Modelling and optimal design in railway applications

Sangkla Kreuawan

N° d'ordre : 82

Ecole Centrale de Lille

THESE

Présentée en vue
d'obtenir le grade de

DOCTEUR

en

Spécialité : Génie Electrique

par

Sangkla Kreuawan

DOCTORAT DELIVRE PAR L'ECOLE CENTRALE DE LILLE

Titre de la thèse :

Modelling and optimal design in railway applications

Soutenue le 24 novembre 2008 devant le jury d'examen :

Rapporteur	J.-L. Coulomb	Professeur	G2ELAB, Grenoble
Rapporteur	C. Marchand	Professeur	LGEP, Paris
Examinateur	E. Lomonova	Associate Professor	TU Eindhoven
Directeur de thèse	P. Brochet	Professeur	L2EP, Lille
Co-encadrant	F. Gillon	Maître de conférences	L2EP, Lille
Invité	L. Nicod	Upstream activities manager	Alstom Transport
Invité	S. Brisset	Maître de conférences HDR	L2EP, Lille

Thèse préparée dans le Laboratoire L2EP à l'Ecole Centrale de Lille

Ecole Doctorale SPI 072 (Lille I, Lille III, Artois, ULCO, UVHC, EC Lille)

Modelling and optimal design in railway applications

Sangkla Kreuawan

24 November 2008

Résumé

La conception d'un système électrique de transport ferroviaire est une tâche complexe qui fait appel simultanément à des experts de domaines de compétence différents. Les constructeurs ferroviaires gèrent cette complexité ce qui leurs permet de fabriquer des équipements performants. Néanmoins, dans un marché global, tout gain méthodologique peut se traduire en avantage concurrentiel.

La conception systémique optimale de composant électrotechnique est abordée dans cette thèse. Une chaîne de traction électrique est choisie comme exemple représentatif d'un système complexe. La démarche et les outils sont mis en oeuvre sur deux applications: la conception d'un moteur de traction et la conception simultanée de plusieurs composants clés.

Pour concevoir un moteur de traction, le cycle de fonctionnement et le comportement thermique transitoire sont primordiaux. La bonne adaptation du moteur à sa mission permet de réduire considérablement sa masse. L'approche multidisciplinaire est utilisée pour gérer les interactions entre modèles de disciplines différentes au sein d'un même processus d'optimisation. Suivant la méthode employée, le temps d'optimisation peut être réduit grâce à la répartition des taches par domaine physique et d'en paralléliser l'exécution. Des optimisations multiobjectif ont également été appliquées. Des fronts de Pareto sont obtenus malgré l'utilisation d'un modèle précis mais coûteux, le modèle éléments finis.

L'approche décomposition hiérarchique de la méthode "Target Cascading" est appliquée au problème de conception de la chaîne de traction. Le système et ses composants sont conjointement conçus. Cette méthode est bien adaptée à la démarche de conception optimale des systèmes complexes, tout en respectant l'organisation par produit de l'entreprise.

Mots-clés

Conception optimale
Chaîne de traction ferroviaire
Modèle de substitution
Optimisation multidisciplinaire
Approche multi-niveaux

Abstract

The design of traction systems is a complex task, which needs experts from various fields. Train manufacturers can manage this complexity and produce high performance rolling stock materials. However, any improvement in design methodology can lead to a competitive advantage in a global market.

This thesis focuses on the optimal design methodology of complex systems such as a railway traction system. The design process and tools are demonstrated via two applications: the design of a traction motor and the concurrent design of several key components.

The load cycle and transient thermal behaviour are essential in the design of a traction motor. The adaptation of a motor to its load cycle reduces significantly its mass. The multi-disciplinary design optimization approach is used to manage interactions between various discipline models in the optimization process. The optimization time can be reduced through a task distribution and a parallel computing. The multi-objective design optimizations are also applied. Pareto fronts are obtained despite the difficulty in using the high fidelity but expensive in computation time such as Finite Element Analysis model.

The hierarchical decomposition approach: the Target Cascading method is applied to the traction system design problem. The system and components are designed simultaneously. This method is suitable for implementing the complex system optimal design process while respecting the product development structure of the company.

Keywords

Optimal design
Railway traction system
Surrogate modelling
Multidisciplinary Design Optimization
Multilevel approach

To my family

Acknowledgements

This work has been carried out as a part of a research project called FuturElec 3, in the "Laboratoire d'Electrotechnique et d'Electronique de Puissance de Lille" (L2EP) at Ecole Centrale de Lille with the support of the French government, the "Region Nord-Pas de Calais" authorities and Alstom Transport.

Firstly, I would like to express my deepest gratitude to my supervisor, Professor Pascal BROCHET for his valuable guidance and assistance. I was able to work in his team in the perfect environment.

Also, I am very grateful to my co-supervisor, Dr. Frédéric GILLON, Associate professor. His constant encouragement and advice help me overcome the most difficult periods while I was working on the thesis.

This work would not have been finished without the contribution of project team members, Dr. Stéphane BRISSET, Associate professor HDR, Professor Souad HARMAND and Fouzia MOUSSOUNI, my colleague in this project.

I would like to thank Professor Jean-Louis COULOMB, Professor Claude MARCHAND and Dr. Elena LOMONOVA, Associate professor, for spending their valuable time to be the jury of my thesis.

I would like to express my acknowledgement to Mr. Mike SCROOBY, Product development director, Alstom Transport for his support and advices. I also wish to thank Mr. Laurent NICOD, Upstream activities manager, Mr. Marc DEBRUYNE, Traction system expert, Mr. Olivier GIACOMONI, Mr. David CYPERS, Mr. Frédéric PORCHER, Mr. Frédéric BODLET and Mr. Michel NOVELLA, engineers at Alstom Transport for their valuable remarks, advices and for sharing their experiences during my work.

I would never forget all the staff members and professors at Ecole Centrale de Lille. I would like to mention in particular Professor Michel HECQUET, Xavier CIMETIERE, Simon THOMY for their help and support.

Big thanks also go to my colleagues at L2EP, Victor MEŞTER, TRAN Tuan Vu, Arnaud VIDET, Jean LEBESNERAIS, David MARIN, LI Peng, ZHOU Tao, Amir AHMIDI, LU Di, François GRUSON, Souleymane BERTHE and PENG Ling.

Finally, I am particularly indebted to my parents and my sister. They have truly always been there for me, and without them, none of this would have been even possible.

Contents

List of Figures

List of Tables

Introduction

Railway transportation has proved itself as a competitive transportation mode for short distance urban trips and medium distance journeys. Big capital cities have metro or tram to move their people around the city, whereas, regional and intercity trains operate in suburban and city links. Modern rail vehicles emphasize their image as regards sustainable development as they are typically equipped with electric propulsion. The emission gas due to the energy conversion process is centralized at the power plant and can be controlled easily and efficiently.

The traction system design process is a very complex task. These days, train manufacturers have sufficient knowledge to design and manufacture good quality rolling stocks. However, due to their concern in global market and environment, they ask themselves **whether** and **how** they can do better in terms of product performances, cost and development time.

This thesis focuses on the **complex system design** methodology and its application in electric railway traction systems. A lot of efforts have been made by aerospace industries in the field of complex system optimal design [30, 89, 102]. The application domain resides mainly in structural and aerodynamic design. The recent literature extends to a more global point of view and system design by taking into account, for example, the aircraft mission, the fuel consumption, the environmental impact [3, 4, 70]. A similar reflection is initiated in railway traction system design. It seems that the methodology is quite general and can be applied with success to electrical system design.

This thesis is divided into 2 parts, Tools and Applications. Chapter 1 presents the complex system design approach in general and more particularly, in railway industry. The **optimal design** (also called design optimization) methodology helps engineers in design and decision-making phases. Railway traction systems and design tools are presented. Potentials and needs to optimise such systems are discussed.

The modelling technique is considered in Chapter 2. The physical-based modellings are firstly presented. These traditional techniques are widely used by design engineers to derive the high fidelity model. Secondly, the Surrogate modelling approach is discussed. The surrogate model replaces the high fidelity model in many situations. The last sections present the non-conventional sampling and surrogate modelling techniques. They intend to introduce the reader to the next chapter.

According to the design methodology presented in Chapter 1, once the models are constructed, the optimization process can be launched. Chapter 3 provides the reader with an overview of optimization techniques and main single and multi-objective optimization algorithms. The surrogate model can be used in these algorithms not only to replace the high fidelity model but it also to **integrate** into the optimization algorithm. This is referred to as

the “Surrogate-assisted optimization algorithm”. In addition, the Multi-disciplinary Design Optimization (MDO) and Target Cascading (TC) approaches are presented. The optimal design of decoupled complex systems can be achieved via these methods.

From Chapter 4 onwards, the applications to railway traction system are presented. This chapter investigates the design methodology of traction motors. A multidisciplinary semi-numerical model is used in the comparative study of design approaches and then in the multi-criteria optimal design. This multidisciplinary model is also used to illustrate the contribution of the MDO methodology. In addition to the semi-analytic model, a high fidelity tool such as Finite Element Analysis is incorporated in the optimal design problem directly thanks to two surrogate-assisted optimization algorithms. Such algorithms allow obtaining the optimal design in a quick time.

Chapter 5 goes up to the top of system view. The optimal design of the control parameters of a tram traction system is presented in the first section. Thereafter, the design scope covers the top system level and goes down to the component level. A decomposition approach and TC method are used to achieve the optimal design of the tram traction system and its components. The conclusion and further research perspectives are discussed in the last chapter.

The main contribution of this work consists in demonstrating the advantages of optimal design process in the industry’s design problem. The two main expected advantages are: (i) a shorter development time and (ii) optimal products. The methodology is applied to many real-world cases. A proprietary corporate tool is used as simulation tool and integrated into the optimal design process. The optimal design can take place in any phases e.g. preliminary and detail design phases. Various tools help engineers to solve the design problem in an efficient way. With promising approaches such as surrogate modelling and surrogate-assisted algorithm, design engineers can obtain optimal solutions within a short computation time. The trade-off problems can be solved by using multi-objective optimization. The Pareto optimal solution set generated by optimal design process is used as a support for decision-making process. In a large and well established organisation, the Target Cascading method, explored in this work, helps teams to formulate an optimal design problem in a manner not far from what they have done in the past. With this innovative approach, the optimal design should gain higher acceptance from industries.

Part I

Tools

Chapter 1

Optimal design of complex systems

Nowadays, due to the global market, strict government regulations and standards, products or electrical systems become more and more complex as they involve many engineering fields as well as other non-technical issues such as marketing, psychology. The objective of any company is to design profitable products within a reasonable development time and at minimum cost. The product development or design processes reflect the vision of each company and can vary.

Designing such complex systems requires a lot of experience. Many product characteristics must be determined, subject to a large number of constraints. Any system or product is designed to be the best in terms of performances, cost, reliability, etc. The analysis capacity of human is limited to a certain level of complexity. Therefore, engineers hardly achieve optimal design of a complex system.

According to [2], a complex system represents:

> *An assembly of interacting members which is difficult to understand as a whole.*

The author also emphasized that an interaction between members exists if the state of one member affects how the system responds to changes made in another member.

Several technical and non-technical aspects are involved in a complex system description. In [49], the authors gave the expression of a complex engineering system as:

> *A complicated product or device requiring a variety of knowledge from several different engineering disciplines.*

In this research, a complex system is referred to as a system consisting of several interacting subsystems and components. This system evolves many technical and non-technical disciplines such as marketing, management and user satisfaction. The system is operated in a certain environment by users.

Three sources of complexity can be identified:

- the interaction between subsystems and/or components constituting the system.
- the interaction between disciplines describing physical or non-physical phenomena.
- the interaction between the system and the environment in which the system operates.

In the context of this research, a rolling stock is a good example of a complex system. A train is composed of many subsystems e.g. traction system, carriage, air conditioning system and communication system. Each subsystem includes lower-level subsystems as well as components. For example, a traction system is an integration of traction motors, converters, transformers, etc. Several disciplines are required to describe a train e.g. mechanical, electrical, thermal. The train is operated by immediate customers (railway operating companies) and provides services to final customers (passengers). As the train system is a complex system, many interactions can be observed e.g. the structural and aerodynamic design of the carriage affects the required power of the traction motor; the traction motor itself implies many physical disciplines, hence interactions between them. Moreover, both train and traction motor designs are strongly linked to their environment such as customer satisfaction, energy consumption. Figure 1.1 illustrates this example.

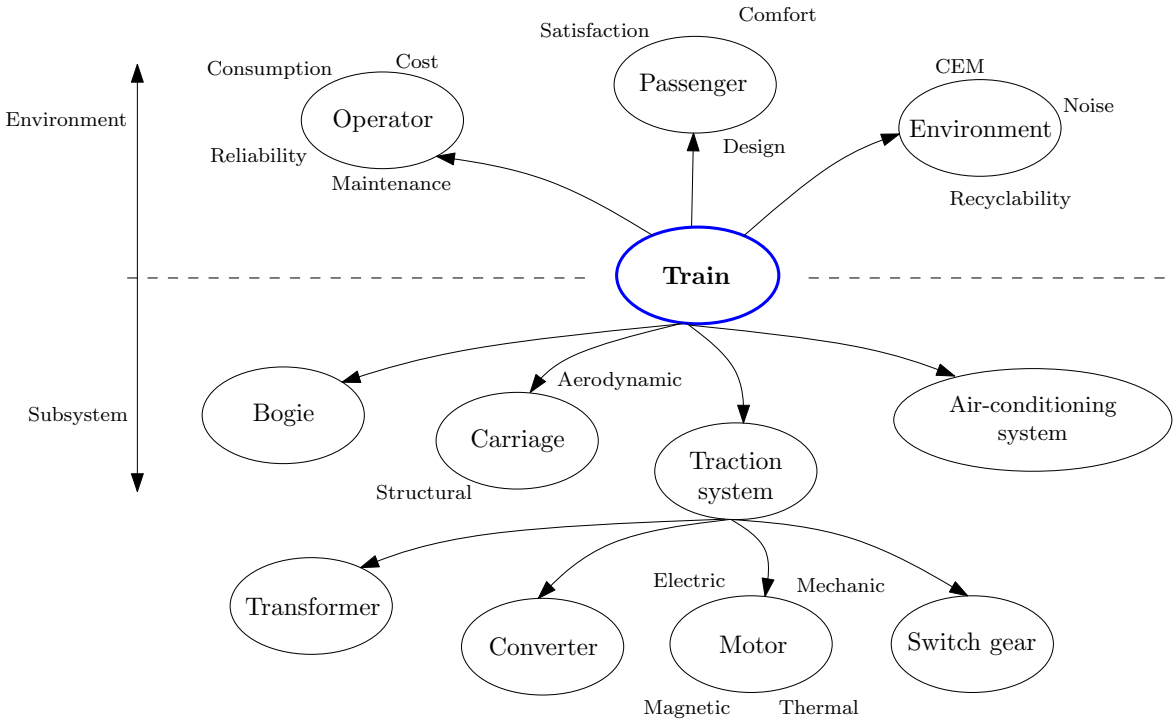


Figure 1.1: Illustration of a train system and its interactions

This chapter intends to give the background ideas on how the complex systems should be designed and how optimization techniques should be used in this design process. The last section focuses on the objective of this research i.e. the rolling stock optimal design process.

1.1 Complex system design methodology

The goal of product design is to obtain an optimal product which can make profit in a reasonable development time and at minimum cost [100]. When developing a complex system, the product development process can be very complicated since the complex system itself is difficult to understand. Designers require a design methodology and tools in order to accomplish this task. Various aspects must be considered in the design process. The following sections provide basic knowledge of design methodology as well as optimal design process.

1.1.1 Conventional approach

The typical system design process named V-cycle is composed of 2 main phases: design and validation [64]. The design process is shown in Figure 1.2. Starting from market research, a specification is defined and a feasibility study is performed. The system, subsystems and components are designed thereafter. The system design phase allows determining the subsystem specification and the subsystem design phase defines the component specification. These design phases are called "Top-down". Once component prototypes have been manufactured, they are firstly tested and then integrated to the subsystems and system. Each "Bottom-up" phase allows validating the corresponding design phases.

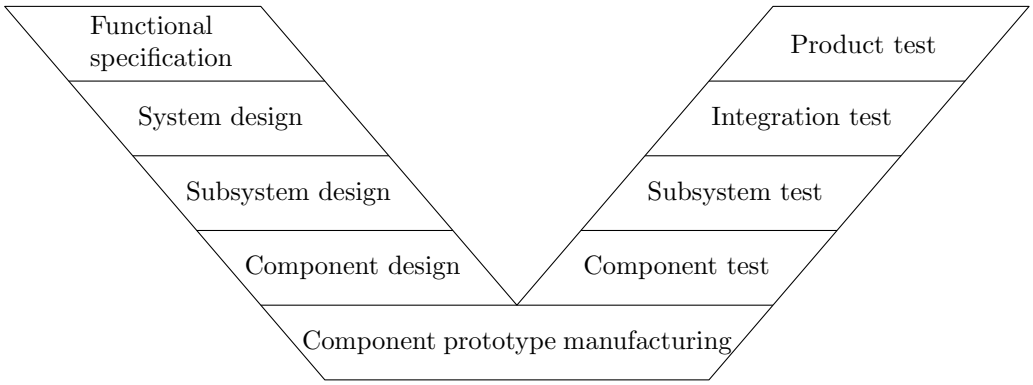


Figure 1.2: V-cycle design process

1.1.2 Design problem as inverse problem

In the field of engineering design, Computer Aided Design (CAD) and Computer Aided Engineering (CAE) are used in design and manufacturing industries. A number of computer modellings have been made to simulate behaviour of real systems. These models are sometimes called Virtual prototypes as they are used by design engineers as a tool to verify their design instead of building a real prototype. With the knowledge of the components/system characteristics e.g. their size, shape and material (cause), these models compute system performances (effects). A problem of this kind can be referred to as a "direct problem"[73] and the model which simulates system performances can be called a "behaviour model". However, one may realise that, in design problem, one want to design a complex system (i.e. to determine the size and shape of the system) which performs the desired performances.

Therefore, system performances are defined as specification and system characteristics are being sought. This kind of problem is called "inverse problem" as opposed to direct problem [73]. Figure 1.3 shows input and output of direct and inverse problems.

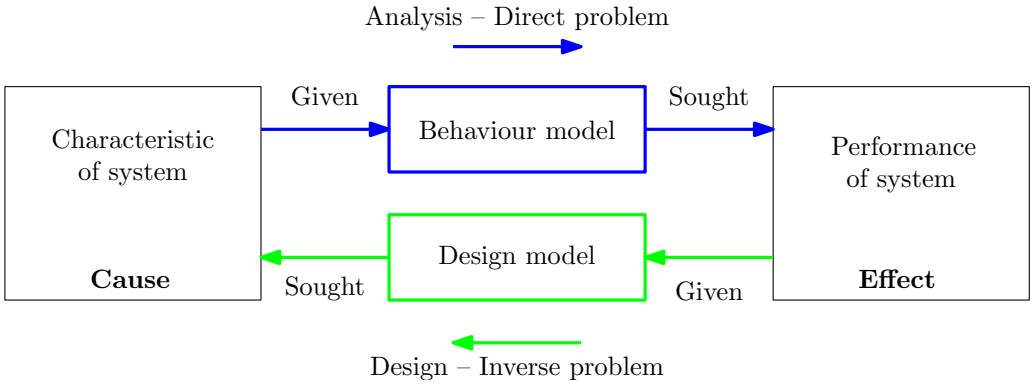
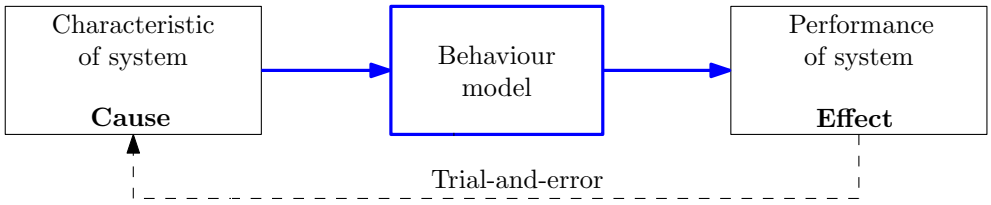
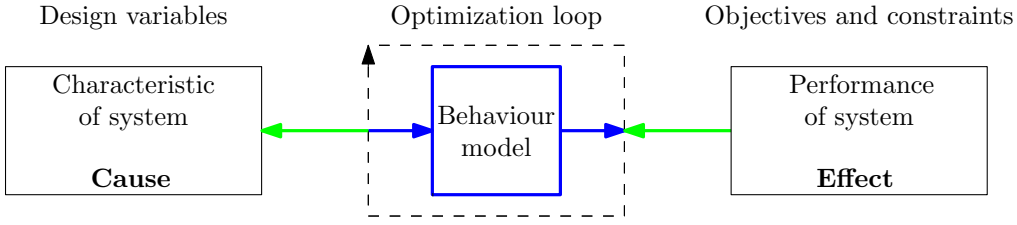


Figure 1.3: Input and output of direct and inverse problems

In order to use behaviour models in design problem, an iterative method is necessary. In the conventional design process, based on his experiences or a base-line design, designer makes a first guess at the system characteristics (design variable), runs simulation model and observes output performances. If the performances do not satisfy constraints and design criteria, the designer adjusts the design variables and makes the simulation again. This Trial-and-error process is shown in Figure 1.4a. The manual iteration can be replaced by an optimization technique, which ensures optimality and allows efficient problem solving. In this case, the behaviour model included in the optimization loop operates like a "design model" [87] as shown in Figure 1.4b. Note that the behaviour model is always present but can be considered as a design model from the designer's point of view, outside of the optimization loop. This optimal design process is presented in Section 1.1.6.



(a) Manual iteration



(b) Behaviour model and optimization technique as design model

Figure 1.4: Use of behaviour model to solve design problem

The most effective approach is to build the model in a design-oriented manner named

"inverse model" or "design model" as shown in Figure 1.3. By giving the desired performances, the model computes system characteristics. In this case, an optimization loop is not required. The design problem solving is faster. Unfortunately, this kind of model is not always available. A model intended to be used in design problem should be constructed in a design-oriented manner as much as possible.

1.1.3 Complex system decomposition

In large and complex system design problems, there are many design variables, constraints as well as several design criteria. A large number of engineers and experts gets involved in the design process. The decomposition approach is naturally used for large and complex system design. Designers usually decompose complex system design problems into subproblems due to cognitive limitations of human [35, 98]. Smaller problems yield a limited number of design variable and a smaller design space. Problem solving can be more efficient since experts in each corresponding field make decision on each subproblem. Global system design criteria can be achieved by allocating some design targets to subproblems. Subproblems are solved with respect and subject to their own design variables and constraints. The system engineer is responsible for coordinating design tasks done by engineering teams as well as ensuring compatibility between subproblem designs.

A decoupled design problem has several advantages:

- the design problem is understandable and easier to solve;
- the subproblems can be solved by an engineering team or an external supplier. The coordination and the design process remain the same;
- the specialized engineering teams own and maintain their models.

Several decomposition approaches are used, depending on companies. Object-based and discipline-based decompositions are natural. These decompositions reflects the managerial structure of the company and how company develops its products. Usually, one approach in itself cannot describe the system in a proper manner. For example, a train can be decomposed into subsystems and components with help of an object-based approach. Furthermore, depending on the problem, the same system can be decomposed in different ways e.g. some subproblems may be simplified or promoted to an upper level.

1.1.4 Global system design

In complex systems, there are interactions between subsystems, components and disciplines. The integration of optimised subsystems cannot ensure the optimality of a complex system [73]. Subsystems or components should not be developed individually without taking into account the interactions from other subsystems. Figure 1.5 depicts the example of a system composed of 2 components. It shows the efficiency of 2 separately designed components as well as the global efficiency of the system. It can be observed that the peak efficiency of the system is not at the same position as that of the subsystems (different working point).

Moreover, an increase in the efficiency of the subsystem (as in $\eta_1^{'}$) does not necessary lead to an improvement of the system and can even reduce the system global performances.

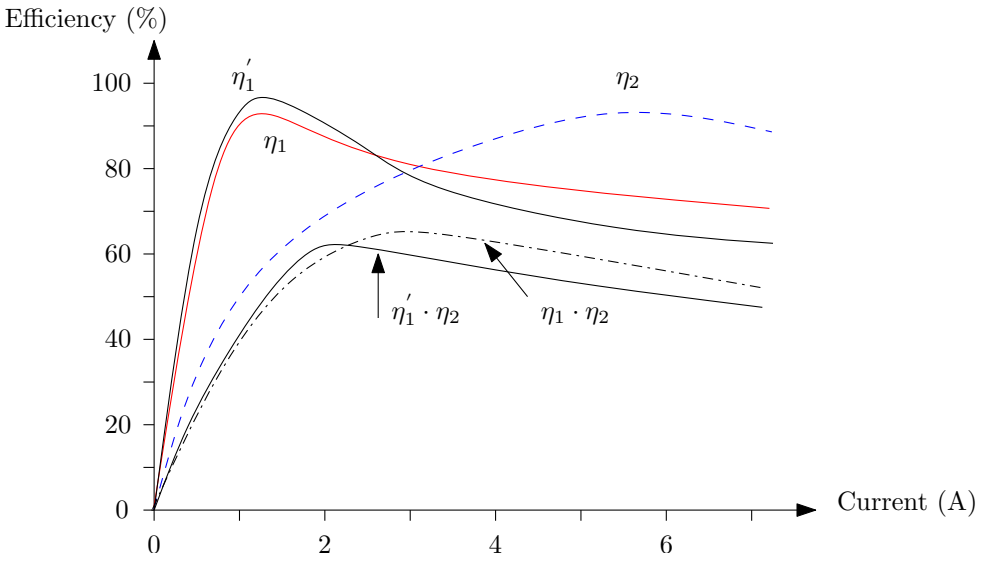
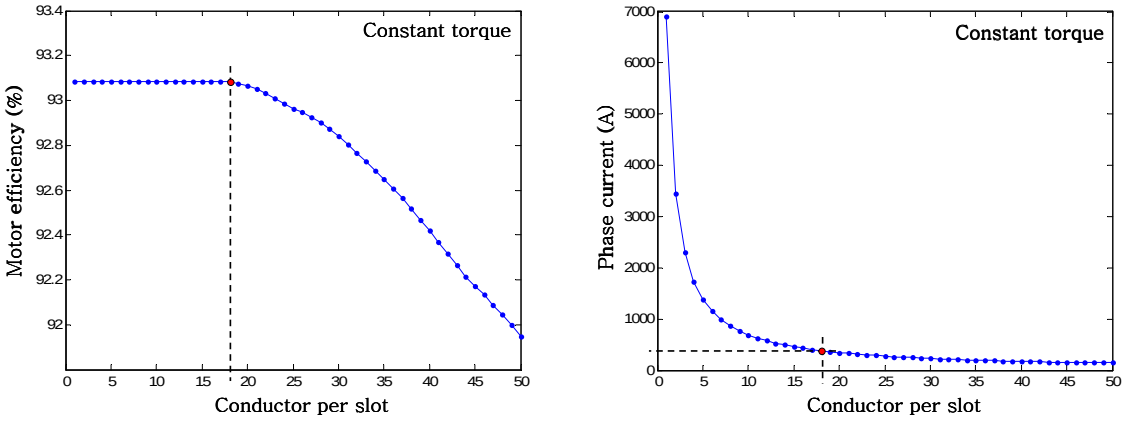


Figure 1.5: Efficiency of a 2 component system

In electrical engineering systems, an interaction between the traction motor and the inverter is very common. Figure 1.6 plots the motor efficiency and the inverter current against the number of conductor. They are obtained from the motor model presented in Appendix B. For the same torque, the motor efficiency and inverter current are computed for various numbers of conductors. The other dimensions of the motor are fixed as a constant. Any increase in the number of turns increases the total flux and reduces the current required to generate the same torque. This results in a low inverter current and low inverter losses. However, a high number of turns leads to low base speed. The motor needs to operate with a flux-weakening strategy and generates high losses in high speed zone. In this example, 18 conductors is the best compromise between motor and inverter efficiency (showed as a red point in the figures). It can be clearly concluded that the motor and the inverter must be designed simultaneously in order to take into account their interaction and to obtain the maximum efficiency of the global system.

In decoupled complex system designs, targets and constraints can be allocated to each subproblems. Subsystems can be designed by experts and by coordinating information between engineering teams so that the system is simultaneously designed. The Target Cascading method presented in Chapter 3 integrates the optimal design in such a complex system design process.



(a) Efficiency – Number of conductors per slot (b) Phase current – Number of conductors per slot

Figure 1.6: Interaction between motor and inverter

1.1.5 Multi-criteria design problem

Real-world design problems always deal with several conflicting criteria. The designer searches for solutions, which satisfy all criteria simultaneously e.g. low cost, high performances, low environment impact, high reliability. Unfortunately, such a "perfect" solution does not exist. In both engineering and marketing points of views, it is clear that these criteria are conflicting and that a compromise must be found. For example, if the high performances criterion is preferred, a trade-off at cost level would be inevitable.

In the preliminary design phase (see Section 1.1.1), the knowledge of the system is generally insufficient to pose a well-defined design problem. Unknown characteristics may be considered as criteria so that the designer obtains a wide range of information at the end of the design process. After having made a decision, some of these criteria are considered as constraints in the detailed design phase.

The decision maker or design engineer makes a decision based on a group of optimal solutions. In multi-criteria problems, optimal solutions satisfy Pareto optimality conditions (see Section 1.1.5.1 below) so that the group of optimal solutions can be presented by a Pareto front[1].

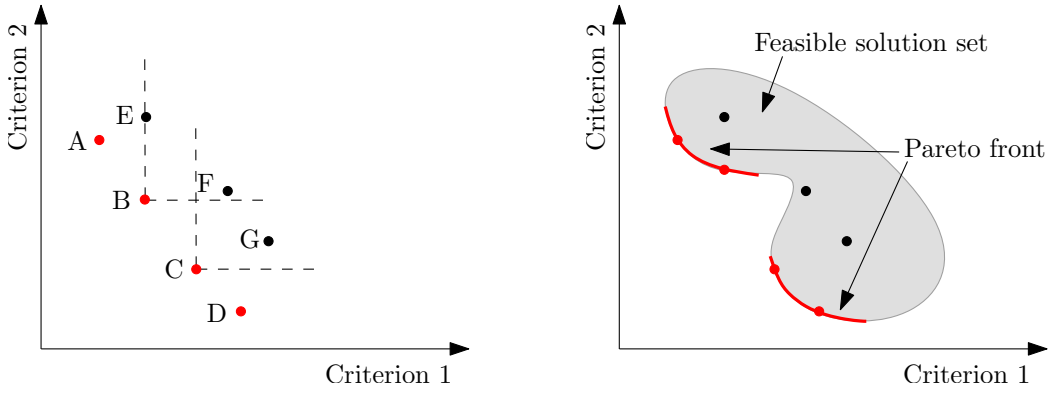
1.1.5.1 Pareto optimality definition

A Pareto front is a group of solutions, which satisfy Pareto optimality conditions. Based on the definition given in [78], a solution S^* is Pareto optimal if there does not exist another feasible solution such as at least one criterion is better than that of S^* and all another criteria are at least equal to - if not better than - those of S^*. This solution S^* is also called non-dominated solution [78].

In Figure 1.7a, The B solution is better than the E and F solutions for at least one criterion. Therefore, the B solution dominates the E and F solutions. Similarly, the C solution dominates the F and G solutions and the D solution dominates the G solution. As a

[1] also called Non-dominated front, trade-off curve or surface

result, the A, B, C and D solutions are non-dominated and satisfy Pareto optimality. These non-dominated solutions are treated as equivalent regarding the two criteria. Any solution is better than any other. Figure 1.7b shows these solutions together, the feasible region and the true Pareto front being depicted as bold lines. In this case, the Pareto front is nonconnected. It can be also nonconvex.



(a) Illustration of domination (b) Pareto front

Figure 1.7: Pareto optimality

In traditional design process, the Pareto front may include a small number of "improved" solutions compared with a base-line solution. For optimal design processes, a larger amount of well-distributed optimal solutions can be generated by using an optimization algorithm. The Pareto front provides a good support for decision-making. It can be used for negotiation in the design process.

1.1.6 Optimal design

Most models used in engineering designs are not design models. They are constructed for the simulation purposes and used in performance evaluation. As mentioned earlier, design aims at determining the characteristics, geometry or parameters of systems/components while satisfying desired performances (criteria). Therefore, the inverse design model is needed. In most cases, the system model cannot be totally built in a design-oriented manner. This means that an iteration must be performed. In conventional approaches, engineering designers carry out this process manually. It is difficult to obtain optimum solutions, the system being complex and many design variables and many constraints being involved.

This lack of human ability to solve the complex problems results in the necessity to integrate optimization techniques in the design process. An optimization technique is a tool, which helps designers to systematically explore the design space. It proposes intermediate or candidate results to designers. This helps the designer to better understand the complex design problem, to clarify his ideas and to make decision. Instead of performing trial-and-error method, optimization technique offers a more complete view. Based on its algorithm, an optimization technique uses information from models to find out how design variables should be varied in order to achieve optimal solutions while satisfying a number of constraint. However, optimization techniques do not have any intelligence. They only take in

charge repetitive automated tasks. Designer remains responsible of formulating a well-posed optimization problem, observes what optimization does during problem solving process and analyses given results.

Figure 1.8 shows an optimal design process. It is composed of 6 phases. Each phase is subject to being modified if it cannot yield satisfied results. Note that the specification definition phase is considered as the input to the problem and can be modified only once the decision is made in the final phase.

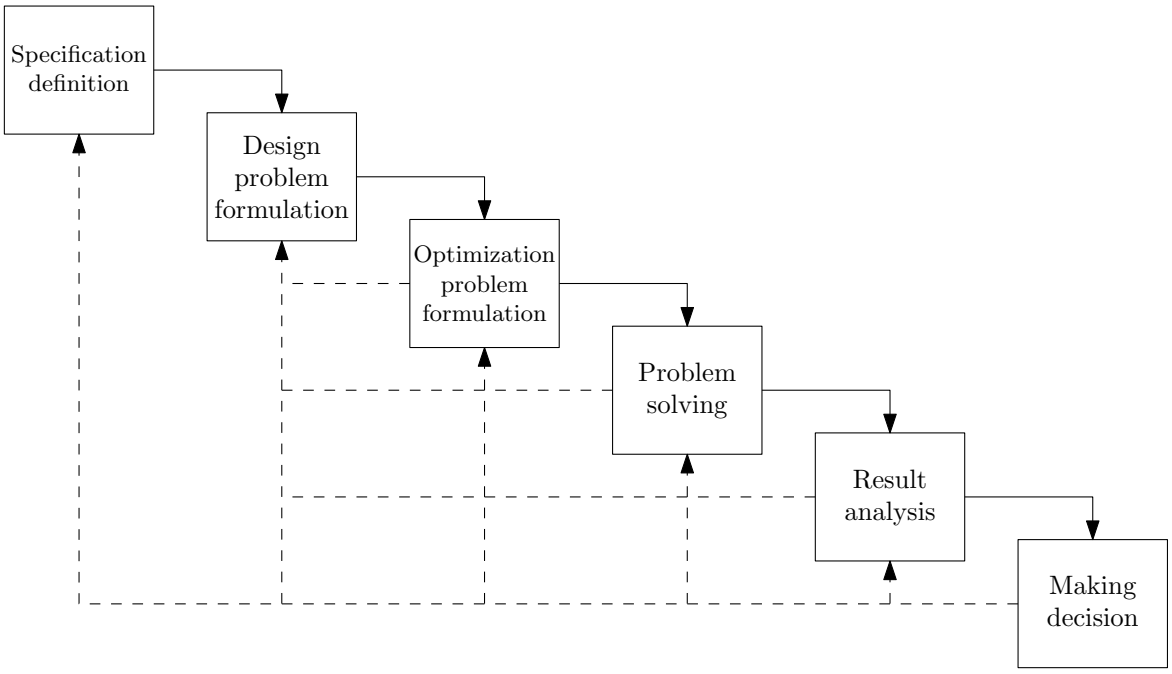


Figure 1.8: Design optimization process

1.1.6.1 Specification definition

This phase consists in collecting requirements and needs. Specification extends from non-technical to technical issues. This phase is essential for the design process because all design teams (engineering, marketing, etc.) are involved and must work together to elaborate all needs, requirements and design objectives.

1.1.6.2 Design problem formulation

After the specification analysis phase, the design problem can be posed. Several design criteria are selected. The other specifications are defined as constraints. These design objectives and constraints are as function of design variables and design parameters. The design variables are controlled by designers during the design process in order to achieve the desired performances expressed by the design criteria. The design parameters are assigned to specific values. They can be determined on the basis of the prior knowledge of the system. At the early design stage, the design problem may consist of a large number of design variables but of very few design parameters since the system is not well known. The more the knowledge of the system

the design gains, the more design variables are fixed. Other interacting systems may also impose some design parameters or constraints to the studied system.

Design problem formulation should not be confused with optimization problem formulation. A design problem can be solved by using different techniques. Optimization is only one of them. As it will be discussed later, a multi-criteria design problem can be formulated as a single- or multi-objective optimization problem. Each technique may lead to a different level of solution as depicted in Figure 1.9. Based on pre-designed or past experiences, a simple table or graph gives rapidly a preliminary design. This solution is hardly optimal. More advanced tools such as CAE can be used to simulate the behaviour of the system. An improved design can be obtained after a few iterations by designer. An optimization algorithm is preferred in order to ensure the optimal solution. In a single-objective optimal design, an algorithm gives only one optimal solution. In a multi-objective case, the solution is given as a Pareto front containing a set of optimal results.

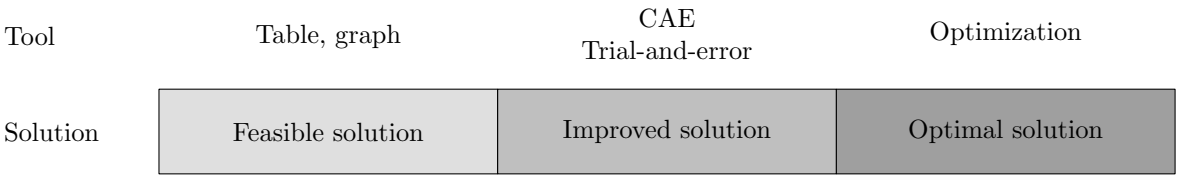


Figure 1.9: Level of optimality

For complex systems, a global system design problem can be decomposed into small linked subproblems by using the decomposition approach. Interactions between subproblems are identified. Design variables and constraints are assigned to subproblems. Therefore, each subproblem deals with fewer design variables and constraints. This helps engineering teams to understand the problems and to solve them more efficiently.

This phase also involves in developing or selecting models. In the complex system design, the model construction process should be done with the optimization purpose in mind. This issue is discussed in Section 3.2.1.

1.1.6.3 Optimization problem formulation

The mathematics formulation of the general optimization problem (referred to as O) is defined as in (1.1)

$$\begin{array}{lll} O: \min\limits_{X} & f_o(X) & o = 1, \ldots, m \\ \text{subject to} & g_i(X) \leq 0 & i = 1, \ldots, n_g \\ & h_j(X) = 0 & j = 1, \ldots, n_h \\ & lb_k \leq x_k \leq ub_k & k = 1, \ldots, n_v \end{array} \tag{1.1}$$

$X = [x_1, x_2, \ldots, x_{n_v}]$ denotes the design variable vector. The V design space is defined by lower and upper bounds for each variable, lb_k and ub_k, respectively. The optimization minimises several f_o objective functions, subject to g_i inequality constraints and h_j equality constraints. The S^* optimal result set, representing group of X^* non-dominated solution, is the subset of the S feasible solution set. These non-dominated solutions in design space

form the NF non-dominated front (Pareto front) in objective space. Figure 1.10 shows the optimization process from design space to non-dominated front.

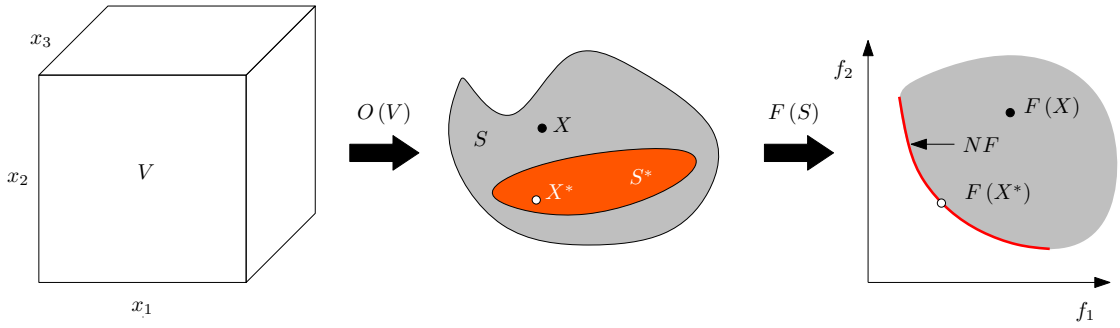


Figure 1.10: Optimization process; from design space to non-dominated front

Usually, engineering design problems are multi-criteria problems. Two approaches can be used to formulate this kind of problems:

- The a priori approach: This approach formulates multi-objective design problems as single-objective optimization problems. The objectives are usually transformed into equivalent objective functions by using transformation techniques e.g. weighted-sum, epsilon-constraint. The designer defines trade-off between objectives before solving the optimization problem. For example, an electrical motor design problem has two conflicting objectives: mass minimisation and efficiency maximisation. By using the weighted-sum approach, the mass weight coefficient may be given 10 and the efficiency weight coefficient is given 1, on the basis of the past experience of the designer. When using the epsilon-constraint approach, the problem is to minimise motor mass while having efficiency superior to 90%. One may realise that it is not very convenient as it requires quite a lot of experience to reach this compromise. However, the *a priori* approach is very useful (i) when refining the existing or well-studied solutions, (ii) to investigate some solutions selected from Pareto front and (iii) for optimal design in detail design phase.

- The a posteriori approach: Optimal solutions on Pareto front are located by an optimization algorithm. Based on this information, the designer or the engineering team selects some interesting solutions to further the investigation. With the *a posteriori* approach, no decision is made before solving optimization problem. It is more practical to make a decision based on trade-off after knowing the Pareto front. For the motor design example above, the optimal results are non-dominated solutions on the Pareto front, which is a trade-off curve between the motor mass and efficiency.

Chapter 3 presents various optimization techniques. Designers should select an appropriate algorithm to solve the optimization problem. It depends on:

- the number of objective functions;

- the unimodal or multimodal problem;

- smooth or noisy functions;
- the type of design variable i.e. continuous, discrete, category (non-classifiable) or mixed;
- the size of the problem i.e. number of design variables. This factor plays an important role for the robustness and convergence properties of the algorithm.

1.1.6.4 Problem solving

The formulated optimization problems are launched. Various parameters of the optimization algorithm have to be adjusted. Optimization algorithm development trends focus on the parallel and distributed computing. Genetic algorithms can benefit from this approach by distributing objective function evaluations to multiple processors or computers [19]. In the Target Cascading method described in Section 3.2.2, a large-scale problem is divided into small subproblems. These optimization subproblems can be solved in parallel, leading to a distributed optimization.

1.1.6.5 Result analysis

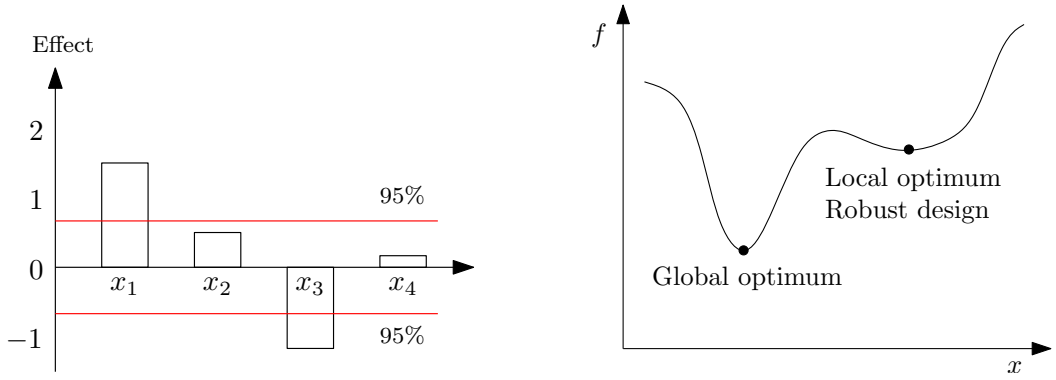
Once the optimization problem is solved, designers should analyse the results. Optimization failures can occur and lead to non valid results. Result design vector and constraint values must be verified against boundary and constraint limits. Sometimes this unvalid results can be suspected from the fact that the system behaves in a non-physical manner. Therefore, designers must have a good knowledge of the studied system.

The optimization technique gives only "intermediate" results to designers. This means that engineers must analyse the given results and make a decision in order to obtain the "final" results. Optimization can provide several forms of results e.g. table, graph, Pareto front.

A Pareto front is a common representation of multi-criteria optimal design results. As discussed earlier, it provides information about trade-offs between criteria. A two criteria Pareto front is comprehensible. A three criteria Pareto front can be represented with 3 projections between any two criteria. They may be given together with a 3 dimension interactive graph i.e. engineers can select the view, orientation and zoom as desired. When more than 3 criteria are involved, a representation may be difficult to achieve. A Pareto front shows optimal results in an objective space i.e. a relationship between objectives. However, other information such as optimal values of design variables and constraint are also necessary. Since there are excessive information, they are usually given as a data table for only several optimal results. Nevertheless, all results obtained during the optimization process must be collected in a database and ready to use if needed.

Sensitivity analysis is a useful tool in the result analysis phase. This analyse is based on the well-known Design Of Experiment (DOE) methodology [29, 50]. By applying sensitivity analysis at optimum design solutions, it provides local relationship between design variables and objectives and constraints. Figure 1.11a shows a bar graph of sensitivity analysis. Design variables with a greater effect than certain statistic thresholds are considered as significant

factors[2]. In some systems, a design less sensitive to changes in the design variable/parameter value[3] may be preferred. Figure 1.11b depicts a graphic representation of local robust and global non-robust optima. A robust system ensures its best operation even in a bad or uncontrolled environment.



(a) Effect of design variable

(b) Robust solution

Figure 1.11: Sensitivity analysis

In DOE, the sensitivity analysis is also used in problem formulation phases. It is called "Screening". Based on a base-line design, the local effect of design variables or parameters is studied. Only significant factors are selected as design variables and used in optimization problems. Other factors are fixed as design parameters. This allows reducing the problem dimensions hence its complexity.

1.1.6.6 Decision making process

In case of a multi-objective problem, optimal results are often provided in the form of a Pareto front. Some of these non-dominated solutions are selected for further study. The chosen solutions can be analysed with high fidelity tools. To select these solutions, one can use other criteria, which were not defined or cannot be expressed as objective functions in the optimization problem.

On the basis of optimal results, designers can also simplify the optimization problem e.g. reduce the study domain, relax the constraint limit and fix one or several criteria as constraint. If a low fidelity tool is used in the optimal design process, the optimization can be therefore re-run by using high fidelity tools in order to verify the obtained results.

[2]In design of experiment methodology, the term "factor" is referred to as "design variable".
[3]Robust design

1.2 Optimal design in the railway industry

In the railway industry, optimal design has been recently used in the development of components [55, 60]. Interaction problems draw attention to extend optimal design to a larger application domain as well as more global systems. Traction system design is emphasized here in this research. This section gives an overview of railway traction systems. A corporate design process is then introduced and analysed in order to discuss on the possibility to integrate optimal design into actual design process.

1.2.1 Overview of railway traction systems

Nowadays, more than 20% of railroads around the world are equipped with electrical supply [21]. Electrical rolling stocks are widely used in many applications such as high speed trains, locomotives, trams, and metros. In a general point of view, a traction system includes all components from pantograph to wheels. Figures 1.12 and 1.13 show typical schema of a railway traction system. Pantograph draws electricity from power supply. Electricity pass through DC supplier, whose main function is to supply a DC source to inverters at the desired voltage level. The inverter then transforms the DC source into a variable frequency AC source and provides it to the traction motors. Mechanical energy is sent to the wheels via gearbox.

