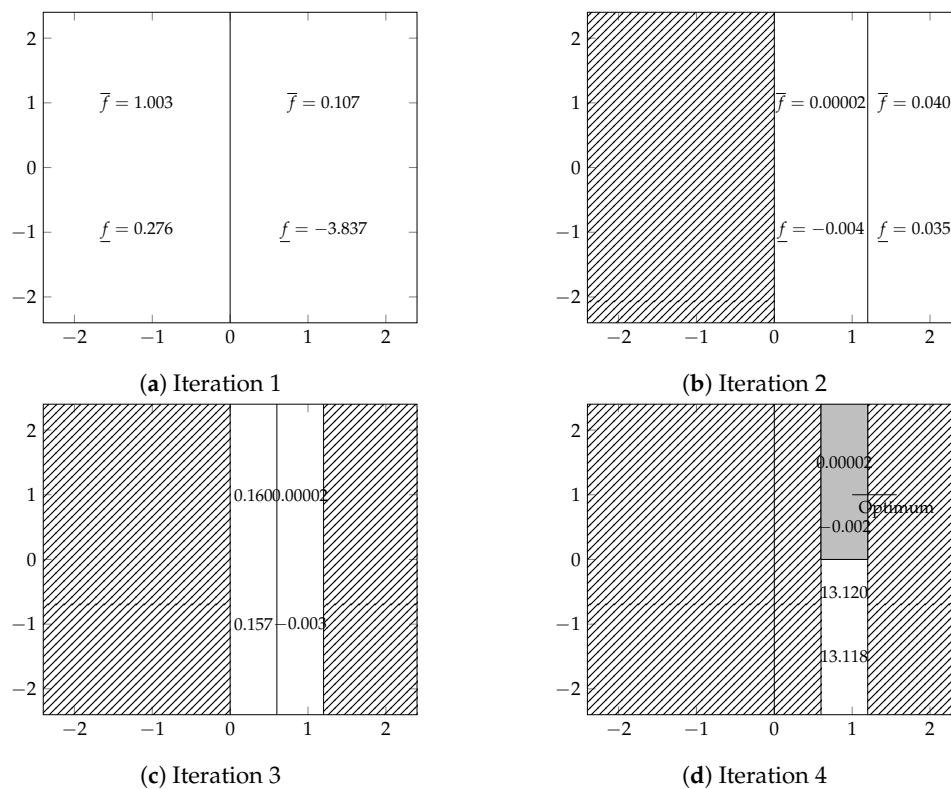


Figure 2. Space exploration by the algorithm.

### 3. Expensive Model Approaches

In these approaches, we can distinguish two categories:

#### 3.1. Derivative-Free

Derivative-free optimization is a discipline in mathematical optimization that does not use derivative information in the classical sense to find optimal solutions. Many gradient-free algorithms are designed as global optimizers and thus can find multiple local optima while searching for the global optimum. Various gradient-free methods have been developed. We investigate some of the most commonly used algorithms: genetic algorithm (from Matlab) and DIRECT algorithm [11].

A genetic algorithm is a stochastic optimization method based on randomness inherent in its operators: selection, mutation, and crossover.

The DIRECT method is one of the strategies for deterministic global optimization. It performs a systematic search of the design space using a hyperdimensional adaptive meshing scheme. Then it explores all the design space to find the optimum.

#### 3.2. Derivative-Based

In design optimization, the gradient enables computing the search direction for gradient-based algorithms, e.g., sequential quadratic programming, active set. When dealing with FEM models, the gradient information is not always available.

##### 3.2.1. Finite-Difference

The gradient can be approximated using finite-difference. However, numerical noise due to remeshing makes this approximation highly sensitive and requires adjusting the step size used for the computation [3]. Thus, the use of such an algorithm is not beneficial and a gradient of good quality need to be computed.