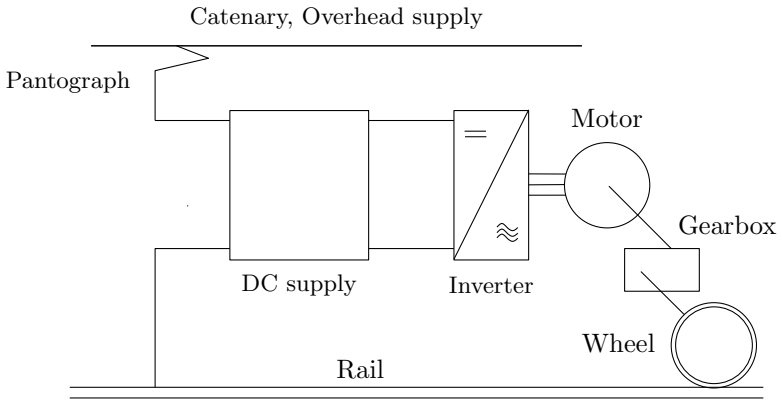


Figure 1.12: Schema of railway traction system

1.2.1.1 Power supply

According to countries and applications, there are many type of power supply as shown in Figure 1.14. High voltage AC sources are suitable for long route and high power trains such as EMU (Electric Multiple Units) and VHST (Very High Speed Trains)[4] due to low transmission losses in the overhead supply. On the other hand, lower voltage DC sources are more common in short route and low power applications such as trams and metros. This allows to reach a favourable compromise between energy losses and traction system cost.

[4]Very High Speed Train (VHST) or Train à Grande Vitesse (TGV) in French.

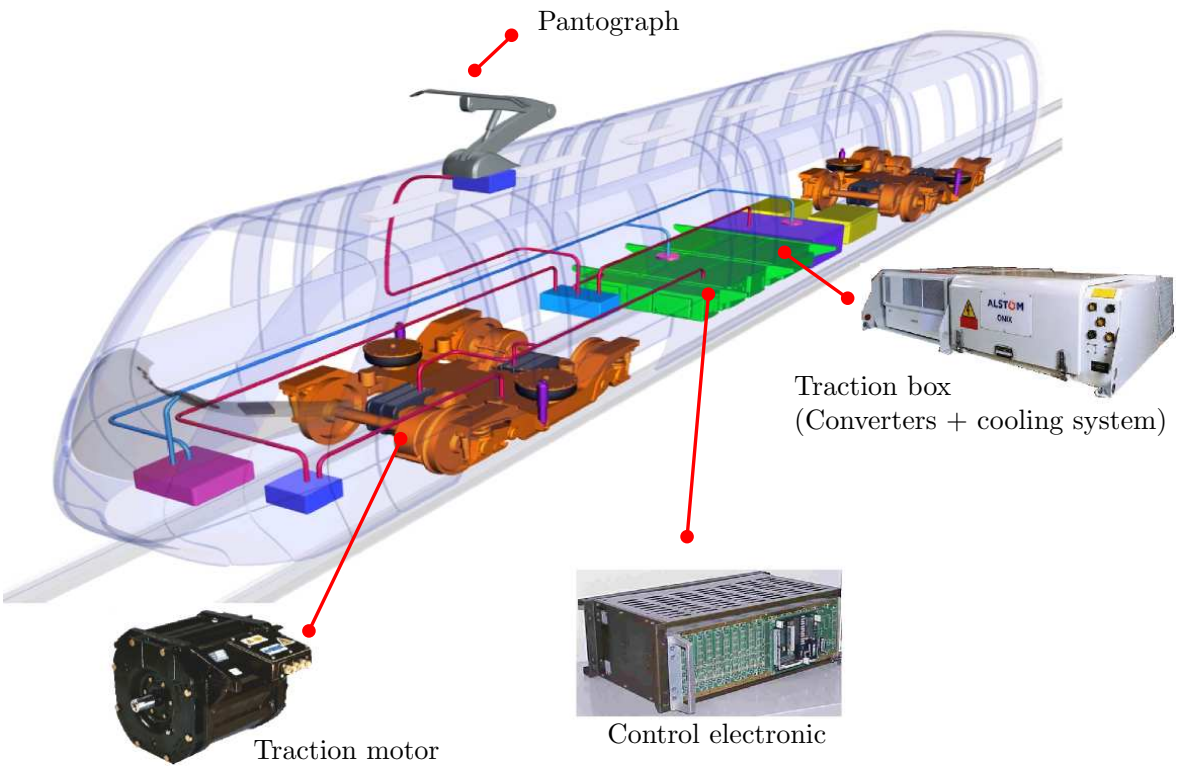


Figure 1.13: Main subsystem/component of railway traction system

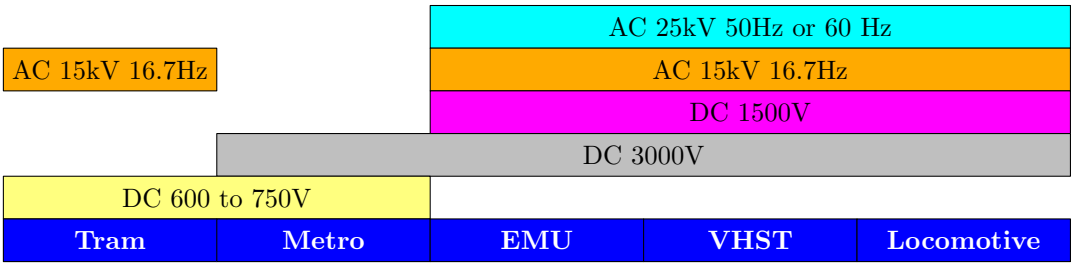


Figure 1.14: Power supply

1.2.1.2 DC supply

The inverter requires a DC source with a certain level of voltage as input. For the DC overhead supply, a step-up or step-down chopper is used in order to change the voltage level of the DC source into the voltage required by the inverter. A DC filter is also an important component preventing harmonic currents from the DC link from going back to the power network. For the AC supply, a transformer and a rectifier are required. In multi-source or multi-voltage traction systems, the transformer can also be used as the harmonic filter when the train is operated with a DC source.

1.2.1.3 Electric braking

There are two types of electric braking: regenerative braking and rheostat braking. The regenerative braking is more interesting because it lowers the energy consumption. However, a braking rheostat resister must always be onboard for safety reasons. The rheostat braking

system consists of a chopper connected to a DC bus and a braking resistance with a cooling fan.

1.2.1.4 Variable speed converter

In recent traction systems, AC traction motors have been used. A DC/AC converter or 3-phase inverter is necessary in order to vary the motor speed. In the past, switching components such as Gate Turn Offs (GTO) and Silicon Controlled Rectifiers (SCR) were preferred in high power applications. They are now obsolete and replaced by Insulated Gate Bipolar Transistors (IGBT), which are of high efficiency and fast switching [66].

To evacuate losses and to maintain an acceptable operating temperature of the IGBT, inverters are always equipped with a cooling system. Depending on the customer requirements and power losses, the cooling medium can be natural air, forced-air with fan, or water.

1.2.1.5 Traction motor

Squirrel cage induction motors are used in many traction systems. This type of motor offers a high reliability and simplicity in terms of manufacturing and maintenance. New development trends are leaning towards permanent magnet synchronous motors. Their high power density and high efficiency are very attractive in traction application. They are the best solution for distributed traction systems, where the space in bogie is very restrained. However, the cost of these motors is still higher and their industrialization phase is more complex due to the lack of experience in railway applications.

1.2.2 Decomposition of train systems

According to the business unit management organization of the company, a train design problem is decomposed into several problems as in the example shown in Figure 1.15. They are coordinated by system engineers at the train level. The system engineering team decides on the allocation of several targets such as reliability, mass, volume and cost. These targets are assigned to subsystem engineering teams including e.g. the bogie, carriage and traction system. One department is in charge of the traction system design and manufacturing. Inside this department, the traction system design problem is also decomposed into subsystem[5] and component design problems (traction box, motor, transformer, etc.). At a lower level, a traction box contains several converters (rectifier, inverter, chopper) and their cooling system. Each subsystem/component design problem is solved by several engineering units and suppliers. The decomposition shown in Figure 1.15 is object-based. However, another approach can be simultaneously used depending on disciplinary expertise.

One may observe that in traditional design processes, communication and data exchange (i.e. target allocation, constraint specification) between the different levels are done iteratively and thanks to meetings of the different design engineers. Usually, few iterations should lead to an "improved solution". For example, the traction system department at the subsystem level imposes mass target to transformer suppliers at the component level. If the supplier

[5]In this case, it means subsystem at lower level than its parent.

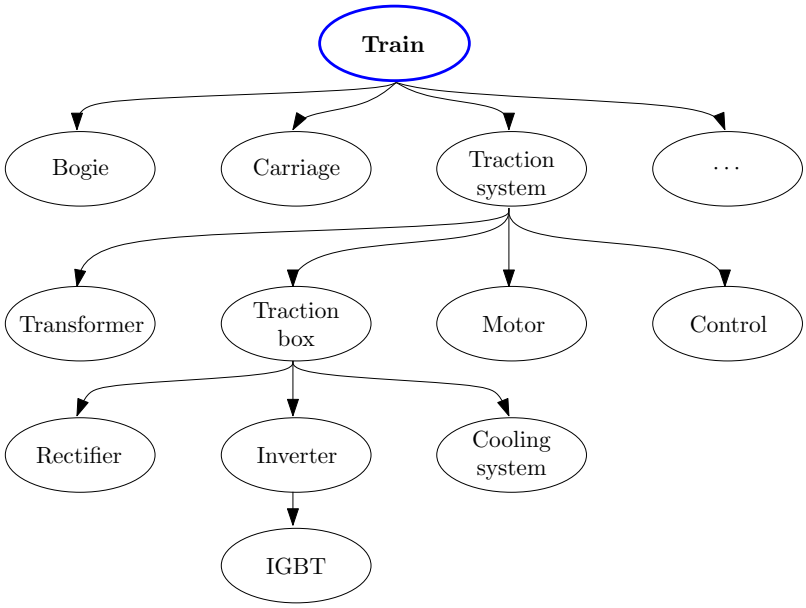


Figure 1.15: Object-based train decomposition

cannot meet this requirement, a compromise will be sought e.g. modifying the traction system design in order to tolerate the constraint; the supplier reviews its transformer design while the customer consults other suppliers. Actually, the problem is much more complex. A concession made with a supplier can affect the design of other components. Interaction problems sometimes lead to high development time and cost.

1.2.3 Corporate design process

The traditional design process of a rolling stock company is composed of three main phases as shown in Figure 1.16. The technology phase concerns the research, observation, selection and validation of new technologies e.g. IGBT semiconductors, permanent magnet motors. In the second phase - the product phase - systems and subsystems are designed and validated. These designs are based on state-of-the-art technologies and new technologies validated during the first phase. The result of this phase is for example an IGBT-based inverter, a tram traction system using such inverter and permanent magnet motors. The systems developed in this phase is added to an internal product catalogue. The third phase is called the contract phase (applicative project phase). It starts when the company concludes a contract with a customer, concerning, for example, an intercity train for a railway operator. The applicative project team designs a train, using the products in the catalogue as much as possible. However, some components must be customized in order to meet some specific requirements of the customer.

For each main phase, the company uses the V-cycle process, as described in Section 1.1.1. Figure 1.17 shows V-cycle for product development phase. Each phase is finalised with a Gate Review (GR). GR is the decision-making stage in which design teams review the design and have to decide whether or not to go on with the next phase.

For traction system development, the main duties of the traction system department are to design and manufacture standardised traction systems, which are mainly in the second

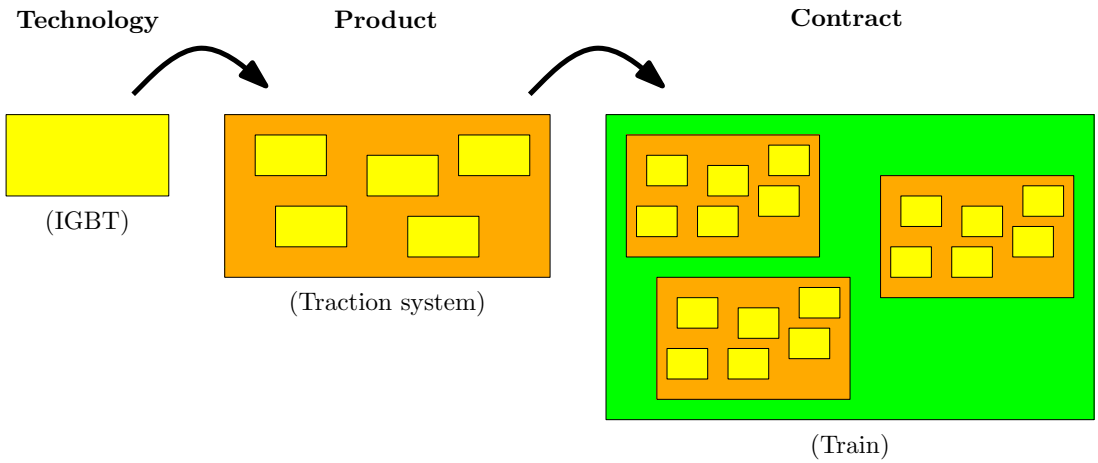
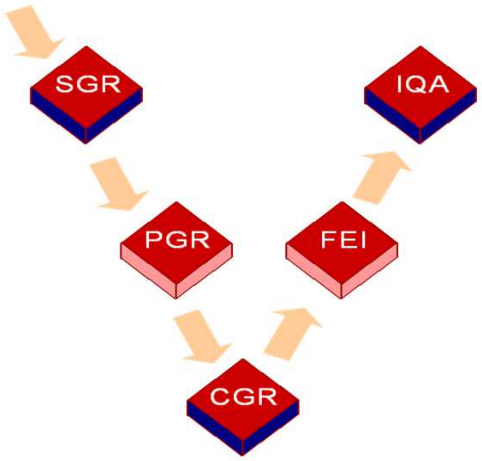


Figure 1.16: The three main design phases of a company



- Specification Gate Review (SGR)
- Preliminary Gate Review (PGR)
- Critical Gate Review (CGR)
- First Equipment Inspection (FEI)
- Initial Quality Approval (IQA)

Figure 1.17: V-cycle product development phase

phase (Product development phase) as well as in the first phase (Technology phase). The input data for traction system development are provided by marketing department. Previous contracts and potential projects are analysed in order to determine the specifications and constraints. Several applicative projects are chosen as reference projects and improving targets are defined. As suggested by their names, each GR concerns different objectives. SGR intends to determine specifications, needs and constraints of the traction system and components. During the PGR phase, preliminary designs are carried out. Practically, all design parameters involving in the main function of the traction system, as well as all interactions between the systems are determined in this phase. More detailed designs are completed during the CGR phase. In the last phases, prototypes are produced and tested in order to confirm the design.

Figure 1.18 shows the Degree Of Freedom (DOF) and modification cost against time. Product and contract phases appear in the figure. In order to obtain an optimal system with a low cost, the system should be well developed and take into account all possible customer needs during the contract phase. Once the subsystems and components are added into the catalogue, it is inappropriate to modify them, due to cost constraints. This consideration

leads to the standardisation approach presented in the following section.

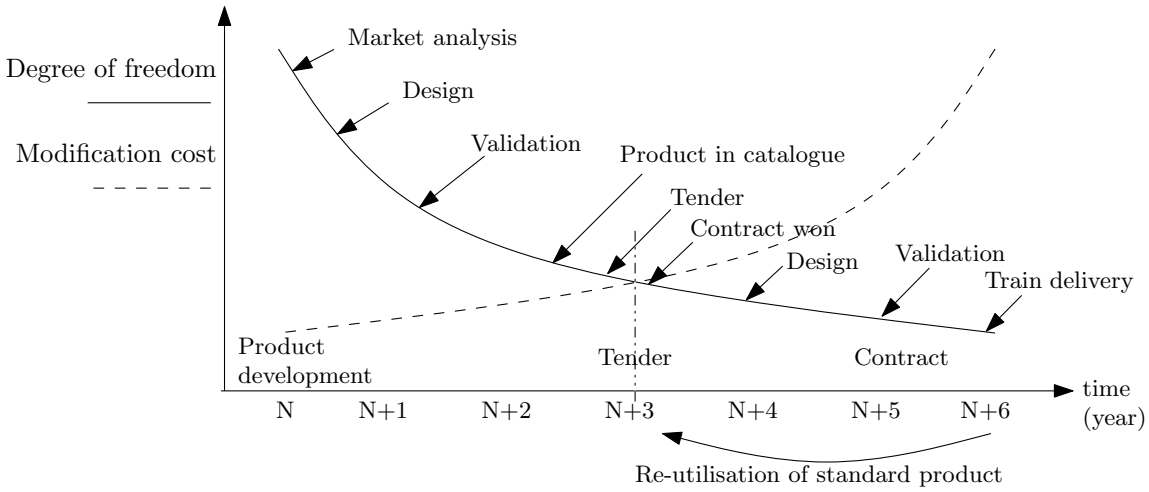


Figure 1.18: Degree of freedom and modification cost

1.2.4 Standardisation

In the context of global market, customized design for each product leads to non-competitive cost and a long time to market. Many industries are considering platform or family product design approaches [100]. A product platform is a group of products, which shares subsystems and components. Even if many subsystems or components are common for the platform, options can be selected and some components are designed specifically for each product in order to satisfy the requirements of customers.

In the rolling stock business, three main platforms can be defined: the carriage platform (tram, metro), the bogie platform and the traction system platform. Platform standardisation can be applied in two aspects:

(i) Standardisation for applicative projects: A platform developed for an applicative project is re-used in other applicative projects in the same product range/platform. For example, a tram traction system can be used in many tram applicative projects.

(ii) Standardisation for platforms: This is done at the component level. A component used for a platform can be used in other platforms. For example, an inverter/power module can be used in the traction system of a metro and EMU.

Figure 1.19 shows cost, development time and optimality level compared to the level of standardisation.

Three examples of this in traction system designs are:

(i) **Low level of standardisation:** This approach was used in the past. A unique traction system is developed for each applicative project. Only few components are re-used later in other applicative projects. A number of components is designed for a particular project and never used in the following projects.

Advantages:

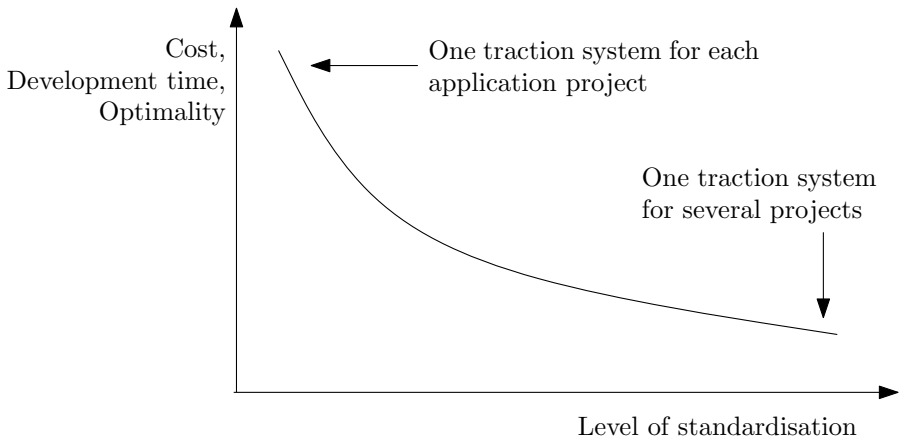


Figure 1.19: Standardisation

- The traction system is highly "optimised" for each project.

Drawbacks:

- High development cost and time
- High development time
- High product cost due to low quantity manufacturing
- High reliability cannot be guaranteed

(ii) **Medium level of standardisation:** This level represents today state-of-the-art. Traction systems are developed for each rolling stock product range i.e. tram, metro, EMU, VHST and locomotive. According to the applicative projects, each traction system can be reconfigured and adapted in order to satisfy specifications made by customers. For example, the architecture of the power scheme and the choice of power converter can be adapted, on the basis of each project. Some components such as transformers and harmonic filters are newly designed to meet the specifications for the projects. The traction system power schemes can be of different AC or DC, voltage level, depending on the countries, the power supply networks. Moreover, the harmonic constraints are also different. This leads to customized designs of the transformer and harmonic filter for each project.

Advantages:

- Lower development cost
- Shorter time to market because the most part of traction system design is already done during the early design phase.
- Lower product cost, due to a higher quantity production
- More reliability. The Return of experience from previous projects allows improving the product reliability. Note that, traction systems are permanently improved, along their life cycle.

Drawbacks:

- A traction system cannot be "optimal" for every project. It is oversized for some projects.

(iii) **High level of standardisation:** An ideal traction system is suitable for all rolling stock product ranges. A more practical idea may be a traction system for several product ranges. Different power or effort requirements can be achieved with redundancy. A design must take into account as many constraints and requirements specified by customers as possible. This approach has been applied by Alstom in AGV (Automotrice à Grande Vitesse) and Pendolino train, which share the same traction system. A great number of components are identical. Based on the modular approach, the traction power can be adjusted according to the project needs. Moreover, a transformer can be used as a DC filter when the train operates in DC power supply. It is designed by taking harmonic constraints for almost any network into account. It then becomes a "standard" component and is not a "customised" component anymore, as in the previous example.

Advantages:

- The development cost is split on several projects.
- Shortest time to market because most designs are already finished before the tender phase. Only small adjustments are to be done before the delivery.
- Lowest component cost due to a scale effect
- More reliability. The Return Of Experience from the previous projects allows improving the product reliability.

Drawbacks:

- A traction system cannot be "optimal" for every project. It is oversized for some projects. However, profit is maximised at global company level.

1.2.5 Design criteria

The traction system design problem is a multi-criteria design problem. Some of the design criteria can be stated:

- Life cycle cost (LCC): It is composed of the purchase cost and possession cost e.g. maintenance and energy consumption costs. LCC should be an important criterion for customers. Regarding the energy consumption cost, it is an implicit function of another criteria. This allows simplifying multi-criteria problems to single-criterion. Minimisation of energy consumption can be achieved through efficiency maximisation and mass minimisation (reduction of resistance effort of train).
- Development cost: The maximisation of the standardisation level allows minimising the development cost.
- Mass: As mentioned earlier, the minimisation of the mass leads to minimisation of the energy consumption or the maximisation of passenger capacity if the axle weight remains the same (e.g. 17 tons for the VHST).

- Volume: The minimisation of the traction box and motor volumes is very important for small rolling stocks such as trams as well as full size vehicles such as metros or EMUs. For AC traction systems, a small transformer is also preferred. Any volume minimisation allows increasing the passenger capacity.

- Reliability: Reliability is defined in terms of failure rate. For trains, the service reliability[6] is usually expressed. The failure of a traction box or a traction motor should not lead to the immobility of train. The train reliability can be improved via a component redundancy as well as the reliability of each component.

- Noise: The main source of noise in traction systems is the forced-air cooling system (for the inverter, rheostat resistance, transformer, filter). The actual solution is to stop the ventilation at the stations. A more effective solution might be a loss minimisation allowing the use of less noisy cooling fans. The inverter and traction motor are also noise sources. An optimal inverter control strategy [107] and an optimal geometry of the motor can minimise the audible noise.

- Electromagnetic compatibility (EMC): According to the rail network (country, railway operator), the harmonic currents rejected to supply network are limited. This ensures that the communication network and the signalisation system will not be affected.

1.2.6 Model and analysis tools

To design a traction system, a company uses a wide range of models and software. For electric and thermal phenomena at the traction system level, several rules are used in the preliminary stage in order to determine the initial variables. The preliminary results are then verified, with the help of an in-house traction system simulation program called CITHEL. This tool allows simulating the cinematic, electric, and thermal behaviour of a train operating on the track e.g. a round-trip of train. It uses a library of standard components and focuses mainly on the traction system. More detailed models of component are available depending on the speciality of each engineering team. Figure 1.20 shows CITHEL's user interface. Section 5.1.2 presents its features in a more detailed manner.

Design engineers also use standard commercial software as well as Finite Element Analysis (FEA) tools in mechanical and thermal analysis e.g. ANSYS®, CATIA®, FloTHERM®. They are used at different stages as shown in Figure 1.21.

Some design criteria are not associated to any analysis tool. Reliability is computed, on the basis of estimated value (from previous experiences), the Return Of Experience (ROE) or value given by supplier. The FMECA (Failure Mode, Effects, and Criticality Analysis) and the Reliability Block Diagram and Fault Trees Analysis are also deployed.

In this work, a model of permanent magnet motor has been developed. It is an semi-analytical model with link to geometry (see Appendix B). Its exploitation will be shown in Part II.

[6] Service reliability is described by additional time used to accomplish a trip.

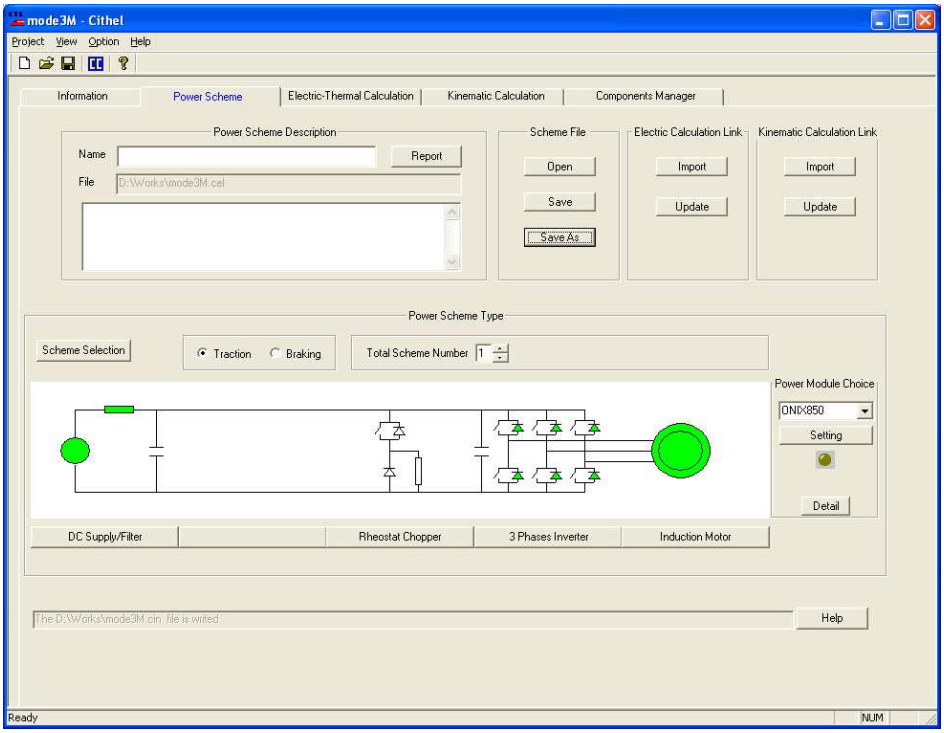


Figure 1.20: CITHEL's user interface

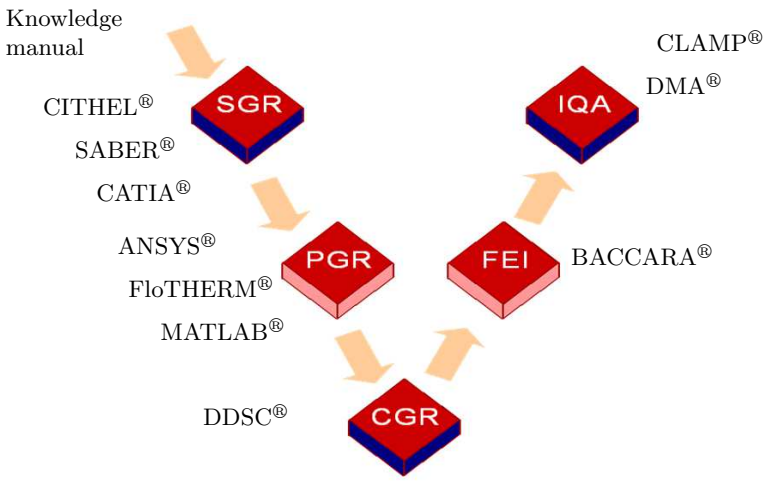


Figure 1.21: Various tools in different design phases

1.2.7 Optimal design consideration

In order to apply and integrate an optimal design process into the corporate design process, several points may be discussed.

For design problems of decoupled parts of traction systems, the information exchange between subproblems should be carried out more or less automatically in order to achieve an optimal global system. Subproblems must be integrated and coordinated by a hierarchical optimization method. This issue will be discussed in Chapter 3. To integrate all optimal design subproblems into a global hierarchical one, various models must be compatible and must be launched from an optimization supervisor. Some difficulties may be encountered in reality e.g. the different teams might not be on the same site or, worse, they work in different countries; the suppliers do not want to communicate their models; or design software are not compatible. In any cases, surrogate models are an efficient communication tool (see Chapter 2).

Regarding the whole design process, it is clear that an optimization can operate in any phase. In the contract phase, it can be used to adjust some parameters, and this allows satisfying customer's requirements. However, it is more effective in the early design phase, where the problem is more complex, a large number of design variable are accessible and interaction problems are not well studied. Optimization can be applied to low fidelity models. the results from this preliminary optimization provide sufficient information to understand trade-off between several design criteria, to fix some variables as parameters and to simplify the problem.

The optimal design of standardised traction systems is of main interest, particularly if a high level of standardisation is aimed at. It is obvious low cost, hence high profit. The problem should be well defined since it is more complex than usual. Traction systems must meet the requirements of a wide range of customers as well as different operating scenarios.

1.3 Conclusion

The design problem of complex systems has been presented in this chapter. A complex system is a system integrating several interacting subsystems and components. In the "business as cause" approach, the design of the system is decomposed into several smaller design subproblems. Once subsystems and components are defined, they are coordinated and integrated back by the system engineer.

Models and software packages used in design are mainly simulation tools providing behaviours. Therefore, iterations are required to solve the design problem.

Usually a complex system design problem deals with a large number of design variables and constraints. The usual trail-and-error method is not suitable as engineers cannot deal with such complex problems. An optimal design is then required to help engineers to obtain results better and faster. However, the engineer will stay to drive the process, devoting the repetitive tasks to the computer and keeping in mind the decision-making process. With a well-formulated optimization problem, a suitable algorithm and a trained designer, the optimization technique can ensure optimal results. An optimal design should be superior to a conventional design method in terms of design time delay. The conventional trail-and-error method hardly gives optimal results and design problems are always simplified. By using an optimization technique, the design problem can be more complete and meet the requirements of the designer in a better way.

The optimal design process has been presented. It takes place in a V shaped design cycle. The design problem is firstly formulated. Modelling or model selection, if possible, should consider the fact that the models will be used in an optimal design. The optimization problem is formulated and solved using various optimization techniques. Results may be given in several forms. A Pareto front is common for a multi-criteria design problem. It allows obtaining a trade-off curve between objectives. The table gives more information e.g. design variables and constraint values and is suitable for investigating several solutions in detail.

In railway applications, an overview of traction systems has been given. The main components are a DC supply (transformer, rectifier, chopper), a line filter, an inverter and a motor. Then, a corporate design process has been presented. Some interesting points regarding optimal design and its place in the product development cycle have been discussed. Design criteria and simulation tools have also been reviewed.

Following the chapter, Chapter 2 presents several modelling techniques and Chapter 3 Optimization techniques. Finally, Part II describes some applications of optimal design to both railway traction systems and components.

Chapter 2

Modelling technique

Computer Aided Design (CAD) and Computer Aided Engineering (CAE) are used in many industrial domains. They allow simulating the behaviour of physical systems. Therefore, computer simulations are widely used by engineers when designing products. Development cost can be significantly decrease since the application of the trail-and-error process has a lower cost, when using these computer simulations as virtual prototypes.

In the last decades, scientists and engineers have put a lot of effort to model engineering systems based on the knowledge of basic physical phenomena e.g. magnetic, electrical and mechanical laws. Experiments are necessary in order to identify the model parameters, verify modelling hypotheses, and assess the model accuracy. The modelling can be very simple or very complex depending on details needed and the complexity of the system itself i.e. the level of interaction between the physical domains and the number of components to be simulated simultaneously. The obtained models are used afterwards to simulate or study the behaviour and performances of the studied systems.

In some cases, mathematics modelling techniques replace physical-based modelling, for example when:

- Physical phenomena are very difficult to understand or not well established.
- The system being studied is very complex.
- Experiments are available.

Mathematics modelling techniques are closely related to the statistics field. A well-known technique is the Design Of Experiment (DOE) methodology. It allows modelling a real system based on a limited number of experiments. The DOE is popular and has proved its effectiveness in a wide range of applications such as in chemical and industrial processes [9]. In statistics literature, another technique can also be stated such as the Bayesian network or the Kriging method [13]. Outside the field of statistics, famous mathematics modelling techniques are for example the Neural network [65] and the Splines [42].

Although mathematics modelling techniques are initially developed for "Real-world" experiments, they can also be used to model "computer simulation". When doing so, the term of **Surrogate model** or **Metamodel**[1] are employed. In this research, we emphasize our

[1] Model of model

interest in the modelling of experiments done on a computer (i.e. simulation). The modelling technique and mathematical formulation are shared between the Mathematics modelling and the Surrogate modelling. The main difference is that the Mathematics model is applied to real experiments and the Surrogate model is applied to computer simulations obtained by using a physical-based model, which can be reproduced without an experimental error. The idea behind the Surrogate model approach is discussed in Section 2.2. This reflects our main concern on CAD/CAE and optimal design processes. As computer technology has been improved on a daily basis, engineers tend to develop more and more complex physical models, which take into account many phenomena. However it is not very practical to include these models in the optimization process because the computations are usually time-consuming. Therefore, an approach using surrogate models is necessary. This kind of model is constructed using information from a physical-based high fidelity model. It replaces the high fidelity model in the optimization process and its sufficiently low computation cost allows to speed-up the process.

Figure 2.1 depicts a summary of the idea introduced earlier. It also gives a summary of various physical-based modelling techniques usually employed in electrical engineering domains as well as several surrogate modelling techniques presented in the following sections.

2.1 Physical-Based Modelling

The physical-based modelling technique is certainly the most common approach used by engineers. A system or process is modelled based on the knowledge of physical phenomena[2] and equations describing them. The experiment is always carried out to identify some model parameters, to validate the hypothesis posed and to verify the accuracy of the model.

In electrical engineering and related fields, three main techniques are used frequently; analytical, numerical, and semi-numerical modelling. Each technique has got its own advantages and drawbacks. There is no technique that is better than another. It depends on how it is used and what application is made of it. For example, an analytical model of motor is very suitable for pre-sizing and a numerical model such as the Finite Element Analysis (FEA) is used for detailed design phase.

2.1.1 Analytical model

Analytical models are expressed explicitly in terms of analytical equations. These analytical equations can be solved without any difficulty. The analytical model can be represent a system at different detail levels. This depends on the applications and needs. Again, there is not any model better than another. A model is more or less suitable in a context. For example, a motor can be very low detailed with a circuit model as used in control application and more detailed with an electromagnetic model with linked to geometries in a sizing application. Because of their nature, explicit models are very cheap regarding their computation time. Therefore, they can be integrated successfully in an optimal design process [60].

[2]This should not be limited to physical phenomena. It can also be economics or else depending on the theory describing the system being investigated.

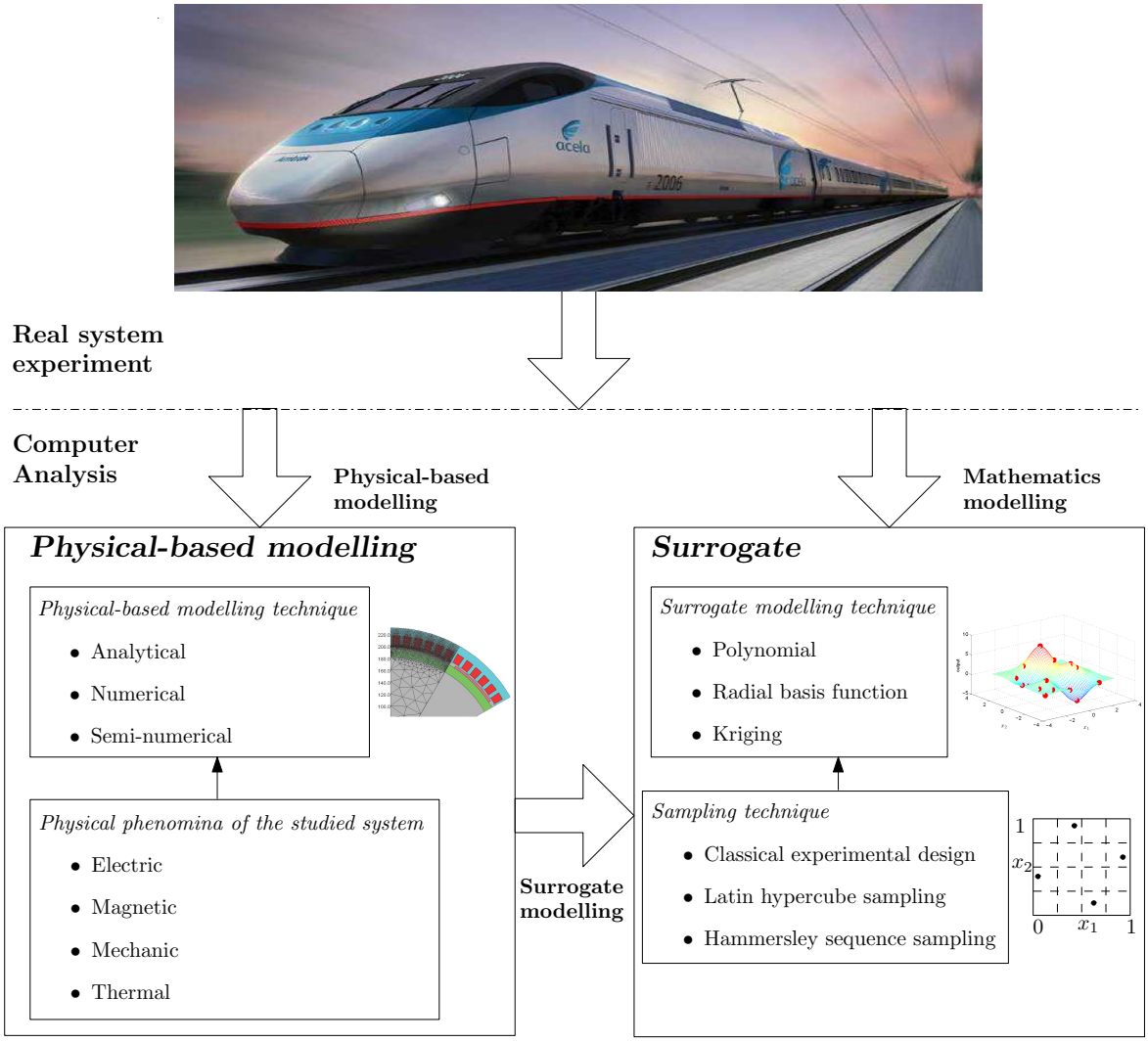


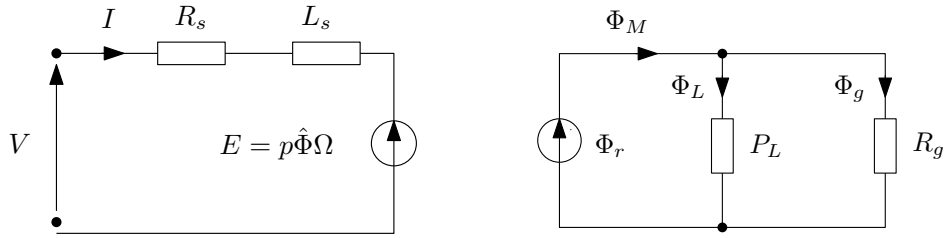
Figure 2.1: Summary of the presented modelling techniques

An example in electrical engineering applications, especially in motor designs, is presented in Figure 2.2. This example shows an analytical model with two interconnected sub-models. A simple circuit model shown in Figure 2.2a represents the electrical domain and an equivalent magnetic circuit model with linked to the geometry (Figure 2.2b) describes the magnetic domain. Both models are used together to simulate performances of a motor. Many hypotheses are considered e.g. the magnetic flux is non-saturated, the phase current and flux density in the air gap are sinusoidal and the leakage flux is introduced via an empirical coefficient.

2.1.2 Numerical model

To solve the nonlinear characteristics of a numerical model, an iterative method is required. They are for example, numerical, semi-numerical and time-domain simulation models.

Many numerical techniques are used nowadays in our activity field: the Finite Element Method (FEM) [115], the Boundary Element Method (BEM) [91] and the Finite Volume Method (FVM) [25]. The FVM is mainly applied in the field of Computation Fluid Dynamic



(a) Electric model (b) Magnetic model

Figure 2.2: Analytical model of a permanent magnet motor

(CFD). The FEM is commonly used in the application of magnetic field computation (magnetic domain) and stress analysis (mechanical domain). These methods consist in dividing an object into small elements via a mesh generation process. Governing partial differential equations are then discretized over this mesh and solved numerically. Such governing equations are for example Maxwell equations in magnetic field computation [11], the elasticity theory for stress analysis [16] and the Navier-Stoke in CFD [110]. Numerical models allow taking into account nonlinear characteristics of materials and many phenomena, which are neglected in explicit models. This increases the precision and accuracy of the analysis results. They can also give local information on the system e.g. the magnetic saturation in a small zone of tooth. Furthermore, the combination of several physical domains in the same solver is possible i.e. solving the coupled magnetic, thermal and mechanical domains simultaneously [24, 68, 90].

The development of numerical software packages seems to be a demanding task. Fortunately, many commercial software packages are available on the market. Some of them can be listed: Opera, Flux, ANSYS, FLUENT, FloTHERM, CosmosWorks and COMSOL. In the point of view of the user/designer, using such software to model a system can reduce significantly their modelling effort, while no deep knowledge of the theory is needed. However, the computation time can be very high. Figure 2.3 shows modelling and analysis results of a motor using FEM package Vector Field Opera 3D [22].

Unlike numerical models, a semi-numerical model is expressed by using analytical equations with some nonlinear characteristics, which are needed to be solved iteratively [11]. Such a model offers a good compromise between computation time and precision because it does not require a huge number of assumptions as in an explicit model (due to its nonlinear property) nor provide a lot of local information. Well-known semi-numerical models are the permeance network (magnetic field computation) and the lumped parameter thermal model (thermal, heat transfer). Figure 2.4 shows the permeance network of a synchronous generator taken from [88].

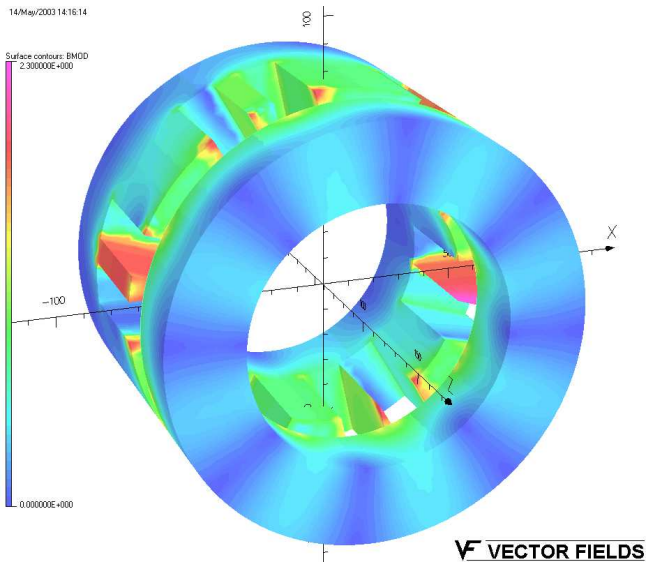


Figure 2.3: Finite element model of a motor

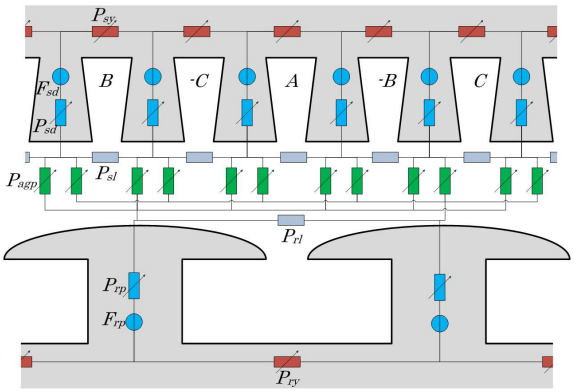


Figure 2.4: Permeance network of a synchronous generator

2.1.3 Summary

The model development is always a very long and delicate task. It requires a lot of time, financial resources, knowledge and experiences in the application domain. This is always an experts job. For analytical models and semi-numerical models, this development phase is realised by experts. The models are usually proprietary or in-house coded. For numerical models, designers commonly use software packages already developed by the software company.

Four criteria are compared in Figure 2.5:

(i) **Modelling details**: The more the model is described in detail, the easier it is to access the local information. It is supposed that if the local information is available, the global one is also obtained.

(ii) **Accuracy**: The model can be build with different levels of accuracy.

(iii) **Computation time**: The model should be selected on the basis of the needs of the designer. The model with too much information may not be suitable for certain applications.

(iv) **System description**: The ability of each component to react and to interacts with the system and the other components around it.

The numerical technique is more accurate but it does not describe the interactions with the system very well. The analytical technique is rapid and able to describe the system. However, it hardly provides detailed local information. The developing trend is to combine these techniques in order to obtain both high accuracy and description of the system. Several examples of hybrid model can be listed:

- The transient finite element analysis, which uses an electric circuit model and mechanical dynamic model (both are time-domain simulation models) and the FEM to simulate the performance of a motor in transient state.
- A system of two components. One is described by an analytical model and the other by a numerical model.

The integration of modelling techniques can be done directly and transparently in one software package or indirectly via an external process such as multidisciplinary design analysis (see Section 3.2.1).

A prototype is also compared in Figure 2.5. It allows validating the modelling due to the fact that all physical phenomena and all interactions can be observed. However, development time and cost are the main drawbacks.

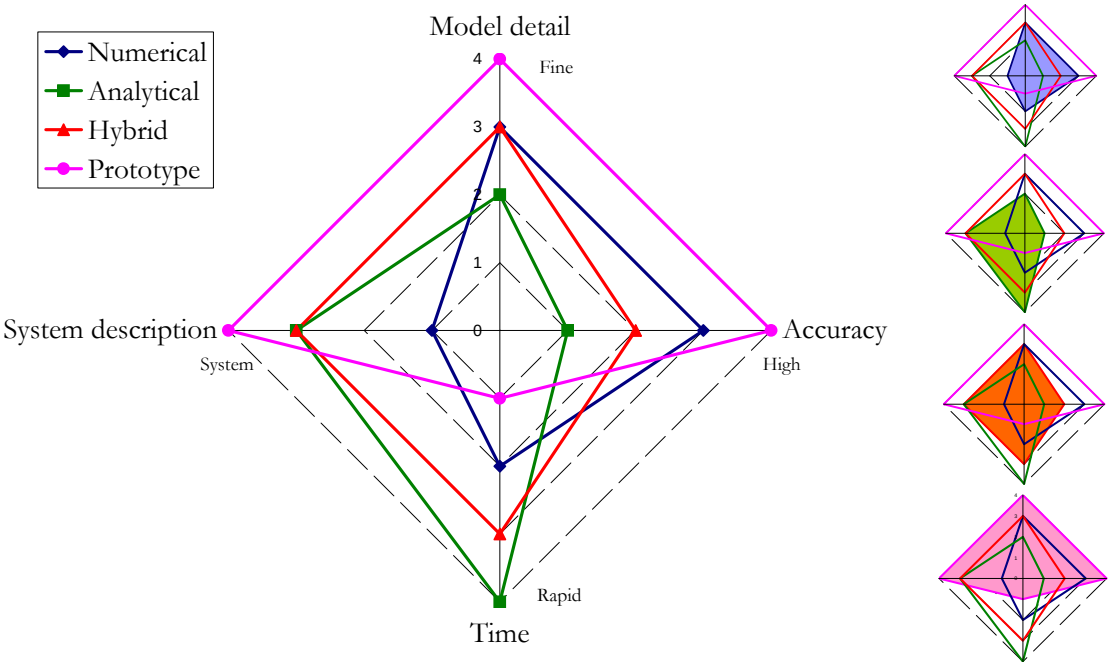


Figure 2.5: Summary of physical-based modelling technique

2.2 Surrogate modelling

2.2.1 Overview

In spite of the increase of computers speed, engineers always find it difficult to integrate the high computational time and high fidelity simulation software constraints in the optimal design process. Many researchers are working on the optimization algorithm to decrease model evaluation number requirements. For example, space mapping techniques allow combining 2 models, high and low fidelity [104], while hybrid algorithm allows taking advantage of global evolutionary algorithms and local gradient algorithms [61]. Others focus on the development of an approximation method called the Surrogate model [67, 92, 99, 112]. The most popular techniques involve the Response Surface Methodology (RSM), the Radial Basis Function, the Neural Network and Kriging.

A Surrogate model or metamodel is simply a model of a model. It is constructed using the sample data computed by a high fidelity software (or fine model). It replaces the time-consuming fine model in many circumstances. The use of a Surrogate model in engineering design can be described as following:

- **Design space exploration:** CAD/CAE tools or physical-based models allow engineers to obtain the relationship between design variables and system performances. However, when designing a new or even well-known system, engineers do not have a perfectly clear idea of how performances would change if design variables are modified. A common approach is to perform a sensitivity analysis around an interesting point. To capture the global idea of design space, engineers can use surrogate models to visually explore the design domain. Design space can be plotted in the way that they need and as many times as they want because working on a surrogate model is very cheap regarding the computation time.

- **Problem formulation:** When designing a system, engineers may have specifications or requirements, but actually they may not know how to formulate the optimization design problem. At this stage, using a high fidelity model is very time-consuming. By using a surrogate model, many formulations can be tested. A problem may be formulated as multi-objective problem at first and then reduced to a single-objective problem and vice versa. Some non-active constraints may be removed and some forgotten constraints may be added as well. Another potential is using a surrogate model to perform screening or sensitivity analysis. This allows selecting the most influential design variables of objective functions.

- **Data exchange and communication:** In the context of multidisciplinary and concurrent design, many groups of engineers work together toward the same global goals. Interactions between groups are unavoidable. A group may need information or a model from another group in order to perform its design. A Surrogate model offers a cheap and confidential means to exchange data and models. A company can avoid giving a proprietary code to its partner or subcontractor.

- **Optimization:** The use of a Surrogate model in an optimization process can reduce the computation time compared with a direct optimization using a high fidelity model. Concerning the accuracy issue, in most cases, the designer performs sequential optimization manually in order to obtain high accuracy results. Such a process will be discussed later in this section. A more advanced optimization algorithm called surrogate-assisted optimization, allows automatically refining a surrogate model during optimization. It is introduced in Chapter 3.

  In the context of Multidisciplinary design optimization, one can profit from the advantages of the surrogate model not only its low computation time but also its compatibility issue. MDO involves many tools and software. Sometimes they are not on the same operating system platform. A designer with no computer/technical skills can construct surrogate models based on various tools running on different platforms and use them to perform an MDO on his preferred computer environment.

Regarding the purpose of optimization, Figure 2.6 shows the typical flowchart of an optimization process using surrogate models. This optimization approach begins with the selection of sample points. The expensive models are then evaluated at these points. The responses or outputs are therefore used to construct surrogate models. These surrogate models replace the expensive models in the optimization process. The optimal results are verified with the high fidelity model. It leaves to the designer the task of reducing the design domain and adding some additional sample points. This manual sequential process runs until the errors between two models are corrected. Finally, the optimization results are analysed and validated.

The optimization approach using a surrogate model is typically used in a large-scale optimization problem such as MDO [39, 102]. In electrical engineering optimization design, the surrogate-modelling approach has been used in many applications. Maruyama et al. [69] applied the Kriging model in the reliability design of electromagnetic devices. An application to a permanent magnet motor design was presented in [32].

A great advantage of the Surrogate model is the reduction in the computation time. However, the main inconvenience is the trade-off between the accuracy and the computational time. Once the Surrogate model is built, it should be validated. The assessment strategies of the surrogate model can be found in [72]. A statistical method called “Leave-k-out” is used to verify if the building of a surrogate model is sufficiently accurate. The surrogate model denoted “Reference model” is firstly built using information from all sample points. The k sample points are randomly left out from the sample point set. A surrogate model denoted “Leave-k-out model” is constructed using the remaining sample points. The “leave-k-out” model is verified against the reference model. An assessment vector such as Root mean square error, Maximum error are computed. By doing this many times, average values of assessment vector can be considered as an approximate error.

The goal of a surrogate model construction is to obtain a cheap-to-evaluate model representing accurately the fine model. It depends on three main factors; (i) how the sample points are placed in the design space, (ii) how many they are, and (iii) what modelling method is used. These three issues will be discussed in the following sections.

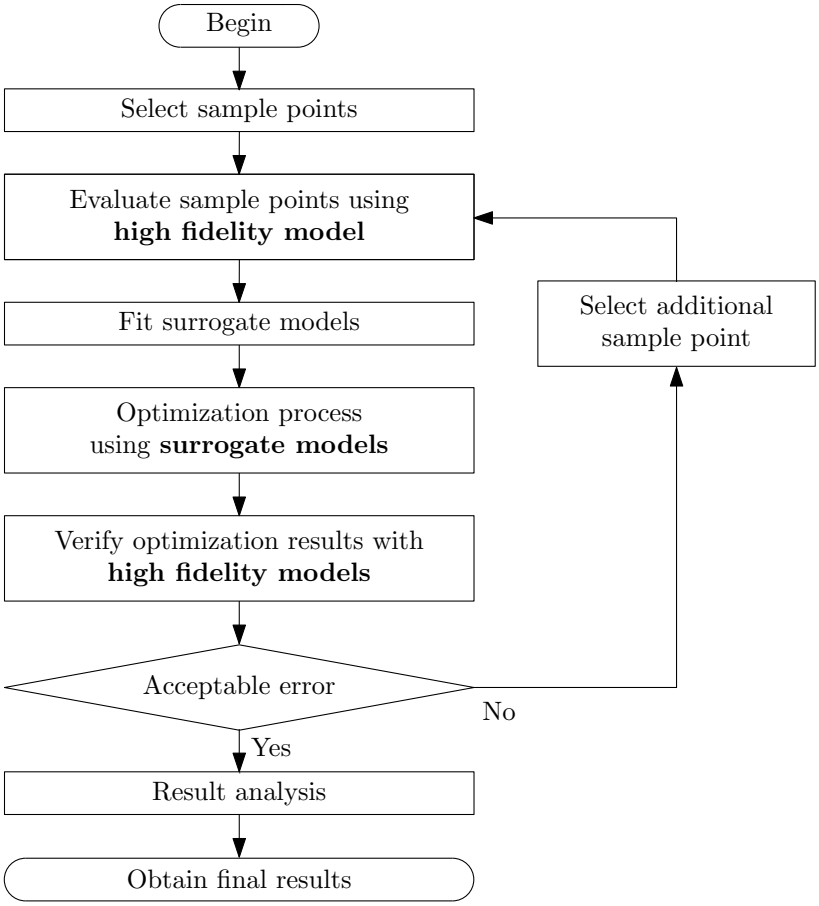


Figure 2.6: Optimization using a surrogate model

2.2.2 *Sampling technique*

In statistical fields, the classical DOE theory combines several sampling techniques, the RSM and other statistic tools such as screening. The aim of the DOE is to place a number of sample design point in the design space in order to minimise the influence of the random error. As a result, the sample points are placed mainly near the boundaries of the design domain. Some designs associating with higher-order polynomial models may place few sample points inside the design domain. Even the classical DOE is more or less perfect when used together with screening and the RSM to estimate a local trend of a function. However, some information in the centre may be lost when using the same experimental design along with other modelling techniques to predict the global trend.

In modern DOE, typically applied to computer or simulation-based experiments [31], two assumptions are made. Firstly, the random error does not exist i.e. two simulations with the same design point give the same response. Secondly, the trend of the fine model is not known beforehand. For this reason, the modern DOE uses "space filling designs" that treats all regions of the design space equally and tend to place sample points inside the design space. This allows minimizing bias errors i.e. the discrepancy between a surrogate model and a fine model[3].

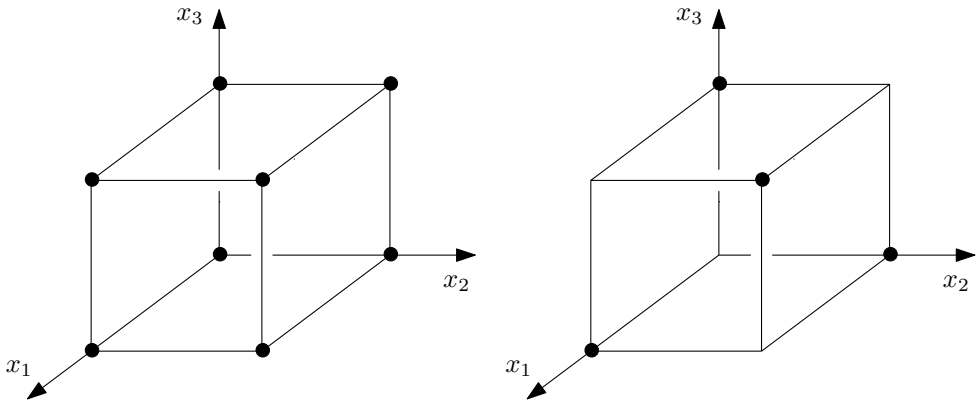
[3]The bias errors is not only at the sample sites but also in the overall design space. This cannot be identified

In the following subsection, a variety of classical and modern DOE is described. A comparative study of modern sampling techniques together with modelling techniques is presented in Section 2.2.4.

2.2.2.1 Classical experimental design

The most basic design is a two-level full factorial design [112]. It places sample points at all combination of lower- and higher-level of each factor. This results in 2^{n_v} experiments. Figure 2.7a shows how sample points are placed on the hypercube at two levels and in a three factors full factorial design. By using a full factorial design and a screening, the main effects and interactions between factors can be studied.

Using a two-level factorial design in a quadratic polynomial RSM is not suitable, as it cannot capture information inside the design domain. Various designs are possible for this purpose. Some of them can be stated: n-level factorial design (n^{n_v} sample points) and central composite design ($1+2n_v+2^{n_v}$). One may realize that the number of a sample point increases rapidly as the number of the factor increases. This problem is known as *the curse of dimensionality* [53]. For high dimensional design spaces, the fractional design can be used. The number of sample points is decreased to n^{n_v-r} and $1+2n_v+2^{n_v-r}$ for fractional factorial and fractional central composite designs, respectively, where r is a reduced order. Figure 2.7b shows a 2^{3-1} fractional factorial design.



(a) 2^3 factorial design (b) A 2^{3-1} fractional factorial design

Figure 2.7: A two-level full and fractional factorial design for three factors. With 2^{3-1} fractional design, the number of experiments is reduced from 8 to 4.

2.2.2.2 Monte Carlo Sampling

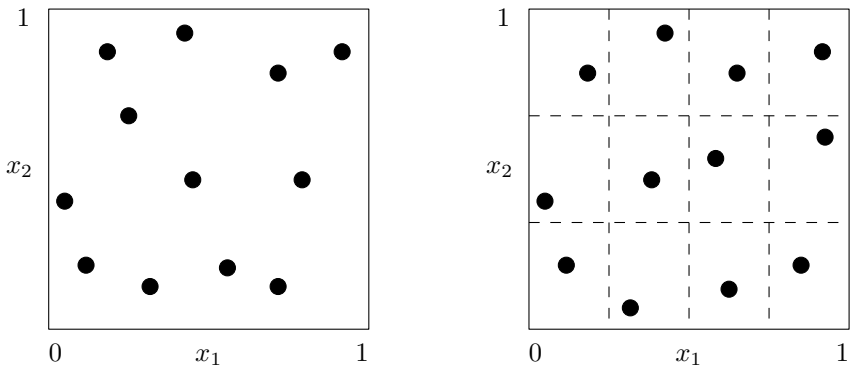
Monte Carlo Sampling (MCS), also called random design, is commonly used in the field of robust or reliability design [14]. It intends to mimic a random natural process[4]. MCS is very easy to implemented when using a programming language in which a *random* function

without any knowledge of the true function.

[4]Actually, computers cannot generate true random numbers. Therefore, a Monte Carlo sampling generated by computer is not a *real* one. A predefined table is classically used. This is why a prefix "Pseudo" is sometimes used.

is available. The position of a sample point in the dimension n^{th} can be located by selecting a random number in an interval of design variables in the dimension n^{th}. Therefore, a MCS point for an n-dimensional design space is represented by a vector of n_v random numbers in the interval $[lb,\ ub]^{n_v}$.

MCS points may not well cover all design spaces due to the fact that each sample point or even each dimension of a sample point is selected independently. Many improved MCS methods have been developed such as Stratified Monte Carlo sampling [54]. The design space is divided into bins. Only one sample point is placed in each bin. This provides more "space-filling" properties than the original MCS. Figure 2.8 shows an example of a MCS and a stratified MCS design. A stratified MCS design provides more uniformly distributed sample points.



(a) MCS (b) Stratified MCS

Figure 2.8: Monte Carlo sampling for a two-dimensional design space with 12 sample points

Hammersley Sequence Sampling (HSS) is in the class of Quasi-Monte Carlo sampling methods[5]. It uses a deterministic algorithm to generate sample points so that the points are uniformly distributed in the design space [45]. It provides better uniformity property than LHS. A ten point HSS of two dimensions are illustrated in Figure 2.9.

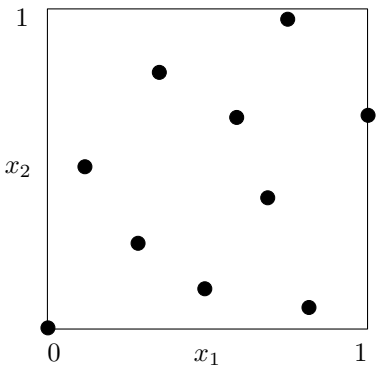


Figure 2.9: Hammersley sequence sampling of a two-dimensional design space with 10 sample points

[5]Also called low-discrepancy sampling

2.2.2.3 Latin Hypercube Sampling

Latin Hypercubes Sampling (LHS) is very popular in computer experiments. It was firstly introduced for computer experiments by McKay et al. [71]. It is widely used with Kriging in Design and Analysis of Computer Experiments (DACE) introduced by Sacks et al. [92]. As MCS, LHS is also a modern DOE or space-filling design. Sample points are scattered throughout the whole design space.

By giving a number of sample points, n_s, the design space is divided into n_s intervals of equal probability for each dimension. This results in $n_s^{n_v}$ hypercubes or bins. LHS requires that sample sites meet the following criteria:

- A bin can contain only one point, which is placed randomly in the bin.
- When projecting a design space in any two dimensions, there are n_s points and bins. Only one bin is selected in each row and column.

LHS are represented using a LHS matrix of n_s rows and n_v columns. This matrix is quite easy to generate. It requires a random permutation of n_s levels for each column. Each column of the LHS matrix is defined as:

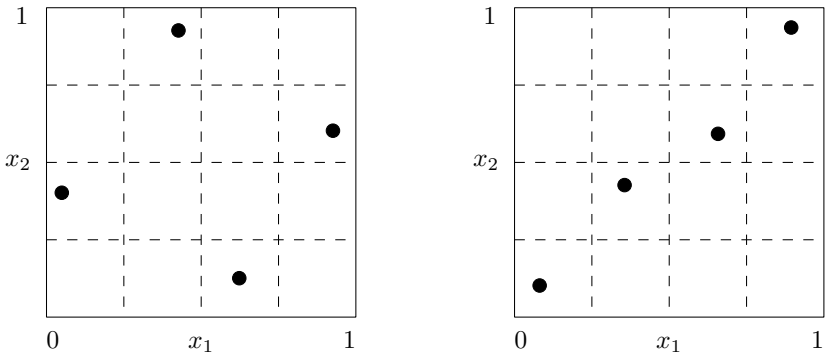
$$LHS_j = \frac{\pi_j - U_j}{n_s} \tag{2.1}$$

where π_j is the uniformly distributed random permutation of the integers from 1 to n_s, U_j is $[0,\ 1]$ uniformly distributed random number vector of n_s elements. For example, the LHS design in Figure 2.10a is constructed from a LHS matrix:

$$\begin{bmatrix} X^{(1)} \\ X^{(2)} \\ X^{(3)} \\ X^{(4)} \end{bmatrix} = \frac{1}{4}\left[\begin{pmatrix} 2 \\ 3 \\ 4 \\ 1 \end{pmatrix} - \underbrace{\begin{pmatrix} 0.35 \\ 0.52 \\ 0.24 \\ 0.81 \end{pmatrix}}_{\text{Random number}} \quad \begin{pmatrix} 4 \\ 1 \\ 3 \\ 2 \end{pmatrix} - \underbrace{\begin{pmatrix} 0.28 \\ 0.47 \\ 0.62 \\ 0.41 \end{pmatrix}}_{\text{Random number}}\right] = \begin{bmatrix} 0.41 & 0.93 \\ 0.62 & 0.13 \\ 0.94 & 0.60 \\ \underbrace{0.05}_{x_1} & \underbrace{0.40}_{x_2} \end{bmatrix}$$

In fact, there are many possibilities to select a LHS that satisfies the criteria. Figure 2.10a shows an example of a well-distributed two-dimension LHS. Figure 2.10b shows an example of a poor LHS. In this case, four sample points are positioned diagonally, leaving the two other extrema unexplored.

To avoid this poor design, many strategies can be used. For example, maximizing the minimum distance between any pair of points (maximin LHS) and minimizing the maximum distance (minimax LHS). To do so, one might simply re-sample the points until a criterion such as maximum number of trial is met. A more advanced user may use an optimization algorithm as it was done in [40]. In [40], the authors used a Simulated annealing algorithm to find the maximin LHS. However, in this research, we use the maximin criteria. The best design is simply selected from a set of trials.



(a) Well distributed LHS (b) Poor distributed LHS

Figure 2.10: Latin hypercube sampling for a two dimensional design space with 4 sample points

2.2.2.4 Number of sample point

As mentioned above, the number of sample points is one of the main criteria in order to achieve an accurate surrogate model. Theoretically, one might put as many points as possible. This idea is always right though it is not practical. Firstly, a lot of computation effort must be paid. This is usually prohibited. Secondly, when using a surrogate model for an optimization purpose, if the number of sample points should be very high, it is more reasonable to perform the optimization directly on the fine model and the surrogate model is not necessary.

In a classical DOE, the number of sample points depends on the dimension of the design space. When using an RSM, the minimum sample point should be the same as the number of polynomial coefficients (see Section 2.2.3.1). In a high dimension design space and high-order polynomial approximation, n_s can be very high and the most important is that it is not controllable by the engineer. This may cause a problem when computation time matters.

In a space-filling design or modern DOE, n_s can be chosen freely by the engineer. It is not tied to the number of design variables. It may be given to meet available computer resources. Jin et al. [42] tested the accuracy of surrogate modelling techniques constructed from sample sets with different sizes. They confirmed that a higher number of sample points yields a more accurate model.

2.2.2.5 Summary

The classical DOE techniques are good sampling techniques suitable for RSM. They place the sample points in the design space so that random errors caused by experiment are minimised. Each design provides special characteristics and depends on the chosen RMS model. The number of sample points is tied with the number of design variable.

The space-filling designs e.g. MCS and LHS have been more investigated and used in the field of surrogate modelling. Their main characteristic is to provide well-distributed sample points in the design space. This is more suitable for describing the large design space due to the fact that the information inside the design space can be captured by its well-distributed sample point property. The number of sample points does not obligatorily depend on the

number of design variables. This gives a degree of freedom to the designer to determine the number of sample points based on criteria such as computer resources, problem complexity and problem nonlinearity. A comparative study of the sampling techniques and the modelling techniques is presented in Section 2.2.4.

2.2.3 Surrogate modelling technique

The objective of a surrogate model is to give an accurate approximation of a fine model. The ideal case would be a surrogate model that represents perfectly the fine model in a short computation time.

A fine model describes the relationship between input: design variables vector x and output: response y.

$$y = f(x) \tag{2.2}$$

With a surrogate modelling technique and the knowledge of sample data: $x_s^{(i)}$ and $y_s^{(i)}$ where i is the number of sample data points from 1 to n_s, the estimated response $\hat{y}$ can be defined as:

$$\hat{y} = \hat{f}(x) \tag{2.3}$$

where $\hat{f}$ is the surrogate model of f.

Three surrogate modelling techniques are presented in the following sections. General principles and important considerations are detailed. These modelling techniques can be classified into two types: a regression model (Polynomial) and an interpolation model (Radial Basis Function and Kriging). The different is that the interpolation model passes exactly through the sample points whereas the regression model does not. Therefore, the regression model is always associated with residual or fitting errors i.e. a difference between a regression model and the observed response.

Two analytic examples are used in this section for an illustrative purpose. The position of sample points is selected in order to show the specific properties of the method. Therefore, these examples are for illustrative purpose and they do not intend to give any comparative information between each technique. A comparative study of surrogate modelling and sampling techniques is presented in Section 2.2.4.

- One-dimensional test function (modified from [94]):

$$\begin{aligned} &f_1(x) = \sin(x) - \exp\left(\frac{x}{20}\right) + 10 \\ &\text{where } x \in [0,\ 10] \end{aligned} \tag{2.4}$$

- Two-dimensional test function (Peaks function):

$$\begin{aligned} f_2(x_1, x_2) =& 3(1-x_1)^2 \exp\left(-x_1^2 - (x_2+1)^2\right) \\ &- 10\left(\frac{x_1}{5} - x_1^3 - x_2^5\right)\exp\left(-x_1^2 - x_2^2\right) - \frac{1}{3}\exp\left(-(x_1+1)^2 - x_2^2\right) \end{aligned} \tag{2.5}$$

where $x \in [-3,\ 3]$ for all dimensions

One-dimensional function is shown in Figure 2.11a. Figure 2.16a shows a two-dimensional test function. It can be observed that both functions are multi-modal. For a two-dimensional problem, there is a large flat area.

2.2.3.1 Polynomial

Polynomial approximations are the most common techniques for constructing a surrogate model. In the statistic field, the combination of DOE, Polynomial-based approximation and Analysis of variance (ANOVA) is known as Response Surface Methodology (RSM) [9, 50, 112]. It is originally developed to build a smooth response surface of observed data from physical experiments, which are always subject to empirical or random errors. Therefore, a regression model[6] is very suitable for this kind of problem. RSM is very popular because it provides a compact and explicit relationship between a true function (so called *response*) and input variables (called *factor*). Each regression coefficient has a meaning in itself[7]. Moreover, the least-square method used to construct the models is inexpensive and simple to implement.

A polynomial model requires a specific number of experiments or sample points depending on the model order and the polynomial term requirements. It is usually associated with the DOE method in order to select appropriated experiments. ANOVA allows to predict the uncertainty of a model by using information from the discrepancy between fine and polynomial models at sample sites.

A relationship between the observed response y and the factors X can be predicted by a polynomial model.

$$y = \hat{y} + \epsilon \tag{2.6}$$

ϵ denotes a fitting error or noise. It is actually an error between the observed response and the predicted response. It is assumed to be normally distributed (Gaussian) and with zero means. This term is called *residual*.

An approximated response $\hat{y}$ is defined as:

$$\hat{y} = X\hat{\beta} \tag{2.7}$$

where $\hat{\beta}$ is a polynomial coefficient vector to be estimated. The number of polynomial coefficient terms n_t depends on the order of the polynomial model. For example, the linear model requires a constant term b_0 and n_v main effect coefficients. The generalized $\hat{\beta}$ is expressed as [112]:

$$\beta = [b_0,\ b_1, b_2, \ldots, b_{n_v},\ b_{11}, b_{22}, \ldots, b_{n_v n_v},\ b_{12}, \ldots, b_{1n_v}, \ldots]^T \tag{2.8}$$

Note that the order and specifications of the polynomial model depends on the knowledge of the fine model. An engineer should have an idea on the system behaviour and interactions between factors. This can be obtained from their knowledge of studied physical phenomena or by performing a screening or a sensitivity analysis [29]. However, if the design space is too large, these information might be inaccurate or completely wrong in the worst case.

[6] The regression model often refers to a model obtained using the least-square method. This kind of model is non-interpolation i.e. the response surface does not pass through the observed data.

[7] Each coefficient represents the level of correlation between factors. They are known as *main effects* and *interactions*.

By giving the sampled design vector x_s and the observed response y_s at observed sites, the unique least square solution to (2.6) is [9]:

$$\hat{\beta} = \left(X_s^T X_s\right)^{-1} X_s^T y_s \tag{2.9}$$

where X_s is a sample design variable matrix of n_s row and n_t column:

$$X_s = \begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} & \cdots & x_{n_v}^{(1)} & \left(x_1^{(1)}\right)^2 & \cdots & \left(x_{n_v}^{(1)}\right)^2 & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots \\ 1 & x_1^{(n_s)} & x_2^{(n_s)} & \cdots & x_{n_v}^{(n_s)} & \left(x_1^{(n_s)}\right)^2 & \cdots & \left(x_{n_v}^{(n_s)}\right)^2 & \cdots \end{bmatrix} \tag{2.10}$$

and

$$y_s = \left[y^{(1)}, y^{(2)}, \ldots, y^{(n_s)}\right]^T \tag{2.11}$$

In spite of its popularity and simplicity, the low-order polynomial model suffers from its accuracy when dealing with a highly nonlinear fine model over a large region [99]. It is efficient only for local approximations over limited design spaces. The increase of the number of sample points does not necessarily lead to increased model accuracy. This is depicted in Figure 2.11. Linear and quadratic polynomials are used to approximate a highly nonlinear function. Figure 2.11a plots models constructed from 5 poor distributed sample points. It can be observed that a second order polynomial model cannot capture the high nonlinearity of the true function. In Figure 2.11b, a sample point is added so that the sample points are better distributed in the design space. Even an additional point is added, both models are still far from the true function. This example confirms that a polynomial model is not suitable for approximating a highly nonlinear fine model. A polynomial model should not be applied without prior knowledge of fine model trends.

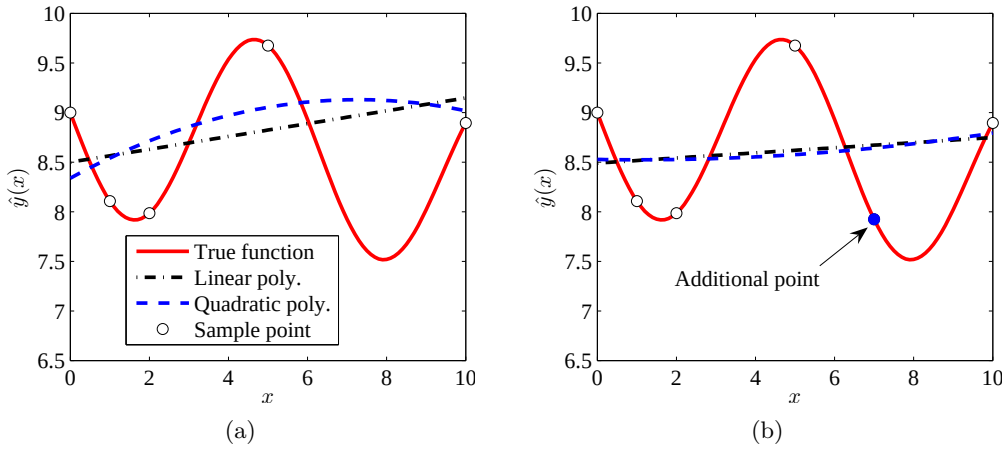


Figure 2.11: One-dimension example illustrating linear and quadratic polynomial approximations.

2.2.3.2 Radial basis function

A RBF model is a weighted-sum of basis functions usually expressed in terms of the euclidean distance $\left\|x - x^{(i)}\right\|$ of a design vector x and a centre, which is usually an observed point $x^{(i)}$.

$$\hat{y}(x) = \sum_{i=1}^{n_s} w_i \varphi\left(x, x^{(i)}\right) \tag{2.12}$$

where w_i is the weight to be estimated and $\varphi\left(x, x^{(i)}\right)$ is the radial basis function. The basis functions depend on the distance between the input vector x and the centre of the basis function. Possible choices for the basis function are [43, 52]:

$$\begin{aligned}
\varphi\left(x, x^{(i)}\right) &= \left\|x - x^{(i)}\right\| && \text{(linear)} \\
\varphi\left(x, x^{(i)}\right) &= \left\|x - x^{(i)}\right\|^3 && \text{(cubic)} \\
\varphi\left(x, x^{(i)}\right) &= \left\|x - x^{(i)}\right\|^2 \ln\left(\left\|x - x^{(i)}\right\|\right) && \text{(thin plate spline)} \\
\varphi\left(x, x^{(i)}\right) &= \sqrt{\left\|x - x^{(i)}\right\|^2 + \gamma^2} && \text{(multiquadric)} \\
\varphi\left(x, x^{(i)}\right) &= \exp\left(-\tfrac{1}{\gamma^2}\left\|x - x^{(i)}\right\|\right) && \text{(Gaussian)}
\end{aligned} \tag{2.13}$$

where $\gamma > 0$ is the width parameter controlling how spread the basis function is. This parameter is given by the user.

In matrix form, $\hat{y}$ can be written as:

$$\hat{y} = AW \tag{2.14}$$

By substituting the observed data pair y_s and x_s, the weight matrix W can be solved and a unique solution is obtained:

$$W = A^{-1} y_s \tag{2.15}$$

where y_s defined as in (2.11) and

$$A = \begin{bmatrix}
\varphi\left(x^{(1)}, x^{(1)}\right) & \varphi\left(x^{(1)}, x^{(2)}\right) & \cdots & \varphi\left(x^{(1)}, x^{(n_s)}\right) \\
\varphi\left(x^{(2)}, x^{(1)}\right) & \varphi\left(x^{(2)}, x^{(2)}\right) & \cdots & \varphi\left(x^{(2)}, x^{(n_s)}\right) \\
\vdots & \vdots & \ddots & \vdots \\
\varphi\left(x^{(n_s)}, x^{(1)}\right) & \varphi\left(x^{(n_s)}, x^{(2)}\right) & \cdots & \varphi\left(x^{(n_s)}, x^{(n_s)}\right)
\end{bmatrix} \tag{2.16}$$

Figure 2.12 shows a one-dimensional example. The RBF model is constructed from 5 sample points. One can observe that the RBF model passes though the sample points. In this research, the thin plate spline is used, as it does not require spread parameter.

2.2.3.3 Kriging

Kriging was first developed by D. Krige, a mining engineer. It was used with success in the field of geological statistics to estimate mineral concentrations over areas of interest, given a set of sampled sites [13]. In the field of computer science and engineering, it was introduced by Sacks et al. as Design and Analysis of Computer Experiments (DACE) [92]. The Kriging method was also used with success in many engineering applications, including electromagnetic device design [63] and traction system design [61].

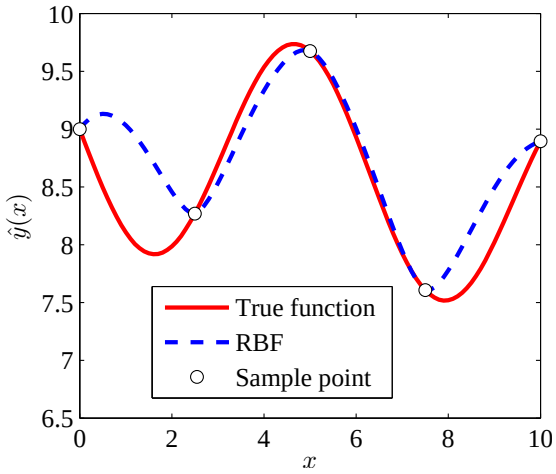


Figure 2.12: One dimensional example – RBF model constructed from 5 equally distributed sample points

In the Kriging modelling approach, an unknown true function can be written in the following form:

$$y(x) = B(x) + Z(x) \tag{2.17}$$

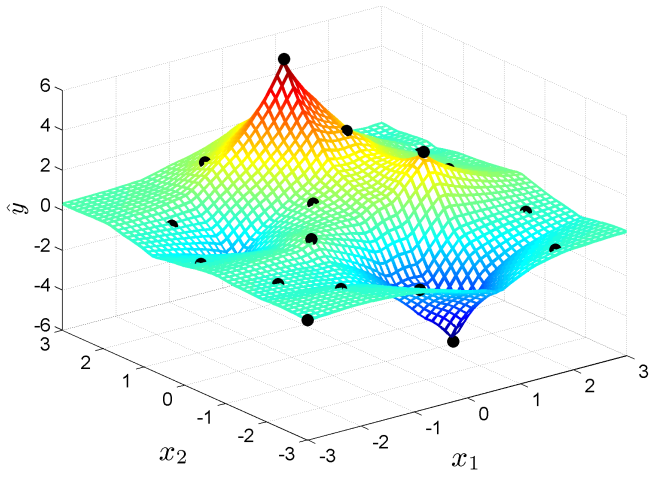
The first term $B(x)$ is a regression or polynomial model representing the global trend of the function. The second term $Z(x)$ gives the localized deviations from the global trend. This term corresponds to the residual term ϵ in polynomial approximation (see Section 2.2.3.1). $Z(x)$ is a model of a stochastic process with zero mean, variance of σ^2 and covariance defined by:

$$Cov\left[Z\left(x^{(i)}\right), Z\left(x^{(j)}\right)\right] = \sigma^2 \mathbf{R}\left[R\left(x^{(i)}, x^{(j)}\right)\right] \tag{2.18}$$

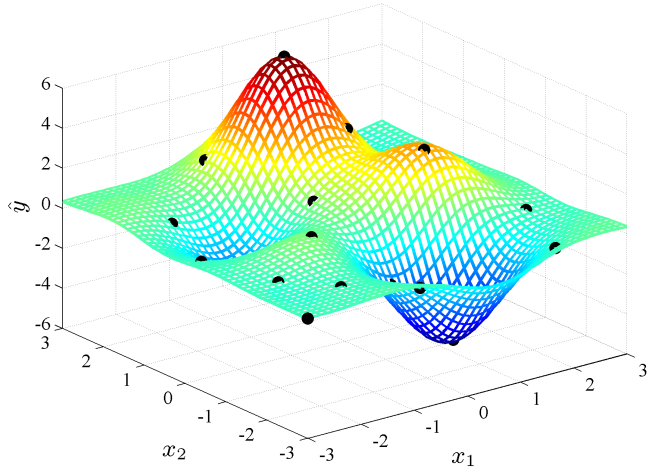
where $\mathbf{R}$ is the correlation matrix, R the correlation function, i and j are the sample points (from 1 to sample point number n_s). The choice of the correlation function controls how the surrogate model fits the data. Various correlation functions are given in [92]. The Gaussian function is the most common.

$$R\left(x^{(i)}, x^{(j)}\right) = \exp\left[-\sum_{k=1}^{n_v} \theta_k \left|x_k^{(i)} - x_k^{(j)}\right|^{p_k}\right] \tag{2.19}$$

where n_v is the number of design variables, θ_k is the unknown correlation function parameter ≥ 0 determining how rapidly the correlation is lost in the k^{th} design variable (large values implying rapid loss in correlation), and p_k the smoothness parameter $0 < p_k \leq 2$. The value near 2 gives smooth functions and value near 0 gives rough functions. Figure 2.13 depicts this effect. Note that $p_k = 2$ is commonly used and gives good results in most cases. One can also notice that at the sample point $\left|x_k^{(i)} - x_k^{(j)}\right|$ is zero and $R = 1$. The Kriging predictor passes exactly through the sample points i.e. is an interpolation model. However, a non-interpolating model may be preferred when the fine model is highly noisy e.g. FEA. A derivative non-interpolating version of Kriging can be used for this purpose. It uses an additional parameter called nugget in the correlation function to help smoothing the data. This thesis does not investigate this Kriging version. Readers should refer to [51, 80, 94] for more detail.



(a) $p_k = 1$



(b) $p_k = 2$

Figure 2.13: Effect of the smoothness parameter p_k. The models are constructed from 20 data points sampled by Hammersley sequence sampling

While (2.17) defines the true response value y, the Kriging model predicts the estimated response value $\hat{y}$. The Mean Square Error (MSE) is the expected value of difference between the true response and the estimated one.

$$MSE\left(x\right) = E\left(\left(y\left(x\right) - \hat{y}\left(x\right)\right)^2\right) \tag{2.20}$$

Since Kriging interpolates the data, the MSE is zero at the sample points. At the other unknown points, the MSE is supposed to be minimum in order to obtain a good approximation. When MSE is minimised, the Kriging model becomes:

$$\hat{y} = \mathbf{f}\hat{\beta} + \mathbf{r}^T\mathbf{R}^{-1}\left(y - \mathbf{f}\hat{\beta}\right) \tag{2.21}$$

and the MSE (2.20) can be rewritten as

$$MSE = \sigma^2\left[1 - \mathbf{r}^T\mathbf{R}^{-1}\mathbf{r} + \frac{\left(1 - \mathbf{f}^T\mathbf{R}^{-1}\mathbf{r}\right)^2}{\mathbf{f}^T\mathbf{R}^{-1}\mathbf{f}}\right] \tag{2.22}$$

where $\hat{\beta}$ is the estimator of regression model defined as in (2.27), $\mathbf{r}$ the correlation vector between a new location x to be estimated and the sampled locations, $\mathbf{f}$ a unit vector with length of n_s.

$$\mathbf{r}(x) = \left[R\left(x, x^{(1)}\right), R\left(x, x^{(2)}\right), \ldots, R\left(x, x^{(n_s)}\right)\right] \tag{2.23}$$

$$\mathbf{R} = \begin{bmatrix} R\left(x^{(1)}, x^{(1)}\right) & R\left(x^{(1)}, x^{(2)}\right) & \cdots & R\left(x^{(1)}, x^{(n_s)}\right) \\ R\left(x^{(2)}, x^{(1)}\right) & R\left(x^{(2)}, x^{(2)}\right) & \cdots & R\left(x^{(2)}, x^{(n_s)}\right) \\ \vdots & \vdots & \ddots & \vdots \\ R\left(x^{(n_s)}, x^{(1)}\right) & R\left(x^{(n_s)}, x^{(2)}\right) & \cdots & R\left(x^{(n_s)}, x^{(n_s)}\right) \end{bmatrix} \tag{2.24}$$

Note that $\mathbf{R}$ is symmetric $R\left(x^{(i)}, x^{(j)}\right) = R\left(x^{(j)}, x^{(i)}\right)$ with ones along diagonal, $R\left(x^{(i)}, x^{(i)}\right) = 1$ [67].

$\mathbf{R}$ and $\mathbf{r}$ depend on θ_k, which can be found using the Maximum Likelihood Estimation (MLE).

The likelihood function (L) is defined as:

$$L = \frac{1}{\sqrt{\left(2\pi\hat{\sigma}^2\right)^{n_s}\left|\mathbf{R}\right|}}\exp\left(\frac{-\left(y - \mathbf{f}\hat{B}\right)^T\mathbf{R}^{-1}\left(y - \mathbf{f}\hat{B}\right)}{2\hat{\sigma}^2}\right) \tag{2.25}$$

The log-likelihood function (LL) is generally used in order to simplify (2.25).

$$LL = \frac{-\left(y - \mathbf{f}\hat{B}\right)^T\mathbf{R}^{-1}\left(y - \mathbf{f}\hat{B}\right)}{2\hat{\sigma}^2} - \frac{n_s\ln\left(2\pi\hat{\sigma}^2\right) + \ln\left(\left|\mathbf{R}\right|\right)}{2} \tag{2.26}$$

$\hat{\beta}$ and $\hat{\sigma}^2$ can be estimated as:

$$\hat{\beta} = \left(\mathbf{f}^T\mathbf{R}^{-1}\mathbf{f}\right)^{-1}\mathbf{f}^T\mathbf{R}^{-1}y \tag{2.27}$$

$$\hat{\sigma}^2 = \frac{1}{n_s}\left(\left(y - \mathbf{f}\hat{\beta}\right)^T\cdot\mathbf{R}^{-1}\cdot\left(y - \mathbf{f}\hat{\beta}\right)\right) \tag{2.28}$$

Substituting $\hat{\beta}$ and $\hat{\sigma}^2$, the log-likelihood function (2.26) becomes:

$$LL = -\frac{n_s \ln\left(\hat{\sigma}^2\right) + \ln\left(|\mathbf{R}|\right)}{2} \tag{2.29}$$

By solving the MLE optimization problem (2.30), θ_k can be determined (p_k is fixed to 2). In this research, a genetic algorithm is used to solve the MLE problem. Once θ_k is determined, the response at any design point x can be estimated using (2.21).

$$\begin{aligned} \max_{\theta_k} \quad & -\frac{n_s \ln\left(\hat{\sigma}^2\right) + \ln\left(|\mathbf{R}|\right)}{2} \\ \text{subject to} \quad & 0 < \theta_k < \infty \end{aligned} \tag{2.30}$$

Jin et al. [42] showed that the Kriging method works slightly better than other surrogate models. However, Kriging needs more computational time to solve the MLE optimization problem in the model fitting process. This time could be very high in large-scale problems with large sample size. In the real implementation, it does not cause any trouble due to the fact that the surrogate model is built only one time, moreover the use of this model in an optimization process is very rapid compared to the high fidelity model.

2.2.4 Surrogate model accuracy assessment

In the above sections, several sampling and surrogate techniques were presented. This section intends to assess performances of the methods. Six test functions are selected. They are comprised of 4 two-dimensional, 1 four-dimensional and 1 six-dimensional functions. These functions are listed in Appendix A.

2.2.4.1 Assessment methodology

This test aims at assessing the performances of surrogate and sampling techniques in the case of a highly nonlinear function. Three sampling techniques, three sample size rules and two surrogate techniques are tested. They are listed below.

- Sampling techniques: LHS, HSS and MCS
- Size of the sample sets: The sample sets are scaled by the number of design variables. They are set to 5, 10 and 20 times the number of design variables for small, medium, and large sample set, respectively.
- Surrogate techniques: RBF and Kriging. Note that the polynomial model is omitted.

This yields a total of 18 combinations for each test function.

The accuracy of a surrogate model can be assessed by comparing the surrogate model with a true function. In this example, the validate points are selected randomly in the design space, 1000 points for each test function. Two standard methods are used: Normalized root mean squared error (NRMSE) and Normalized maximum absolute error (NEMAX). The normalized errors are used because they allow comparing different test functions i.e. they

are not sensitive to the design space of each test function. NRMSE represents global errors while NEMAX represents local errors.

$$NRMSE = \sqrt{\frac{\sum_{i=1}^{n_s}\left(y^{(i)} - \hat{y}^{(i)}\right)^2}{\sum_{i=1}^{n_s}\left(y^{(i)}\right)^2}} \tag{2.31}$$

$$NEMAX = \frac{\max_{i=1:n_s}\left|y^{(i)} - \hat{y}^{(i)}\right|}{\sqrt{\frac{1}{n_s}\sum_{i=1}^{n_s}\left(y^{(i)} - \bar{y}\right)^2}} \tag{2.32}$$

where $y^{(i)}$ is the true response, $\hat{y}^{(i)}$ the estimated response both at the sample site $x^{(i)}$ and $\bar{y}$ the mean of the true function at the n_s sample sites.

2.2.4.2 Results

The average NRMSE and NEMAX values of all test functions are plotted in Figure 2.14 and Figure 2.15, respectively. Full results are given in Appendix A. At the first sight, one can rapidly conclude that the MCS gives the lowest accuracy and the HSS gives slightly better results than the LHS. Regarding the size of sample point sets, the larger the sample set, the higher the accuracy. Kriging performs better than the RBF except for the high dimensional problem, the RBF gives slightly better NRMSE than Kriging. However, the NEMAX for the RBF is higher than that of Kriging (see Appendix A).

The Kriging technique seems to be a good interpolation model. It is applied to a traction system re-design problem as can be found in [61].

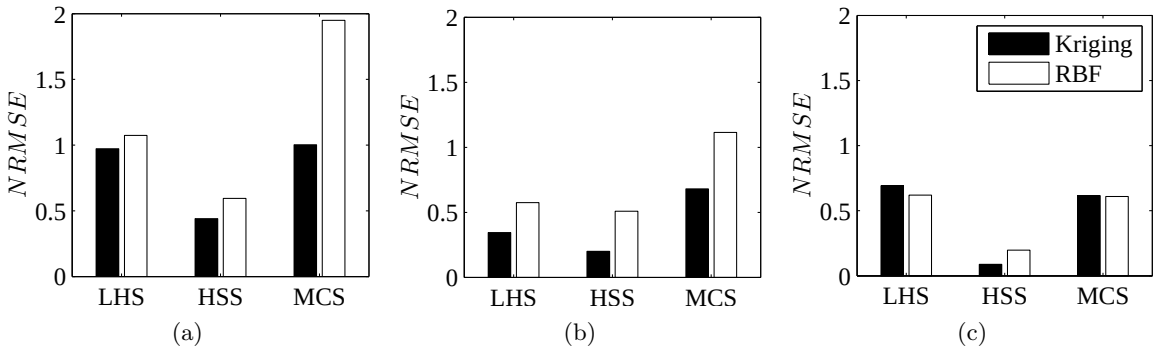


Figure 2.14: Normalized root mean squared error (NRMSE) for a (a) small sample set, (b) medium sample set and (c) large sample set

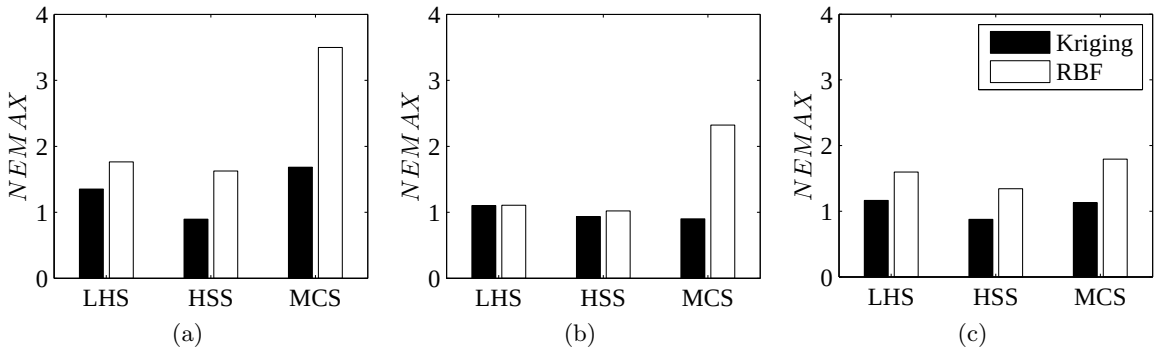


Figure 2.15: Normalized maximum absolute error (NEMAX) for a (a) small sample set, (b) medium sample set and (c) large sample set

2.2.5 Non-conventional sampling and modelling technique

This section presents two non-conventional sampling and modelling techniques. The first one, the Global response surface methodology, improves the traditional RSM in the case of highly nonlinear function modelling. The second one, the Kriging-assisted sampling technique, adds the additional sample points to the initial Kriging model according to the predicted error information. It gives the possibility to control the number of sample points *just needed*.

2.2.5.1 Global response surface methodology

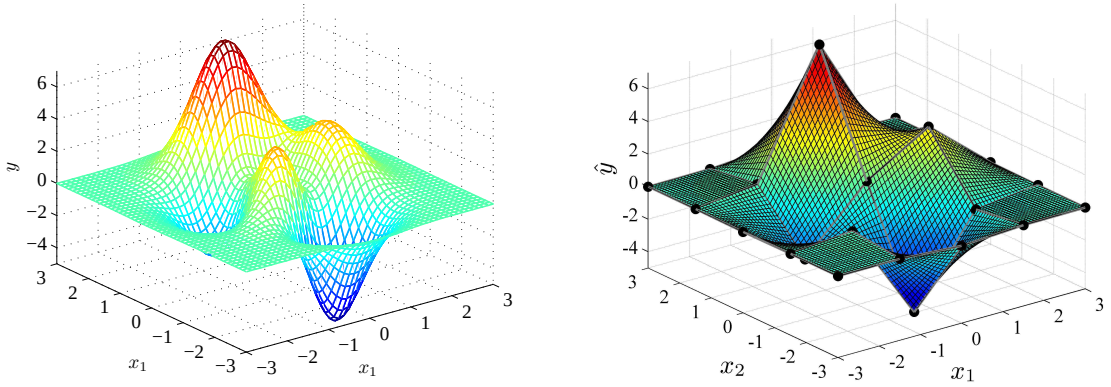
One of the drawbacks of the polynomial modelling technique is that it is suitable for a local approximation not for a global one. In order to obtain a global model, one can divide the design space into multiple sub-domains. Each polynomial approximation is associated to only one sub-domain. The linear or quadratic polynomial approximates a smaller design space. Therefore, it is more accurate than modelling of the entire design space. S. Vivier [112] presented some variants of a classical DOE, which reuse sample points when performing a global RSM.

Figure 2.16 shows a global RSM example. The entire design space is divided into 16 sub-domains. 16 linear polynomial models form the global RSM model. These models are constructed from 25 sample points in total.

2.2.5.2 Kriging-assisted sampling technique

One useful property of a Kriging surrogate model is its error estimator (MSE or standard error). This information indicates that the Kriging model gives the exact value at the sample points (interpolation model) and may present some errors at other design vectors. One-dimensional example is showed in Figure 2.17. The Kriging model is constructed using 5 sample design points. One can observe that the standard error is zero at the sample point and higher in the sample point gap.