### 3.2.2. Adjoint Variable Method

The adjoint variable method (AVM) enables to efficiently compute the gradient from FEM code. Thus, no numerical noise due remeshing is observed as opposed to finite-difference. We implemented a 2D FEM code and an AVM code based on the discrete approach [12] for the gradient computation. The implementation of the AVM code is cumbersome since it requires intrusive manipulation of the FEM code itself. Nevertheless, it remains particularly appealing in terms of the computational cost, as the time for computing the gradient is equivalent to evaluating FEM code once whatever the number of variables. This can be very advantageous when treating large-scale problems, e.g., eight variables and more.

### 3.2.3. Globalization Strategy

Gradient-based algorithms suffer from the convergence to local optima that depends on the initial point. To handle this issue, a multistart strategy with uniformly distributed initial points is considered. By doing so, we increase the probability of attaining the global optimum.

Therefore, we define a convergence rate (C.R.) of the method. The C.R. is the proportion of optimizations that led to the best solution. Two solutions are considered the same if the distance between them is less than a small value. In this study, we took 1% of the design range.

## 4. Comparison Protocol

The comparison of the performance of different algorithms will be based on two criteria: the cost and the quality of the solution.

The quality of the solution stands for the value of the objective function found. The smaller, the better.

On the other hand, the optimization cost stands for the number of evaluations needed to attain the solution with a given probability. As we compare different approaches, we propose a probabilistic metric to measure the cost.

The convergence rate (CR) is the percentage of the number of runs that have converged to the best solution. The convergence rate concerns mainly SQP and GA since we perform multiple runs and the number of runs needed is unknown beforehand. To compute CR, we look among the solutions of each run for the best one (smallest value), then we count the number of runs that converged to that solution within a tolerance of 1% of the design range.

Once the convergence rate is determined, we assess the probability that at least one optimization leads to the best solution is

$$P = 1 - (1 - \text{C.R.})^n$$

where  $n$  is the number of optimization runs to be considered.

If we want to consider a high probability ( $P = 99.73\%$ ) of finding the best solution, we can compute  $n$  the expected number of runs needed.

$$n = \frac{\ln(1 - P)}{\ln(1 - \text{CR})} \quad (6)$$

Then, the expected total number of evaluations (Expected # FEM evals) is  $n$  times the total average number of evaluations performed.

$$\text{Expected \# FEM evals} = n \text{ average}(\text{\# FEM evals}) \quad (7)$$

It should be noted that the cost of computing the gradient using the adjoint variable method is equivalent to evaluating the FEA once; thus, the # FEM evals are doubled for the SQP method.



## 5. Algorithms Settings

In this section, we present the configuration of the optimization algorithms. A set of parameters are defined based on our experience of using each algorithm.

### 5.1. B2M2 Algorithm

The branch and bound based metamodel (B2M2) algorithm is detailed in Section 2. Some options need to be configured, mainly the depth level and the number of FEA evaluations needed for fitting the metamodels.

**Initial design:** The initial design of each subspace is created using a space-filling LHS of size two times the number of variables. This may seem few, but as the algorithm progresses, the sample points of a parent space will be added to the initial design of its subspaces. Thus, the sample will grow as the subdivision of the space continues.

**Infill points:** Additional sample points are added in two different manners; the first one aims to explore the design space further and increase the metamodel precision by adding points where the prediction error is high. In contrast, the second one enables exploiting the best solution by adding new sample points that maximize the EI criterion.

The number of sample points added by each ISC is set to twice the number of variables. However, for EI, it may be stopped before this limit if there is no improvement in additional sampling points.

Globally, in each subspace, there will be sampled at most six times the number of variables.

**Stopping criteria:** The algorithm stops if a predefined depth level is reached. The maximum depth level is set to  $l = 10$ .

### 5.2. SQP Algorithm Assisted by Adjoint Variable Method

We use the SQP algorithm from the Matlab Optimization toolbox with the gradient of the quantities of interest computed using the adjoint variable method.

**Multistart:** To cope with the local search of the SQP method, we use a multistart strategy. We perform 100 runs with different initial points; these points are sampled using an LHS design to obtain a uniform distribution on the whole design space.

**Stopping criteria:** The stopping criterion is based on the StepTolerance option, meaning the algorithm stops when the relative change in the solution is less than  $10^{-8}$ .

### 5.3. DIRECT

We use DIRECT implementation of Finkel [11]. The default options are maintained except for the number of iterations.

**Constraints:** The constraints are handled by the penalty method by minimizing the penalized objective function

$$OF(p) + \sum_i \lambda \max(0, g_i(p))$$

where  $\lambda$  is the penalty factor and has been chosen to be equal to 10.

**Stopping criteria:** The algorithm does not have an appropriate stopping criterion except the number of iterations and the number of evaluations. We stop the algorithm manually when there is no improvement in the objective function in 50 successive iterations.

### 5.4. Genetic Algorithm

We use the GA from the Matlab Global Optimization toolbox [10].

**Multirun:** As the inherent randomness in the genetic algorithm, meaning solving twice the same problem may not lead to the same solution, we adopted a multirun strategy by performing 50 runs of the algorithm.

**Population:** The population options are set as follows



**Table 1.** Bounds of design variables for TEAM problem 25.

$p$	R1	L2	L3	L4
min	5.0	12.6	14	4
max	9.4	18	45	19

The results are summarized in Table 2.

**Table 2.** TEAM Workshop problem 25 optimization results.

Approach	SQP	GA	DIRECT	B2M2
R1	7.31	7.51	7.31	7.28
L2	14.21	14.64	14.20	14.17
L3	14.11	14.39	14.08	14.06
L4	14.37	14.44	14.39	14.34
W	$7.62 \times 10^{-5}$	$12.44 \times 10^{-5}$	$7.61 \times 10^{-5}$	$9.73 \times 10^{-5}$
# FEM evals	280	10100	24255	2575
CR	34%	2%	100%	100%
Expected # FEM evals	3585	2,146,501	24,255	2575

The genetic algorithm (GA) could not get reliable results; only one optimization converged to the solution shown in the table. Furthermore, this one is not competitive compared to other solutions. B2M2 can locate the optimal solution but with a higher cost, which is expected since it is based on branch and bound algorithms that enable to get reliable results. SQP and DIRECT algorithms gave the best solution, while the former outperforms all the others in terms of computational cost.

TEAM Workshop Problem 25 has a convergence rate (CR) of 34% when using the SQP method. This suggests that the optimization problem is not highly multimodal.

In terms of solution quality, SQP and DIRECT methods lead to the best solutions (smaller value of  $W$ ), while SQP has a cost of almost seven times less than DIRECT.

B2M2 approach led to a competitive solution, even though the value of the objective function is a little higher. However, the design variables are very close to the ones found by SQP (difference of less than 1% of the design range). Most importantly, B2M2 cost is the most affordable.

## 6.2. TEAM Workshop Problem 22

The Superconducting Magnetic Energy Storage (SMES) device in Figure 4 consists of two concentric superconducting coils fed by currents that flow in opposite directions [15]. The inner coil is used for storing magnetic energy  $E$ , while the outer one has the role of diminishing the magnetic stray field  $B_{stray}$  computed on *line a* and *line b*.

The optimization problem's goal is to find the design configurations (eight design variables) that give a specified value of stored magnetic energy and a minimal magnetic stray field while satisfying some constraints. Mathematically, this is formulated as

$$\min_p OF(p) = B_{stray}^2(p)/B_{norm}^2 + |E(p) - E_{ref}|/E_{ref} \quad (9)$$

$$s.t. \quad |J| + 6.4|B| - 54 \leq 0 \quad (10)$$

$$R_1 - R_2 + \frac{1}{2}(d_1 + d_2) < 0 \quad (11)$$

where  $E_{ref} = 180$  MJ,  $B_{norm} = 200$   $\mu$ T and  $p$  are the design variables and their bounds are shown in Table 3.