This MSE predictor can be used as sampling point selection criterion. A Kriging-assisted sampling algorithm is shown in Figure 2.18. By starting with a very small number of sample experiments, the initial Kriging model is fitted. The initial Kriging model is usually associated



(a) True function (b) Global RSM

Figure 2.16: Global RSM example

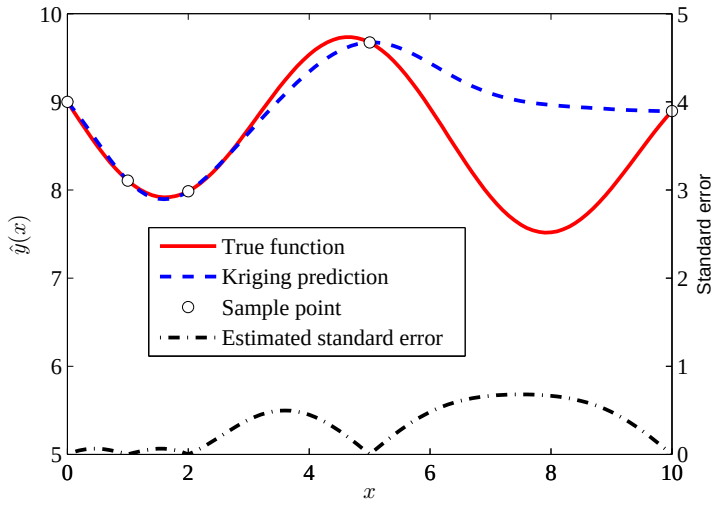


Figure 2.17: One-dimensional illustration example showing the true function, the Kriging prediction and its estimated standard error. The Kriging model is constructed with the use of 5 sample points.

with high errors. The MSE tends to be very high at non-sampled design vectors. One may realise that the most beneficial action is to add extra points to where the MSE value is the highest. A single-objective optimization algorithm can be used for this purpose. Once the most promising design vectors have been located, the high fidelity model is then evaluated at these points and the Kriging model is refitted. Stopping criteria can be a maximal allowable number of design points, the maximal allowable time or an error assessment such as the "leave-k-out" method [72].

An example is shown in Figure 2.19. An initial Kriging model is constructed using 10 initial sample points (LHS). Figure 2.19b shows the Kriging estimated MSE for the initial model. The algorithm described above is run for 10 iterations. At each iteration, a maximisation of the MSE is performed. Note that the $-MSE$ is plotted in Figure 2.19b. In this case, a minimum of $-MSE$ surface is sought. After 10 iterations, Figure 2.19c and 2.19d

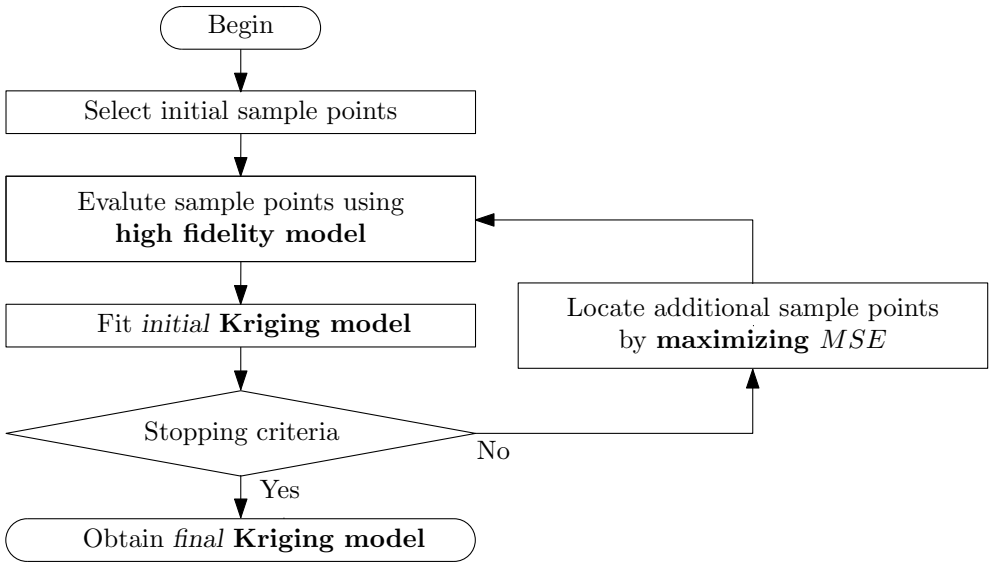
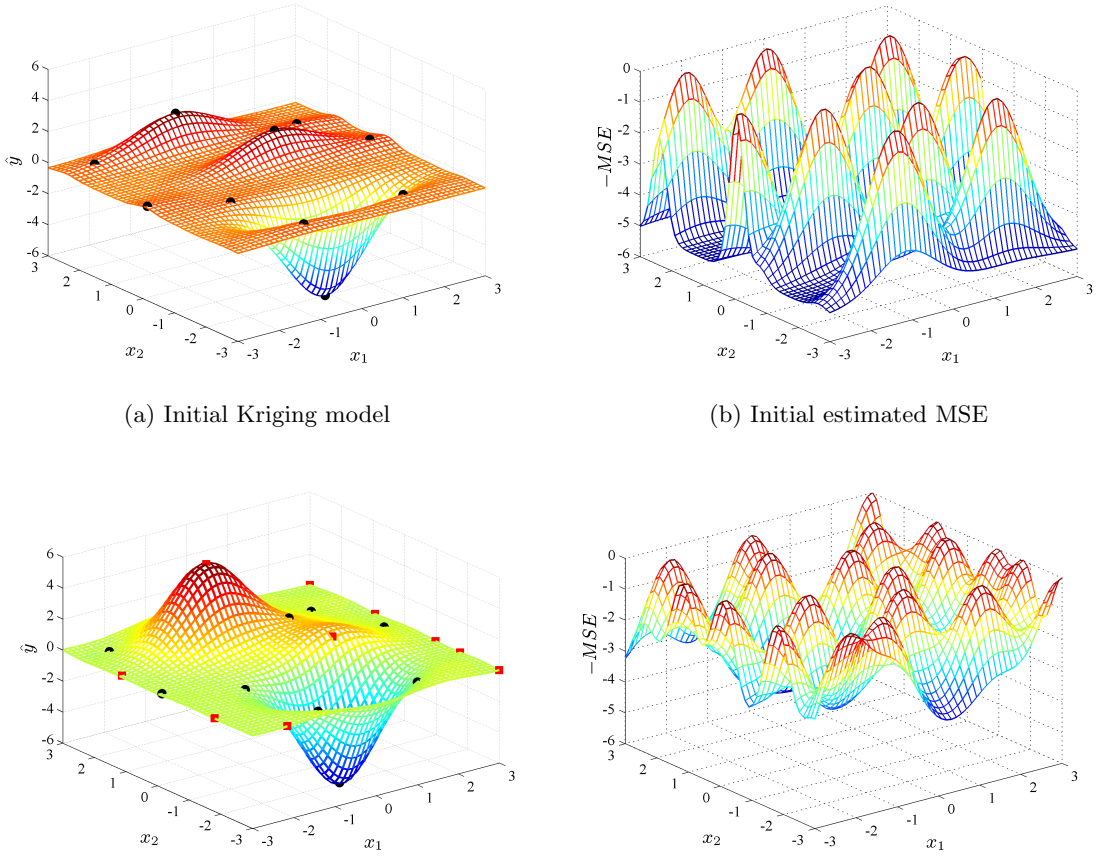


Figure 2.18: Flowchart of a sequential technique based on Kriging MSE information

present the final Kriging model and its associated MSE. The algorithm adds 8 points near the boundary and only 2 points inside the design space. This extrapolation keeps on being minimum.

Although the additional points are added sequentially, this is not limited to only one point per iteration. Several additional points can be added at each iteration if we use an optimization algorithm capable of locating multiple optima [41]. It is possible to apply the parallel and distributed computation. This avoids spending too much time on the surrogate model building process.



(a) Initial Kriging model

(b) Initial estimated MSE

(c) Final Kriging model

(d) Final estimated MSE

Figure 2.19: Analytical example of Kriging-assisted sampling technique. The initial model was constructed from 10 initial points (black circle). Then other 10 points were added iteratively by an algorithm. (additional points are shown as red squares)

2.3 Multimodel approach

Engineers endeavour to develop more and more complex models, which take into account many physical domains and components. This is due to the needs for high accuracy results in the system context. Outputs of a sub-model may act as inputs to others and vice versa. Many modelling techniques can be mixed together into the same model. For example, the model of a motor can combine three sub-models:

- On the basis of the motor geometry, a FEM is used for magnetic field computation.
- The output from the FEM is then fed to analytical electric and control models. Losses are then computed.
- The semi-analytical thermal model uses losses as heat sources and allows calculating the temperature of the motor.

A problem might occur if an output coming from the model A (to be used as input for the model B) is outside the validity domain of the model B. These circumstances may lead to an error in the computation code or, even worse, give wrong results without any notice to the designer. The robustness of the model is very important when used in an optimization process. Two boundary constraints should be added to any linking output to ensure that these values are inside the validity domain.

2.4 Conclusion

This chapter presented various modelling techniques arising in electrical engineering. The physical-based modelling techniques are commonly used. These include analytical, numerical and semi-numerical models. The analytical model is very cheap to evaluate with a limited accuracy due to the fact that many hypotheses must be assumed. On the contrary, the numerical model is more general and needs less modelling assumptions. It provides a high accuracy results. However, these advantages are obtained with high computation time. The semi-numerical model like lumped-mass model is a good compromise between accuracy and computational cost.

Another modelling approach is the surrogate modelling technique. It is a mathematics model constructed by using information i.e. computation results from physical-based model. It extracts a black-box model from a physical-based model. The main advantage is that the designer does not need to have a strong knowledge to model a system. This is not the case when using physical-based modelling techniques. Surrogate models can replace high fidelity models for various purposes, not only for data visualization but also for communication, problem formulation and optimization. They allow a rapid estimation of the relationship between the input and the output of a fine model. However, some precision must be lost. An analytical example has illustrated how Kriging and RBF models behave on multi-dimensional test functions. The influence of the sampling techniques and the size of sample points have also been tested.

Two non conventional sampling and modelling techniques have been presented. (i) The global response surface methodology improves traditional RSM in the case of a highly nonlinear function. (ii) The Kriging-assisted sampling technique uses the estimated standard errors to select iteratively the added sample points. In this chapter, they were presented only for the modelling purpose. However, these methods intend to offer an opening towards the next chapter – Optimization technique. One must realize the potential of the extension of these techniques to perform optimization. For example, instead of dividing a design space equally and a priori, it can be panned and zoomed into the zone where the optimum might be found. A more or less similar principle can be applied to the surrogate-assisted technique. The combination of model accuracy and surrogate model information offers exploration and exploitation possibilities, both at the same iteration. These techniques are detailed in Chapter 3.

New developments focus on the system and multidisciplinary approach. In such approaches, several components are modelled and optimised simultaneously. More attention is needed when working with a multimodel. The most important issue is the validity domain of each model, concerning the input and the output. The out-of-domain input can cause a fatal error of the computation code or unexpected results. This affects the robustness of the model when used in an optimization process. Chapter 3 presents the Multidisciplinary Design Optimization approach that takes into account all these considerations.

Chapter 3

Optimization technique

An optimization technique is established in any discipline and application domain. It is a general tool that helps designers to manage the complexity of design problems in order to produce at low cost and efficiently. The *Trial-and-error* conventional method cannot fulfil the needs of industries any more. This Chapter intends to give an overview of the optimization technique used to solve complex system design problems.

For less complex system, an optimization technique is applied to the design problem as a whole. This approach is called the Global System (GS) approach [73]. It is commonly used in optimal design. Several conventional optimization techniques as well as a more advanced technique using the surrogate model in the algorithm are presented. These techniques can apply to single- and multi-objective optimization problems. This approach is not recommended due to the too high amount of work it requires.

For most complex system, complexity is mainly due to the existence of interactions between disciplines and between physical or functional subsystems. Such interactions are usually difficult to manage because of the increased number of design parameters. In the aeronautic community, the optimal design of such a system is referenced under the concept of Multidisciplinary Design Optimization (MDO) [102]. MDO associates different aspects: analysis tools, optimization algorithm and problem formulation. The problem formulation concept does not only refer to the definition of the objective function, constraints and design variables as in single disciplinary optimization but also corresponds to the way of solving interactions between disciplines [102].

In most cases, the design problems are broken-down into several smaller independent problems. This issue is already discussed in Chapter 1. To optimise such a decoupled system, several subproblems need to be brought back together. A multilevel hierarchical coordination and optimization algorithm called Target Cascading (TC) gives a straightforward way to mange such a hierarchical breaking down. The coordination process is more or less similar to the organisation of the company.

MDO and TC are considered as complex optimal design problem formulations. They ensure consistency of the whole system via coordination and information procedures between models. The optimization algorithm used in the global system approach is thus used to solve the subproblem. Figure 3.1 summarizes the three optimization techniques for the complex

system design detailed in the following sections.

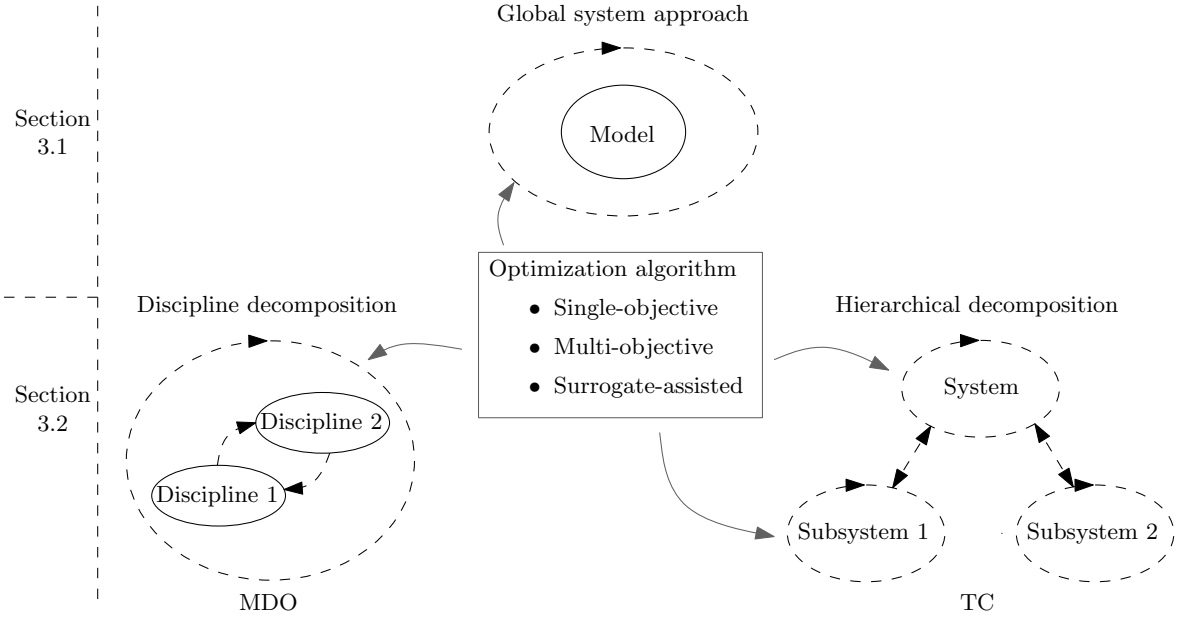


Figure 3.1: Summary of optimization techniques

3.1 Global system approach

The global system approach considers the whole system as a single unit. The system modelling may consist of several models[1], which are launched sequentially or solved using a numerical method. In any case, the system model is considered as a whole black-box model. The global system design problem is formulated as a single optimization problem. Therefore, the whole system is optimised simultaneously. This is a traditional approach commonly used in engineering optimal designs [73].

This section presents optimization algorithms used to solve nonlinear and constrained optimization problems, which are common for engineering design problems. Several conventional optimization algorithms for both single-objective and multi-objective optimization are firstly presented. Thereafter, the surrogate-assisted algorithm, an extension of the surrogate modelling concept from Chapter 2, will be discussed.

3.1.1 Single-objective

This section intends to give practical information on single-objective optimization algorithms to solve a multimodal problem. Two popular algorithms are presented. The first one is Sequential Quadratic Programming (SQP) and the second one is Genetic Algorithm (GA). GA is a global optimization algorithm and SQP is a local method. A multistart approach is used with SQP in order to improve the SQP performances in the multimodal optimization

[1]For each disciplinary or sub-system

problem. Each algorithm presents specific advantages and drawbacks. The hybridization of 2 algorithms allows making use of their advantages.

3.1.1.1 Gradient method

In the field of deterministic optimization, gradient information is commonly used in nonlinear optimization methods[2]. The Sequential Quadratic Programming (SQP) algorithm is one of the best gradient algorithms. It allows solving nonlinear constrained optimization problems. SQP includes directly objective and constraint functions into its optimum solution search process. The optimum solution satisfies Kuhn-Tucker conditions [109]:

$$\begin{aligned} \nabla f(x^*) + \sum_{m}^{i=1} \lambda_i^* \cdot \nabla g(x^*) &= 0 \\ \lambda_i^* \cdot \nabla g(x^*) &= 0 \qquad i = 1, \ldots, m \\ \lambda_i^* &\geq 0 \qquad i = m_e + 1, \ldots, m \end{aligned} \tag{3.1}$$

At each iteration, the SQP solves Quadratic Programming (QP) problem. The Hessian of Lagrange function is included in the QP problem. This allows taking into account direct constraints.

The gradient of objective and constraint functions are needed in the Hessian estimation process performed by the BFGS method. The gradient and derivative can be computed by symbolic math when it is possible. This increases the robustness of the algorithm. However, this symbolic derivative calculation may not always be possible. The finite difference method can be used but it increases significantly the number of function calls. The SQP method is efficient in the following conditions:

(i) The problem is not of a too high dimension, because a high dimensional problem becomes highly multimodal and the SQP is then trapped into local optima. Moreover, the number of function evaluations increases rapidly in high dimensional problems if finite difference is used to compute the gradient;

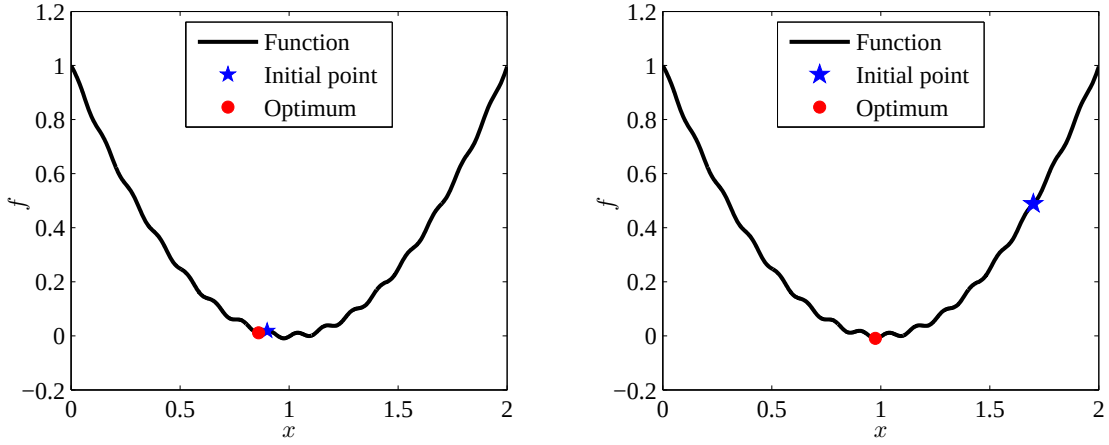
(ii) Computation of a gradient can be obtained with a rather high precision. The convergence speed depends on the gradient precision;

(iii) The problem is smooth and scaled. Design variables must be scaled in order to ensure a correct operation of the algorithm.

As SQP uses gradient information, it encounters some difficulties to find the global optimum if the problem has several optima. A similar problem is stated in the case of a noisy function with one or several *true* optima[3]. Figure 3.2 shows one-dimensional noisy example as expressed in (3.2). Optimal points can be stuck in local optima (caused by noise in this case). The global optimum is found at $x = 0.9758$ and $f = -0.00937$

$$\begin{aligned} \min_x \quad & f = (x-1)^2 + \underbrace{0.01 \sin(50x)}_{\text{Noise term}} \\ \text{subject to} \quad & 0 \leq x \leq 2 \end{aligned} \tag{3.2}$$

[2]Nonlinear Programming (NLP) problem refers to an optimization problem whose models are nonlinear.
[3]In a noisy function, the true optimum is referred as the optimum found in the function without noise.



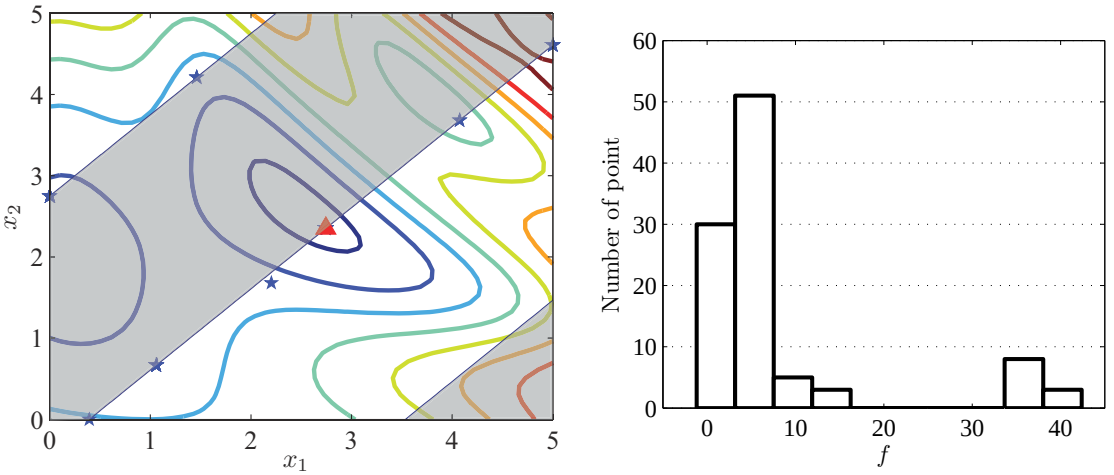
(a) Initial point $x = 0.9$ (b) Initial point $x = 1.7$

Figure 3.2: SQP example with Noisy function

The same problem occurs in a multimodal optimization problem. An analytical optimization problem taken from [95] illustrates this issue. The optimization problem is to minimise a multimodal objective function with respect to a nonlinear constraint:

$$\begin{aligned} \min_{x} \quad & f = 2 + 0.01\left(x_2 - x_1^2\right)^2 + (1 - x_1)^2 + \\ & \qquad 2\,(2 - x_2)^2 + 7\sin(0.5x_1)\sin(0.7x_1x_2) \\ \text{subject to} \quad & g_{exp} = -\sin(x_1 - x_2 - \pi/8) \leq 0 \\ & x_1 \in [0,\ 5],\ x_2 \in [0,\ 5] \end{aligned} \tag{3.3}$$

Figure 3.3a shows the contour plot and different optimal results found by 100 SQP runs using random initial points. The global optimum is shown by a red triangle. One can observe from Figure 3.3b that only 30 runs converge to the global optimum, which is $x_1 = 2.7450$, $x_2 = 2.3523$ and $f = -1.1743$.



(a) Contour plot (b) Different optima

Figure 3.3: SQP Multimodal example

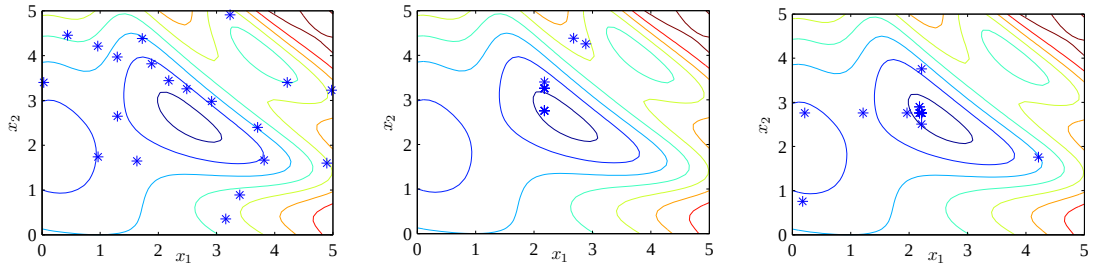
When dealing with the problem with multiple optima (multimodal), the "multistart" strategy is very useful in order to obtain the global optimum. However, multiple SQP runs can require high computation time.

3.1.1.2 Genetic algorithm

Genetic algorithms (GAs) are part of the most popular Evolutionary Algorithm (EA). GAs are stochastic algorithms whose search methods are based on Darwin's theory of natural selection [23, 77].

Like other EAs, a genetic algorithm applies selection operators and variation operators or search operators (called genetic operators in GAs' case) to a set of *individuals* (design vector). A *Fitness* value is assigned to each individual. The fitness depends on the objective function and represents how good the individual is. The set of individuals or *population* is treated simultaneously and improved from the actual *generation* (iteration) to the next one. Two classical genetic operators are used: mutation and crossover. The mutation operators transform an individual. The crossover operators use two or more parents to create a child (called *offspring*) for the next generation.

GAs are known as global optimization algorithms as they are less sensitive to noisy functions and able to solve multimodal problems. Figure 3.4 shows a two-dimensional example (3.3) without constraint. A GA from Matlab "Genetic Algorithm and Direct Search Toolbox" is used in this example. Individuals are plotted for the initial population, intermediate and final generations. One can observe that individuals progressively converge to the global optimum. In the intermediate generations, the populations are grouped at several design locations. Most of them, therefore, are placed nearby the global optimum at the few last generations.



(a) Initial population (b) Generation 10 (c) Generation 20

(d) Generation 30 (e) Generation 40 (f) Generation 51 (final)

Figure 3.4: GA convergence over iterations

One of the main advantages of GAs is that they can handle discrete and non-classifiable variables [105]. This is not possible with an algorithm using gradient information. However, some of GAs drawbacks can be stated:

(i) GAs are not likely to be able to find high accurate results compared with a deterministic algorithm such as SQP, since GAs cannot ensure optimality [77]. The stopping criterion is usually the maximum number of generations or the number of "stall" generations[4].

(ii) GAs usually require a high number of function calls due to their global search characteristics. However, since GAs work with population, they can be easily parallelized. This reduces significantly the computation time.

To overcome these drawbacks, a hybrid algorithm combining GA and SQP is presented in the next section.

3.1.1.3 Hybrid algorithm

As mentioned above, SQP and GAs have their own advantages and drawbacks. To summarize, SQP requires a smaller number of objective and constraint function calls than GAs. It can also find accurate optimum results as it is a deterministic algorithm. However, due to the fact that SQP uses gradient information in its search algorithm, it tends to be trapped in the local optimum and suffers from noise in objective or constraint functions. In contrast, GAs search more globally and have more chance to find a global optimum. J. H. Holland suggested [37] that GAs should be used to perform the initial global search. The results are used to guide the local search.

In order to benefit the global search ability of a GA and the accurate local search of a SQP, they are used as a complement of each other [36, 62]. To do so, the GA stopping criteria are set so that the GA would stop prematurely e.g. with a low generation, a low population or a high tolerance. It is assumed that the GA should find its optimal results near the true global optimum. The GA results are therefore used as an initial point for the SQP algorithm. The SQP proceeds the local search and find its local optimum, which is the global optimum sought.

Table 3.1 shows optimization results for a multimodal 2D problem (3.3) obtained using a SQP, a GA and a hybrid GA and SQP algorithm. 100 runs were performed for each algorithm. The GA parameters are 100 for the population and 100 maximal generations. In the hybrid algorithm, the population is reduced to 20. The hybrid algorithm gives more accurate results than the GA. 87 runs of the hybrid algorithm give the results with a relative error lower than 0.01% compared to only 4 runs of the GA and 26 runs of the SQP. It can be observed that the GA performed better for higher relative error (5%). This is due to the fact that the GA uses a higher population than the hybrid algorithm. This allows the GA to serach more globally and not to be trapped into the local optima. Regarding the number of function evaluation, the SQP used the lowest function evaluation. However, the SQP must be launched many times, using multi-start strategy and needs a higher function call number. The GA required

[4]Number of successive generations with no improvement in the fitness value.

the highest number of evaluations. The hybrid algorithm offers a compromise between SQP and GA. In this example, if only one optimization is launched, the hybrid algorithm has the highest probability to find the accurate global optimum.

This kind of comparison is also performed for a traction system design problem. The reader should refer to [61] for details.

Table 3.1: Comparison between 3 optimization algorithms

Method	Relative error			Number of function evaluation		
	$\leq 0.01\%$	$\leq 1\%$	$\leq 5\%$	min.	avg.	max.
SQP	30	30	30	9	23	97
GA	4	66	92	1969	5593	10657
Hybrid GA,SQP	87	87	87	667	1261	1971

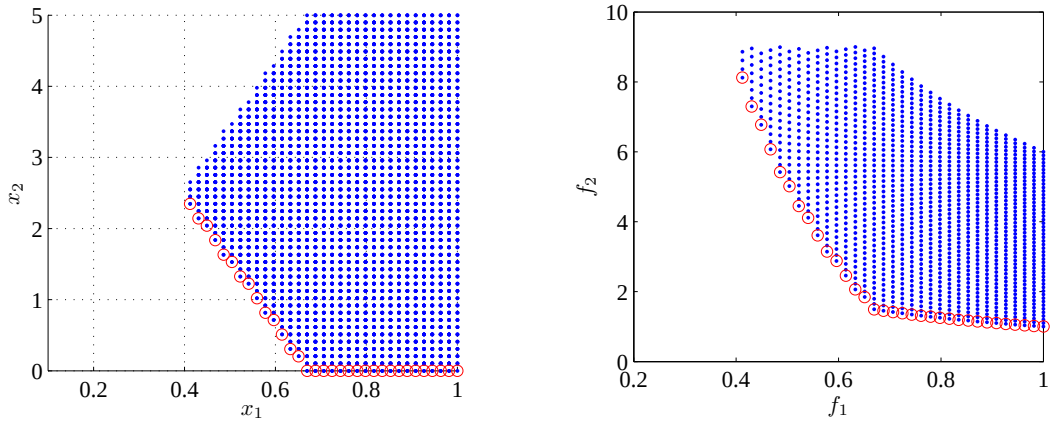
3.1.2 Multi-objective

One should be realise that in real life, most of engineering design problems are multi-criteria or multi-objective problems. To optimise such design problems, the engineer may simplify the multi-objective "design problems" and formulate a single-objective "optimization problem" due to the limited performance of early developed optimization algorithms. Only the most important criterion is selected as a sole objective function. The others are fixed as parameters or constraint limits. It is important to note that this process is done during the optimization problem formulation phase. This approach is called the "*a priori method*". The transformation techniques are introduced in Chapter 1 and will be detailed here since these techniques can also be used to generate a Pareto front, which is known as the "*a posteriori method*". This section intends to explain how to solve the "multi-objective design problem" using the a posteriori approach.

To illustrate the algorithms, two test problems are chosen from the literature. These examples will be used all along this Chapter. Both problems have 2 design variables and 2 objectives to be minimised. The first problem is a constrained problem "CONSTR" retrieved from [20]. It is expressed as following:

$$\begin{aligned} &\min && f_1 = x_1 \\ &\min && f_2 = (1 + x_2)/x_1 \\ &\text{subject to} && g_1 = x_2 + 9x_1 \geq 6 \\ & && g_2 = -x_2 + 9x_1 \geq 1 \\ & && x_1 \in [0.1,\ 1] \text{ and } x_2 \in [0,\ 5] \end{aligned} \tag{3.4}$$

Figures 3.5 are obtained from 50 step grids (2500 points in total). Only feasible designs are plotted in the design space as shown in Figure 3.5a and in the objective space as in Figure 3.5b. The Pareto optimal solutions are depicted by the red circles.



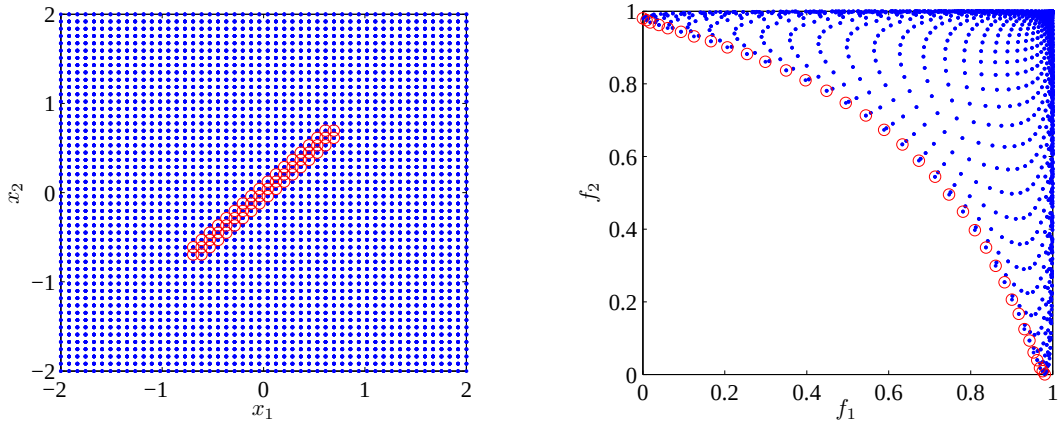
(a) Design space (b) Objective space

Figure 3.5: CONSTR test problem

The second problem "VLMOP2" from [108] is without constraint. It is expressed as:

$$\begin{aligned} &\min && f_1 = 1 - \exp\left(-\sum_{i=1,\ldots,n}\left(x_i - \frac{1}{\sqrt{n}}\right)^2\right) \\ &\min && f_2 = 1 + \exp\left(-\sum_{i=1,\ldots,n}\left(x_i - \frac{1}{\sqrt{n}}\right)^2\right) \\ &\text{subject to} && x_1 \in [-2,\ 2] \text{ and } x_2 \in [-2,\ 2] \\ & && n = 2 \end{aligned} \tag{3.5}$$

Figure 3.6 shows 2500 sample points obtained from a grid design. Since there is no constraint, the design points are spread throughout the design space. The Pareto front is concave.



(a) Design space (b) Objective space

Figure 3.6: VLMOP2 test problem

3.1.2.1 Transformation

In the early era of optimization, there were only single-objective optimization algorithms. However, the real-world design problems are multi-criteria problems. As mentioned above, engineers simplify or transform the design problem into an equivalent design problem that can be solved by using a single-objective optimization algorithm. This approach is not very suitable for industrial design processes since the decision (e.g. importance of each criterion, allowable limit) must be made before obtaining optimal design solutions and the optimization process has to be rerun if the results are not satisfying due to the lack of problem knowledge. The a posteriori approach is preferred as it gives the possibility to generate the Pareto front (see Section 1.1.5) and to shift the decision to perform after the optimization phase. Meanwhile, researchers are forced to develop transformation techniques. A multi-objective design problem can be solved i.e. generate the Pareto front, by using multiple runs of single-objective optimization. The objectives are transformed into one objective using transformation technique such as weighted-sum and epsilon-constraint techniques [15, 78]. The equivalent problem is then optimised using single-objective optimization algorithm. The optimization is rerun with different transformation parameters so that an optimal solution set is located.

3.1.2.1.1 Weighted-sum

Several objectives are transformed into one objective with weighted-sum formulation. The weighted-sum optimization problem is expressed as:

$$\begin{aligned} \min_{x} \quad & f_{ws} = \sum_{i=1}^{m} w_i f_i \\ \text{subject to} \quad & g \leq 0 \\ & h = 0 \end{aligned} \tag{3.6}$$

where w_i is the weight for the i^{th} objective function. For the *a priori* approach, weights are given by an engineer using his experiences and knowledge of design problems. The equation (3.6) results in a unique final solution.

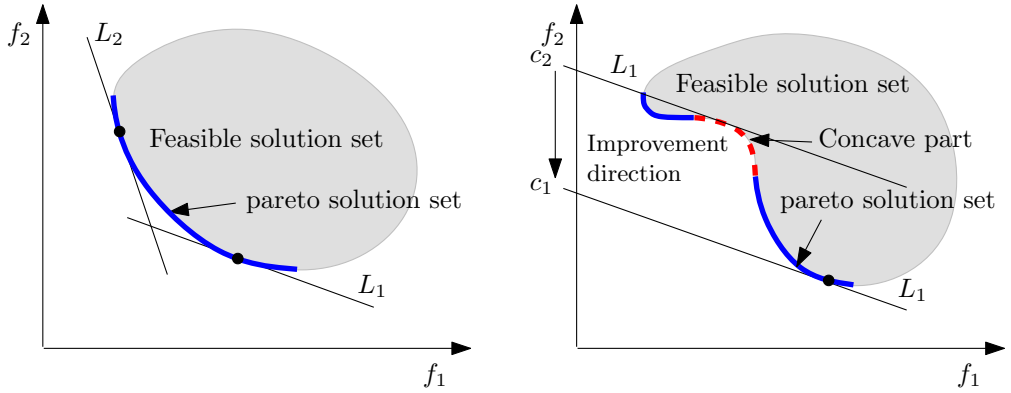
This transformed objective function can be represented as a straight line in an objective space as shown in Figure 3.7. L_1 and L_2 show the slope of f_{ws} with different weight. The minimisation of f_{ws} is equivalent minimizing c in:

$$f_2 = -\frac{w_1}{w_2} f_1 + c = L \cdot f_1 + c \tag{3.7}$$

In the figure, c is minimised. Therefore, Pareto solutions can be located.

One weight vector and one optimization yield a Pareto optimum solution. The weighted-sum optimization problem is then run with different weight parameters in order to find the Pareto front. This approach is the *a posteriori* approach mentioned above. Usually, $w_i \geq 0$ for $i = 1, \ldots, m$ and

$$\sum_{i=1}^{m} w_i = 1 \tag{3.8}$$



(a) Convex feasible solution set (b) Nonconvex feasible solution set

Figure 3.7: Weighted-sum transformation

Even the weighted-sum method is known as an efficient approach [84], this method cannot ensure the non-dominated solutions in the concave part of Pareto front as shown in Figure 3.7b. With the same slope L_1, it tends to find a solution in the convex part where the constant value c is minimised.

3.1.2.1.2 Epsilon-constraint

In the epsilon-constraint method, only one criterion is chosen as objective function. Other criteria are considered as constraints. The epsilon-constraint optimization problem is defined as following:

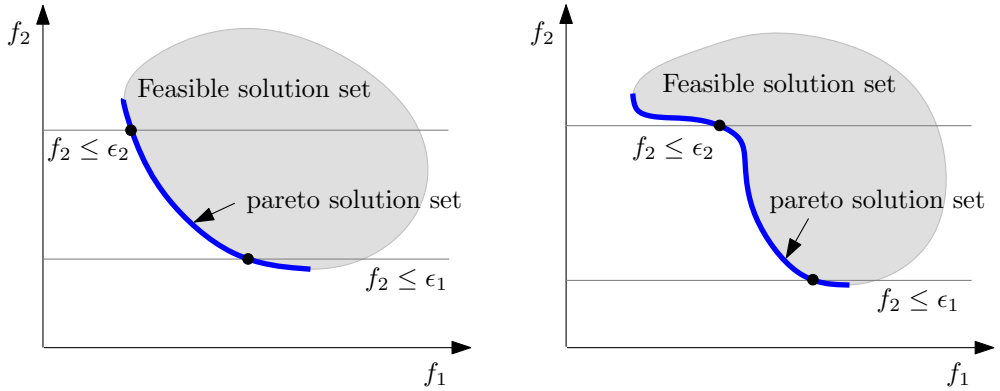
$$\begin{aligned} \min_{x} \quad & f_1 \\ \text{subject to} \quad & g \leq 0 \\ & h = 0 \\ & f_i \leq \epsilon_i \qquad i = 2, \ldots, m \end{aligned} \tag{3.9}$$

where ϵ parameter is an additional constraint value.

By varying ϵ_i for each optimization, the Pareto front can be found. This method can located the whole Pareto front even in the nonconvex zone.

3.1.2.1.3 Illustration example

As already stated in the previous section, the Weighted-sum technique cannot find the non-convex Pareto front. Therefore, it is tested only with the "CONSTR" problem. For solving the "CONSTR" problem, 11 weight parameters are given (10 steps from 0 to 1) and 11 optimizations are launched independently. A SQP with multistart is used for each optimization in order to ensure a global optimum. Table 3.2 shows 11 weight parameters and optimization results i.e. objective functions and constraint functions. Figure 3.9a shows a non-dominated front. One can observe that the weight parameters of point 1 to point 6 lead to the same non-dominated solution. The same behaviour occurs for points 8 to point 10. This is because the change in the slope of the Pareto front is not continuous. Figure 3.9b



(a) Convex feasible solution set (b) Nonconvex feasible solution set

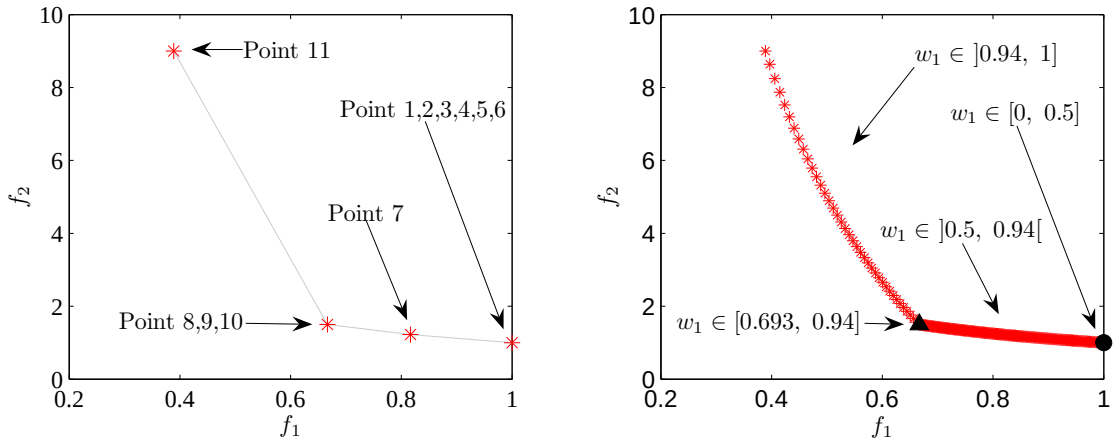
Figure 3.8: Epsilon-constraint transformation

shows the non-dominated front plotted from 1000 weight parameter steps. The large part of non-dominated front on the left side is obtained from very small variation of w_1 from 0.94 to 1.

Table 3.2: Weighted-sum results for "CONSTR" problem

Number	Weight		Design variable		Objective		Constraint	
	w_1	w_2	x_1	x_2	f_1	f_2	g_1	g_2
1	0.00	1.00	1.00	0.00	1.00	1.00	-3.00	-8.00
2	0.10	0.90	1.00	0.00	1.00	1.00	-3.00	-8.00
3	0.20	0.80	1.00	0.00	1.00	1.00	-3.00	-8.00
4	0.30	0.70	1.00	0.00	1.00	1.00	-3.00	-8.00
5	0.40	0.60	1.00	0.00	1.00	1.00	-3.00	-8.00
6	0.50	0.50	1.00	0.00	1.00	1.00	-3.00	-8.00
7	0.60	0.40	0.82	0.00	0.82	1.22	-1.35	-6.35
8	0.70	0.30	0.67	0.00	0.67	1.50	0.00	-5.00
9	0.80	0.20	0.67	0.00	0.67	1.50	0.00	-5.00
10	0.90	0.10	0.67	0.00	0.67	1.50	0.00	-5.00
11	1.00	0.00	0.39	2.50	0.39	9.00	0.00	0.00

For the epsilon-constraint technique, anchor points are sought in the first step by minimizing each objective function. This allows determining epsilon parameters according to the variation in the objective function. Ten epsilon steps are defined for "CONSTR" problems. This yields 11 optimization runs in total. Figure 3.10a plots non-dominated fronts. The front with stars is the case in which the objective function f_1 is minimised and the objective function f_2 is considered as a constraint. Table 3.3 lists design variables and objective and constraint values for each epsilon ϵ_2. The part situated on the right side of this front is not explored. The fixed epsilon constraint step is larger than the variation of the objective f_2 in this part. To overcome this problem, one can minimise f_2 and use f_1 as epsilon constraint



(a) 10 weight steps (b) 1000 weight steps

Figure 3.9: Non-dominated front for "CONSTR" problem obtained from weighted-sum technique

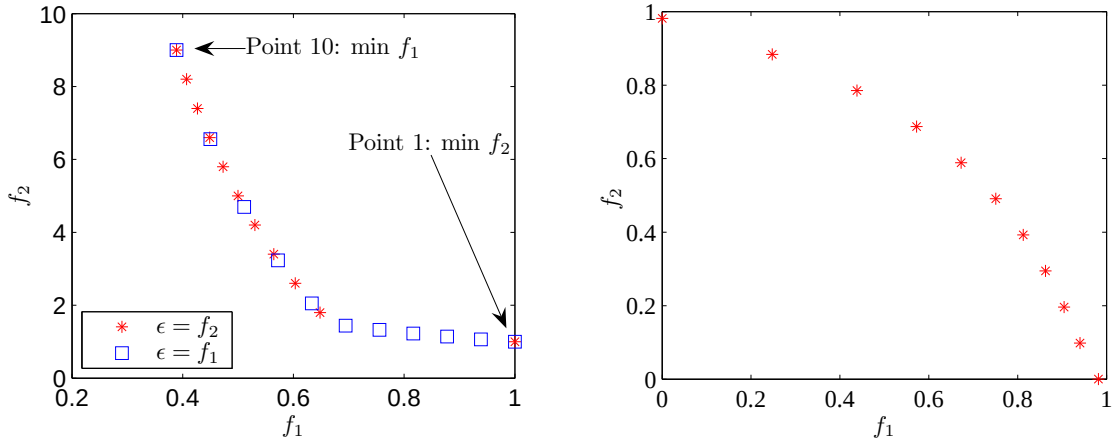
instead. The results are depicted by squares in the same figure. Non-dominated solutions are better spread throughout the front. Figure 3.10b shows Pareto optimal front for the "VL-MOP2" problem obtained by using epsilon-constraint technique. It shows that this technique can locate nonconvex Pareto fronts.

Table 3.3: Epsilon-constraint results for "CONSTR" problem

Number	Epsilon	Design variable		Objective		Constraint	
	ϵ_2	x_1	x_2	f_1	f_2	g_1	g_2
1	min f_2	1.00	0.00	1.00	1.00	-3.00	-8.00
2	1.80	0.65	0.17	0.65	1.80	0.00	-4.67
3	2.60	0.60	0.57	0.60	2.60	0.00	-3.86
4	3.40	0.56	0.92	0.56	3.40	0.00	-3.16
5	4.20	0.53	1.23	0.53	4.20	0.00	-2.55
6	5.00	0.50	1.50	0.50	5.00	0.00	-2.00
7	5.80	0.47	1.74	0.47	5.80	0.00	-1.51
8	6.60	0.45	1.96	0.45	6.60	0.00	-1.08
9	7.40	0.43	2.16	0.43	7.40	0.00	-0.68
10	8.20	0.41	2.34	0.41	8.20	0.00	-0.33
11	min f_1	0.39	2.50	0.39	9.00	0.00	0.00

Both transformation techniques allow using a state-of-the-art single-objective optimization algorithm to generate a Pareto front. The application of a weighted-sum technique to the optimal design in electrical engineering can be found in [55].

However, the examples above show some drawbacks of the method. The weighted-sum technique encounters the duplicated solution problem when the slope of the Pareto front is



(a) CONTRE problem

(b) VLMOP2 problem

Figure 3.10: Non-dominated fronts obtained from epsilon-constraint technique

not continuous. Moreover, it cannot solve the problem with non-convex fronts. The epsilon-constraint works slightly better since it can locate nonconvex front. Furthermore, it solves the problem in a more intuitive manner.

3.1.2.2 Multi-objective algorithm

Some recently developed multi-objective algorithms can generate a Pareto front in a single run. These algorithms fall into the class of evolutionary algorithms. They deal with a group of feasible solutions concurrently. They allow finding a set of optimal solutions i.e. non-dominated solutions from a single optimization. Moreover, evolutionary algorithms can easily cope with discontinuous or non-convex Pareto fronts [20]. This section introduces one of the most popular and best performances multi-objective evolutionary optimization algorithm: Non-dominated sorting genetic algorithm II (NSGA-II) [20].

NSGA-II uses an elitist approach and a Pareto optimal concept. It combines two sorting procedures: non-dominated sorting and crowding distance sorting. From the previous population P_t, an offspring population Q_t of the same size N_{pop} is created by using binary tournament selection, recombination and mutation operators. A combined population R_t is sorted based on the nondomination and each solution is assigned a rank as shown in Figure 3.11a. The rank 1 solutions (denoted as F_1) are the best solutions. For a constrained optimization, the definition of domination is modified in order to take into account the constraint issue. According to Deb et al. [20], a solution i "constrained-dominates" a solution j if any of these three conditions is true:

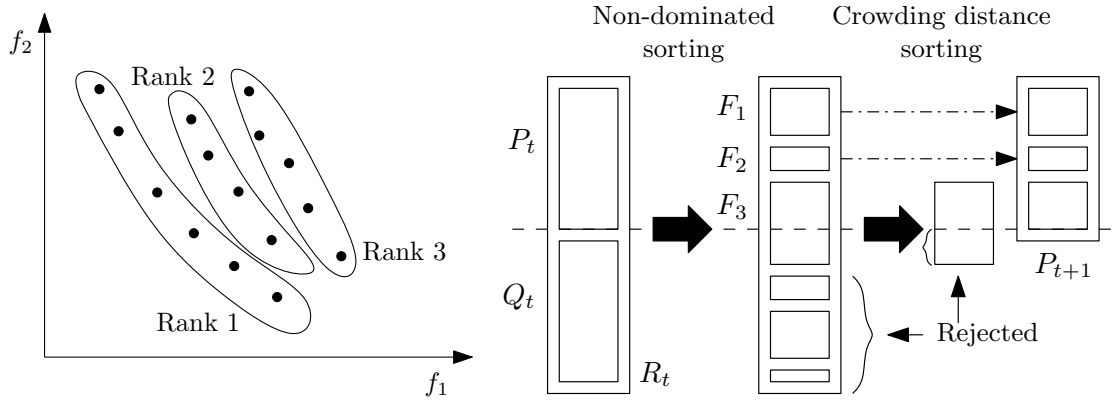
(i) The solution i is feasible and the solution j is infeasible.

(ii) The solution i and j are infeasible but the solution i generates a smaller constraint violation.

(iii) The solution i and j are feasible and the solution i dominates the solution j.

Any feasible solution is always given a better rank level than any infeasible one. Any infeasible solution with a smaller constraint violation is better than a solution with a higher constraint violation. The comparison between feasible solutions is not changed. Their ranks depend on the level of domination according to (iii). By using the ranks, the algorithm converges to non-dominated and feasible solutions.

To construct a new population for the next generation P_{t+1}, the lower rank sets are selected in priority i.e. rank 1 then rank 2 and so on. Assume that F_l is the last non-dominated set that can fit into P_{t+1}. If the total number of solutions from $F_1, F_2, \cdots, F_l$ is larger than N_{pop}, only the best solutions of F_l can be selected. The crowding distance is used for this purpose. The crowding distance allows estimating the density of solutions around a solution in the population. The higher the crowding distance, the more sparsely solutions on the front[5]. The crowding distance is computed for the last rank F_l. The solutions with higher crowding distance are selected into P_{t+1} until the total number of population reaches N_{pop}. Figure 3.11b from [20] depicts this sorting procedure.



(a) Non-dominated Ranking (b) Sorting procedure

Figure 3.11: NSGA-II

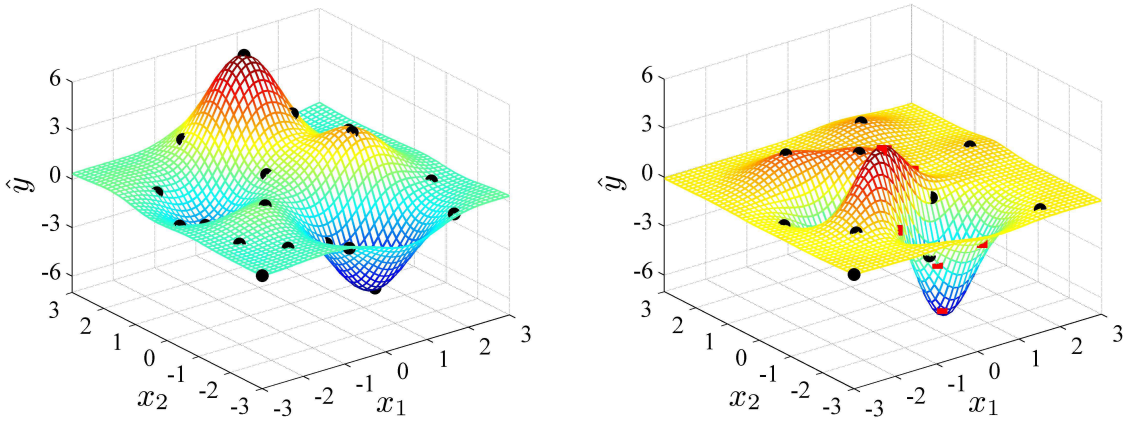
As GAs for single-objective problems, NSGA-II can handle the mixed variable problem. The Pareto fronts for two test problems "CONSTR" and "VLMOP2" are given in Section 3.1.3.7.2. Figure 3.22 shows the results from NSGA-II and a surrogate-assisted algorithm presented below. The application of NSGA-II to the design of traction motors can be found in [60].

[5]The crowding distance assignment process can be found in [20]

3.1.3 Surrogate-assisted optimization algorithm

3.1.3.1 Overview

Chapter 2 has presented several surrogate modelling techniques. The surrogate model constructed on the basis of these techniques can be used for many purposes as mentioned in Section 2.2.1. One of the main purposes is to replace the fine model in the optimization process in order to decrease optimization time. The evaluation of surrogate models is fast during an optimization process, however, the final solution remains the approximated one. This approximated solution must be validated with high fidelity models. This validation may not be satisfying if a surrogate model presents low accuracy characteristics. If the number of sample points is low, the surrogate model will not be accurate. On the other hand, if the number of sample points is too high, this yields a high accuracy model, but some sample points can be wasted if they are not close to the sought optimum. When using a surrogate model in the optimization process, the most important area is the area where the optimum is located. The sample points should be dense in this zone. In other areas, the sample points can be more sparse. The surrogate model gives the global trend and guides the optimization algorithm towards the optimum. Figure 3.12 shows two surrogate models constructed from well distributed sample point sets (Figure 3.12a) and sample points concentrated near the optimum with the same amount of sample points (3.12b). The optimal results are found at $\hat{y} = -4.6742$ for the former case and -6.2504 for the latter case. The true optimum is at $f = -6.5511$. Even the form of the model constructed from well-distributed sample points is similar to the true function (see Figure 2.16a) but the optimal result is not correct. The model built from sample points concentrated near the optimum gives a better result. In this case, the overall error is higher but the model is more accurate near the optimum sought.



(a) Distributed sample point set

(b) Sample points concentrated near optimum

Figure 3.12: Surrogate model for an optimization purpose

Surrogate model refinement can be performed manually by the designer or even better, by an optimization algorithm. For example, the simplest algorithm can iteratively add sample points where the minimum of the surrogate model is located until the error between the fine

model and the surrogate model is sufficiently small. Such optimization algorithms are in the class of Surrogate-assisted optimization algorithm. Some of them can be stated: Efficient Global Optimization (EGO) [97], SuperEGO [94], RBF based global optimization [33]. A taxonomy of surrogate-assisted algorithms can be found in [43]. Most of them have a common general algorithm. The difference is how the infill criteria are described.

The following sections give the general algorithm and the explanation for the single-objective case. Then the particular information concerning the multi-objective case are given. As the general principle is common between single and multi-objective case, the information given for the single-objective case is also useful for the multi-objective case if it is not stated otherwise.

3.1.3.2 Main algorithm

Figure 3.13 shows the general flowchart of a surrogate-assisted optimization algorithm (single or multi-objective). Some differences may be found, depending on the algorithm.

The algorithm can be described step-by-step as following:

(i) The algorithm starts with sampling a set of initial points using sampling technique[6].

(ii) The initial sample points are evaluated by using a high fidelity model.

(iii) Initial surrogate models are fitted individually for each objective function and constraint.

(iv) An infill point is located by maximizing the infill criteria. Most of the surrogate-assisted algorithms use a Kriging surrogate model and an estimated standard error ($\hat{s}$) associated with Kriging in their infill criteria. This step allows selecting the most promising design vector to evaluate high fidelity models such as FEA at the next iteration.

(v) Once a design vector of infill point is located, the high fidelity model is evaluated.

(vi) The infill high fidelity results are validated against acceptance rules. In a single-objective optimization, it may be to simply check if the objective function of the infill point is lower than the best objective function known so far. For constraint optimization, it is important that the constraint value of the infill point does not violate the constraint condition (within a small tolerance). It will be included in the improvement solution set if these conditions are verified.

(vii) In any case, the new infill point is added to the sample point set.

(viii) At each iteration, the algorithm checks if the stopping criteria are met. Various stopping criteria can be used. They include the maximal number of iterations, the maximal time limit or the minimal expected improvement. The minimal expected improvement is particularly used for EGO algorithm. The maximal number of iterations and time limit are more practical as the engineer can control his allowable computer resources.

(ix) Return to (iii). The Kriging model is then reconstructed and the process continues until the stopping criteria are met.

[6] DOE or space-filling design. See Section 2.2.2

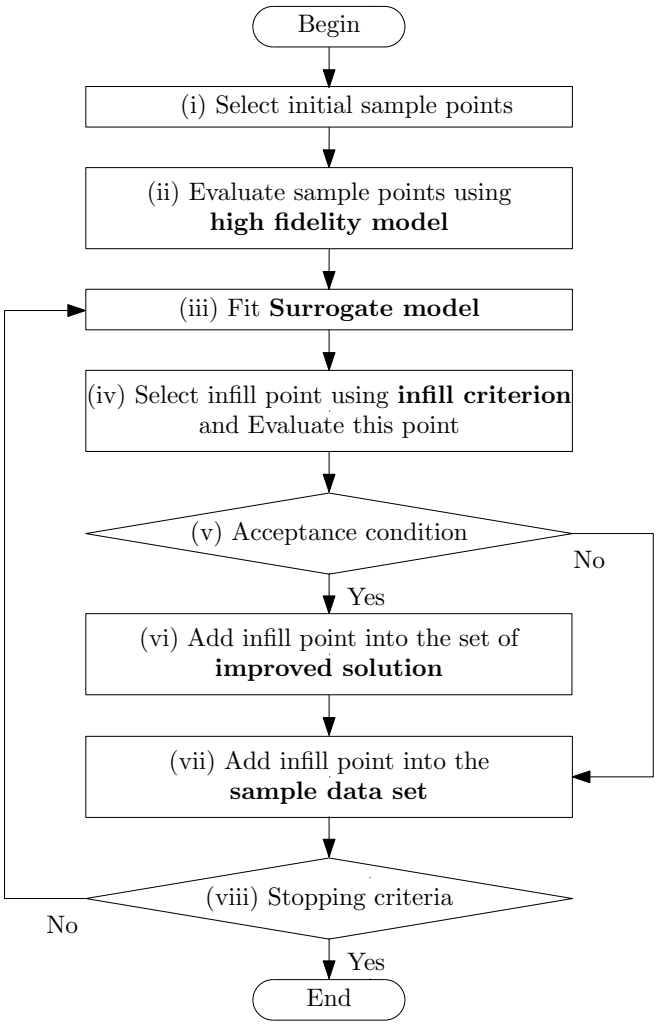


Figure 3.13: Flowchart of a Surrogate-assisted algorithm

3.1.3.3 Initial point selection

The location of initial sample points should be well distributed over the design space. The sampling techniques presented in Section 2.2.2 can be used. Space filling designs are reported to perform better with a small number of data set [42].

The number of initial points $n_{s_{init}}$ is not well studied in the literature and remains arbitrary. D. Jones et al. suggest $10{\cdot}n_v$ initial points [44] for a single-objective optimization. In constraint optimization, the number of initial points may depend on the number of design variables (n_v) and on the inequality and equality constraint (n_g, n_h). From previous experiences, the rule described as in (3.10) gives good results. For a multi-objective optimization, $n_{s_{init}}$ is also scaled according to the number of objective m. This issue is not investigated in this research. Further study of it would however be interesting.

$$n_{s_{init}} = 5m \cdot (n_v + n_g + 2n_h) \tag{3.10}$$

It should be noted that some of these initial points must be feasible. In hard constrained optimization problems, the feasible sample point search procedure may be performed before doing the infill point search.

3.1.3.4 Infill criteria for single-objective optimization

In single-objective optimization, in order to find an optimum of the true function using the surrogate model and predicted model error, one should evaluate a new sample point where:

(i) the surrogate model is minimal and the error is low. This is called exploitation.

(ii) the error is high. This is called exploration.

The combination of 2 criteria of selection is known as Infill Criteria (IC). The infill criteria are usually based on statistic approaches like, for example, the probability of improvement [43], the expected improvement [44] and the generalized expected improvement [97]. The infill criterion has to be maximised using a conventional single-objective optimization algorithm. Due to its multimodal properties, a global optimization algorithm is more suitable. A genetic algorithm or a gradient-based algorithm with multi-start can be used.

In this research, the Generalized Expected Improvement (GEI) is used. It was proposed by M. Schonlau et al. [97] in EGO algorithm. It is the probability that the estimated response ($\hat{y}$) is smaller than the current minimal feasible objective function (f_{min}) and the uncertainty associated with Kriging model ($\hat{s}$) is high. It can be explained as how the objective will be improved if we add a candidate point into the sample point set. Mathematical proof of GEI can be found in [96].

GEI can be expressed as:

$$\text{GEI}=\text{E}\left[I^g\left(x\right)\right]=\hat{s}^g\sum_{k=0}^{g}\left(-1\right)^k\left(\frac{g!}{k!\left(g-k\right)!}\right)u^{g-k}T_k \tag{3.11}$$

where

$$u\left(x\right)=\frac{f_{min}-\hat{y}\left(x\right)}{\hat{s}\left(x\right)} \tag{3.12}$$

$$T_k=-\phi\left(u\right)u^{k-1}+\left(k-1\right)T_{k-2} \tag{3.13}$$

$$T_0=\Phi\left(u\right) \tag{3.14}$$

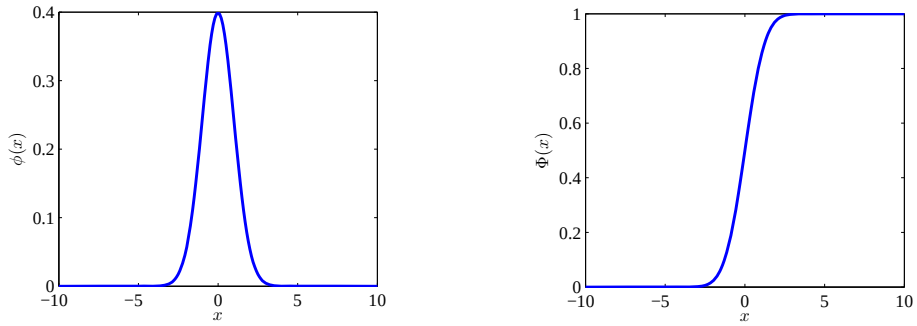
$$T_1=-\phi\left(u\right) \tag{3.15}$$

$$\hat{s}\left(x\right)=\sqrt{MSE\left(x\right)} \tag{3.16}$$

Φ and ϕ are normal cumulative distribution and normal probability density functions, respectively (see Figure 3.14).

The value of the non-negative integer g allows controlling the behaviour of GEI. If g is high, GEI is high in the uncertain region. But if g is small, the peak of the GEI function moves to the region where the probability of finding a better response is high (see Figure 3.15). The GEI with $g=1$ becomes the expected improvement (EI) given as in (3.17) [44]. In this case, two parts can be observed clearly. The first part performs local search i.e. $\hat{y}$ is smaller than f_{min}. The second part allows to search more globally in low accuracy areas i.e. high $\hat{s}$.

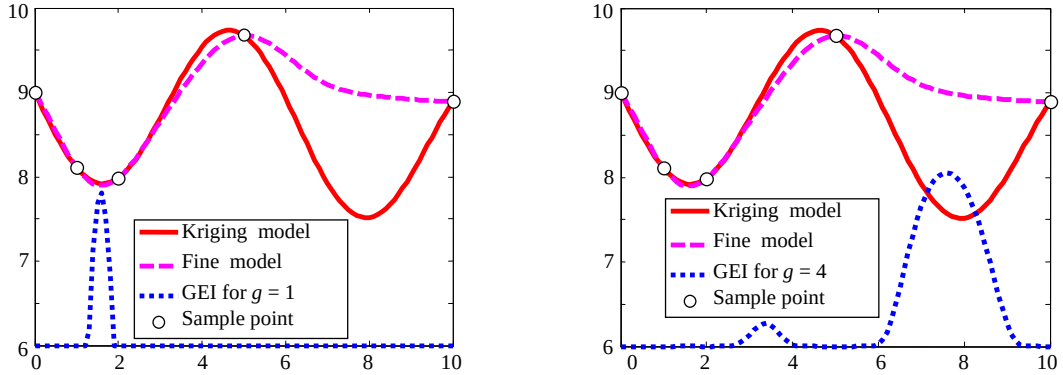
$$\text{EI}=\text{E}\left[I\left(x\right)\right]=\underbrace{\left(f_{min}-\hat{y}\right)\cdot\Phi\left(\frac{f_{min}-\hat{y}}{\hat{s}}\right)}_{\text{Local search term}}+\underbrace{\hat{s}\cdot\phi\left(\frac{f_{min}-\hat{y}}{\hat{s}}\right)}_{\text{Global search term}} \tag{3.17}$$



(a) ϕ Normal probability density function

(b) Φ Normal cumulative distribution function

Figure 3.14: Normal cumulative distribution and normal probability density functions



(a) $g = 1$

(b) $g = 4$

Figure 3.15: Generalized Expected Improvement for (a) $g = 1$ and (b) $g = 4$

M. J. Sasena et al. proposed the cooling schedule [95], which starts with high g and then decreases. In this work, g has been fixed to 5 for the first iterations. This allows the algorithm to search globally. It is then decreased to 1 when reaching higher iteration numbers in order to improve local search. Other schemes can be used depending on the problem considered.

3.1.3.5 Constraint handling

The constraints can be grouped into 2 types: inexpensive constraint (g_{inexp}) and expensive constraint (g_{exp}). An inexpensive constraint can be evaluated rapidly using analytic model for example mass calculation or geometry constraint. It is then used directly in the infill optimization process (infill sub-optimization problem). It is handled from the constraint optimization algorithm being used. On the contrary, an expensive constraint is evaluated by using the high fidelity model requiring high computation time. In addition to the objective function, the expensive constraint is modelled individually, by using a surrogate model.

Two constraint handling methods are investigated. The probability method transforms a constrained optimization problem into an unconstrained one. The second method uses a direct constraint approach. A particular equivalent constraint function allows taking into

account the uncertainty of a surrogate constraint model. The two methods are presented firstly and an example with 4 cases shows the particular characteristics of each technique.

3.1.3.5.1 Probability method

In order to take into account constraints, the Infill Criteria (IC)[7] is multiplied by the probability that the interesting point is feasible. The infill maximisation problem with constraint using the probability method is described as [96]:

$$\begin{aligned} &\max_{x} && IC\left(x\right) \cdot \prod P_{exp}\left(x\right) \\ &\text{subject to} && g_{inexp} \leq 0 \\ &\text{where} && P_{exp}\left(x\right) = P\left(\hat{g}_{exp}\left(x\right) \leq 0\right) = \Phi\left(\frac{0 - \hat{g}_{exp}\left(x\right)}{\hat{s}_{g_{exp}}\left(x\right)}\right) \end{aligned} \tag{3.18}$$

P_{exp} is the probability that the estimated expensive constraint ($\hat{g}_{exp}$) is lower than or equal to 0 i.e. the interesting point is feasible.

This probability method allows sampling the infill points in the infeasible zone, which is necessary for increasing the accuracy of a Kriging constraint model. However the probability method impacts on the infill criterion function too strongly near the border of the constraint [95]. After a number of iterations when the constraint model is sufficiently accurate, it is suggested to incorporate a constraint surrogate model ($\hat{g}_{exp}$) directly in the sub-optimization problem as already done with inexpensive constraints. This prevents choosing too many infill points in the infeasible zone. This method is described in the following section.

3.1.3.5.2 Direct integration into sub-optimization problems

This method uses a Kriging constraint model directly in the infill sub-optimization problem. The infill optimization maximises the infill criterion and points out an infill design vector, which is feasible according to the information from surrogate models. Since the Kriging models are not sufficiently accurate at the first iterations, it is possible that it is not feasible and fails when evaluating with the fine model and validating against the acceptance condition. This is always true even at high iteration numbers, when the surrogate models are sufficiently accurate. For these reasons, the surrogate models should not be treated as if they were correct.

In order to avoid this problem, we propose a novel approach [59], which allows taking into account the uncertainty of a surrogate constraint model instead of directly using the surrogate model of expensive constraints. The surrogate equivalent of an expensive constraint ($\hat{g}_{exp_{eq}}$) is expressed as:

$$\hat{g}_{exp_{eq}} = \hat{g}_{exp} + \alpha \hat{s}_{g_{exp}} \tag{3.19}$$

where $\hat{g}_{exp}$ and $\hat{s}_{g_{exp}}$ are the Kriging prediction and the estimated standard error of expensive constraint[8], respectively. α is positive adaptive coefficient.

[7] It can be any infill criteria mentioned in 3.1.3.4. In this research, the GEI is considered.

[8] $\hat{g}_{exp}$ and $\hat{s}_{g_{exp}}$ are the same scale and use the similar unit

The infill maximisation problem with constraint using direct integration method is described as:

$$\begin{aligned} \max_{x} \quad & IC\,(x) \\ \text{subject to} \quad & g_{inexp} \leq 0 \\ & \hat{g}_{exp_{eq}} \leq 0 \end{aligned} \tag{3.20}$$

This method allows taking into account the estimated errors of the Kriging constraint model. As $\hat{s}_{g_{exp}}$ is always positive, $\hat{g}_{exp_{eq}}$ is higher than $\hat{g}_{exp}$ at any unknown design site[9]. This acts as a "factor of safety". One can start with high α so that the infill optimization is more conservative i.e. the infill point is selected only in the feasible zone. The adaptive coefficient is then decreased gradually when the Kriging model is more and more accurate, as the number of iterations and sample points increase. Therefore, the infill points are placed near or at the constraint border.

Drawbacks of this surrogate equivalent of expensive constraint ($\hat{g}_{exp_{eq}}$) are that the infill selection process is conservative and that the constraint model in the infeasible zone is hardly improved. For this reason, it is more efficient to use the probability method at the first iterations and switch to the direct constraint integration approach when the constraint model is sufficiently accurate as suggested above.

3.1.3.5.3 Constraint handling illustration example

A one-dimensional mathematics test problem is described in (3.21). This example intends to show the behaviour of constraint handling methods via four illustration cases. The objective function is the same as (2.4) in Chapter 2. Figure 3.16 shows objective and constraint functions. A coordinate of three local optima are shown in Table 3.4.

$$\begin{aligned} \min_{x} \quad & y = \sin(x) - \exp\left(\frac{x}{20}\right) + 10 \\ \text{subject to} \quad & g_{exp} = (0.0106x^3 - 0.0943x^2 + 0.0822x + 8.0447) - y \leq 0 \\ & 0 \leq x \leq 10 \end{aligned} \tag{3.21}$$

Table 3.4: Three optima of a one-dimensional test problem

Number	x	Objective
1	1.1430	8.0313
2	1.8482	7.9414
3	7.2662	7.7297

[9] At an unknown design site, $\hat{s}_{g_{exp}}$ is of positive value. $\hat{s}_{g_{exp}}$ is zero at a known sampled site.

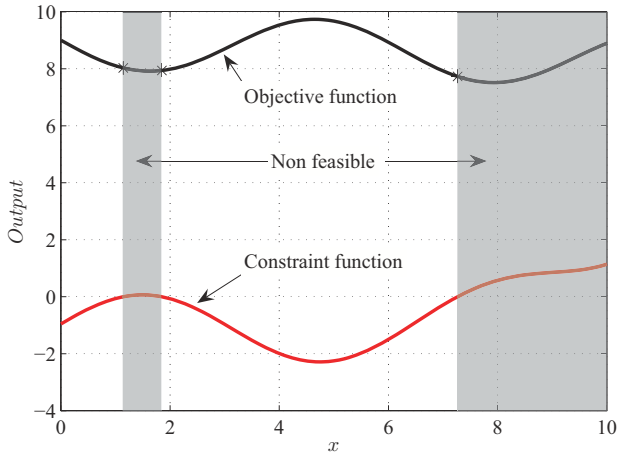


Figure 3.16: One dimensional test optimization problem

Figure 3.17a and 3.17b show Kriging models constructed from five sample points (the five sample points and corresponding objective and constraint values are shown in Table 3.5). Lines represent true functions and dotted lines depict Kriging models. An error between true functions and Kriging models can be observed. Figure 3.17b also shows the surrogate equivalent constraint $\hat{g}_{exp_{eq}}$ and Kriging constraint model $\hat{g}_{exp}$. It is constructed for $\alpha = 1$.

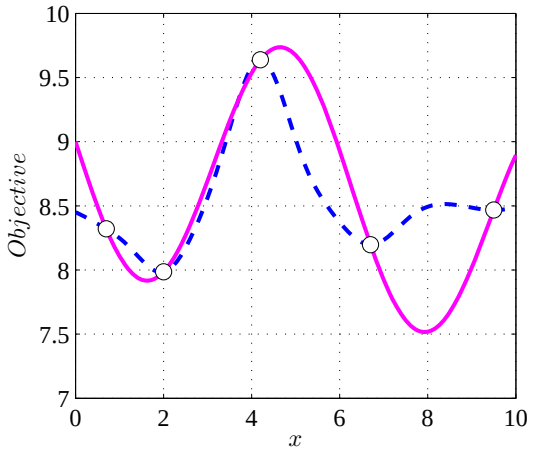
Table 3.5: Sample point for one dimensional test problem

Point	x	Objective	Constraint
1	0.7	8.3202	-0.2605
2	2	7.9855	-0.0688
3	4.2	9.6379	-2.1261
4	6.7	8.1972	-0.6468
5	9.5	8.4671	0.9361

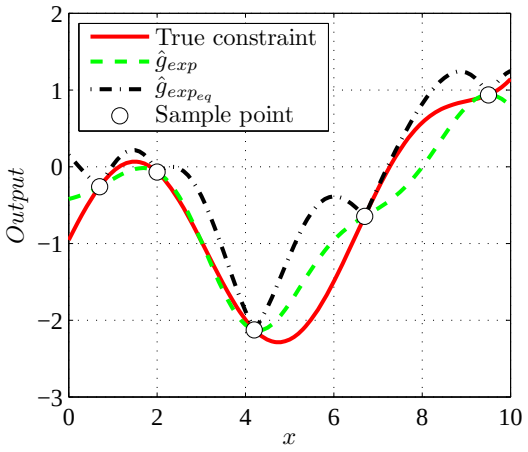
Figure 3.17c plots infill criteria based on GEI with and without probability constraint handling. To compute GEI (3.11), the control parameter g is set to 4 and f_{min} is the minimum objective value among 5 sample points i.e. 7.9855. One can observe that the infill criteria differ from each other. The maximum value of infill criteria moves from the right side to the left side in the case of a GEI with probability constraint where the feasible probability is high.

The following cases give an example of when the algorithm is at the first iteration. Four cases correspond to different ways of treating objective and constraint models in the infill sub-optimization problem. Figures 3.18 shows infill criteria and the constraint model for each case (The comparison of different models is already shown in Figure 3.17 above.) and Table 3.6 gives a summary of the model, the infill location and the results during the validation step.

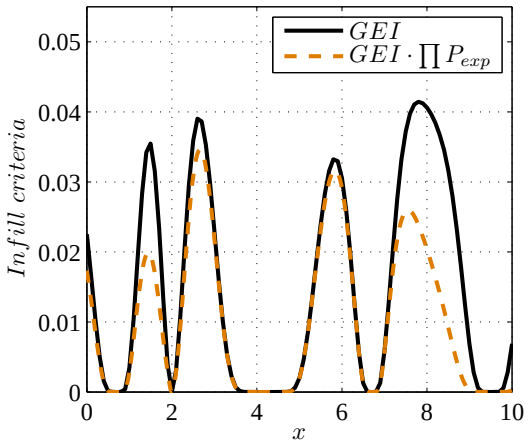
Case 1: This is the simplest surrogate-assisted algorithm. The Kriging objective model is



(a) Objective function

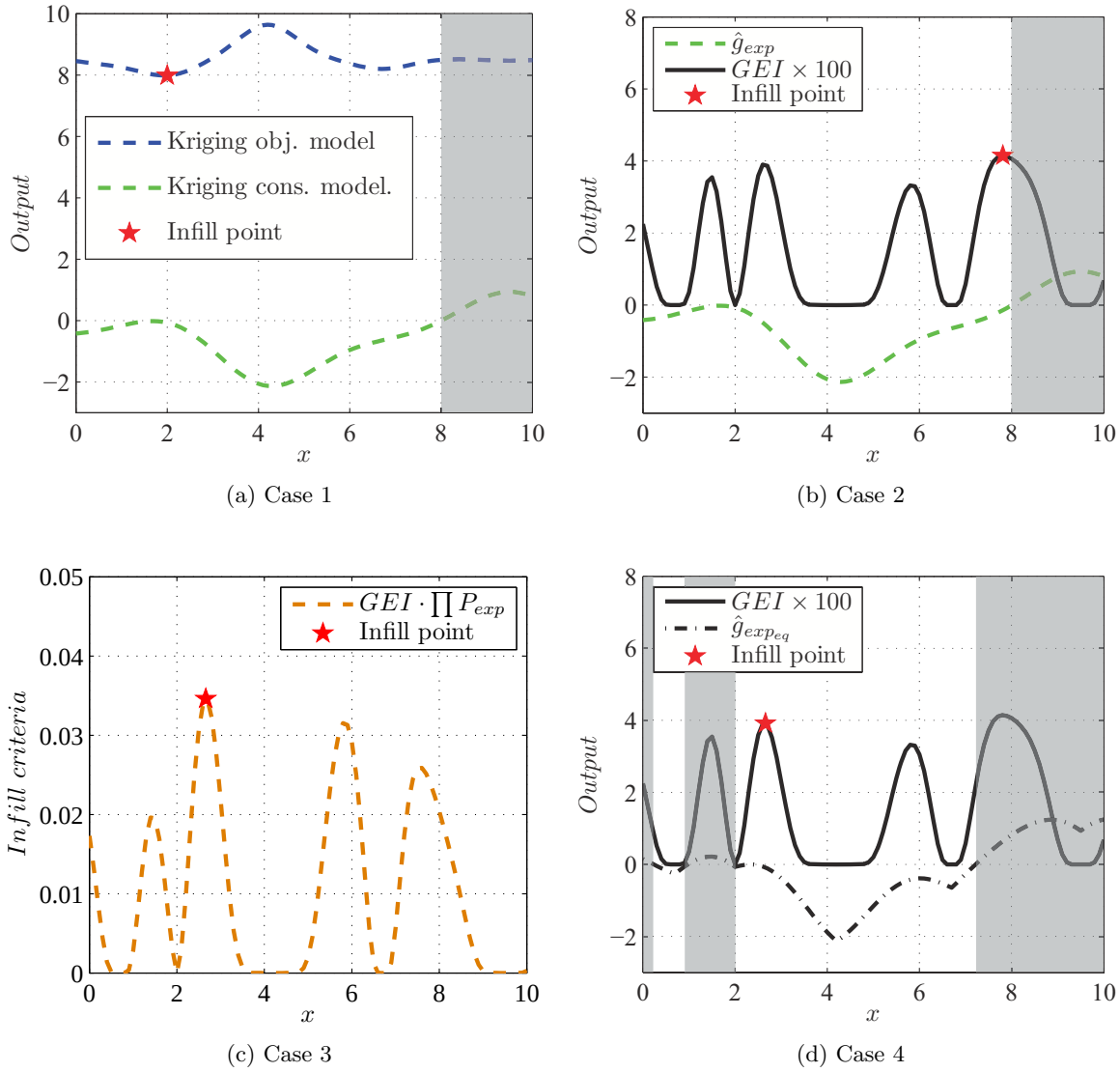


(b) Constraint function



(c) GEI with probability constraint handling method

Figure 3.17: Different models used in one-dimensional examples



(a) Case 1

(b) Case 2

(c) Case 3

(d) Case 4

Figure 3.18: One-dimensional example: Objectives and constraints for each case

minimised and Kriging constraint models are used. This is done without taking into account the inaccuracy of the models. The optimal result is then verified with the fine model (true function in this case) and if the error between two models is higher than a predefined allowable error, the Kriging model is updated with this optimum point and optimization is rerun. As a result, The minimum is found when $x = 2$ for the first iteration. It is clear that the result is not improved, even when the additional point is added at this minimum point, because the Kriging model and the true function give the same result. Therefore, the algorithm will stop at this iteration.

Case 2: Case 2 uses GEI as the infill criteria and the Kriging constraint model is taken into consideration. By maximising GEI with a constraint using the Kriging constraint model[10], the maximum is found at the rightmost peak ($x = 7.81$). This infill point will not be accepted as the improve result when validating against the true models due to the

[10]The location of the maximum infill criterion is used as the new design vector for the next iteration. See Equation (3.18) and Figure 3.13 for the algorithm

fact that this sample point violates the true constraint value ($g_{exp} > 0$). This is because the Kriging constraint model for x between 6.7 and 9.5 gives an incorrect prediction. The constraint model is underestimated. The first two cases confirm that the constraint model must be treated with the use of a more efficient method.

Case 3: Case 3 demonstrates the case of the probability method. The GEI with a feasible probability is given as the infill criterion. It is maximised without any constraint in the infill subproblem (3.18). The infill point is found when $x = 2.66$. This infill point can be validated with the true functions.

Case 4: This case illustrates the direct constraint integration method. It uses GEI as the infill criteria and the surrogate equivalent constraint $\hat{g}_{exp_{eq}}$ instead of the Kriging constraint model as in case 2. The surrogate equivalent constraint shows a more pessimistic constraint prediction compared to the Kriging constraint model. As a result, the infill maximisation with constraint (3.20) gives a result when $x = 2.64$, which is near to that of case 3. This infill point is feasible when validating against the true constraint function.

At this iteration level, these infill points have not yet allow to improve the results. The infill point is added into the sample data set, the Kriging models are refitted and the algorithm proceeds to the next iteration.

Table 3.6: Summary of the model used in four cases and 1^{st} iteration infill point location

Case	Infill criteria	Cons. model	1^{st} iter. infill point $(x,\ y)$	True function validated
1	$\hat{y}$	$\hat{g}_{exp}$	(2, 7.99)	yes
2	GEI	$\hat{g}_{exp}$	(7.81, 8.46)	no
3	$GEI\,(x) \cdot \prod P_{exp}\,(x)$	-	(2.66, 8.28)	yes
4	GEI	$\hat{g}_{exp_{eq}}$	(2.64, 8.27)	yes

3.1.3.6 Multi-objective optimization

In this section, the information specific to multi-objective optimization using surrogate-assisted algorithms are given. The infill criteria and the acceptance conditions are different from those of single-objective optimization.

3.1.3.6.1 Infill criteria for multi-objective optimization

In the explanation below, a trial point represents an objective vector (in objective space) evaluated by using the Kriging model at the trial design vector (in design variable space) given by a single-objective optimization algorithm used in the infill optimization process. An infill point represents an objective vector evaluated with the high fidelity model and the infill design vector, which is the optimal result from the above mentioned optimization process.

For multi-objective optimization, the infill criterion is different. In this work, the new infill criterion, named pseudo distance is proposed. It is based on the non-dominate concept. The pseudo distance comprises two terms: the dominate distance D_d and the neighbour distance

D_n given in (3.22)–(3.24). It intends to give rise to the most promising design vector, which has the following two properties:

(i) It dominates the maximum number of existing non-dominated solutions with the furthest distance. This improves the optimality of non-dominated front.

(ii) It fills the largest gap between two existing non-dominated solutions. This gives the uniformly spread non-dominated front.

These two properties are common for multi-objective optimization algorithms using non-dominated concepts. However, a unique characteristic of surrogate-assisted algorithms is that they take into account the uncertainty of surrogate models. Therefore, the pseudo distance infill criteria search for:

(i) an accurate solution that offers the longest euclidean distance between itself and the existing non-dominated solutions that are dominated by it.

(ii) a less accurate solution that is located on the non-dominated front and that fills the largest gap between its two nearby existing non-dominated solutions.

The pseudo distance infill criteria are expressed as:

$$D_{pseudo}(x) = D_n(x) + D_d(x) \tag{3.22}$$

$$D_n(x) = \sum_{i=1}^{m} \left(\left(\frac{f_i^{(s+)} - \hat{f}_i(x)}{f_{i_{max}} - f_{i_{min}}} \right) \hat{s}_i(x) \right) \tag{3.23}$$

$$D_d(x) = \sum_{j=1}^{n_{dom}} \sum_{i=1}^{m} \left(\left(\frac{f_i^{(s_j)} - \hat{f}_i(x)}{f_{i_{max}} - f_{i_{min}}} \right) \frac{1}{\hat{s}_i(x)} \right) \tag{3.24}$$

where m corresponds to the number of objective functions, n_{dom} to the number of existing non-dominated points dominated by the trail point, $f_{i_{min}}$ and $f_{i_{max}}$ minimum and maximum of the i^{th} objective function in the existing non-dominated front (NF), $\hat{s}_i$ to the predicted standard error (3.16) of the i^{th} objective function Kriging model, $\hat{f}_i$ to the Kriging prediction of the i^{th} objective function and $f_i^{(s+)}$ to the i^{th} objective function of the point next to the trial point in the i^{th} objective space.

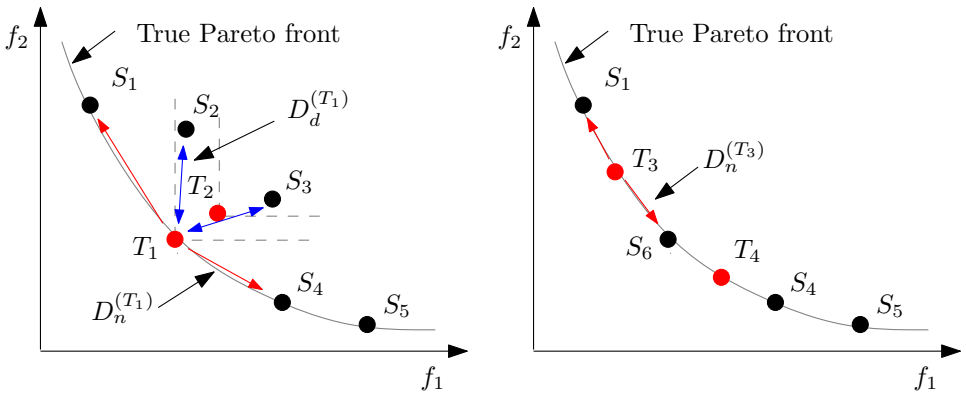
D_d seeks for a design vector with the lowest estimated standard error (low $\hat{s}$) i.e. an accurate Kriging prediction, and dominates the existing NF with the greatest distance. It tends to push the infill point to the global utopia point if there is an accurate solution that dominates all existing non-dominated solutions.

Contrarily, D_n focuses only on neighbours of the trial point. It gives a design vector with the highest standard error (high $\hat{s}$) i.e. a non-accurate Kriging prediction, and the highest distance between the trial point and its neighbour non-dominated points. D_n will try to push the infill point to the local utopia point if the solution is not an accurate one. And compared to another pairs of neighbour points, D_n will simultaneously fill out the largest gap between two existing non-dominated points. The high standard error allows improving the Kriging models and doing global search in the design variable domain.

To explain the idea of the pseudo distance, an illustration example is given in Figure 3.19. Design solutions are shown in an objective space. A set of existing non-dominated front $NF_{init} = [S_1\ S_2\ S_3\ S_4\ S_5]$ is obtained from the initial sample point set. At iteration 1, many trial design vectors are tested through an infill criteria optimization process depending on the optimization algorithm used. For simplification reasons, only two trial points T_1 and T_2 are shown in Figure 3.19a. T_1 dominates 2 existing non-dominated points S_2 and S_3. A trial non-dominated front is determined $TNF^{(T_1)} = [S_1\ T_1\ S_4\ S_5]$. Similarly, T_2 dominates only S_3 and $TNF^{(T_2)} = [S_1\ S_2\ T_2\ S_4\ S_5]$ can be identified. D_n and D_d for each trial point are then computed according to (3.23) and (3.24), respectively. Note that the modelling accuracy is considered as equal throughout the entire Kriging models in order to simplify the explanation. In this example, T_1 gives higher D_{pseudo} than that of T_2 and is chosen as the infill point. The fine model is evaluated using the design vector of T_1. The Kriging model is then rebuild. This infill point (T_1) is validated against the acceptance conditions (see Section 3.1.3.6.2 below). It is accepted as the improved solution S_6. Therefore, a new non-dominated front can be located $NF_1 = [S_1\ S_6\ S_4\ S_5]$.

At iteration 2, two trial points T_3 and T_4 are tested. These trail points do not dominate any existing solution since all existing solutions are on the true Pareto front. This gives $D_d = 0$. In this case, only D_n influences the choice of the infill point. The infill criteria optimization process will try to improve the distribution property of non-dominated front by filling the largest gap between any two points on the non-dominated front. Therefore T_3 is selected as the infill point.

In reality, when the Kriging model accuracy is not equal and several constraints are taken into account, the infill optimization may select the infill point in a different manner. It reaches a compromise between the distance and the accuracy of models.



(a) Iteration 1 (b) Iteration 2

Figure 3.19: Illustrating example of the Pseudo distance concept

3.1.3.6.2 Acceptance condition for multi-objective optimization

Once the infill design vector is located by the infill optimization subproblem, the high fidelity model is evaluated. The acceptance condition is the verification that the new infill point improves the existing NF and satisfies the constraints. Based on the non-dominated

concept, a solution improves the existing non-dominated front if one of the two following criteria is met:

(i) It dominates at least one existing non-dominated solution.

(ii) It is not dominated by any existing non-dominated solution.

If the candidate infill point respects the acceptance condition, it will be added to the existing non-dominated front. In any case, it is added to the set of sample points. Kriging models are then refitted with the new set of sample points.

3.1.3.7 Mathematical test example

3.1.3.7.1 Single-objective noisy function optimization

The first example intends to demonstrate the capability of a surrogate-assisted algorithm (EGO in this case) to cope with noisy functions. This is a very important issue in the design of electrical machines using Finite Element Analysis (FEA) [56]. The application to electric machine designs will be presented in Chapter 4.

The originally proposed problem expressed in (3.3) is multimodal and with a constraint. It deals with smooth objective and constraint functions. In order to imitate the numerical noise in FEA, the problem has been modified. A noise term is added into the constraint function. The optimization problem is expressed as following:

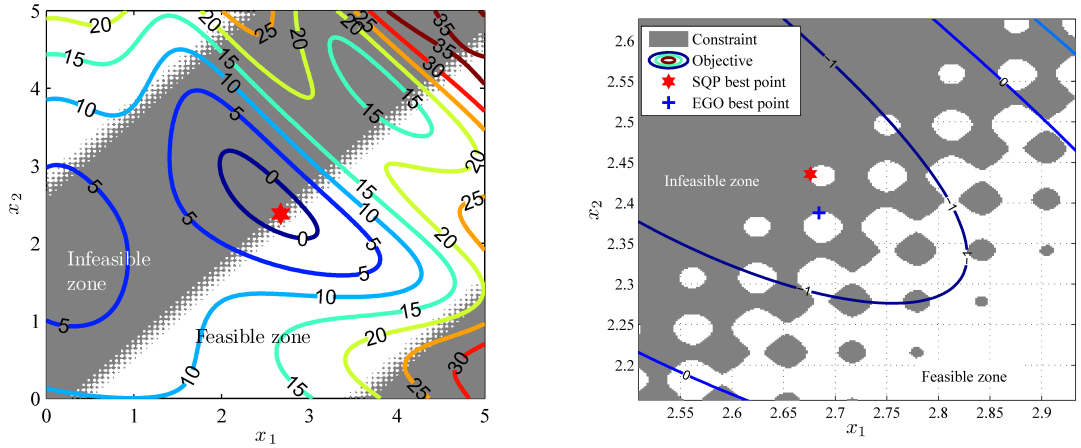
$$\begin{aligned} \max_{x} \quad & f = 2 + 0.01\left(x_2 - x_1^2\right)^2 + (1 - x_1)^2 + \\ & \qquad 2\,(2 - x_2)^2 + 7\sin(0.5x_1)\sin(0.7x_1x_2) \\ \text{subject to} \quad & g_{exp} = -\sin(x_1 - x_2 - \pi/8) + \underbrace{0.1\left\{\sin(100x_1) + \sin(100x_2)\right\}}_{\text{Noise term}} \leq 0 \\ & x_1 \in [0,\ 5],\ x_2 \in [0,\ 5] \end{aligned} \tag{3.25}$$

Figure 3.20 shows the contour plot and feasible area of analytical test problem. Multiple local optima can be observed especially on the border of constraint limits. The global optimum is presented by the red star at the design vector $x = [2.6759\ 2.4356]$, which gives $f = -1.323$. This reference global optimal solution is found by using an SQP multistart with 1000 random initial points.

In the EGO algorithm, both the objective and constraint functions are considered "expensive". Therefore, they were modelled by two independent Kriging models. The constraint is handled by the probability method. Ten tests were carried out with different sets of initial sample points selected by using the Latin Hypercube Sampling (LHS) technique. EGO stops when the GEI value is lower than 0.01. This is done with an average of 80 function calls. Figure 3.21a shows initial sample points and infill points in the design space. Figure 3.21b shows the EGO infill history. One can observe that EGO evaluates the points with a high objective function or infeasible points in order to improve the Kriging model accuracy and the global search.

Despite the noisy constraint function, the worst EGO results (from 10 tests) are found with a relative error of 13% on the optimal objective function value[11] and 10% on the Euclidian

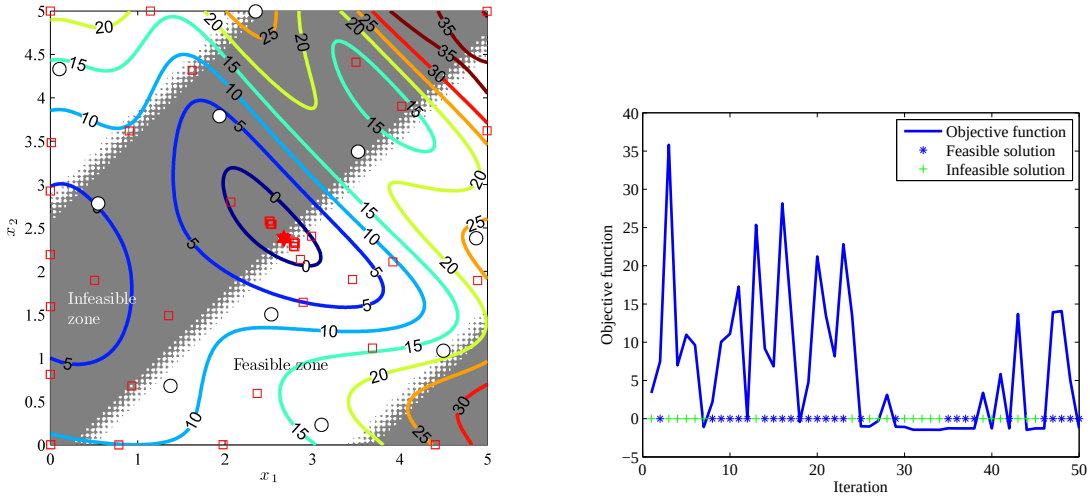
[11] compared with the reference global optimum objective value



(a) entire design space (b) zoom

Figure 3.20: Contour plot of an analytical test function and constraint for the entire design space (a) and with a zoom (b)

distance[12]. The best EGO results are within 2% for both criteria. However it cannot find the exact global optimum.



(a) initial and infill points in design space (b) objective function value at each iteration

Figure 3.21: Infill point history. (a) Initial and infill points in design space. Circles are initial points and squares are infill points (b) Objective function value at each iteration

Compared with a traditional algorithm Sequential Quadratic Programming (SQP) method with multistart (100 starts with random initial points), each optimization required an average of 145 function evaluationss. As a result, 3 initial points can precisely find the global optimum. But only 25 out of 100 initial points lead to the same precision as or better precision than EGO worst case (i.e. 13% on the objective function and 10% on Euclidian distance). This proves that SQP is not very suitable for this kind of problem.

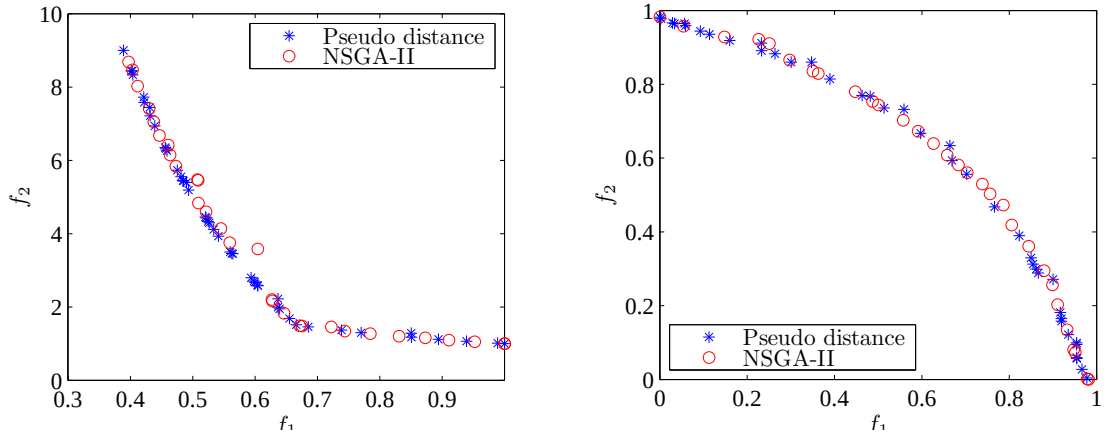
[12] compared with the reference global optimum design vector

This example shows the capability of EGO to find the solutions around the optimal point in the global optimization problem with noise in the constraint function. EGO does not have the capability to reduce noise in noisy functions. But it overlooks this noise by using the Kriging model. Thus, it improves the convergence towards a global solution. However, since the Kriging model is an interpolation model, there is always noise in the global optimum.