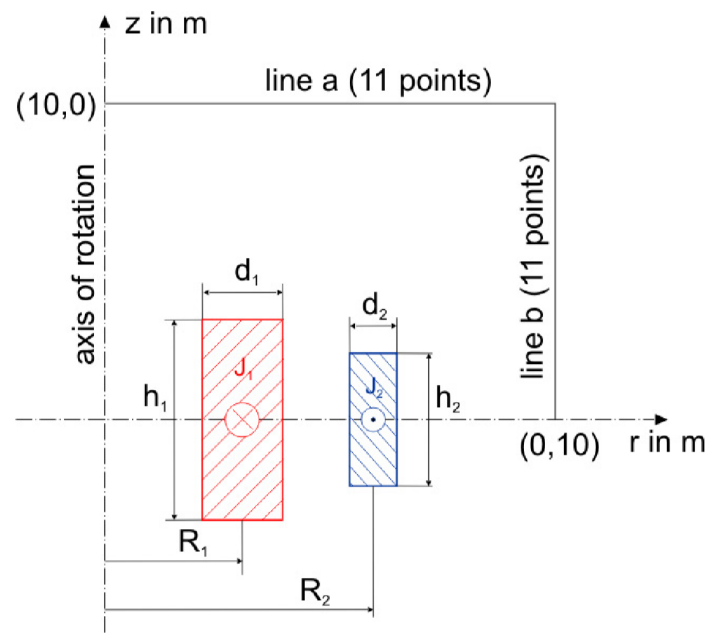


Figure 4. SMES Device [15].

Table 3. Bounds of design variables for TEAM problem 22.

$p$	$R_1$	$R_2$	$h_1/2$	$h_2/2$	$d_1$	$d_2$	$J_1$	$J_2$
min	1.0	1.8	0.1	0.1	0.1	0.1	10	-30
max	4.0	5.0	1.8	1.8	0.8	0.8	30	-10

The constraint (10) aims to limit the maximal flux density ( $|\mathbf{B}|$ ) in the coils to ensure the quench condition [15]. The maximal values of  $|\mathbf{B}|$  are located on the boundaries of the coils and specifically on coordinates  $P_1 = (R_1 - d_1/2, 0)$ ,  $P_2 = (R_1 + d_1/2, 0)$  and  $P_3 = (R_2 - d_2/2, 0)$ . Thus, we replace this constraint by three (13)–(15).

The second constraint aims to prevent both coils from overlapping. Unfortunately, optimization algorithms can sometimes violate some constraints, and this implies taking some special care of this issue that depends on the type of algorithm used. If this constraint is violated, the model is not valid physically; thus, other quantities such as the objective function cannot be computed. Some optimization algorithms can handle this issue if the model returns a high value of the objective function, but others cannot deal with this kind of constraint.

As we compare different approaches, we chose to define another optimization problem equivalent to the initial one that avoids the issue mentioned above. We define a variable  $A_2$  as

$$A_2 = R_2 - R_1 - (d_1 + d_2)/2$$

that replaces  $R_2$ . Using interval arithmetic, we compute the bounds of this new variable  $[-3, 3.9]$ . By using the constraint (11) preventing the two coils from overlapping, we impose  $A_2$  to be higher than 0.001. Then, the interval of variation of  $A_2$  is  $[0.001, 3.9]$ .