3.1.3.7.2 Multi-objective problem

Two test problems "CONSTR" and "VLMOP2" presented in Section 3.1.2.1.3 are used to demonstrate the capability of surrogate-assisted algorithm using the pseudo-distance infill criterion. An application of the algorithm to an electrical machine design is presented in Section 4.5.

For each problem, LHS is used to select 20 initial sample points. 80 other infill points are added according to the pseudo-distance infill criterion. In total This yields only 100 fine model evaluations. For the "CONSTR" problem, the constraint is managed through a direct integration method with a surrogate equivalent constraint. Non-dominated fronts for both problems are shown in Figure 3.22a and 3.22b. The algorithm founds 40 and 46 non-dominated solutions for "CONSTR" and 'VLMOP2" problem, respectively.



(a) CONSTR problem

(b) VLMOP2 problem

Figure 3.22: Non-dominated fronts obtained from NSGA-II and surrogate-assisted algorithm using pseudo distance infill criterion

For comparison purposes, an optimization using NSGA-II is also performed. The NSGA-II parameters are defined as following: crossover probability = 0.9, population number = 30 and generation number = 100. This yields 3000 fine model evaluations in total. One can observe that two algorithms achieve more or less similar results. For both problems, non-dominated fronts from NSGA-II are slightly better distributed than those of the surrogate-assisted algorithm. However, NSGA-II requires a very high number of function evaluations in order to obtain a good result. For surrogate-assisted algorithm, similar results can be achieved with considerable fewer function calls (30 times lower than that of NSGA-II for these examples).

3.2 Decomposition approach

Complex systems are referred to as systems composed of several subsystems and components. Such systems also imply several disciplines. They can be physical/engineering disciplines or non-technical disciplines. In conventional design processes, such complex system design problems are always decomposed into several subproblems, as already discussed in Chapter 1. In this section, the Multidisciplinary Design Optimization (MDO) and Target Cascading (TC) methodologies are presented.

3.2.1 Multidisciplinary Design Optimization

3.2.1.1 Overview

Complex systems always deal with many engineering disciplines. The interaction between such disciplines is mainly concerned in MDO. As mentioned earlier, a complex system design problem is usually decomposed into subproblems. In the MDO case, the decomposition is based on the discipline. This comes from the traditional design methodology in aeronautic field as MDO was firstly introduced by the aeronautic and aerospace community. MDO formulations allow coordinating various design/simulation tools and solving interacted problem among such tools. Moreover, MDO integrates optimization tools in modelling phase. Both processes are done simultaneously and not sequentially anymore as it used to be in the past.

The MDO implies many formulations, which can be grouped into 2 families: single level and multilevel formulations. In this research, only single level MDO formulations have been investigated: Multidisciplinary Feasible (MDF), Individual Discipline Feasible (IDF), All-At-Once (AAO) [1, 18, 103]. Multilevel MDO formulations can be found in [1, 10, 101].

The aim of MDO formulation is to manage and coordinate calculation, interaction i.e. feedback and feed-forward dataflow from one analysis tool to another. Some single level formulations manage the interactions, so that analysis tools can be run in parallel.

In the multidisciplinary design context, many analysis tools are employed to evaluate the performance of the system. These tools usually correspond to the different disciplines. Since there are interactions between disciplines i.e. an input of a discipline is evaluated using another discipline tool as:

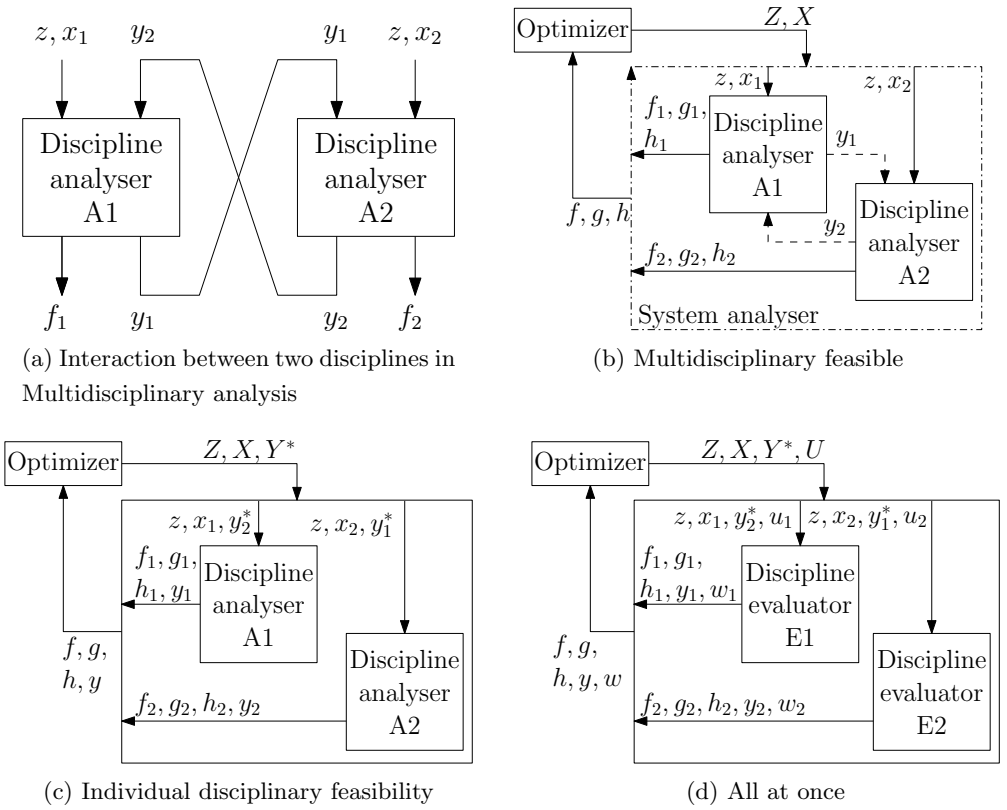
$$\begin{aligned} &output_1\left(z, x_1, y_2\left(z, x_2, y_1\right)\right) \\ &output_2\left(z, x_2, y_1\left(z, x_1, y_2\right)\right) \end{aligned} \tag{3.26}$$

where $output_1$ and $output_2$ are calculated from disciplines 1 and 2, respectively and can be the objective f, inequality or equality constraint (g or h) and y coupling variable. $output_1$ is as a function of their local design variables x_1, global or share design variable z and coupling variable y_2, which, itself, is an output from discipline 2 and as a function of z, x_2, and y_1. Figure 3.23a presents this explanation. The resolution of these interactions can be performed by a numerical method such as the Newton-Raphson or fixed point method [17, 102]. This kind of problem is called Multidisciplinary Design Analysis (MDA). Two MDA disciplines can be written in a compact form as:

$$MDA = SA\left[A_1 \rightarrow A_2\right] \tag{3.27}$$

Two discipline analyses A_1 and A_1 are nested within a System Analysis (SA) and are evaluated sequentially (denoted as $\rightarrow$).

The optimization of such problems can be achieved by using several approaches. The following subsections explain different single level MDO formulations. They give the possibility to perform the optimal design of a multidisciplinary system in efficient and systematic ways by using an optimization process in the connection of the analysers or the evaluators between disciplines.



(a) Interaction between two disciplines in Multidisciplinary analysis

(b) Multidisciplinary feasible

(c) Individual disciplinary feasibility

(d) All at once

Figure 3.23: Structure and dataflow of a multidisciplinary design problem

3.2.1.2 Multidisciplinary feasibility

The simplest way to perform a MDO is to directly put the MDA within an optimization process. This is called a Multidisciplinary Feasible (MDF). This formulation ensures the feasibility of the whole system at each optimization step. Figure 3.23b shows the structure of the MDF formulation. The system analyser solves a problem with interaction. The optimizer solves the optimization problem in which the design variables only remain design variables of the design problem. Multidisciplinary feasibility is achieved for each design vector. A MDF formulation can be shown as in (3.28). The MDA (3.27) is nested within a System Optimiser (SO).

$$MDF = SO\left[MDA\right] = SO\left[SA\left[A_1 \rightarrow A_2\right]\right] \tag{3.28}$$

The MDF optimization problem is expressed as:

$$\begin{array}{ll} \min & f\left(Z,X\right) \\ \text{with respect to} & Z,X \\ \text{subject to} & g\left(Z,X,Y\left(Z,X\right)\right)\leq 0 \\ & h\left(Z,X,Y\left(Z,X\right)\right)=0 \end{array} \tag{3.29}$$

where SA forces $Y^{k+1}=Y^{k}$ at each optimization iteration, k being the SA internal loop iteration.

A MDF is the simplest MDO formulation. However, some drawbacks can be stated. The fixed point iteration method commonly used in MDF is not always very efficient. If the fixed-point method fails to supply a solution to the MDF, the optimization may not be convergent. Moreover, the fixed-point method is a sequential method. The parallelised analysis is not possible.

3.2.1.3 Individual disciplinary feasibility

IDF allows decoupling the discipline analysers. There is not any direct interaction between disciplines anymore. The optimization gives an initial guess of the coupling variables via additional design variables Y^{*}. IDF ensures only the feasibility of each discipline (i.e. each discipline performs discipline analysis). In order to obtain the multidisciplinary feasibility solution at the convergence point, multidisciplinary feasibility constraints are added to ensure consistency between the estimated coupling variable Y^{*} given by the optimizer and the value Y evaluated by the discipline analyser. This yields a more complicated optimization problem than a MDF. The IDF optimization problem can be described as:

$$\begin{array}{ll} \min & f\left(Z,X,Y^{*}\right) \\ \text{with respect to} & Z,X,Y^{*} \\ \text{subject to} & g\left(Z,X,Y\left(Z,X,Y^{*}\right)\right)\leq 0 \\ & h\left(Z,X,Y\left(Z,X,Y^{*}\right)\right)=0 \\ & y_{i}^{*}-y_{i}\left(z_{i},x_{i},y_{i}^{*}\right)=0 \end{array} \tag{3.30}$$

where i and j are different disciplines.

The architecture of IDF can be written as in (3.31). The discipline analyser A_1 and A_2 can be launched in parallel (depicted as $\|$) and are nested within the SO.

$$IDF=SO\left[A_1\|A_2\right] \tag{3.31}$$

In Figure 3.23c, it can be observed that two different discipline analyses do not carry out any direct information exchange. The discipline analyses can be evaluated in parallel.

A derivative of the IDF formulation called Sequenced IDF [18] corresponds to the case where only some of the coupling variables are treated by way of optimization and others are the actual computed output. Some disciplines have to be sequenced and this allows forming groups of sequenced discipline, which can be parallelised. Sequenced IDF can be used if there

are too many coupling variables or if the optimization algorithm encounters convergence difficulties. It is also useful in order to balance the load of parallelised computation. For example, two low time-consuming disciplines can be sequenced and parallelised with one high time-consuming discipline analysis.

3.2.1.4 All at once

AAO is also named Simultaneous Analysis and Design (SAND). The optimizer solves simultaneously the design problem, the coupling variables and the governing equations. The state variables are added to the design variable set and the residuals are treated as constraints. Both multidisciplinary and discipline feasibility are obtained only at the optimization convergence point. The state variable may be referred to as a numerical solving in finite element analysis (FEA) or any nonlinear characteristic or implicit variables such as a nonlinear B-H curve in electromagnetic field computation and a thermal characteristic of material depending on the temperature in thermal discipline. In AAO, the code acts as an evaluator, not as an analyser. It computes a set of equations in a one-pass manner rather than solving those equations iteratively i.e. there is no internal iteration loop.

Figure 3.23d shows an AAO structure. The codes of the discipline are shown as the "Discipline Evaluator". The AAO formulation can be expressed in a mathematical equation as in (3.32). Two discipline evaluators are used in parallel and are nested in a system optimiser. As with IDF, disciplines are completely decoupled and can be parallelised.

$$AAO = SO\left[E_1 \| E_2\right] \tag{3.32}$$

The AAO optimization problem can be described as:

$$\begin{array}{ll} \min & f\left(Z, X, Y^*, U\right) \\ \text{with respect to} & Z, X, Y^*, U \\ \text{subject to} & g\left(Z, X, Y\left(Z, X, Y^*\right), U\right) \leq 0 \\ & h\left(Z, X, Y\left(Z, X, Y^*\right), U\right) = 0 \\ & y_i^* - y_i\left(z_i, x_i, y_i^*, u_i\right) = 0 \\ & W\left(Z, X, Y^*, U\right) = 0 \end{array} \tag{3.33}$$

where U is a state variable and W a residual of the state variable of the disciplines.

AAO solves analysis and design problems simultaneously. This reduces the computation time because the disciplinary and multidisciplinary feasibilities are not sought when the actual solution is far from an optimum. However, the optimization algorithm may encounter difficulties due to a very high number of auxiliary design variables and equality constraints getting involved. Furthermore, in order to perform this formulation, one must have access to state variables. Therefore, commercial software cannot be used since the code modification is not possible.

3.2.1.5 Mathematical example

In this section, the MDF and IDF formulations are illustrated via a mathematical example. The example consists of two disciplines A_1 and A_2 with one share design variable z and 2 coupling variables y_1 and y_2 characterising the interaction between both disciplines as shown in Figure 3.24. The discipline A_1 and A_2 are expressed by y_1 and y_2, respectively. These functions are constructed from a certain number of sample points by using the regression least-square method.

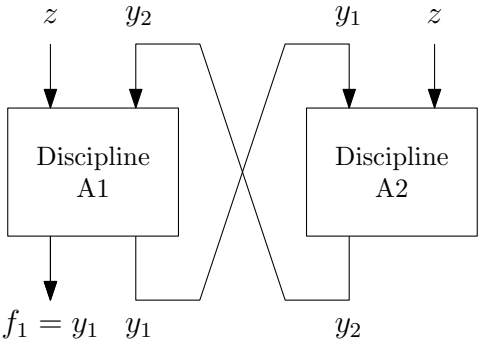


Figure 3.24: Interaction between two disciplines

The problem is described as:

$$\begin{array}{lrcl}
\text{min} & f(z) & = & y_1 \\
\text{with respect to} & z & & \\
\text{where} & y_1(z, y_2) & = & \left(4.2667y_2^4 - 1.0667y_2^3 - 5.8667y_2^2 + 2.6667y_2 + 0.7\right) \times \\
 & & & \left(4.2667z^4 - 1.0667z^3 - 5.8667z^2 + 2.6667z + 0.7\right) \\
 & y_2(z, y_1) & = & y_1 - 0.76923z + 0.76923
\end{array} \tag{3.34}$$

The optimization problem above is formulated using the MDF formulation. The system analysis uses the fixed point iteration method for solving interaction problems. This problem can be formulated by using IDF as:

$$\begin{array}{lrcl}
\text{min} & f(z, y_1^*, y_2^*) & = & y_1 \\
\text{with respect to} & z, y_1^*, y_2^* & & \\
\text{where} & y_1(z, y_2^*) & = & \left(4.2667{y_2^*}^4 - 1.0667{y_2^*}^3 - 5.8667{y_2^*}^2 + 2.6667y_2^* + 0.7\right) \times \\
 & & & \left(4.2667z^4 - 1.0667z^3 - 5.8667z^2 + 2.6667z + 0.7\right) \\
 & y_2(z, y_1^*) & = & y_1^* - 0.76923z + 0.76923 \\
\text{subject to} & y_1 & = & y_1^* \\
 & y_2 & = & y_2^*
\end{array} \tag{3.35}$$

The problem is solved by using MDF and IDF formulations. For the MDF formulation, the interaction problem is solved through the fixed point method with a tolerance of 10^{-4}. The convergence problem can be observed. To cope with this problem, a maximum iteration

limit is fixed at 300 iterations. The initial design vector for the SQP algorithm is set to $z = 0.5$, $y_1 = 0.2$ and $y_2 = 0.2$. Note that for the MDF, only z is required.

Both formulations give the same results: $z = 0.7756$, $f = y_1 = 0.2419$ and $y_2 = 0.4145$. The summary of the function evaluation is shown in Table 3.7. As the MDF has only 1 design variable, SQP converges to the optimal result faster than in the IDF formulation with 3 design variables and 2 constraints. However, the MDF needs more discipline function evaluations than the IDF because of its fixed point iteration solver. For each function objective call, the MDF requires an average of 76.7 discipline function evaluations to solve its interaction problem between y_1 and y_2. This results in the higher total computation time i.e. 997 discipline evaluations are needed in total. Whereas, IDF does not need any internal iteration. Only one discipline function evaluation per objective function call is required i.e. 67 discipline evaluations.

Table 3.7: Function evaluation for MDF and IDF examples

	MDF	IDF
Design variable	1	3
Constraint	0	2
Average discipline evaluation per objective function call	76.7	1
Number of objective function call	13	67
Total evaluation for each discipline	997	67

3.2.1.6 Conclusion

In complex system design problem, multiple design tools/models are involved. These tools represent, in general, disciplines. Traditionally, interactions between disciplines or models are solved through iteration methods such as the fixed point method. This ensures multidisciplinary feasibility at each model evaluation. When solving such design problems using an optimization technique, this group of models is wrapped with optimization layers and forms the simplest MDO formulation, the multidisciplinary feasibility (MDF). The IDF formulation treats the design problem and the multidisciplinary interaction problem simultaneously. This speeds up the optimization process since multidisciplinary feasibility is ensured only at the final optimal results level. In extreme cases, AAO optimization treats the internal iteration loop (discipline analysis) in addition to the interdisciplinary loop. IDF and AAO require additional design variables and constraints in order to take into account these iteration processes. Depending on the level of interaction and the performance of the optimization algorithm used, one can be faced with a very high dimension optimization problem and this can result in a higher computation time than that of the MDF formulation (conventional approach). However, IDF and AAO problems can be parallelised. This is done at the model level and does not depend on the optimization algorithm.

It can be concluded at this stage that MDO gives a new point of view. When the model is meant to be used in an optimization process, the modelling process can be different from a model used only for simulation purpose. A carefully developed model taking into account its final purpose - in our case optimal design purpose - can result in a more robust, more efficient optimal design problem. The application of the MDO methodology to the traction motor design problem can be found in Chapter 4.

3.2.2 Target cascading optimal design

3.2.2.1 Overview

Designing of a complex system is usually a delicate task and involves many engineering domains. Furthermore, a complex system is always composed of several subsystems and components, which are sometimes designed concurrently. In traditional design methodology, complex system design problems are decomposed into several subproblems (see Figure 3.25) in order to understand the problem better and, after all, to be able to solve it. Subproblems are assigned to design teams. They provide optimal results to the system engineer, who coordinates and ensures the consistency between these subsystems. This is very common way of working, widely used by industrials.

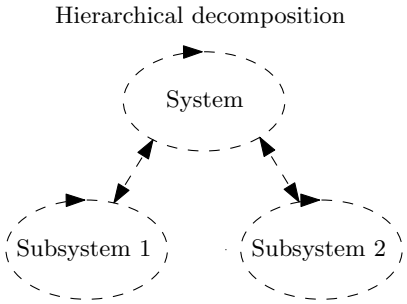


Figure 3.25: Hierarchical design problem

In order to apply an optimization technique to solve such already decomposed complex system design problems, two approaches can be considered

(i) To discard any already decomposed system and reformulate the problem as a global system optimization problem, which can be solved using the optimization technique presented in Section 3.1. If there are interactions between subsystems (models), the MDO methodology can be applied.

(ii) To continue working with hierarchical decomposed problems and find a suitable method to solve this kind of problem.

The former approach requires the reformulation of the design problem. This can require significant time and resources as the design problem formulation can affect the company structure in the case of a big company in which engineering units are well structured for the conventional design methodology. Therefore, the latter approach is considered as a better solution because the only change is the way the problem is *solved* and not the way it is *formulated*. Moreover, specialized engineering teams can only focus on the design subproblem in the domain they are expert in. Such a hierarchical optimization problem can be solved through multilevel optimization methods. In the literature, one can find multilevel MDO formulations such as Collaborative Optimization (CO) [10], Bi-Level Integrated System Synthesis (BLISS) [101] and Target Cascading method (TC) [47]. TC is the only presented in this research. TC is a multilevel coordination formulation. It has been used in many applications such as aircraft design [3] and automotive vehicle design [48]. It allows solving multilevel hierarchical

optimization problems. It is assumed that the problem is already decomposed. Figure 3.26 shows a decomposed optimization problem consisting of 3 levels: the system level, the subsystem level and the component level. Each optimization subproblem has its own local design variables and local constraints (not shown in the figure). A child subproblem is connected to its parent subproblem via targets, responses and linking variables[13]. One can also observe that there is no direct link between two subproblems at the same level. The interactions are managed by their parent subproblems via linking variables. Subproblems can be launched in parallel, by being decoupled.

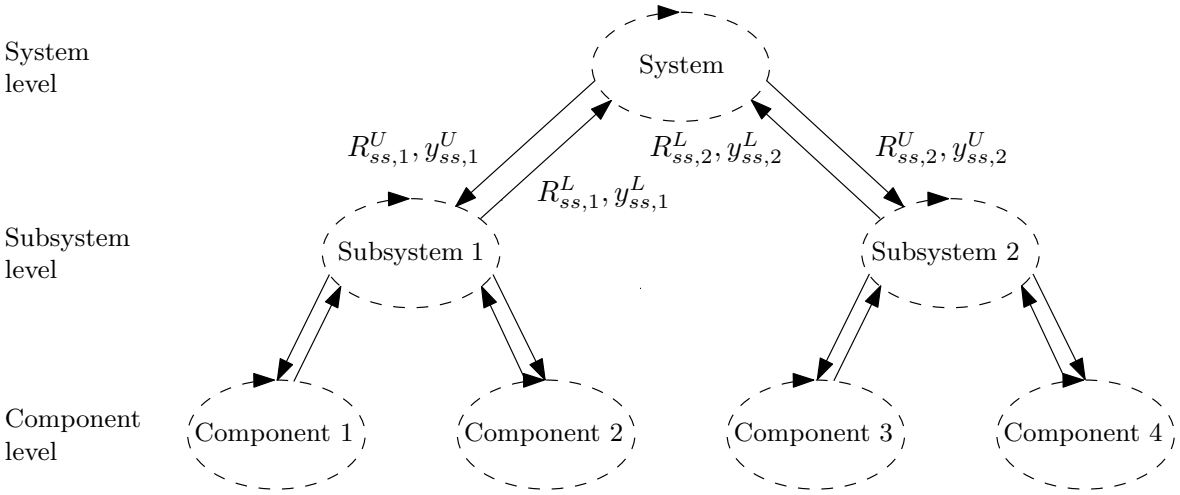


Figure 3.26: Target cascading hierarchical optimization problem

3.2.2.2 TC formulation

TC solves the target setting problem at the system subproblem and propagates the subsystem targets to the subsystem subproblems and so on. A target setting problem at the system level is to minimise discrepancy between set targets and computed responses [47]. This optimization subproblem at the system level consists in finding local design variables and also linking variables and sublevel targets. Once the system level optimization has finished, the sublevel targets and linking variables found at the system level are then cascaded to the sublevel problems. At the sublevel, the sublevel problems minimise errors between sublevel targets and sublevel responses and between upper level and sublevel linking variables. The process continues until reaching the bottom level. The information from bottom level are then returned back to the upper level. The optimizations at the upper level are rerun. TC completes one iteration whilst the optimization returns back to the system level.

[13]The linking variable characterises the share variable and the interaction between subproblems.

System level (top level): The optimization problem at the system level is to minimise errors between targets T_{sys} and system responses R_{sys}. It is defined as:

$$\begin{aligned}
P_{sys}: \quad & \min && \|R_{sys} - T_{sys}\| + \epsilon_R + \epsilon_y \\
& \text{with respect to} && x_{sys}, y_{ss}, R_{ss}, \epsilon_R, \epsilon_y \\
& \text{where} && R_{sys} = f\left(R_{ss}, x_{sys}\right) \\
& \text{subject to} && \sum_{k \in C_{sys}} \left\|R_{ss,k} - R^L_{ss,k}\right\| \leq \epsilon_R \\
& && \sum_{k \in C_{sys}} \left\|y_{ss,k} - y^L_{ss,k}\right\| \leq \epsilon_y \\
& && g_{sys}\left(R_{ss}, x_{sys}\right) \leq 0 \\
& && h_{sys}\left(R_{ss}, x_{sys}\right) = 0
\end{aligned} \tag{3.36}$$

It can be observed that R_{sys} is a function of the system local design variables x_{sys} and subsystem responses R_{ss}. Additions to the x_{sys}, optimization problem must search for the y_{ss} linking variables of the R_{ss} subsystem and the ϵ_R and ϵ_y coordination tolerances to be minimised. These tolerances are used as constraint values to ensure consistency between system and subsystem levels for each child subproblem of the system $C_{sys} = \{k_1, k_2, \ldots, kc_{sys}\}$. Note that $R^L_{ss,k}$ and $y^L_{ss,k}$ are the optimal results sent back from the subsystem level. The L superscript indicates that the value comes from a "**L**ower" level.

Subsystem level (intermediate level): The j^{th} subsystem level problem is to minimise the discrepancy between subsystem targets given by the system (**U**pper) level ($R^U_{ss,j}$) and $R_{ss,j}$ subsystem responses and between the $y^U_{ss,j}$ subsystem linking variables from system level and those from the $y_{ss,j}$ subsystem level. Note that $R^U_{ss,j}$ and $y^U_{ss,j}$ are the optimum values found by the system optimization subproblem.

$$\begin{aligned}
P_{ss,j}: \quad & \min && \left\|R_{ss,j} - R^U_{ss,j}\right\| + \left\|y_{ss,j} - y^U_{ss,j}\right\| + \epsilon_R + \epsilon_y \\
& \text{with respect to} && x_{ss,j}, y_{ss,j}, y_{com}, R_{com}, \epsilon_R, \epsilon_y \\
& \text{where} && R_{ss,j} = f\left(R_{com}, x_{ss,j}, y_{ss,j}\right) \\
& \text{subject to} && \sum_{k \in C_{ss,j}} \left\|R_{com,k} - R^L_{com,k}\right\| \leq \epsilon_R \\
& && \sum_{k \in C_{ss,j}} \left\|y_{com,k} - y^L_{com,k}\right\| \leq \epsilon_y \\
& && g_{ss,j}\left(R_{com}, x_{ss,j}, y_{ss,j}\right) \leq 0 \\
& && h_{ss,j}\left(R_{com}, x_{ss,j}, y_{ss,j}\right) = 0
\end{aligned} \tag{3.37}$$

Like the system level optimization, the subsystem optimization problem searches for its own $x_{ss,j}$ local design variables, $y_{ss,j}$ linking variables and linking variables and y_{com} and R_{com} respectively responses for its child component subproblems.

Component level (bottom level): The j^{th} component subproblem minimises the deviation for component responses ($R_{com,j}$ and $R^U_{com,j}$) and component linking variables (y_{com} and y^U_{com}).

$$\begin{array}{lll} P_{com,j}: & \min & \left\| R_{com,j} - R^U_{com,j} \right\| + \left\| y_{com,j} - y^U_{com,j} \right\| \\ & \text{with respect to} & x_{com,j}, y_{com,j} \\ & \text{where} & R_{com,j} = f\left(x_{com,j}, y_{com,j}\right) \\ & \text{subject to} & g_{com,j}\left(x_{com,j}, y_{com,j}\right) \leq 0 \\ & & h_{com,j}\left(x_{com,j}, y_{com,j}\right) = 0 \end{array} \tag{3.38}$$

3.2.2.3 Mathematical example

A mathematical example is used to demonstrated TC. The problem is composed of 3 subproblems at 2 levels. According to the TC formulation, they are expressed as:

$$\begin{array}{lll} P_{sys}: & \min & \left(R_{sys} - T_{sys}\right)^2 + \epsilon_R \\ & \text{with respect to} & x_1, x_2, R_{ss,1}, R_{ss,2}, \epsilon_R \\ & \text{where} & R_{sys} = f\left(R_{ss,1}, R_{ss,2}, x_{sys}\right) \\ & & = \left(x_1 - 1\right)^2 + \left(x_2 - 1\right)^2 + R_{ss,1} + R_{ss,2} \\ & \text{subject to} & \left(R_{ss,1} - R^L_{ss,1}\right)^2 + \left(R_{ss,2} - R^L_{ss,2}\right)^2 \leq \epsilon_R \\ & & 0 \leq x_1 \leq 2 \\ & & 0 \leq x_2 \leq 2 \\ & & 0 \leq R_{ss,1} \leq 2 \\ & & 0 \leq R_{ss,2} \leq 2 \\ & & 0 \leq \epsilon_R \leq 2 \end{array} \tag{3.39}$$

The system target T_{sys} is set to 0 in order to minimise R_{sys}. The system optimization is firstly launched. The optimal value of sublevel targets: $R^U_{ss,1}$ and $R^U_{ss,2}$ are cascaded to the sublevel problems:

$$\begin{array}{lll} P_{ss,1}: & \min & \left(R_{ss,1} - R^U_{ss,1}\right)^2 \\ & \text{with respect to} & x_3, x_4 \\ & \text{where} & R_{ss,1} = f\left(x_3, x_4\right) = \left(x_3 - 1\right)^2 + \left(x_4 - 1\right)^2 \\ & \text{subject to} & 2 - x_3 \cdot x_4 \leq 0 \\ & & 0 \leq x_3 \leq 2 \\ & & 0 \leq x_4 \leq 2 \end{array} \tag{3.40}$$

$$
\begin{array}{lll}
P_{ss,2}: & \min & \left(R_{ss,2} - R^U_{ss,2}\right)^2 \\
& \text{with respect to} & x_5, x_6 \\
& \text{where} & R_{ss,2} = f(x_5, x_6) = (x_5 - 1)^2 + (x_6 - 1)^2 \\
& \text{subject to} & 0 \leq x_5 \leq 2 \\
& & 0 \leq x_6 \leq 2
\end{array} \tag{3.41}
$$

Note that there are 2 levels for this problem. The subsystem level is considered as the bottom level. The *ss* subscription in (3.41) is used instead of *com* as in (3.38).

Table 3.8 shows the optimal results for each iteration. TC converges in 3 iterations. The $P_{ss,1}$ subproblem is subject to a constraint. At the first iteration, the system level gives the sublevel target that is not feasible for the subproblem 1. Therefore, it returns the best response to the system level. At the second iteration, the system problem increases the sublevel target for the subsystem 1 ($R_{ss,1}$) in order to meet the response returned from the sublevel at the first iteration. As shown in Figure 3.27, TC found the compromise between two levels at the third iteration i.e. relative error between target and response is lower that 1%. Unlike in the subsystem 1, there is no conflict between the target and the response in the subsystem 2. The subsystem 2 problem can always provide the response required by the system level problem.

Table 3.8: Optimal value at each iteration

Problem	Variable	Optimal value at iteration:		
		1	2	3
System	x_1	1	0.99994	1
	x_2	1	0.99994	1
	$R_{ss,1}$	0	0.33975	0.3428
	$R_{ss,2}$	0	0	0
	ϵ_R	0	$1.15 \cdot 10^{-5}$	$1.18 \cdot 10^{-7}$
	R_{sys}	0	0.3398	0.3428
Subsystem 1	x_3	1.4142	1.4142	1.4142
	x_4	1.4142	1.4142	1.4142
	$R_{ss,1}$	0.34315	0.34315	0.34315
Subsystem 2	x_5	1	1	1
	x_6	1	1	1
	$R_{ss,2}$	0	0	0

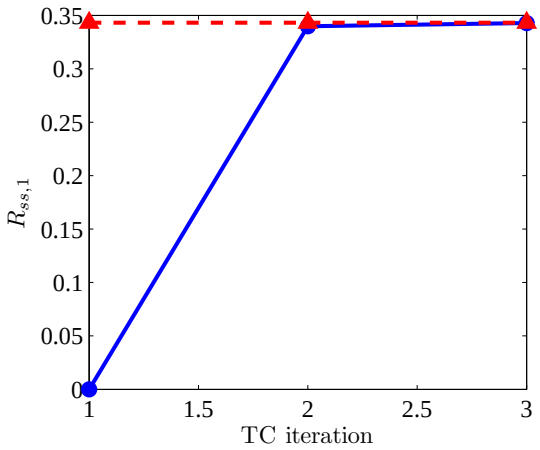


Figure 3.27: Evaluation of $R_{ss,1}$ subsystem 1 target

3.2.2.4 Conclusion

TC allows applying optimization in a large-scale design problem, which is decomposed into several small subproblems. By using the TC formulation, each design subproblem is formulated as an optimization subproblem. TC coordinates the interactions and information exchange of these subproblems. In the global system approach, all design subproblems are treated as a whole optimization problem. The optimization algorithm may suffer from a high number of design variables and constraints. In contrast, TC works with several small optimization problems, which are easier to deal with.

In the traditional design methodology, a system engineer designs the system and imposes targets to subsystem design teams. The subsystem design teams try to match the given targets. Sometimes, the targets cannot be achieved due to the fact that there are several constraints. The results are reported back to the system engineer. System performances are therefore re-evaluated and new targets are redistributed to subsystem design teams. TC works in more or less the same manner as the traditional design methodology. At lower levels, subproblems do the best by selecting their local variables while respecting constraints to achieve the targets imposed by upper levels (e.g. R_{ss}^U and y_{ss}^U). The optimal results are therefore communicated back to the upper level. TC iteration proceeds until convergence i.e. consistency between the system and subsystem levels is obtained.

3.3 Conclusion

In this chapter, three main points have been discussed. The global system approach is commonly used in engineering optimal design. A model representing the whole system is optimised in a single loop. This global system optimization problem searches for all system design variables with respect to all system constraints. Several optimization techniques allow solving this kind of problem. Figure 3.28 summarizes the characteristics of the presented single-objective optimization algorithms. Some main points are emphasized in Table 3.9. SQP is very suitable for the unimodal constrained optimization problem. It can rapidly find a local optimum. For multimodal problems, multistart strategy can be used with SQP. However, a GA may perform better but it requires more function evaluations. A hybrid GA and SQP algorithm combines advantages of GA and SQP e.g. global search ability, low number of function calls and accurate results. A surrogate-assisted algorithm is preferred, while allowing to use a high fidelity model with a very low number of evaluations. It is a global optimization algorithm and also proved to be less sensitive to numerical noise. But it may not be suitable for problems using a fast model due to overhead time for fitting Kriging models and solving infill criteria subproblems.

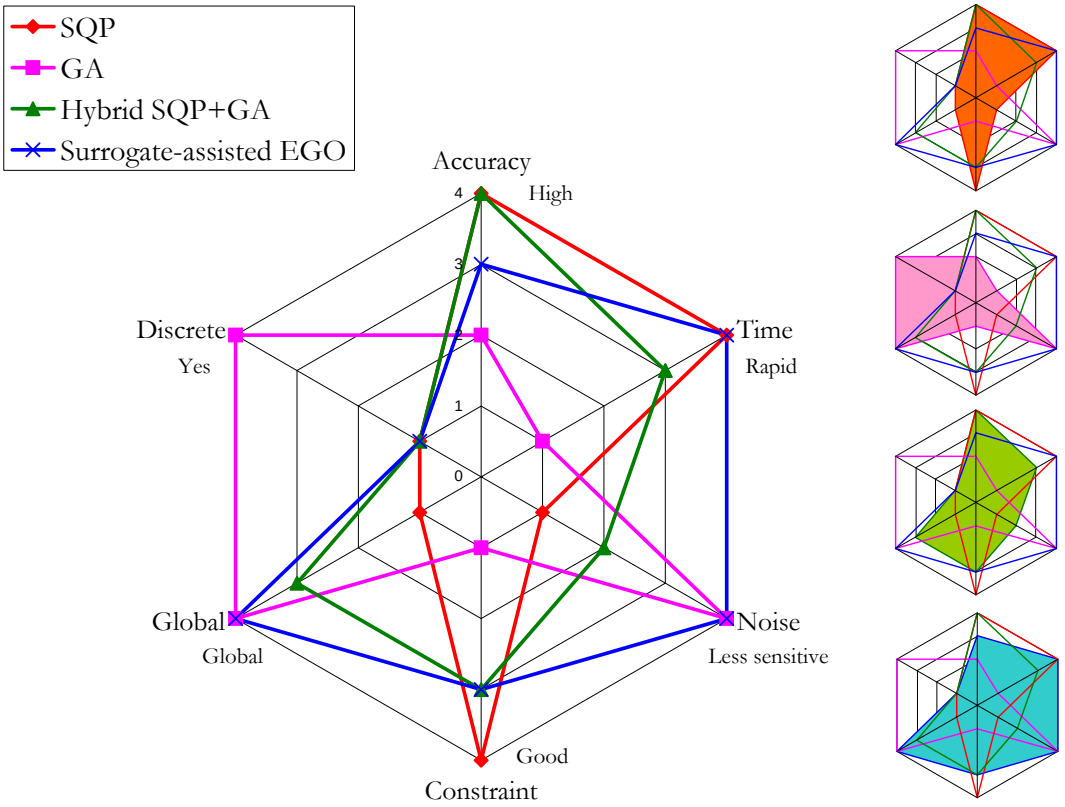


Figure 3.28: Summary of single-objective optimization algorithms

Table 3.9: Advantages and drawbacks of the presented optimization methods

Method	Advantages	Drawbacks
SQP	+ High accuracy + Small number of function evaluations + Performs well with constrained problems	- Sensitive to noise - Local method - Needs gradient information - Continuous variables - Needs initial design point
GA	+ Global method + Robust, no gradient need + Continuous and discrete variables + Parallelisable	- Low accuracy - Large number of functions evaluation
Hybrid GA and SQP	+ Able to find global optimum + High accuracy + Less evaluations than GA alone + GA search process can be parallelised	- Sensitive to noise - Needs gradient information - Continuous variables
Surrogate-assisted EGO	+ Global method + Very small number of function evaluations + Less sensitive to noise + No gradient needed	- High overhead time - Continuous variables

Engineering design problems always deal with several conflicting criteria/objectives. Techniques to generate the Pareto front have been presented. They can be obtained using transformation techniques or via a multi-objective algorithm such as NSGA-II or surrogate-assisted by using pseudo-distance infill criterion. The main drawback of the weighted-sum technique is that it cannot locate a non-dominated front in the nonconvex part. In most cases, weighted-sum and epsilon-constraint techniques operate quite well. Some problems are stated if the slope of the Pareto front is not continuous (see 3.1.2.1.1). NSGA-II is more robust. It is able to find any kind of Pareto front e.g. noncontinuous front, nonconvex front. However, like other GAs, it needs a very large number of function evaluations in order to obtain good results. For the surrogate-assisted algorithm, the same conclusion as for the single-objective may be drawn. It uses a smaller number of function evaluations. Experiment have shown that it is able to locate a nonconvex front and gives similar results to NSGA-II, but with a very small number of function calls. These optimization techniques intend to solve ordinary optimization problems as well as the Global System approach.

The second point discussed in this chapter has been the MDO approach. MDO is a pool of formulations dealing with interactions between models. These interactions are unavoidable in multidisciplinary/multiphysical modelling and optimization context. The MDF is the simplest MDO formulation that puts the MDA into the optimization loop. This formulation is commonly used. The IDF and AAO treat interactions and discipline implicit variables (for AAO case) by way of optimization. This allows parallel and distributed computation. It has

been shown that the MDO offers the possibility to break a system model into disciplinary models and to perform the optimization process in an efficient manner.

For complex systems including many subsystems and components, the design problem is generally broken into subproblems. TC allows formulating the system optimization problem as coordinated optimal design subproblems. Large-scale optimal design problems can then be performed with respect to the organisation of the company. The optimal design subproblems are mapped directly to each corresponding engineering unit. This should make it easier for the company to accept and use the optimal design for their everyday design tasks.

Part II

Applications

Chapter 4

Optimal design of traction motors

A great amount of research and development in permanent magnet synchronous motors allows the railway industry to develop and industrialise this kind of motors in their traction applications. The two main reasons are its high efficiency and its compactness. However, traction motor design is a task for specialists. In a global market and in the context of sustainability, electrical motors must fulfil various requirements, not only physical and technological but also regulatory and environmental. Motors must be better adapted to their usage. Therefore, the design process becomes more complex since many engineering domains are involved. The complexity is mainly due to the existence of interactions between disciplines and between physical or functional subsystems. Such interactions are usually difficult to manage because of the increasing number of design parameters. This problem arises in many engineering situations.

The traction motor design problem implies many interactions, which can be "internal" interactions e.g. material properties are influenced by temperature, and "external" interactions. The main external interactions are the inverter (electrical) and performances (mechanical). When designing a motor without considering the system, these interactions are usually fixed by specifications and given as design parameters. The mechanical performances of a motor are basically expressed as the required torque and speed. They are described by specifications, which must take into account the constraints of the system. The adaptations of component specification is a complex task. The optimal design methodology is used to carry out this task.

The optimal design of such complex systems is under investigation in the engineering field. Four main subjects are discussed in this chapter:

(i) Definition of required torque and speed

(ii) Solving of interaction problems using MDO

(iii) Multi-criteria optimal design

(iv) Use of high fidelity tool in optimal design

First, a multidisciplinary semi-numerical traction motor model is briefly introduced. This model is used in Section 4.2, 4.3 and 4.4. In Section 4.5, a multi-physical traction motor

model including a finite element electromagnetic field analysis and an analytical thermal model is presented and used in the optimal design process.

4.1 Traction motor design methodology

Figure 4.1 depicts a general process of traction motor design. Vehicle performance requirement such as acceleration, maximum speed needed, or route profile are used to determine required torque and speed of motor. Several approaches are introduced in Section 4.2. Then, design problem is formulated i.e. definition of design variable, design parameter, constraint and objectives. Next step is the preliminary design. Optimal design technique or just simple calculation sheet can be performed. Generally, analytical model is used in this design phase. Results from this phase (i.e. dimensions of motor) are sent to the detail design phase and considered as base-line or initial design. High fidelity tools are used in this phase to verify local phenomena and to adjust some geometries such as tooth or magnet shape.

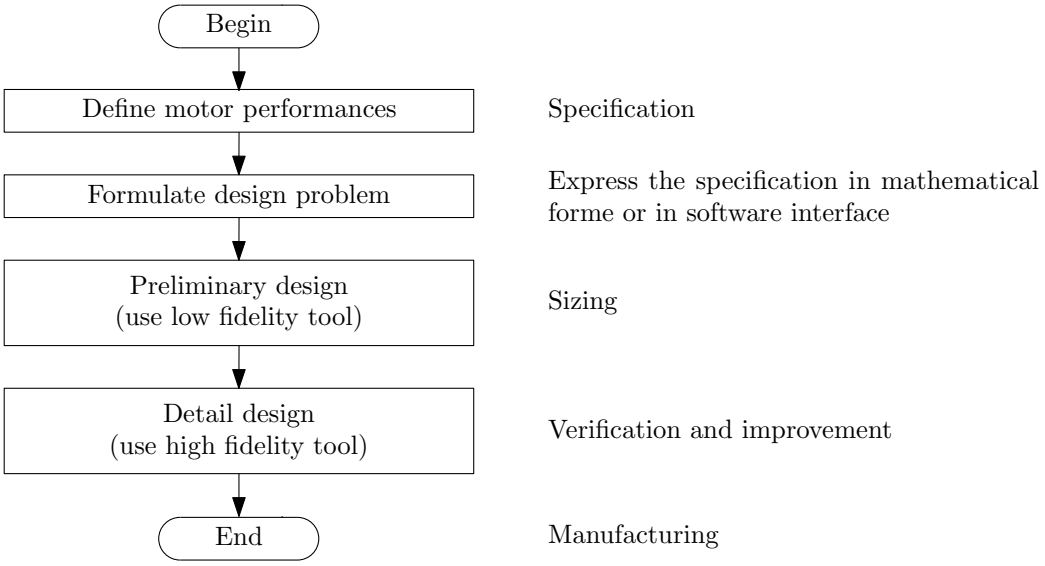


Figure 4.1: Traction motor design methodology

The following sections semi-numerical traction motor model and associating design problem. This model is used in preliminary design phase.

4.1.1 Traction motor modelling

A Surface-Mounted Permanent Magnet (SMPM) motor model is specially developed due to the specific needs of traction system designs. The SMPM motor model's structure, as well as its inputs and outputs, are shown in Figure 4.2. The model is developed using a modular approach. Each module represents a discipline. This modular approach simply allows modifying or replacing modules as needed in terms of accuracy, functionality, etc. Several simulation possibilities can be followed e.g. maximal torque per ampere control strategy or flux-weakening control, steady state temperature or transient temperature simulation. These functionalities are important for a traction motor design. In this section, only important

points are given, Readers should refer to Appendix B for a detailed description of the design model.

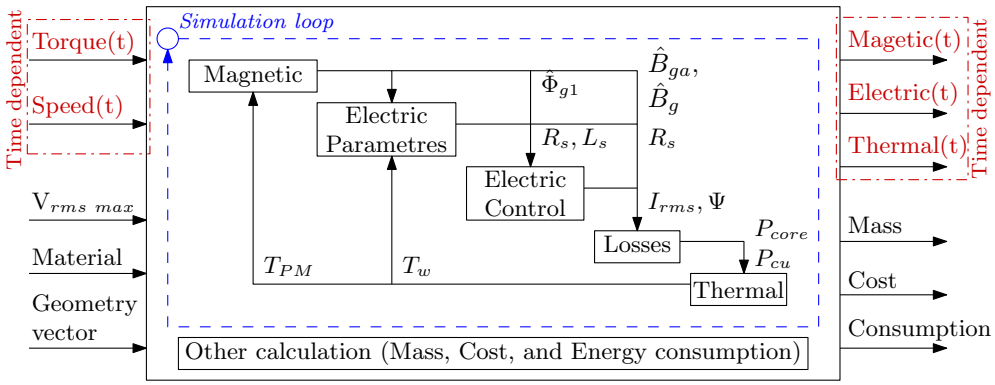


Figure 4.2: Structure of a SMPM motor model

4.1.1.1 Thermal module

The thermal behaviour is a very important constraint in traction motor designs. A lumped parameter thermal network is used to predict the temperature in various parts of the motor, both in transient and steady state. When solving steady state temperatures, thermal capacitors are omitted. Copper losses and iron losses in teeth and yoke are heat sources. Thermal resistances and capacitors are computed from motor geometries and materials. Heat convection is also taken into account in the air gap and on the external surface. The thermal model has 8 nodes and 66 equations in total.

4.1.1.2 Temperature feedback

In the SMPM model, two temperature feedbacks are defined: the winding temperature (T_w) and the permanent magnet temperature (T_{PM}). The winding temperature allows an update phase resistance using:

$$R_s = R_{s20}\left(1 + \alpha_{cu} \cdot (T_w - 20)\right) \tag{4.1}$$

where T_w is in °C, α_{cu} conductivity temperature coefficient of copper and R_{s20} phase resistance at 20°C. For $\alpha_{cu} = 0.00393\ °C^{-1}$, a resistance of 1Ω at 20°C increases to 1.3144Ω at $T_w = 100°C$.