Introducing this new variable is somewhat relaxing the initial optimization problem by allowing the variable  $R_2 = R_1 + A_2 + (d_1 + d_2)/2$  to vary in a more significant interval  $[1.101, 8.7]$  than the one defined in the initial problem  $[1.8, 5]$ . Therefore, two constraints (16) and (17) are added. Finally, the optimization problem becomes

$$\min_p OF(p) = B_{stray}^2(p)/B_{norm}^2 + |E(p) - E_{ref}|/E_{ref} \quad (12)$$

$$s.t. \quad (J_1 - 54)/6.4 + |\mathbf{B}(p, P_1)| \leq 0 \quad (13)$$

$$(J_1 - 54)/6.4 + |\mathbf{B}(p, P_2)| \leq 0 \quad (14)$$

$$(-J_2 - 54)/6.4 + |\mathbf{B}(p, P_3)| \leq 0 \quad (15)$$

$$R_1 + A_2 + (d_1 + d_2)/2 \leq 5 \quad (16)$$

$$-R_1 - A_2 - (d_1 + d_2)/2 \leq -1.8 \quad (17)$$

where  $p$  are the design variables  $p = (R_1, A_2, h_1/2, h_2/2, d_1, d_2, J_1, J_2)$ .

The Table 4 summarizes the optimization results.

**Table 4.** TEAM Workshop problem 22 optimization results.

Approach	SQP	GA	DIRECT	B2M2
$R_1$	1.336	1.457	1.543	1.369
$A_2$	0.027	0.481	0.229	0.054
$h_1/2$	1.011	1.209	0.951	0.888
$h_2/2$	1.452	1.800	1.526	1.394
$d_1$	0.677	0.347	0.374	0.791
$d_2$	0.269	0.121	0.217	0.203
$J_1$	15.579	19.834	22.346	14.099
$J_2$	−15.069	−17.305	−13.441	−18.273
$OF$	0.00197	0.03502	0.04881	0.00510
# FEM evals	1902	160,201	421,995	106,308
CR	1 %	2 %	100 %	100 %
Expected # FEM evals	529,640	94,276,389	421,995	106,308

We notice that SQP is largely outperforming the other approaches in terms of solution quality by leading to the smallest value of the objective function. Furthermore, 17% of the solutions found by this approach have an objective function value less than 0.05 (corresponding to the solution found by DIRECT) while being distinct in terms of variables. This explains the low CR of the method and, thus the high expected number of evaluations.

The genetic algorithm has performed better than DIRECT in terms of quality of the solution, but the former suffers from its inherent randomness, and thus the convergence rate is low.

B2M2 approach was able to get a solution of better quality than GA and DIRECT but less than SQP.

The TEAM Workshop Problem 22 is multimodal (there exist many local minima). The low convergence rate of SQP highlights this fact; indeed, the solutions found tend to have similar values of the objective function while having design variables that are entirely different.

The SQP assisted by the adjoint variable method was able to get the best solution. However, due to multimodality, the cost of obtaining this solution is high when compared to the one found by B2M2, which is of inferior quality and with a lower cost.

Two facts can explain GA and DIRECT failure; first, the multimodal behavior of the problem treated renders locating the best solution very delicate, and secondly, the constraint handling strategy. Indeed, the penalty approach used may not be best-suited, and the penalty factors may not be very adequate. Thus, the tuning of these factors will induce additional cost.

TEAM Workshop Problem 22 is one of the most treated benchmarks for the optimization of electromagnetic devices. Researchers have reported their results in many papers [2,4,15,16]. However, due to the differences in the FE mesh and solver used by each researcher, the comparison to their results is not possible. Table 5 shows a comparison of some results published by other researchers. We compare the value of the objective function  $OF$  between the one reported in the articles (Reported  $OF$ ) and the one evaluated using the same set of variables in our FE model (Calculated  $OF$ ). What one can observe is that the values of the Calculated  $OF$  are different from the Reported  $OF$ . This confirms the

fact that the results are clearly dependent on the method and the parameters used to solve the TEAM Workshop problem.

**Table 5.** Comparison of optima reported on other research papers.

	Ref. [15]	Ref. [16]	Ref. [2]	Ref. [4]
$R_1$	1.296	1.3012	1.432	1.103
$R_2$	1.800	1.8000	2.023	2.318
$h_1/2$	1.089	1.1325	0.784	1.596
$h_2/2$	1.513	1.5422	1.411	0.144
$d_1$	0.583	0.5802	0.787	0.259
$d_2$	0.195	0.1961	0.178	0.734
$J_1$	16.695	16.422	14.290	22.5
$J_2$	−18.910	−18.925	−18.010	−22.5
Reported $OF$	0.0018	0.0020	0.0030	0.0014
Calculated $OF$	0.0092	0.0025	0.0131	25.2714

Therefore, we mainly focused on selecting some available algorithms such as genetic algorithm and DIRECT, while the SQP assisted by the adjoint variable method was developed in our previous work [17]. Thus, the algorithm's selection from each category is based on what is generally used in the literature. The selection of a particular algorithm or implementation is based on what we are or have worked on, i.e., B2M2 and SQP, and also the availability and simplicity of usage. These are usually the challenges that designers are facing when doing optimization.

## 7. Conclusions

The proposed approach is implemented and applied successfully in the context of optimization of electromagnetic devices. The algorithm builds many metamodels, each on a specified region of the design space. Each one of them is relatively easy to fit and to use for the purpose of estimating the upper and lower bounds. Then, iteratively, the algorithm prunes the regions that are not promising. This process can keep a good tradeoff between exploration and exploitation of the design space.

This article also presents a comparison among different approaches for the optimization of electromagnetic devices using FEM. We treated two well-known benchmarks from the literature [14,15]. The choice of the algorithm from each category is based on what is generally used in the literature. The choice of a particular algorithm or implementation is based on what we are working on, i.e., B2M2 and the adjoint variable method, and also the availability and simplicity of usage. Furthermore, we picked one algorithm from different classes of algorithms.

The genetic algorithm (GA) is one of the most used algorithms to deal with noisy data. For both test cases, GA did not perform well. These performances might be slightly improved by doing some parameter tuning or using other implementations.

DIRECT performed very well for the first test case, but it was ineffective for the second one; the constraints handling strategy can explain this; the penalty method implemented in the algorithm may not be the best.

In terms of performances, the SQP approach outperforms other strategies for both test cases; this was possible due to the computation of the gradient using the adjoint variable method. This improves the quality of the solutions drastically, but this comes with the expense of intrusive manipulation of the FEM code that is not trivial or even possible.

The B2M2 algorithm offers the best tradeoff in terms of implementation and solution quality. The developed approach was able to overcome some of the very known issues when using metamodels for optimization. The B2M2 algorithm is easily parallelizable since the subspaces are independent, and it can be further improved in the future mainly to handle multiobjective optimization and optimization under uncertainty.