The permanent magnet temperature effect can be taken into account via:

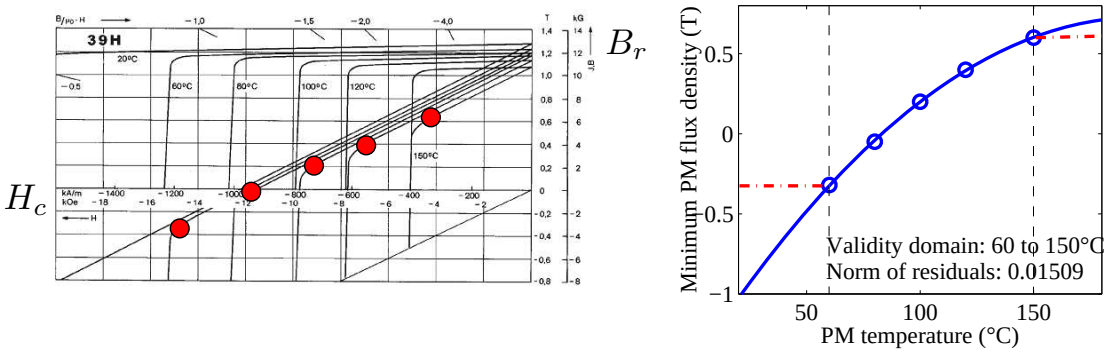
$$B_r = B_{r20}\left(1 + \frac{\alpha_{B_r}}{100} \cdot (T_{PM} - 20)\right) \tag{4.2}$$

where T_{PM} is expressed in °C, α_{B_r} reversible temperature coefficient of the PM remanence flux density (B_r) and the B_{r20} PM remanence flux density at 20°C. The PM relative permeability is considered constant. For NdFeB, α_{B_r} is negative. It means that when the PM temperature is high, B_r is low. For example, for NdFeB 39H, α_{B_r} is -0.1 %·°C^{-1} and B_{r20} is equal to 1.28T. If the temperature increases to 100°C, B_r becomes equal to 1.18T

The PM temperature also determines the demagnetization of the PM. The resultant flux density cannot be lower than the PM flux density at the knee point of a normal demagnetization curve (B_{rmin}) (red points in Figure 4.3a, a PM datasheet is provided in [93]). This relationship can be approximated from B-H curve at various temperatures given in the PM data sheet (see Figure 4.3b). For example, for NdFeB 39H, B_{rmin} can be given as:

$$B_{r_{min}} = -5.58 \cdot 10^{-5} \cdot T_{PM}^2 + 0.022 \cdot T_{PM} - 1.445 \tag{4.3}$$

This relationship is valid from 60°C to 150°C. Out of this domain, constant values corresponding to values at lower and upper boundaries are given. In Figure 4.3b, one can see clearly that at a high PM temperature, the minimum PM flux density to avoid demagnetization is high, hence a poor flux weakening capability. Therefore, this relationship must be taken into consideration in order to avoid the PM demagnetization at high temperatures. A constraint on the PM maximal temperature must be taken into account when using the model.



(a) B-H characteristic from PM data sheet

(b) Approximate model

Figure 4.3: The minimal value of the remanence flux density as a function of temperature can be estimated on the basis of the PM B-H curve

4.1.1.3 Control strategy

The flux-weakening control strategy is very useful in traction applications. It allows operating at constant power in a high speed range and limiting the inverter current in low speed. It is essential to verify whether the motor can successfully operate in a flux-weakening zone not only on an electric point of view but also on a magnetic and thermal point of view, by using the PM demagnetization curve and temperature feedback.

Control module is the inverse model i.e. design model. Its inputs are the torque and speed requirements. A flowchart of flux-weakening control is shown in Figure 4.4. Circuit parameters (phase resistance, phase inductance and flux) computed by electric parameter modules are also needed. Figure 4.5 depicts two operating zones and their electrical properties. The detailed expressions are given in Appendix B.

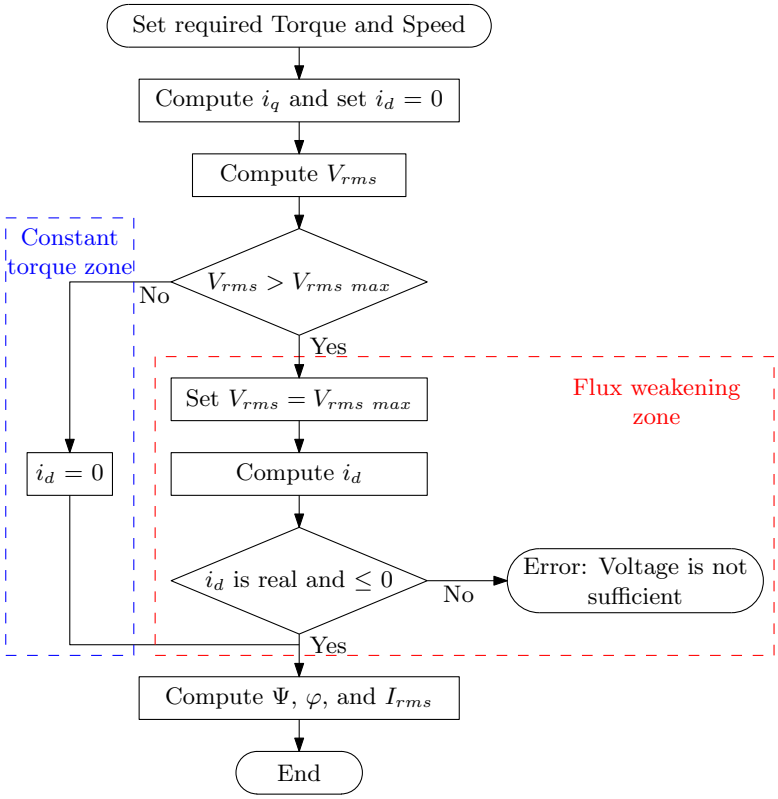


Figure 4.4: Flux weakening control flowchart

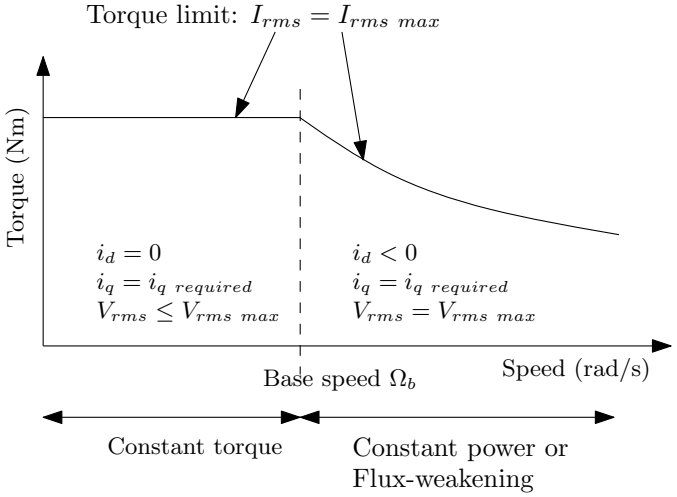


Figure 4.5: Traction motor operating zone

4.1.2 Design problem formulation

In this chapter, several design problems are expressed with different design variables, design parameters, constraints and objectives. Regarding the SMPM motor semi-numerical model, design variables and parameters can be chosen from model input variables, for example the following:

- Motor architecture: number of pole pairs, number of phases, number of slots per pole;
- Motor dimensions: stack length, PM height, slot height;
- Winding characteristics: number of conductors, winding pitch;
- Material and material properties;
- Inverter interface: maximum voltage[1].

Some are fixed as design parameters and some are design variables. Their upper and lower bounds depend on which design problem is being studied.

Design objectives and constraints are defined from model outputs variables, which are the following:

- Model hypothesis: maximum flux density (linear B-H characteristic);
- Material properties: maximum winding and PM temperatures;
- Failure protection: demagnetization of the PM;
- Mechanical interface: external diameter and length, mass, volume;
- Inverter interface: maximum current;
- Motor efficiency: energy consumption, efficiency, losses;
- Other: cost.

[1] In this model, maximum voltage is fixed as design parameter

4.2 Definition of required torque and speed

In electric motor design, a wide range of specification must be defined. They may characterise the interaction to the other components e.g. maximum current and voltage (inverter interaction), external volume, mass (mechanical installation interaction), cooling mean (cooling system interaction). Figure 4.6 shows the interfaces of the traction motor with the system. The system imposes mechanical performances to the motor. The motor interacts with the system via mass, volume, temperatures, current and voltage.

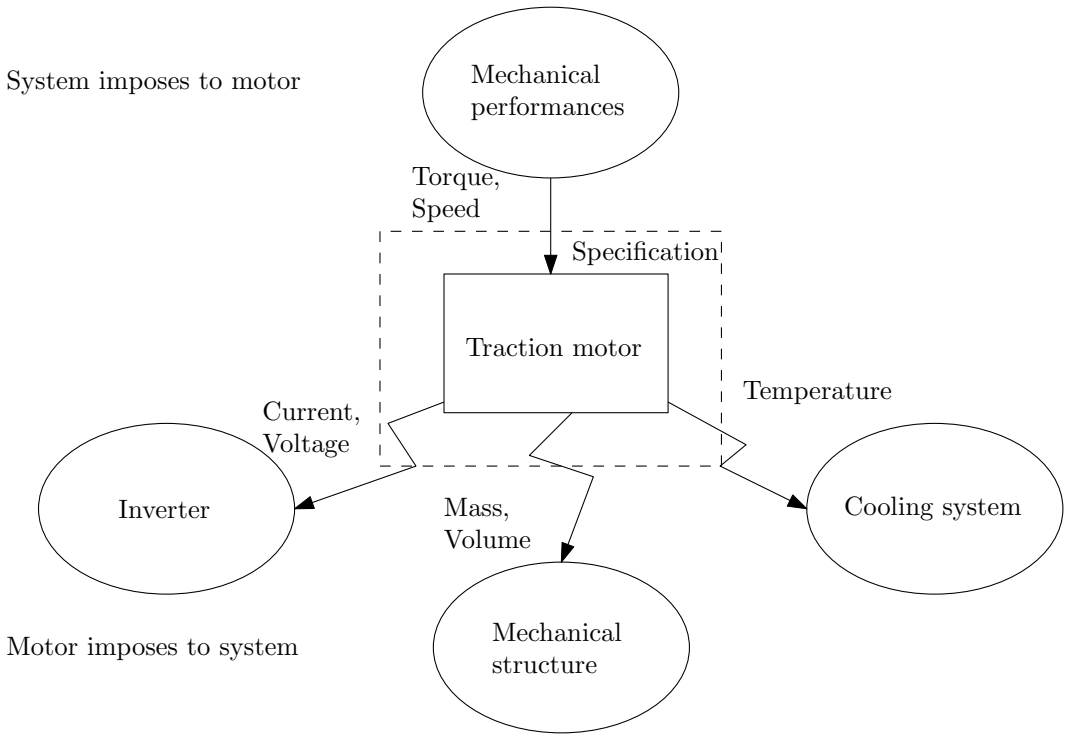


Figure 4.6: Interaction with the traction motor

The most important specifications may be the torque and speed that the motor must produce. Many approaches can be used depending on applications and needs [5, 26, 74]. In industrial applications, a classical design approach using rated torque and speed is commonly used. It can perform quite well due to the fact that the load is constant. However, in traction applications, loads vary according to driving cycles. Different approaches should be employed. This section aims at assessing various electric traction motor design approaches. The "optimal" results obtained using different approaches are compared.

4.2.1 Torque and speed characteristics

The torque and speed requirements depend on the application and operating conditions. Some applications need a fixed speed and a constant torque motor and other applications meet the requirements through the use of the load cycle. The load cycle may be simply described as the amount of **torques** that the motor can produce at a fixed **speed** in a certain period of **time**. This cycle is repeated and allows the motor to operate in the steady-state. In some

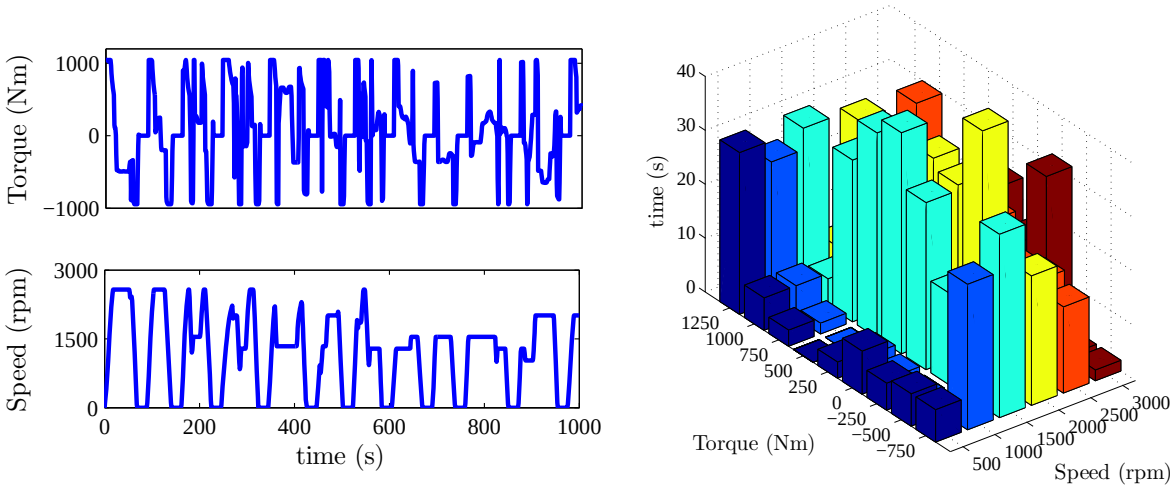
traction motor applications, a more complicated description of load cycle is typically used. The motor load profile can be computed on the basis of the known vehicle characteristics and the requirements of:

(i) the driving cycle (i.e. speed or acceleration as a function of time) in the automotive application.

(ii) the route profile (i.e. maximal speed and climbing slope as a function of distance) and trip time requirements in the rail application.

For rail applications, the customer (train operator) specifies how the train will be operated. This implies giving the track characteristics (such as speed limit, track slope, bending curve), station information (stop time, stop distance) and driving characteristics related to passenger comfort (maximum acceleration, deceleration and jerk) and trip time. All these information and vehicle characteristics allow determining the tractive force and braking force necessary to accomplish the route profile. The motor torque and rotation speed can then be easily deduced from the number of motors and the gear ratio. The reader should refer to [7, 75, 114] for vehicle dynamic and torque, speed calculations.

Figure 4.7a shows the motor torque and speed characteristics in time domain. The motor operates in both traction and braking modes. Figure 4.7b shows time sharing of torque and speed in the load cycle. In this load cycle, the motor operates most of the time in the medium torque and speed zone. It can be stated that the thermal behaviour becomes very essential when a motor works in variable load and variable speed conditions [57].



(a) Time domain

(b) Time sharing in load cycle

Figure 4.7: Torque and speed requirement

The following section provides different torque and speed requirement definitions used as inputs in traction motor design problems.

4.2.1.1 Approach using rated torque and speed

A rated point, also called base point, represents a rated torque and base speed of a motor (see Figure 4.5). At base speed, the inverter reaches its maximal voltage [73]. In the classical design approach, the base point is usually considered as a design point. The motor can operate permanently at this load and speed without excessive heat. This approach works very well with industrial motor designs, in which the load and speed are fixed and the flux-weakening operating mode is not considered. A motor designed by using this approach can be successfully used in variable speed applications and can operate at a speed over the base speed by using an appropriate inverter with flux-weakening control. However, the full operation ability is not ensured at the design stage and is used as it is in variable speed applications.

In traction applications, a question may be asked: How to determine this rated point? The rated speed is easily referred to as the transition speed from the constant torque zone to the constant power zone. The rated torque is more complicated to select. The maximal torque requirement can be used but it might lead to an oversized motor because the motor would never operate permanently at the maximal torque. One might determine the rated torque as a percentage of the maximal torque requirement. However how many percent it should be set to is not always clear. This depends on the application and requires a lot of experience.

In this work, three rated points are defined:

(i) The maximal torque and base speed (case I)

(ii) 75% of the maximal torque and base speed (case II) (75% is chosen arbitrarily.)

(iii) The average torque and average speed (case III) computed using (4.4) and (4.5):

$$T_{avg} = \sqrt{\frac{\Sigma\left(T_i^2 \cdot t_i\right)}{\Sigma t_i}} \tag{4.4}$$

$$\Omega_{avg} = \frac{\Sigma\left(|\Omega_i| \cdot t_i\right)}{\Sigma t_i} \tag{4.5}$$

where T_i and Ω_i denote torque and rotational speed at time period t_i.

For the considered load profile (shown in Figure 4.7), the average torque is 53% of the maximal torque and the average speed is 45% of the maximal speed. This average point can be seen as the operating point of the motor most of the time. For both cases I and II, the base speed is defined by the transition speed from the constant torque mode to the constant power mode in the traction torque-speed curve as shown in Figure 4.8. The figure also plots torque versus speed at each time step (the data are from Figure 4.7a).

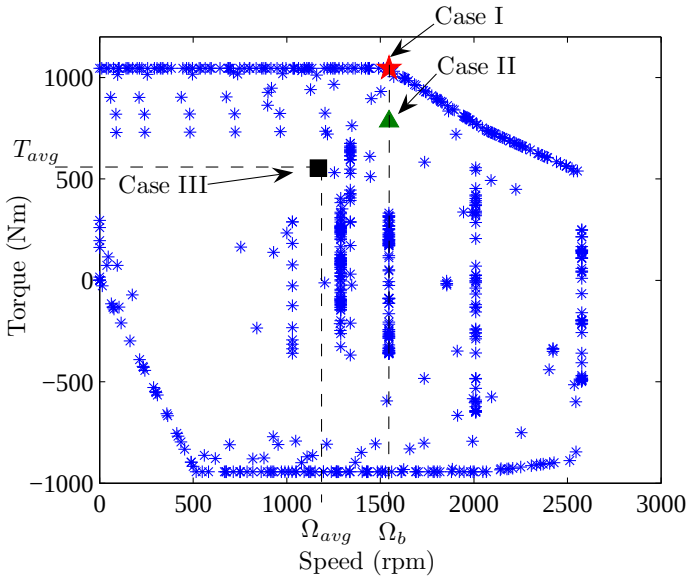


Figure 4.8: Rated point definition

4.2.1.2 Approach using torque versus speed curve

This approach is very similar to the rated point approach. Several operating points on the torque-speed curve are used instead of only one point. This ensures that the motor can operate at all required torques and speeds including the flux-weakening zone [26]. As in the previous approach, a percentage between the rated torque-speed curve and the maximal torque-speed curve can be introduced. The steady state operation is considered throughout the rated torque-speed curve. Four characteristic points are selected (depicted as stars in Figure 4.9). These points represent the maximum torque, maximum speed point in both traction and braking modes (case IV). A 75% torque-speed curve is also examined (case V) and represented by triangles in Figure 4.9.

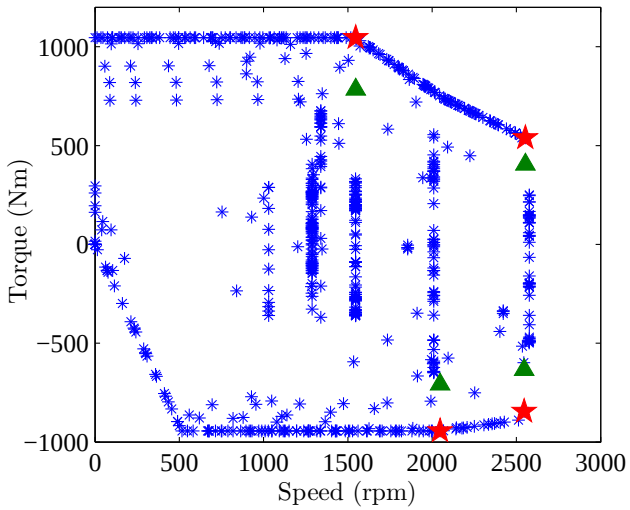


Figure 4.9: Torque-speed characteristic. The stars represent the 4 characteristic points (maximal torque envelop) and the triangles represent 75% of the maximal torque envelop

4.2.1.3 Approach using load cycle

In some applications such as light rails (metro and tram), load profiles are precisely known. It is very interesting to take into account the transient thermal behaviour of the motor in the design [74]. Thermal simulation on load profile is repeated until the temperatures reach their steady state. The steady state operation concerns not only an operating point but also the whole load cycle. This approach ensures that the optimal motor operates normally at each operating point on the predefined load cycle. The load cycle shown in Figure 4.7a is used in this approach (case VI).

4.2.2 Comparative study methodology

A study is performed in order to investigate how different torque speed definition approaches affect the design of a motor. Three comparison metrics are determined: (i) optimality in terms of design criterion, (ii) ability of the motor to operate on the load cycle and (iii) design time (in our case, optimization time). The comparison is based on optimal motors designed by using different approaches (case I – case VI). Each optimal motor is obtained through an optimization technique.

In this work, the motor mass is considered as the objective function. Many constraints are also defined e.g. the maximal temperature, maximal flux density, and maximal current. To solve the optimization problems, a single-objective optimization technique, Sequential Quadratic Programming (SQP) with multistart [83], is used in order to ensure a global optimum.

All design variables, constraints and objective functions are shown in Table 4.1. Some parameters are fixed as a constant due to the fact that SQP cannot deal with integer design variables. They are firstly considered as continuous variables. The optimization results are rounded to integers and fixed as a constant in the comparative study.

To summarize, six optimization cases are carried out. The objective is mass minimisation. The six optimization cases correspond to different design approaches as shown in Table 4.2. Note that:

(i) The N_c number of conductors per slot is fixed to 4 for all cases except case III. The optimization algorithm cannot converge unless the number of conductor is considered as a design variable. In this case, the value of 8.32 conductors per slot was found by optimization. This number was rounded to 8 and fixed as a constant. After that the optimization is rerun.

(ii) The initial temperature condition in simulation (case VI) during the optimization is set to 100°C and the time step is of 10s (instead of a 1s time step from torque and speed input data) in order to accelerate the optimization process.

(iii) The RMS maximum inverter voltage is of 290V in traction mode and of 350V in braking mode.

Table 4.1: List of design variables, constraints, constants, and objective function

Symbol	Quantity	Unit	Remark	Limit
Y	Yoke height	mm	D.V.	[10, 50]
l_m	PM height	mm	D.V.	[1, 15]
g	Air gap	mm	D.V.	[2, 10]
d_s	Slot height	mm	D.V.	[20, 40]
r_a	Armature radius	mm	D.V.	[100, 500]
l_{stk}	Stack length	mm	D.V.	[100, 700]
r_{wt}	Tooth width ratio	-	D.V.	[0.3, 0.7]
k_p	PM span coeff.	-	Constant	0.8
p	Pole pair	-	Constant	3
N_{slot}	Slot/pole/phase	-	Constant	2
N_c	Conductor/slot	-	Constant	4, 8
A	Parallel path	-	Constant	1
V_{rms}	rms maximum voltage	V	Constant	290 (Traction), 350 (Braking)
T_w	Winding temp.	°C	Con.	≤ 200
T_{PM}	PM temp.	°C	Con.	≤ 150
B_{Y_s}	Stator yoke flux density	T	Con.	≤ 1.6
B_{Y_r}	Rotor yoke flux density	T	Con.	≤ 1.6
B_T	Teeth flux density	T	Con.	≤ 1.6
I_{rms}	rms current	A	Con.	≤ 300
Pt_{dem}	PM demag. point	-	Con.	≤ 0
M	Mass	kg	Obj.	min.

Note: D.V. = Design variable; Con. = Constraint; Obj. = Objective function

Table 4.2: Summary of optimal design cases

Case I	Maximal torque and base speed	steady state
Case II	75% of the maximal torque and base speed	steady state
Case III	Average torque and speed	steady state
Case IV	Four characteristic points on the maximal torque-speed curve	steady state
Case V	75% of the torque-speed curve	steady state
Case VI	Load cycle	transient

4.2.3 Results

Table 4.3 shows optimization results. Only the main motor characteristics are shown. The values in brackets are relative values compared with case VI solutions. The comparison between approaches is based on the optimal solutions designed by using the same criterion. Therefore, it only reflects the influence of different design approaches.

The optimal motors, obtained by using optimization, are simulated on the load cycle with time steps of 1s and an initial temperature of 40°C. Their behaviour can be observed in Table 4.4. In this table, the values in brackets are relative values, compared to constraint limits. These simulations ensure that the motors can operate correctly on the load cycle.

Figure 4.10 depicts the electrical limit of the motor torque when the maximal current is supplied without taking into account the thermal limit.

Table 4.3: Optimal motor characteristics for each case

Symbol	Quantity	Unit	I	II	III	IV	V	VI
M	Mass	kg	531	423	296	562	437	364
			(45.9%)	(16.2%)	(-18.7%)	(54.4%)	(20%)	
$Vol.$	Volume	m^3	0.0976	0.0713	0.05562	0.1105	0.0845	0.0632
			(54.4%)	(12.8%)	(-11.1%)	(74.8%)	(33.7%)	
C	Cost	euro	2793	2153	1471	3276	2609	1815
			(53.9%)	(18.6%)	(-19%)	(80.5%)	(43.7%)	
R_{ext}	External radius	mm	217.7	176.3	185.3	235.2	212.2	174.8
l_{ext}	External length	mm	655.4	730.2	520.9	635.6	597.2	659.1
g	Air gap	mm	3.67	2	2	5.93	5.38	3.37
l_m	PM height	mm	11.94	7.17	7.67	15	15	6.61
N_c	Conductor per slot	-	4	4	8	4	4	4
R_s	Phase resistance	Ω	22.9	40.7	107.1	17.3	21.3	37.3
L_s	Inductance	mH	1.148	1.474	3.782	0.794	0.729	1.119
$\hat{\Phi}_{g1}$	Flux	Wb	0.53	0.53	0.69	0.47	0.42	0.39
T_{PM}	PM temp.	°C	150	150	150	150	150	145.3
T_w	Winding temp.	°C	173.4	168.3	183.3	185.8	188.5	200.0
I_{rms}	RMS current	A	219.2	164.1	89.9	246.0	209.9	300
B_{Y_T}	Flux density	T	1.6	1.6	1.6	1.6	1.6	1.6
t_{opt}	Time per	s,	8s	8s	8s	15s	15s	7min
	optimization	min						(10s step)

Note: Percentage relative values in brackets are **compared with case VI**.

4.2.3.1 Case I and II

As expected, the cases using steady state temperature constraints lead to oversized motors as the simulated temperatures on cycle do not reach maximal value. Furthermore, the flux density in the tooth, the stator and the rotor are over the limit for the reason that the

Table 4.4: Transient simulation on load cycle

Quantity	I	II	III	IV	V	VI
T_{PM}	82.4	99.7	-	79.3	90.8	145.7
	(-45.1%)	(-33.5%)		(-47.1%)	(-39.5%)	(-2.8%)
T_w	93.9	117.3	-	91.2	114.1	196.8
	(-53.0%)	(-41.4%)		(-54.4%)	(-43.0%)	(-1.6%)
I_{rms}	245.6	221.4	-	257.8	262.0	300.3
	(-18.1%)	(-26.2%)		(-14.1%)	(-12.7%)	(0.1%)
B_{Y_T}	2.06	2.09	-	1.84	1.79	1.65
	(28.8%)	(30.6%)		(15.0%)	(11.9%)	(3.1%)
Cycle to steady state	26	26	-	28	28	40
Energy per cycle (kWh)	8.85	8.92	-	8.83	8.85	9.08

Note: Percentage relative values in brackets are **compared with constraints**.

steady state temperatures were considered in the optimization and are higher than the initial temperature of 40°C. At 40°C, the PM yields a better performance (Equation 4.2 gives the relationship between the PM remanence flux density and the temperature). Therefore, the flux density is usually higher at a low temperature than at a high temperature. The flux density is always higher than the constraint limit of 1.6T as it can be seen in Figure 4.11a. This shows that the normal operating on load cycle cannot be ensured by the approach using steady state temperatures.

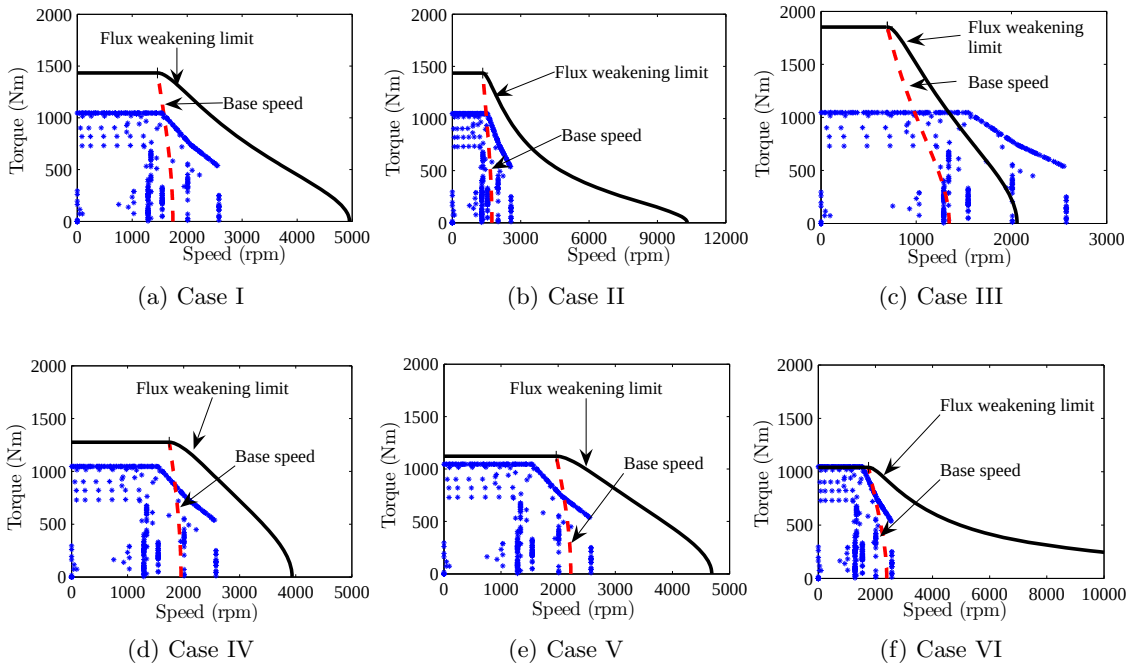
From Table 4.3, one can notice that the motors do not use the maximum current. This is because a higher current would generate higher losses and the motor should then be bigger in order to be able to evacuate these losses.

4.2.3.2 Case III

Case III results in the lightest motor. However, the simulation on the load cycle cannot be achieved with this motor. The flux-weakening range does not cover all the desired speeds as shown in Figure 4.10c. This case confirms that the rated point approach can lead to an inappropriate design because the flux-weakening characteristic is neglected during the optimization design process.

4.2.3.3 Case IV and V

The motor mass for the cases using a characteristic torque-speed curve (case IV and V) is greater than for the cases using a rated point. This is due to the fact that the motor needs to cover a higher speed range (until maximal speed is reached), hence higher flux density and demagnetization issue. Moreover, the motor is subject to high iron losses while operating at a high speed. The heat exchange surface must be greater. This results in bigger and heavier motors than for other cases.



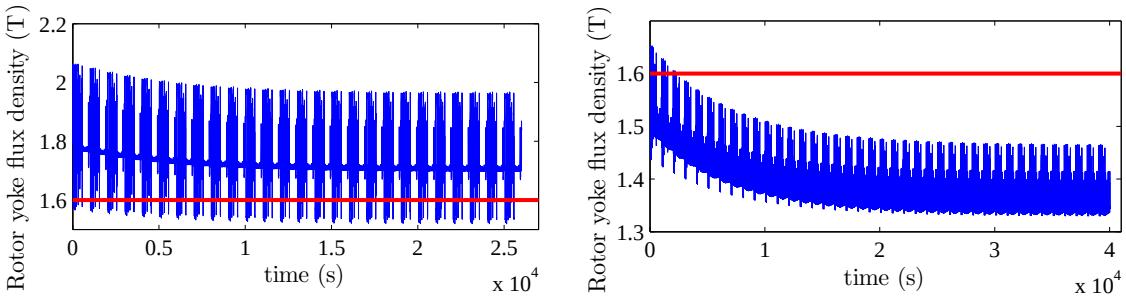
(a) Case I (b) Case II (c) Case III

(d) Case IV (e) Case V (f) Case VI

Figure 4.10: Torque at electrical limits for different cases L_s and $\hat{\Phi}_{g1}$ are given in Table 4.3. The phase resistance is neglected. $I_{rms} = 300\text{A}$, $V_{rms_{max}} = 290\text{V}$

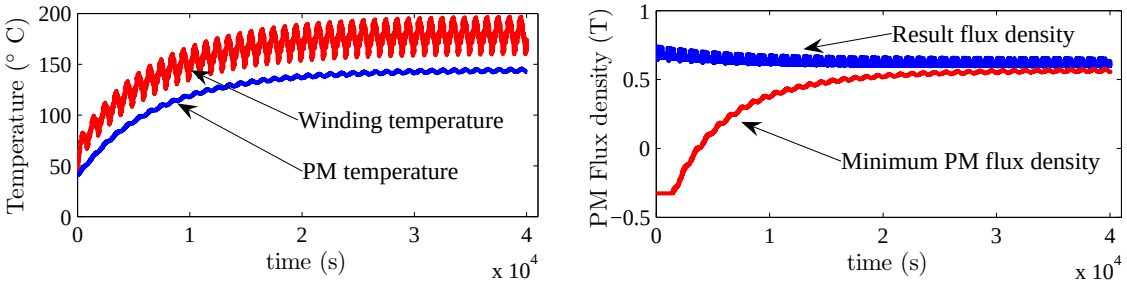
4.2.3.4 Case VI

The approach using the load cycle (case VI) gives the best results among the six cases. The simulation during optimization allows taking the requirements exactly into account. As the time steps during the optimization and the simulation are not the same (10s and 1s), some constraint violations can be observed (see the case VI column in Table 4.4). For example, the rotor yoke flux density constraint is of 1.65T instead of 1.6T, as defined in the optimization problem. Figure 4.11b shows that the flux density is over 1.6T at the beginning and then decreases to 1.5T at a steady state. Figure 4.12 shows the simulation results for the Case VI motor. The temperatures reach a steady state after 40 cycles, instead of 26 or 28 cycles for other cases. In Figure 4.12b, the PM demagnetization does not occur since the result flux



(a) Case I (b) Case VI

Figure 4.11: Rotor yoke flux density simulation results



(a) Winding and PM temperatures

(b) PM flux density

Figure 4.12: Simulation results for the Case VI motor

density curve is always higher than the minimal PM flux density curve obtained while using equation (4.3).

In order to compare this with other solutions, the rated torque and base speed can be found through an optimization problem using the motor geometries as input parameters and searching for the highest torque and speed that satisfy the above mentioned constraints in the steady state operation. As a result, this motor has a rated torque of 626.14Nm and a base speed of 1923rpm (60.15% of the maximal torque and 75% of the maximal speed). It is not obvious to find these values beforehand and use them in a steady state approach.

4.2.3.5 Computation time

Regarding computation time, the steady state approaches have taken much lower time: approximately 8s per optimization for rated point approaches and 15s per optimization for torque-speed approaches. For load cycle transient approaches, each optimization takes 7min. These times cannot be generalized to other design problems.

In this work, analytic models are used in each modules of the SMPM model. If FEA is used for magnetic module, the time discrepancy may be reduced since the time consuming task is now FEA. At each trail optimization design vector, only one FEA is called since FEA only provides circuit parameters (inductance and flux linkage). Other modules such as the control and thermal modules remain analytical equations and are relatively rapid compared with FEA. On the other hand, if more complex thermal models (e.g. model with more node number) or co-simulations (control integrated with FEA) are used, the time difference between the steady state and the load cycle approaches will be much higher since at each time step, a high computation time model is called. However, for preliminary design purposes, a more or less complex analytic model is used. Therefore, the computation time is acceptable. It is suggested to incorporate the load cycle and transient thermal behaviour in the design of traction motors.

Another issue can be stated. The computation time can depend on the load cycle. For example, if the load cycle is long and contains high torque or speed variations, the time step must be smaller. The computation time can be very high. In this case, an inevitable load cycle analysis needs to be carried out to simplify the load cycle or select an important part of it.

4.2.4 Conclusion on the comparative study

The electric motor can be designed, on the basis of various approaches depending on the application and needs. This research intends to compare different approaches for the traction application. The optimal design results deduced from each approach are compared and analysed. The advantages and drawbacks are pointed out.

The approach using a rated point is suitable for the fixed speed and constant torque motors since it allows obtaining the optimal design rapidly. However it cannot ensure an operation in a flux-weakening mode. The case III depicts this inconvenient.

To overcome this problem, several load points, including some points in the flux-weakening zone, can be used instead of the rated point. This is referred to as the torque-speed curve approach. It preserves the advantages of the previous rated point approach while ensuring a proper operation in the flux-weakening zone. Nevertheless, this approach results in extremely oversized motors because the high speed zone must be taken into account in the design.

The approaches mentioned above allow taking only the steady state operation into account. In traction applications, it is strongly recommended to take into account the transient thermal behaviour of the motor. According to the results, downsizings can be obtained. The motor mass can be reduced significantly and the constraints are well respected when operating on the load cycle. In this example, the motor mass can be reduced of at least 16%, as shown in Table 4.3. Despite these advantages, when using load cycle in optimal design, the computation time will be higher than with the rated point approach. However, it is acceptable while using an analytical model. The compromise between the fidelity of model (e.g. analytical or numerical) and the optimization time requires further investigation.

4.3 Multidisciplinary Design Optimization

In complex system design problems, the interactions between subsystems and disciplines are very important issues. As presented in Chapter 3, the Multidisciplinary Design Optimization (MDO) addresses to such problems. The MDO federates different aspects: analysis tools, optimization algorithms and problem formulation. The problem formulation term does not only correspond to the definition of objective function, constraint and design variables as in a single disciplinary design optimization but corresponds also to the way of solving the interactions between disciplines. The MDO has been used successfully in aeronautic and automotive fields. This concept can be applied in electrical engineering and seems to be very helpful for designers facing complex systems or multidisciplinary aspects.

Electrical machine design involves a number of disciplines e.g. magnetics, electrics, control, heat transfer, mechanics, cost analysis, and load cycle. An electrical motor model consisting of several discipline models is presented in Section 4.1.1. The interactions between modules must be solved by using a numerical method. However, when using such a model in an optimal design process, different approaches can be implemented. This section presents an application of MDO techniques to SMPM optimal designs. Three single-level MDO techniques are applied (see Section 3.2.1 for mathematical formulations).

4.3.1 MDO techniques comparison

The design problems for a MDO comparison are the same as in the previous section. A list of design variables and constraints is given in Table 4.1. For these problems, the required torque and speed are defined by a rated point[2] and a steady state thermal behaviour is considered. The rated torque and speed are defined: $T_{rated} = 626.14$Nm and $N_{rated} = 1923$rpm. This rated point was found in Section 4.2.3.4.

The SMPM model presented in Section 4.1.1 presents two thermal interactions as well as 6 implicit variables in the thermal discipline. In the original version, these interactions and implicit variable loops are solved by using fixed point iteration method. According to the MDO techniques used, the model is then re-configured and some auxiliary design variables and constraints are added to the optimization problem [58]. A SQP algorithm is used to solve the optimization problem.

Starting from the same initial design point, all approaches lead to approximately the same final optimal motor. These results will be discussed in the next section. This section intends to observe the impact of each approach on the optimization problem formulation and the behaviour of the optimization process. The assessment is done by comparing the number of design variables and constraints as well as the number of system analysis evaluations (i.e. the number of objective function evaluations and number of evaluations for each discipline. These evaluation quantities are reported in Table 4.5.

[2] see Section 4.2

Table 4.5: Comparison of different MDO cases

Number of	MDF	IDF	AAO
Design variable	7	15	21
Constraint	10	18	24
Function objective evaluation	158	859	1149
Magnetic module evaluation	1745	859	1149
Electric parameters module evaluation	1745	859	1149
Electric control module evaluation	1745	859	1149
Losses module evaluation	1745	859	1149
Mass module evaluation	158	859	1149
Thermal module evaluation	4516	3401	1149

4.3.1.1 MDF case

Figure 4.13a shows the data flow and model architecture of a SMPM motor optimal design problem formulated using an MDF. Design variables are given by optimization (shown as a full line). The interactions between physical domains are presented: feed-forward (depicted as dotted line) and feedback (dashed line). An MDF can be considered as an optimization layer enveloping a multidisciplinary model or a simulation model. The problem of interactions between disciplines must be solved by an iterative method such as the fixed point method. Furthermore, there are internal iteration loops inside the thermal discipline. These loops are solved separately by the fixed point method as well. The SMPM model and optimal design problems are identical to those in Section 4.2[3].

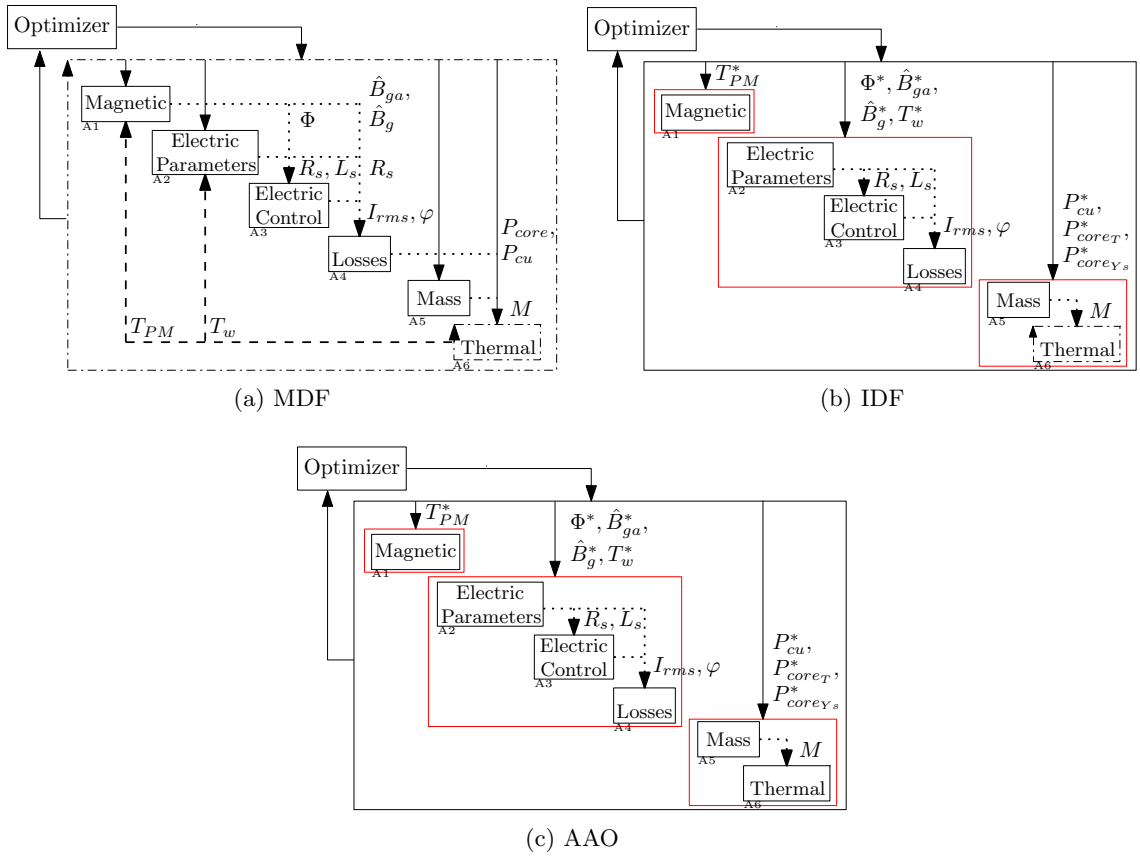
4.3.1.2 IDF case

The sequenced IDF[4] is used in the aim of minimizing the number of coupling variables. The load balancing feature is not investigated since time discrepancy between each disciplinary module of a SMPM motor model is not important. Three groups are identified (depicted as three big rectangles in Figure 4.13b and 4.13c): (i) magnetics, (ii) electric parameters, control, and losses, and (iii) mass and thermal. This results in a total of 8 coupling variables; 2 feedback and 6 feed-forward coupling variables.

The advantage of the IDF over the MDF is that the parallelization capability can be exploited if each module is completely independent i.e. the optimization process takes control over both feedback and feed-forward interactions. Two feedbacks i.e. winding and magnet temperatures (T_w, T_{PM}) and 6 feed-forward variables are removed. Eight design variables and 8 constraints are added into the MDF optimization problem. Table 4.6 shows IDF auxiliary design variables added into the original MDF optimal design problem. The optimization gives

[3]Optimal design problem in Section 4.2 is formulated using MDF approach.

[4]Sequenced IDF is a variant of IDF formulation. Only some coupling variables are treated by optimization. Please refer to Section 3.2.1.3 for more details.



(a) MDF

(b) IDF

(c) AAO

Figure 4.13: Data flow of different MDO approaches

a guess of the temperatures and other linking parameters as the additional design variables. At the end of the optimization process, the outputs computed by each discipline are enforced to be equal to the guessed input via 8 compatibility constraints.

The optimization algorithm needs a high number of function evaluations in order to cope with a large set of design variables and constraints. As a result of the parallelization, the computation time should decrease compared with the MDF case. A compromise between the number of auxiliary variables and constraints and the parallelization capacity must be studied in order to obtain the lowest computation time as possible.

4.3.1.3 AAO case

In an IDF, an assumption is made that the existing or commercial analysis tools are used. In an AAO, the codes are supposed to be modified and integrated directly to the optimization process. All iteration processes are handled by a system optimizer. In this example, the internal loop in the thermal module is removed. The thermal block is represented with a full line in Figure 4.13c. It should be noted that the thermal loop has 6 implicit variables. This adds 6 design variables and 6 constraints to the IDF cases.

The number of function evaluations for an AAO case is higher than that of an IDF case, due to a higher number of design variables and constraints. Due to the fact that the thermal discipline evaluator is called once per objective function evaluation, the number of total

Table 4.6: Auxiliary design variables for an IDF case

Number	Symbol	Quantity	Unit
1	T^*_{PM}	PM temp.	°C
2	Φ^*	Flux	Wb
3	$\hat{B}^*_{ga}$	Armature air gap flux density	T
4	$\hat{B}^*_{g}$	Air gap flux density	T
5	T^*_{w}	Winding temp.	°C
6	P^*_{cu}	Copper loss	W
7	$P^*_{core_T}$	Core loss in teeth	W
8	$P^*_{core_{Ys}}$	Core loss in stator yoke	W

thermal discipline computations is divided by more than two compared with the IDF case.

Figure 4.14 shows the convergence properties of each formulation. The figure plots the objective function i.e. the motor mass against the optimization time. In fact, the real computation time of a SMPM model is very small and the difference between each formulation cannot be captured clearly. Therefore, the computation time for one evaluation of each module is fixed to 1s for demonstration purposes. This computation time is multiplied by the total number of module function calls in order to obtain the total optimization time. For IDF and AAO, the parallelization is also taken into account. The computation time for each case is obtained as follows:

- MDF case: All modules are launched sequentially. $t_{MDF} = 4{\times}1175{+}158{+}4516 = 11654$
- IDF case: There are 3 groups that are launched in parallel. The computation time is the slowest group i.e. the mass and thermal modules. $t_{IDF} = 859 + 3401 = 4260$
- AAO case: The computation time is computed from the slowest group i.e. electric parameters, electric control and losses modules. $t_{AAO} = 3 \times 1149 = 3447$

From 1500s onwards, the objective function value for all formulations is subject to very little changes. This is because the algorithm was in fact tracking for the feasible results. It suffers from a high number of constraints in design problems. On the basis of the figure, it can be concluded that the slowest approach is the MDF. The AAO performs better than other formulations, in this example. The IDF takes a slightly higher time than the AAO. The AAO and the IDF can profit from the parallel computation, which decreases significantly the total optimization time.

4.3.2 SMPM design results

The same optimal motor, illustrated in Figure 4.15a, has been deduced from all the MDO formulation. Table 4.7 shows the characteristics of the optimal motor, such as the motor mass, its volume and material cost. It can be observed that the constraints are respected. Some important constraints are tightly active. For example, the stator yoke, the rotor yoke,

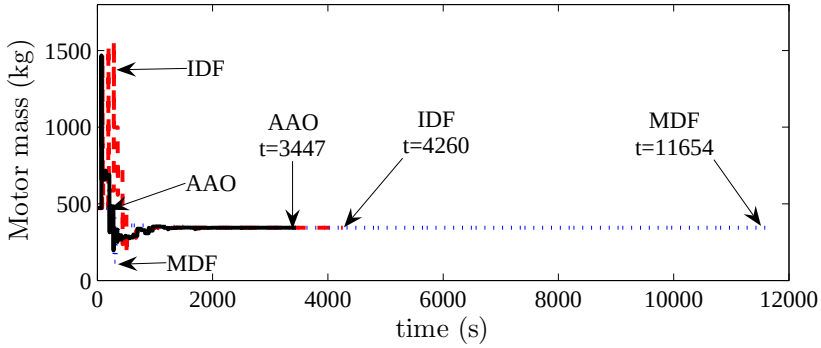


Figure 4.14: Convergence comparison between each formulation. The motor mass is plotted against the optimization time.

and the tooth flux density are of 1.6T. The permanent magnet temperature constraint is very essential in traction motors. In this example, the PM temperature reaches its limit at 150°C. It should be noted that auxiliary constraints are not shown in this table.