**Author Contributions:** Conceptualization, R.E.B., S.B., S.C., F.G. and J.C.M.; methodology, R.E.B., S.B. and F.G.; software, R.E.B.; validation, R.E.B., S.B., S.C. and F.G.; formal analysis, R.E.B.; investigation, R.E.B.; resources, S.B. and J.C.M.; data curation, R.E.B., S.B., S.C. and F.G.; writing—original draft preparation, R.E.B.; writing—review and editing, R.E.B., S.B., S.C. and F.G.; visualization, R.E.B.; supervision, S.B., S.C., F.G. and J.C.M.; project administration, R.E.B., S.B., S.C., F.G. and J.C.M.; funding acquisition, S.B., S.C. and J.C.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Jones, D.R.; Schonlau, M.; Welch, W.J. Efficient Global Optimization of Expensive Black-Box Functions. *J. Glob. Optim.* **1998**, *13*, 455–492. [CrossRef]
2. Alotto, P.; Eranda, C.; Brandstatter, B.; Furntratt, G.; Magele, C.; Molinari, G.; Nervi, M.; Preis, K.; Repetto, M.; Richter, K.R. Stochastic algorithms in electromagnetic optimization. *IEEE Trans. Magn.* **1998**, *34*, 3674–3684. [CrossRef]
3. Berkani, M.S.; Giurgea, S.; Espanet, C.; Coulomb, J.L.; Kieffer, C. Study on optimal design based on direct coupling between a FEM simulation model and L-BFGS-B algorithm. *IEEE Trans. Magn.* **2013**, *49*, 2149–2152. [CrossRef]
4. Li, Y.; Xiao, S.; Rotaru, M.; Sykulski, J.K. A Kriging-Based Optimization Approach for Large Data Sets Exploiting Points Aggregation Techniques. *IEEE Trans. Magn.* **2017**, *53*, 7001704. [CrossRef]
5. Li, M.; Gabriel, F.; Alkadri, M.; Lowther, D.A. Kriging-Assisted Multi-Objective Design of Permanent Magnet Motor for Position Sensorless Control. *IEEE Trans. Magn.* **2016**, *52*, 7001904. [CrossRef]
6. Mohammadi, H.; Riche, R.L.; Durrande, N.; Touboul, E.; Bay, X. An analytic comparison of regularization methods for Gaussian Processes. *arXiv* **2016**, arXiv:1602.00853.
7. Ababou, R.; Bagtzoglou, A.C.; Wood, E.F. On the condition number of covariance matrices in kriging, estimation, and simulation of random fields. *Math. Geol.* **1994**, *26*, 99–133. [CrossRef]
8. Picheny, V.; Wagner, T.; Ginsbourger, D. A benchmark of kriging-based infill criteria for noisy optimization. *Struct. Multidiscip. Optim.* **2013**, *48*, 607–626. [CrossRef]
9. Sasena, M.J.; Papalambros, P.; Goovaerts, P. Exploration of Metamodeling Sampling Criteria for Constrained Global Optimization. *Eng. Optim.* **2002**, *34*, 263–278. [CrossRef]
10. The MathWorks Inc. *Matlab, Global Optimization Toolbox*; The MathWorks Inc.: Natick, MA, USA, 2019.
11. Finkel, D.E. *DIRECT Optimization Algorithm User Guide*; Center for Research in Scientific Computation, North Carolina State University: Raleigh, NC, USA, 2003; Volume 2, pp. 1–14.
12. Lee, H.B.; Ida, N. Interpretation of adjoint sensitivity analysis for shape optimal design of electromagnetic systems. *IET Sci. Meas. Technol.* **2015**, *9*, 1039–1042. [CrossRef]
13. Deb, K. An efficient constraint handling method for genetic algorithms. *Comput. Methods Appl. Mech. Eng.* **2000**, *186*, 311–338. [CrossRef]
14. Takahashi, N. Optimization of Die Press Model: TEAM Workshop Problem 25. *COMPUMAG TEAM Workshop 1996*; pp. 1–9. Available online: <http://www.compumag.org/jsite/images/stories/TEAM/problem25.pdf> (accessed on 10 November 2020).
15. Alotto, P.; Baumgartner, U.; Freschi, F. SMES Optimization Benchmark: TEAM Workshop Problem 22. *COMPUMAG TEAM Workshop 2008*; pp. 1–4. Available online: <http://www.compumag.org/jsite/images/stories/TEAM/problem22.pdf> (accessed on 10 November 2020).
16. Brandstätter, B.; Ring, W. Optimization of a SMES Device Under Nonlinear Constraints. *Citeseer* **1997**. Available online: <https://imsc.uni-graz.at/ring/papers/smes.pdf> (accessed on 10 November 2020)

17. El Bechari, R.; Guyomarch, F.; Brisset, S.; Clenet, S.; Mipo, J.-C. Efficient Computation of Adjoint Sensitivity Analysis for Electromagnetic Devices Modeled by FEM. *Electr. Magn. Fields* **2018**. [[CrossRef](#)]

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).