Table 4.7: Constraint and objective values of an optimal motor

Symbol	Quantity	Unit	Remark	Limit	Optimal result
T_w	Winding temp.	°C	Con.	≤ 200	176.77
T_{PM}	PM temp.	°C	Con.	≤ 150	150.00
B_{Y_s}	Stator yoke flux density	T	Con.	≤ 1.6	1.60
B_{Y_r}	Rotor yoke flux density	T	Con.	≤ 1.6	1.60
B_T	Tooth flux density	T	Con.	≤ 1.6	1.60
I_{rms}	rms current	A	Con.	≤ 300	196.97
Pt_{dem}	PM demag. flag	-	Con.	≤ 0	-0.097
$Pt_{V_{rms}}$	rms voltage flag	-	Con.	≤ 0	-52.41
R_{ext}	External radius	mm	Con.	≤ 700	203.66
L_{ext}	External length	mm	Con.	≤ 1000	552.69
M	Mass	kg	Obj.	min.	343.82
Vol	Volume	m^3	-	-	0.068
$Cost$	Material cost	euro	-	-	1796

Note: Con. = Constraint; Obj. = Objective function

Even though this optimal design problem is solved by using the rated torque approach, the simulation on the load cycle must be carried out in order to confirm that the optimal motor can operate properly on the load cycle. This load cycle is computed on the basis of the vehicle characteristics and required performances. The simulation or design based on the basis of the load cycle is very important and should be performed if the load cycle is known e.g. in the case of a light rail traction motor as discussed in Section 4.2. Figure 4.15b shows the motor temperatures obtained from the simulation. The simulation is run until the temperatures reach their steady-state on the load cycle. It takes 30 load cycles for this motor.

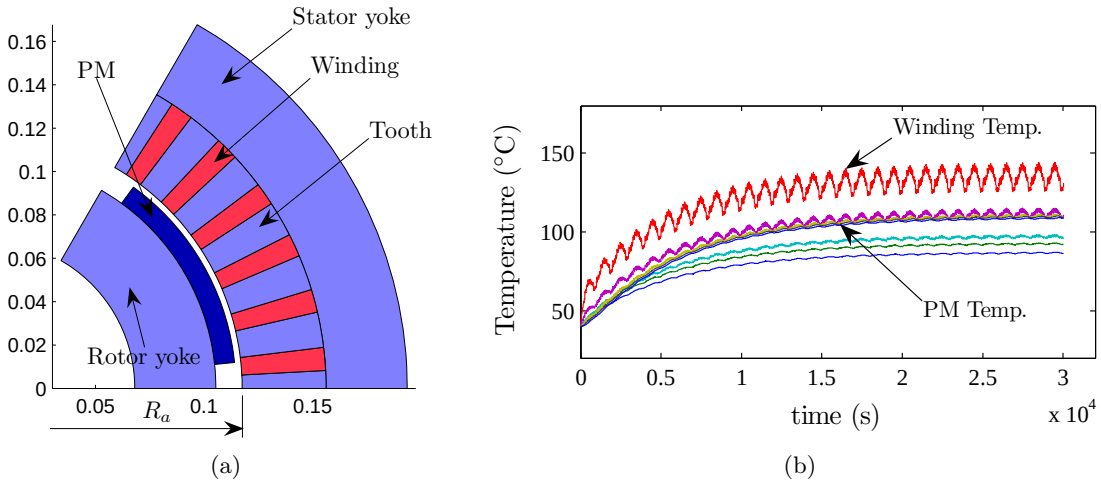


Figure 4.15: (a) Optimal motor, (b) Simulation of an optimal motor on the load cycle. The figure shows various temperatures.

One can observe that the PM and winding temperatures do not reach their constrained limit (150 and 200°C, respectively). This indicates that the motor is oversized for the chosen load cycle. This is one of the drawbacks of the rated point design approach.

4.3.3 Conclusion on the MDO

This section applied MDO formulations in the optimal design process of SMPM motors. These help engineers in formulating optimization problems. Some of these formulations can even be used to determine how the model is constructed. An optimal design of a SMPM motor was performed. It illustrates the application of the method in the light of a rather frequent design problem. Three single level MDO formulation were tested: MDF, IDF and AAO. The IDF and AAO allow decoupling interactions between submodels (disciplinary module in our case). Nevertheless, auxiliary design variables and constraints must be added to the optimization problem. This can be problematic depending on the manner the optimization algorithm is used since the optimization problem dimensions are higher. Due to the fact that these submodels are decoupled, they can be evaluated in parallel. For this problem, the optimization time is, therefore, divided by 2 for the IDF and by 4 for the AAO formulation, compared to the MDF formulation. It should be noted that all MDO approaches lead to similar optimal solutions. A selection of different formulations can be based on the number of implicit and explicit variables, the complexity of the design problem, the code accessibility, and time constraints.

MDO formulations open new approaches to formulate and to solve complex problems. They can be used in optimal design layers (as in MDF) as well as in model construction processes (as in AAO). It is clear that a well developed model, with an optimization purpose in mind, can benefit from a low optimization time and more robust optimization problem solving.

4.4 Multi-criteria optimal design problem

In the sections presented earlier, the design problems are formulated as single-criterion problems. However, the design of a traction motor usually involves trade-offs between several criteria. The design problem should be formulated as a multi-criteria problem. At the preliminary design stage, the multi-criteria design problem helps the design engineer to understand the trade–offs [60]. The Multi-objective Optimization (MO) approach allows constructing the trade-off curve between objectives, called Pareto optimality front. The Pareto front represents the group of the "best solutions", which can be used as supports in the decision making process.

4.4.1 Design problem definition

Several criteria are considered as important for traction motor designs. Three objectives are defined:

(i) minimisation of mass;

(ii) minimisation of cost;

(iii) minimisation of energy consumption.

In this problem, the load cycle approach (see Section 4.2) is used to define the required torque and speed of the motor. In this case, a transient simulation must be performed. The MDF approach is applied to the optimal design problem.

The design variables are the motor geometries. They include 8 continuous variables (D.V.c.) such as the armature radius, the yoke height and 4 discrete (integer) variables (D.V.d.) e.g. the number of pole pairs and the number of conductors per slot. The optimization algorithm must allow to handle such design variables.

The solutions must also satisfy several constraints e.g. the maximal winding and PM temperatures, the maximal flux density in the yokes and tooth, maximal RMS current, etc. These constraints are imposed by the model hypothesis, the material properties and the interaction with other components. The maximal RMS voltage, material properties, and load profile i.e. the required torque and speed as a function of time (time dependent inputs) are considered as design parameters.

The multi-criteria design problem can be graphically represented in Figure 4.16. The design variables are defined on the left side and the objective functions and constraints are shown on the right side. The parameters are on the top. Table 4.8 lists all the objective functions, design variables, and constraints.

A multi-objective optimization algorithm is used to solve this multi-criteria design problem. NSGA-II[5] is one of the most efficient multi-objective evolutionary algorithms using an elitist approach. It generates a Pareto front representing the trade-offs between the different objectives. NSGA-II can handle both continuous and integer design variables [20, 82].

[5] see Section 3.1.2.2

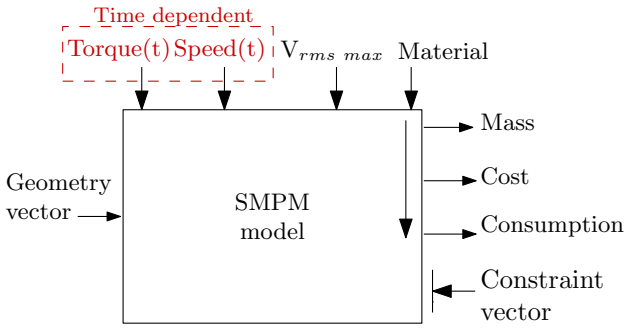


Figure 4.16: Optimization problem

In this test, the following algorithm parameters are used: population size $N = 500$, maximum number of generations $T = 100$, mutation probability 0.1, crossover probability 0.9, and the distribution index for crossover and mutation operators are $\eta_c = 20$ and $\eta_m = 20$, respectively.

4.4.2 Results

The multi-objective optimization takes 8 hours with the NSGA-II algorithm. 3D Pareto front and projections of this front in each direction are shown in Figure 4.17. There are main four solution groups. The solutions in each group share a common trend in design variable values. The solutions in Groups 1 and 2 offer almost the same design with a low conductor number, a low armature radius but a long stack length. Unlike the first two groups, Groups 3 and 4 present a high armature radius but a short stack length. Group 1 has a higher PM height value but a lower PM span coefficient and a thinner air gap compared with other groups.

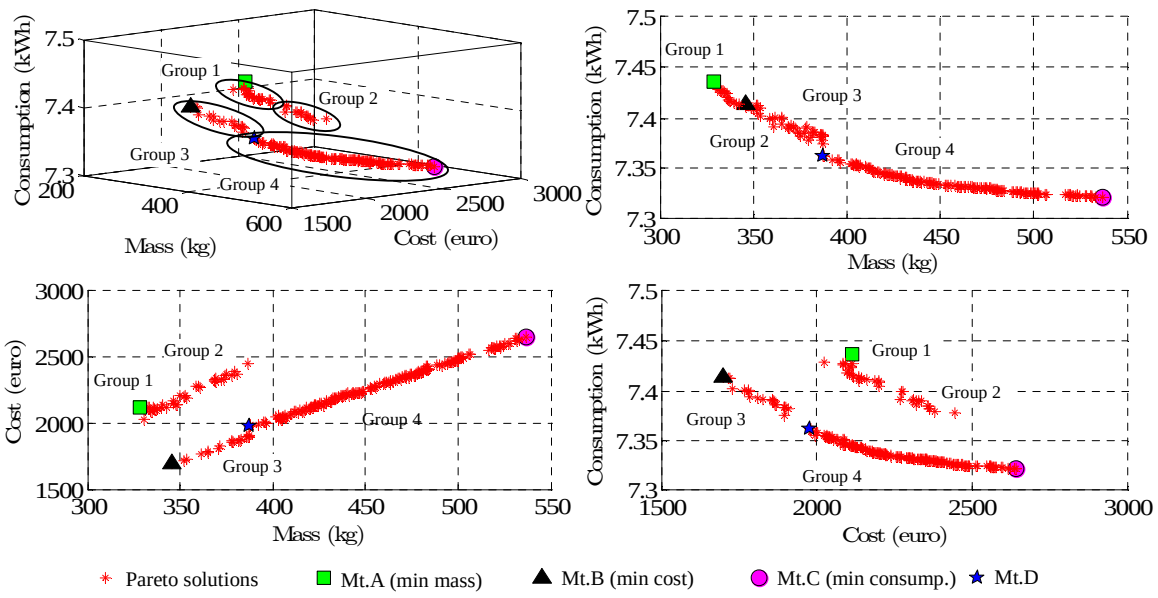
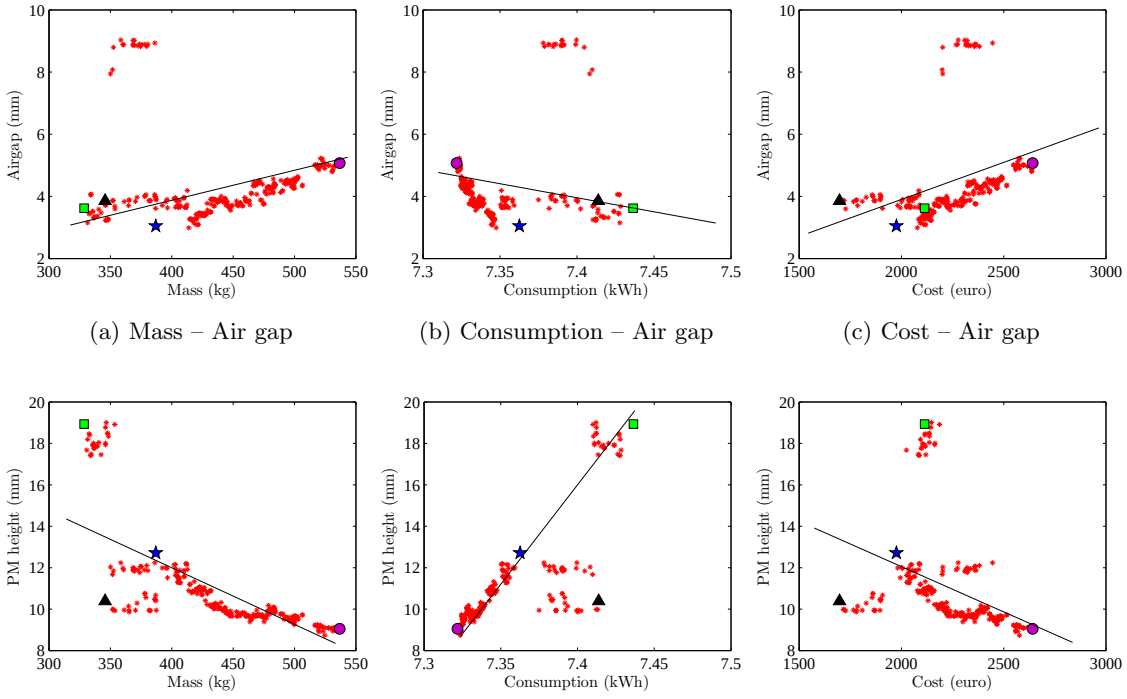


Figure 4.17: Optimization results – Pareto front

The mass and consumption criteria are clearly conflicting. A higher mass motor consumes

less energy than a motor with a low mass. For mass and cost criteria, an increase in mass and cost are in the same direction. However, for motors whose mass is lower than 400kg, a trade-off exists between Groups 2 and 3. Having the same mass, a Group 2 motor is more expensive but its energy consumption is slightly better than that of a Group 3 motor.

The evolutions of the design variable on the Pareto front are depicted in Figures 4.18. It is not obvious to clearly identify design criterion trends. They depend on the interactions between design variables as well as between other design criteria. Nevertheless, the linear global trends are presented in the figures. Without an optimal design, designers may have difficulty to find the optimal solutions. This confirms the advantage of an optimal design.



(a) Mass – Air gap (b) Consumption – Air gap (c) Cost – Air gap

(d) Mass – PM height (e) Consumption – PM height (f) Cost – PM height

Figure 4.18: Evolution of design variables on a Pareto front

In order to make a decision, four design points are selected among the optimal Pareto solutions: Motor A–minimal mass, Motor B–minimal cost, Motor C–minimal consumption, and Motor D–average mass, cost and consumption (shown in Figure 4.17). Any solution in Group 2 has been taken because the cost increase does not significantly improve the energy consumption compared to Group 3, which presents the same mass but at a lower cost.

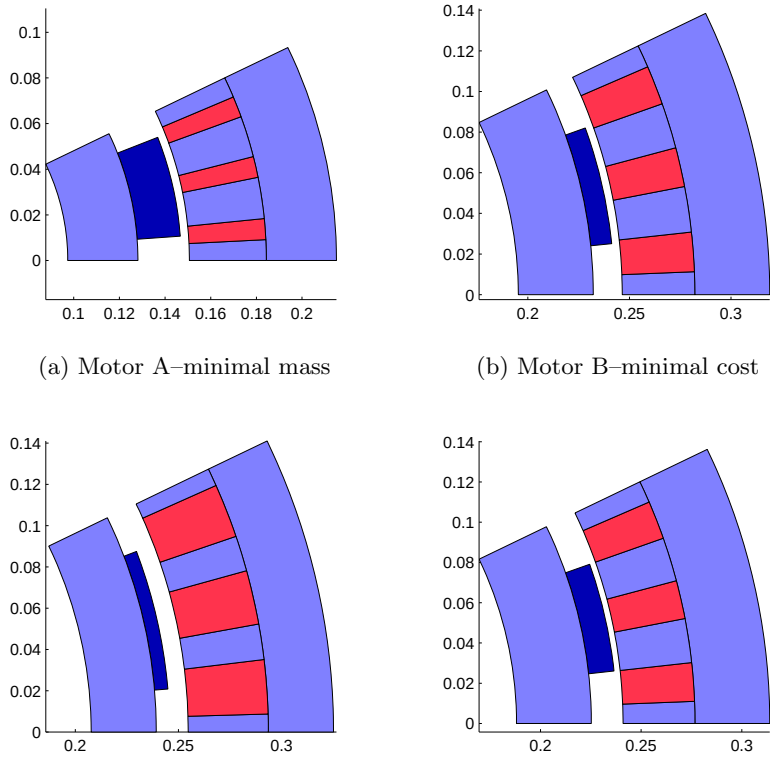
The comparison of the design variable vectors, constraint vectors and objective vectors of these solutions are shown in Table 4.8 and Figure 4.19. All solutions show 7 pole pairs and 1 slot per pole and per phase. Motor C and Motor D are almost of the same design. As Motor C is longer, its heat exchange surface is larger than that of Motor D. The temperatures can be kept lower. As a PM offers a better performance at a low temperature, the current needed to generate the required torque tends to be lower. By combining this effect with a low phase resistance at a low temperature, a high efficiency, hence a low energy consumption can be achieved.

Table 4.8: Optimization results

Symbol	Quantity	Unit	Mt.A	Mt.B	Mt.C	Mt.D	Remark	Limit
Y	Yoke height	mm	30.74	36.89	40.19	37.17	D.V.c.	[10, 50]
l_m	PM height	mm	18.93	10.39	9.05	12.71	D.V.c.	[1, 20]
g	Air gap	mm	3.62	3.86	5.07	3.05	D.V.c.	[1, 10]
d_s	Slot height	mm	33.71	35.73	36.77	35.73	D.V.c.	[20, 40]
r_a	Armature radius	mm	150.6	246.4	237.3	241.0	D.V.c.	[100, 300]
l_{stk}	Stack length	mm	265.9	117.4	202.1	139.7	D.V.c.	[100, 300]
r_{wt}	Tooth width ratio	-	0.663	0.536	0.487	0.535	D.V.c.	[0.3, 0.7]
k_p	PM span coeff.	-	0.676	0.538	0.500	0.510	D.V.c.	[0.5, 1]
p	Pole pair	-	7	7	7	7	D.V.d.	[2, 8]
N_{slot}	Slot/pole/phase	-	1	1	1	1	D.V.d.	[1, 3]
N_c	Conductor/slot	-	8	28	10	11	D.V.d.	[1, 50]
a	Parallel path	-	2	4	2	2	D.V.d.	[1, 4]
T_w	Winding temp.	°C	191.4	200	117.7	142.9	Con.	≤ 200
T_{PM}	PM temp.	°C	150	144.8	103	118.1	Con.	≤ 150
B_Y	Yoke flux density	T	1.17	1.19	0.84	1.23	Con.	≤ 1.6
B_T	Tooth flux density	T	1.60	1.48	1.30	1.59	Con.	≤ 1.6
I_{rms}	RMS current	A	266.5	286.7	286.7	282.0	Con.	≤ 300
$Pt_{dem.}$	PM demag. point	-	0	0	0	0	Con.	≤ 0
M	Mass	kg	329	346	537	387	Obj.	min
C	Cost	euro	2114	1698	2642	1976	Obj.	min
E	Consumption	kWh	7.44	7.41	7.32	7.36	Obj.	min

D.V. = Design variable; c. = continuous variable; d. = discrete variable

Obj. = Objective function; Con. = Constraint



(a) Motor A–minimal mass

(b) Motor B–minimal cost

(c) Motor C–minimal consumption

(d) Motor D–average mass, cost and consumption

Figure 4.19: Shape of optimal motors in two dimensions

Motor A and Motor B are also interesting for traction applications due to their low mass, regardless of their high energy consumption. A rather thick PM (high value of PM height) is used in Motor A. It allows using a lower current compared to other solutions. However, the cost would be higher. One can observe that the flux density in yoke has not reached the constraint limits. In a conventional motor design approach, in order to minimise the mass, the yoke height has to be minimized. This can be obtained by computing it directly on the basis of the allowable flux density. In our case, the temperature constraints are active, instead of the magnetic constraints. This shows the advantages of using of the thermal model. If the conventional approach is performed, the motor could suffer from extremely high temperatures.

The last decision was made. Motor A was chosen using mass and dimension criteria. The energy consumption is a bit higher but this is not crucial because it can be compensated by lower current, hence lower inverter losses. However the cost is much higher than the Motor B. It has to be investigated in detail, by including manufacturing cost.

4.4.3 Conclusion on the multi-criteria optimal design

This application shows an interesting approach to design a traction motor. In real-world design problems, the objective is to satisfy several design criteria. With multi-objective optimal design, multiple criteria can be defined instead of simplifying the design problem to a single-objective as in the previous sections. Designers obtain intermediate results in the form of trade-off curves. This provides useful information to make decision i.e. select the final results.

Multi-criateria optimal design of traction motors was presented. A load cycle approach was used in order to define the required torque and speed of the motor. The SMPM model takes into account transient thermal simulations. Design problems include three objective: mass, cost and energy consumption minimisations. Design variables are of two types: continuous and discrete variables. The NSGA-II optimization algorithm was used to generate a Pareto front. Results show that the mass and cost are not conflicting for big motors (high mass), but for smaller motors, some conflicts may be observed. In contrast, the mass and energy consumption are clearly conflicting. An ncrease in mass leads to a lower consumption. The influence of design variables on objective functions cannot be observed clearly. There are interactions between design variables and between objective functions. Four optimal motors were chosen from Pareto-optimal solutions. A motor was chosen to be further studied by using mass and dimension criteria, regardless of its higher energy consumption.

4.5 Use of high fidelity tools

A classical design process of electrical machines uses an analytical model to obtain preliminary designs. These results are then verified with a more accurate tool: the Finite Element Analysis (FEA). The FEA allows taking into account phenomena neglected in the analytical model and does not require strong hypotheses. Optimization techniques are usually performed using the analytical or semi-analytical model as in the previous sections. The FEA is used in refining problem such as shape design [86, 113]. Such problems are rather simple e.g. the initial design is available, only a few design variables are considered and the design problems are without or only few constraints. Design of experiment method is very suitable for such kind of problems [12, 28]. For full-range problems such as sizing problems, no initial design exists. There are several constraints. Most of time, it is multi-criteria design problem. The use of FEA in sizing problems is more or less prohibited because of a high computational time. However, due to its high accuracy, it is interesting to integrate the FEA in the optimal design process. But some precautions must be taken:

(i) the design space is usually large and the mesh size must be adapted automatically with the geometries of machines in order to avoid inappropriate meshes and to obtain accurate results;

(ii) a numerical error may cause difficulties to some optimization algorithm (e.g. gradient method);

(iii) the computational time is usually high;

(iv) other physical phenomena such as thermal and mechanical must be included in the design process.

To solve the problem of optimization time, two main research tracks are carried out by researchers: the surrogate modelling technique (see Chapter 2) and the optimization technique (see Chapter 3). Section 3.1.3 presented a promising optimization approach assisted by the surrogate model. This kind of optimization algorithm uses the surrogate model and its uncertain properties to determine where to select a new point. The design point set increases and the surrogate model are updated over optimization iterations. The mathematical example in Section 3.1.3 showed the robustness of this algorithm in the case of a noisy constraint function, which is typical in the optimal design using FEA [56, 59].

In this section, surrogate-assisted optimization techniques are applied to two SMPM motor design problems using FEA. The first problem is a single criterion optimal design problem. This problem is solved using an EGO algorithm. The results are compared with the previous results obtained by a SQP algorithm [76].

The second problem is a multi-criteria problem. A new infill criteria called Pseudo distance, presented in Section 3.1.3.6.1, allows generating a trade-off curve between two criteria: the mass and the total losses minimisations.

4.5.1 Multi-physical model

To achieve the optimal design, it is necessary to take into account different physical domains: electromagnetic, thermal, mechanical etc. In this work we consider the two most important phenomena: the electromagnetic and thermal phenomena.

The electromagnetic phenomenon is computed by a commercial FEA software: Vector Field Opera 2d. The FEA simulation uses the nonlinear characteristics of B-H curve. The simulation was performed in only one static position. The torque at this position is calculated by using the Maxwell stress integral method. An automatic mesh generation is implemented. It adapts the mesh size and the node number to the geometry of the motor. This procedure allows obtaining an appropriate mesh, hence good simulation results, for any geometry within the variation domain of the design variables [73, 76].

Figure 4.20 shows a finite element model and mesh created automatically in the FEA software. However, due to mesh changes during optimization, a numerical noise can be observed and causes problem to some optimization algorithms such as the SQP method.

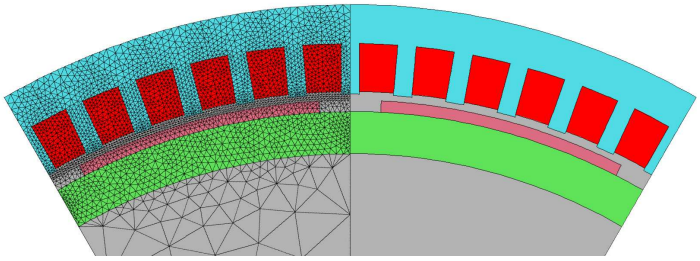


Figure 4.20: PMSM FEM model with and without mesh

The thermal phenomenon is a very important constraint in electrical machine designs. The geometries of the machine, iron losses and copper losses are injected into the thermal model in order to determine the temperature of the various components of the machine. The motor thermal model is a lumped parameter model composed of 8 nodes representing 8 motor elements [76]. The parameters are computed on the basis of the motor geometries and material properties. This thermal model allows computing the steady state temperature.

4.5.2 Single objective problem (SOP)

A single-objective optimal design problem (SOP–FEA) can be represented graphically in Figure 4.21. It consists of 7 design variables, which are the motor geometries and the current density. Some motor data including machine structure, some geometries and material properties are fixed as design parameters e.g. PM width ratio is fixed to 0.8 in order to obtain sinusoidal flux density distribution over pole. Design objective is to minimise motor mass while satisfying 6 constraints as shown in Table 4.9 ("Con." remark). These constraints express the maximum flux density in the various parts of the machine, the maximum allowable winding temperature, the torque requirement and the length of the slot. The last constraint (L_s) ensures that the automatic geometry generation in the FEA software is correct. A detailed description of the problem can be found in [73, 76].

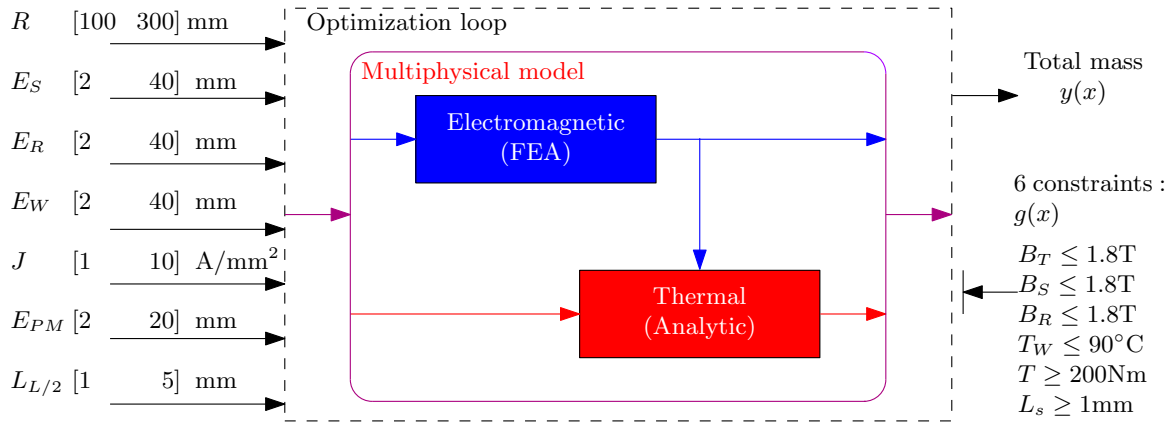


Figure 4.21: Graphical representation of the single-objective optimal design problem

Six constraints and one objective function can be considered as "expensive". Therefore, all the 7 functions are approximated by using surrogate models. LHS is used to generate 41 initial points (= 5 × 7 design variables + 6 constraints), according to (3.10). The number of initial points is important because a too small number of initial points leads to a slow convergence, while if there are too many initial points, some of them are wasted.

The SOP–FEA sizing problem was solved by using multistart SQP and EGO. The result comparison is shown in Table 4.9. The previous SQP results (from [76]) are slightly better than the EGO results as the EGO constraints tolerance is set to be larger than that of SQP. The EGO managed to find the global solution from 140 FEA evaluations. At the obtained optimum, four out of six constraints are active within the relative tolerance of 0.01. This is rather difficult for an algorithm using the surrogate model. The surrogate models of constraints have to be fairly accurate near the feasible limit. This difficulty increases as the number of active constraints is increased. Figure 4.22 shows the FEA of EGO's final result.

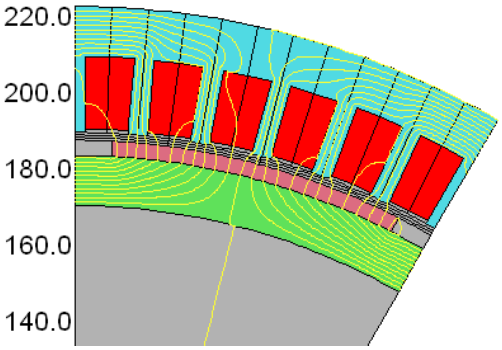


Figure 4.22: Finite Element Analysis of EGO's results

On the contrary, the SQP constraints are tightly active (tolerance of 10^{-6}). The SQP algorithm requires approximately 80 function calls per launch. But it is necessary to provide many different initial points to the SQP in order to guarantee a global solution due to the fact that the SQP is a local optimization algorithm that depends on the initial points. For this reason, its overall function evaluation is much higher than that of the EGO. The given SQP results were obtained from 4 optimizations with random initial design vectors.

Table 4.9: Optimization results for single objective problem

Symbol	Quantity	Unit	EGO result	SQP result	Remark
R	External radius	mm	221.55	231	D.V.
E_S	Stator yoke thickness	mm	13.10	13	D.V.
E_R	Rotor yoke thickness	mm	13.75	14	D.V.
E_W	Winding thickness	mm	19.03	16	D.V.
J	Current density	A/mm^2	1.94	2.16	D.V.
E_{PM}	PM thickness	mm	4.07	3.8	D.V.
$L_{T/2}$	Length of the half-tooth	mm	3.00	3.08	D.V.
M	Mass	kg	98.40	98.3	Obj.
B_T	Tooth flux density	T	1.63	1.57	Con. ≤ 1.8
B_S	Stator yoke flux density	T	1.802	1.8	Con. ≤ 1.8
B_R	Rotor yoke flux density	T	1.802	1.8	Con. ≤ 1.8
T_W	Winding temperature	°C	90.13	90	Con. ≤ 90
T	Torque	Nm	199.20	200	Con. ≥ 200
L_s	Length of the slot	mm	10.4	11.4	Con. ≥ 1

D.V. = Design variable; Obj. = Objective function; Con. = Constraint

Another great advantage of the surrogate-assisted algorithm is that surrogate models of objective function and constraints are available after the optimization process has finished. These models are accurate around the optimal point. They can be used in analysis or communication phases in order to better understand the design space and the optimization results. In this work, we use these surrogate models to perform sensitivity analysis by applying DOE [50]. The design variable domain for each variable is reduced to ±2.5% of the optimal point. The effect of each design variables on the mass, the winding temperature and the torque are studied and shown in Figure 4.23. To interpret this graph, for example, if the external radius R is changed from a low to a high level (from -2.5% to +2.5% of the optimal value) while keeping other factors unchanged, it can be then observed that the mass, the winding temperature and the torque will be increased by 3, 0.5, and 5% of the average value, respectively. This tendency is physically correct because increasing of external radius, thus armature radius, would increase the mass and the torque. Despite a higher torque, as the external radius provides a greater heat exchange surface, the winding temperature would not be affected so much. Moreover, it can be concluded from this graph that the most significant factors on mass are R, E_W, E_R, and E_S. And in the same way, change in R, J, E_W, and E_{PM} affect the torque significantly.

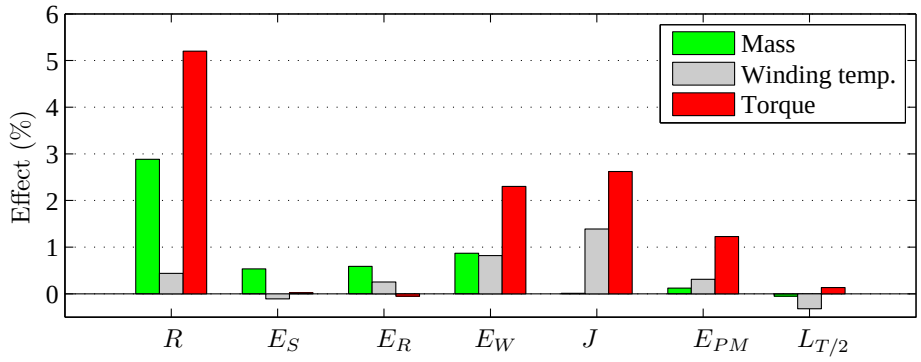


Figure 4.23: Sensitivity analysis results

4.5.3 *Multi-objective problem (MOP)*

For multi-criteria problems (MOP–FEA), two design criteria are defined: the mass and the total losses. The problem consists of 5 design variables. Two design variables are removed from the single-objective problem presented above: the rotor yoke thickness (E_R) and the length of the half-tooth ($L_{T/2}$). The rotor and the stator yoke thicknesses are equal and $L_{T/2}$ is set to $E_S/5$. The other problem and model definitions are the same as in single-objective problems.

Surrogate-assisted using pseudo distance infill criterion was applied to MOP–FEA problems. The non-dominated front shown in Figure 4.24 is obtained from only 250 FEA evaluations in total: 200 infill points and 50 initial points using LHS. The results show that the proposed algorithm is very promising because this problem is a very difficult constrained optimization problem. The design space is quite large because the base-line design does not exist and the feasible zone is very small. NSGA-II has only found infeasible solutions and cannot construct any front with the same number of function evaluations.

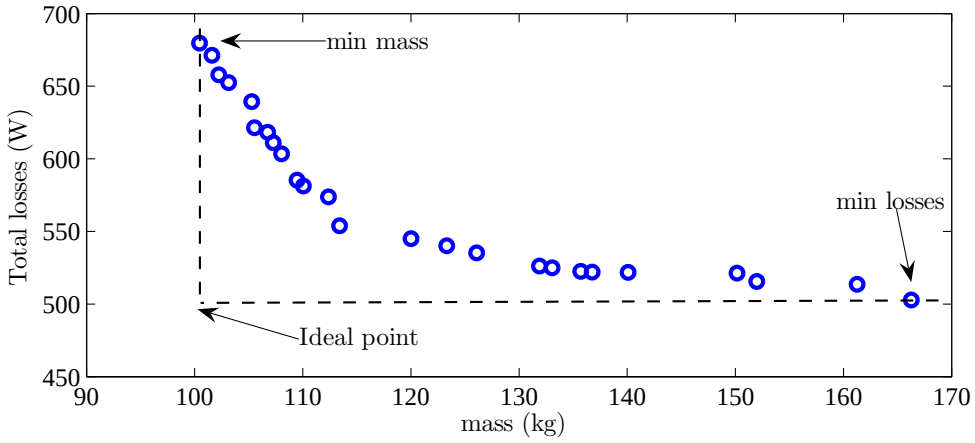
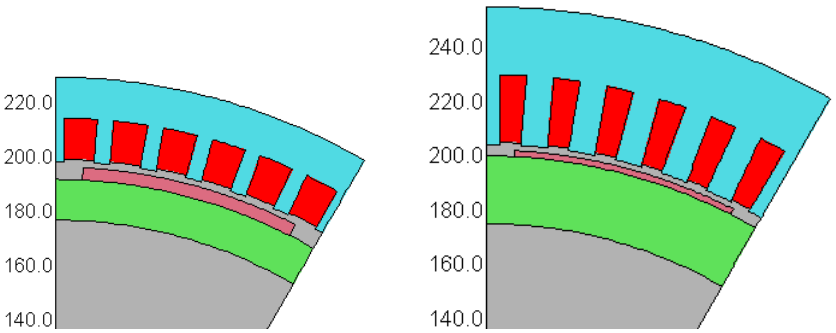


Figure 4.24: Non-dominated solutions

On an engineering point of view, this non-dominated front is very useful for design engineers. The trade-off study between the motor mass and losses can be carried out. Several interesting solutions can be investigated in detail afterwards. For example, two designs with a minimal mass and minimal losses are shown in Figure 4.25.



(a) Minimum mass (b) Minimum losses

Figure 4.25: Optimal motors for the multi-objective problem

4.5.4 *Conclusion on the optimal design using high fidelity tools*

In design of electrical machine, the FEA offers high accuracy analysis results. However, it is still difficult to use FEA directly in optimal design processes due to the fact that a high optimization time is required. Surrogate-assisted optimization algorithms allow using high fidelity tools such as FEA in an optimization process, both single-objective and multi-objective optimal designs depending on the infill criteria used. The results shows that they can cope with the noise characteristics of the FEA and multi-modal design problem. Furthermore, they requires a very small number of FEA evaluations. For the presented design problems, 140 FEAs are need for single-objective problem compared with 320 FEA evaluations by using SQP. For multi-objective problem, it requires 250 FEA calls to generate a Pareto front whereas NSGA-II cannot even find any feasible solution with the same number of function calls.

4.6 Conclusion

In this chapter, the traction motor optimal design methodology was presented. A surface-mounted permanent magnet motor has been chosen as an example of application. Two models were used: the multidisciplinary semi-analytical model and the multi-physical FEA model.

To design an electrical motor, the required torque and speed are needed as inputs of the design problem. Several approaches were compared. The design approach using a load cycle and transient thermal simulation is suitable for the traction application. The thermal behaviour of the motor can be taken into account. This allows downsizing the motor, in comparison with the traditional approach using a rated torque and a steady-state temperature. However, the optimization time is then significantly higher than in approaches taking a steady-state temperature into consideration.

A multidisciplinary motor model is usually composed of several tools or submodels (Each submodel represents a discipline.). There are interactions between these models. MDO formulations allow solving simultaneously these interaction problems and the design problem. The simplest and commonly used formulation is MDF. The interaction between disciplines is solved by an iteration method and forms a system analysis model. This system analysis is then introduced into the optimization loop. IDF and AAO formulations solve the problem in a different manner. They treat an interaction (coupling variable) as an auxiliary design variable. As a result, submodels are decoupled and can be launched in parallel. The SMPM design application shows that the optimization time can be divided by 2 to 4.

The design problem was extended to the multi-criteria problem, which is common in the electrical motor designs. The design problem is to minimise the motor mass, cost and energy consumption. The trade-off curve named Pareto front was found using NSGA-II. The Pareto front represents the group of the best solutions, which can be used as a support in the decision making. Designers must choose the final solution among the solutions suggested by the Pareto front. This is based on other criteria, which did not or cannot be expressed as objective functions in optimal design problems.

The use of FEA in motor design problem may cause convergence problems to some optimization algorithms as there is a numerical noise associated with it. Moreover, a high computation time is required. The last section presented the application of the surrogate-assisted optimization algorithm. It allows using a high fidelity tool such as FEA in the optimal design of a traction motor. The multi-physical motor model combines the electromagnetic FEA model and the analytical thermal model. The surrogate-assisted algorithm can be used in single-objective problems and multi-objective problems. It requires a small number of FEA evaluations.

Chapter 5

Optimal design of railway traction systems

A railway traction system is considered as a complex system. Its design problem is a complicate task requiring many designers and experts as well as an efficient industrial management. In the rolling stock industry, optimization techniques are used in the design process to improve performances of components. But without taking into account the interaction between subsystems, some designs can lead to suboptimal systems. To design a system or even a subsystem, it is essential to consider the system as a whole [74].

In this chapter, two application examples are presented. The first one is a tram traction system re-design problem. In this problem, only the system level is considered. The control parameters are adjusted by the optimization method in order to satisfy the performances and component specifications. The traction system design tool of Alstom is used. The surrogate modelling approach, as well as a hybrid optimization algorithm, are applied.

The second application problem covers the tram traction system as well as its components. The goal is to design a standard tram traction system. In the application example, a simplified problem is studied. The decomposition approach and the Target Cascading method are used to solve the problem. The problem is decoupled into 3 linked subproblems. The coordination and optimization are performed by the Target Cascading method presented previously in Chapter 3.

5.1 Tram traction system re-design problem

As presented in Section 1.2.1 Alstom Transport has developed the standard traction system for tram applications in order to minimise cost and delivery time. Each applicative project has different customer specifications. These components are chosen and adapted according to these specifications in order to satisfy the global performance of trams operating on a predefined track.

In this study, the goal of Alstom is to increase the passenger capacity of the tram. The traction system would be re-designed due to the increase in global weight. The hardware re-design is discarded because the cost is unacceptable. For this reason, only control parameters can be modified. The solution has to meet not only the hardware constraints but also the customers' specifications.

5.1.1 Problem description

Due to a change in passenger capacity, the vehicle weight is increased. In the normal operating mode, the simulation results show that the tram can be operated perfectly without any problem. According to the customer contract specifications, in the faulty operating mode (one traction box is defective) the tram must be able to continue the service until the end of the trip with the remaining 50% motorization (3 motors) with the same performance as in the normal operating mode, except that the maximum speed had been reduced. The simulation results show that the required performance can be achieved [61]. However, some problems are observed:

(i) The temperature of IGBT (Insulate Gate Bipolar Transistor) modules used in VSIs was over the limit of 125°C defined by the semi-conductor manufacturer. The normal operation of IGBT cannot be guaranteed beyond this temperature limit. Figure 5.1 shows the simulation results obtained by using the simulation tools i.e. CITHEL. The maximum temperature reachs 170°C;

(ii) Due to the increased torque, the line current did not respect the line filter specifications.

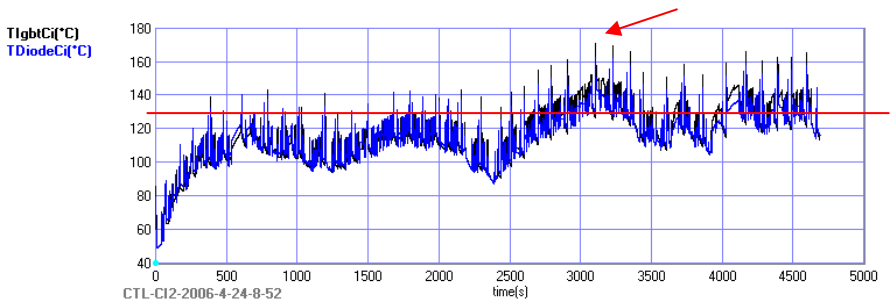


Figure 5.1: IGBT temperature – initial design, 50% motorization mode

5.1.2 Modelling of traction systems

An overview of railway traction systems has been given in Section 1.2.1. A tram is usually equipped with a DC traction system. For low power application, the DC power supply allows decreasing the cost of the traction system, as a transformer and a rectifier are not needed.

A simplified schema of a tram traction system is shown in Figure 5.2. A pantograph collects electricity from the direct current (DC) overhead supply with a nominal voltage of 750V. Electricity passes through a line filter and then through a voltage source inverter (VSI) and finally traction motors. Normally, a regenerative braking is used. However, for security reasons, a braking chopper with a braking resistance is connected to a DC bus, for example in the case of voltage saturation of the DC line.

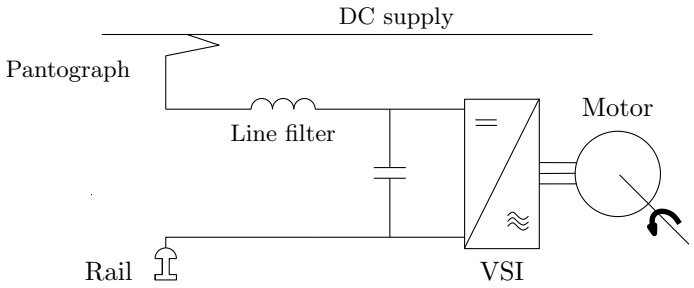


Figure 5.2: Simplified power schema of a tram traction system

In reality, the studied tram is composed of 2 traction boxes. Each traction box contains a line filter, 2 braking choppers and 3 VSIs, which can drive one motor each, and other electrical equipments e.g. breakers, sensors etc.

The traction system model is an in-house code of Alstom Transport–CITHEL. The model is the integration of the train system, the traction system and its components. These models are described with different levels of complexity and form a system model. It allows simulating the behaviour of rolling stocks operating on tracks e.g. a round-trip of tram (mission of more than 1 hour). It requires a lot of knowledge:

(i) the rolling stock characteristics e.g. the aerodynamic properties, the component parameters.

(ii) the track characteristics i.e. the track curve, the speed limit, the stop station, the stop time, the altitude level.

(iii) the driving performance requirements e.g. the maximal acceleration, the deceleration.

The simulation takes into account the transient thermal behaviour of components such as the motor and the inverter. The electrical computation only takes the steady state into consideration. At each time step, it computes the global dynamic values of the train (speed, acceleration, etc.) and the local electrical, thermal, and mechanical values of the components (line current, IGBT temperature, output torque, etc.).

The main role of CITHEL is to define the architecture of traction systems (number of motors, inverters, etc.) and to verify the operation of a train on a track based on the components already developed. The component properties are defined by rather global parameters,

for example, the motor model is a single-phase equivalent circuit model and is not linked to the geometries of the motor. The simulation must be performed in many operating scenarios in order to satisfy the specification requirements.

Usually, several missions (different tracks) and several operation modes including normal operation and various faulty situations, have to be taken into account. However in this problem, the track profile of only one customer and a 50% motorization faulty mode are taken into consideration. Two driving strategies are tested: with and without regenerative braking. This allows verifying the maximum current passing through the line filter and the braking resistance.

The surrogate model approach [61] is used to decrease the computation time. Figure 5.3 depicts the main idea of a surrogate model approach. The surrogate model uses information from the high fidelity model and provides a cheap-to-evaluate model estimating the high fidelity model. It replaces the high fidelity model in the optimal design process. In this work, the Kriging surrogate model [63] is constructed independently for each output.

The traction system design tool is then replaced by the surrogate models during the optimization process. The computation time is then decreased from more than one hour to less than one second per design evaluation.

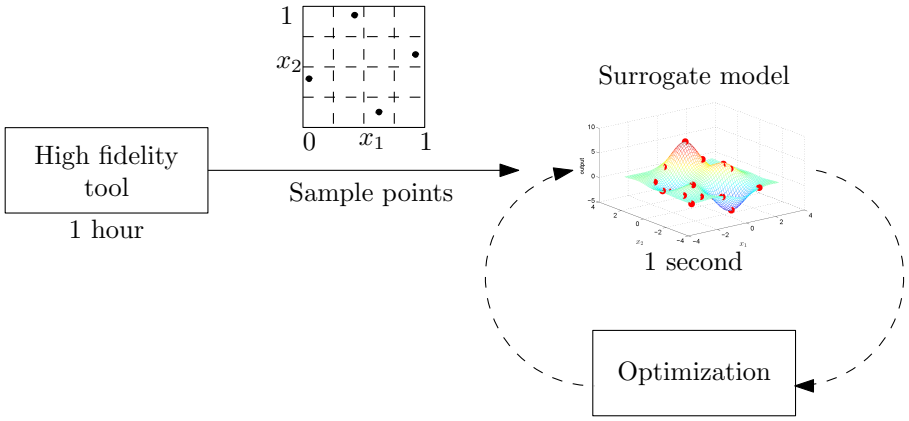


Figure 5.3: Surrogate modelling approach

In this example, 108 sample points have been computed with the high fidelity tool i.e. CITHEL. These sample points are used to build 12 surrogate models of CITHEL outputs.

5.1.3 *Optimal design problem*

The optimal design methodology is applied to this problem. The design criteria is to minimise the maximum IGBT junction temperature (T_j) during the round-trip of the tram in faulty operating mode on the customer route profile.

Four control parameters are selected (see Table 5.1). The first two design variables are the parameters for the Pulse Width Modulation (PWM) used in VSI. Two Asynchronous Space Vector (ASV) schemes [38] with different switching frequencies are used in different speed zone.

Table 5.2 and Figure 5.4 shows the PWM schemes for different VSI output frequencies

Table 5.1: List of design variables

Number	Design variable (unit)	Symbol	Limit
1	VSI switching frequency (Hz)	f_{pwm_low}	[300, 900]
2	Min. ratio $\frac{\text{VSI switching frequency}}{\text{VSI output frequency}}$ (-)	n	[10, 20]
3	Maximum torque limit (Nm)	T	[800, 840]
4	Maximum power limit (kW)	P	[70, 90]

(f_{vsi}). For VSI output frequencies from zero to $f_{pwm_{low}}/n$, the switching frequency is set to $f_{pwm_{low}}$ and from $f_{pwm_{low}}/n$ to $f_{vsi_{max}}$, the switching frequency is $f_{pwm_{high}}$. It should be noted that $f_{vsi_{max}}$ is the maximum frequency generated by the VSI and $f_{pwm_{high}}$ is a constant value.

Table 5.2: PWM scheme parameters

Output frequency (Hz)	Switching frequency (Hz)
0 to $f_{pwm_{low}}/n$	$f_{pwm_{low}}$
$f_{pwm_{low}}/n$ to $f_{vsi_{max}}$	$f_{pwm_{high}}$

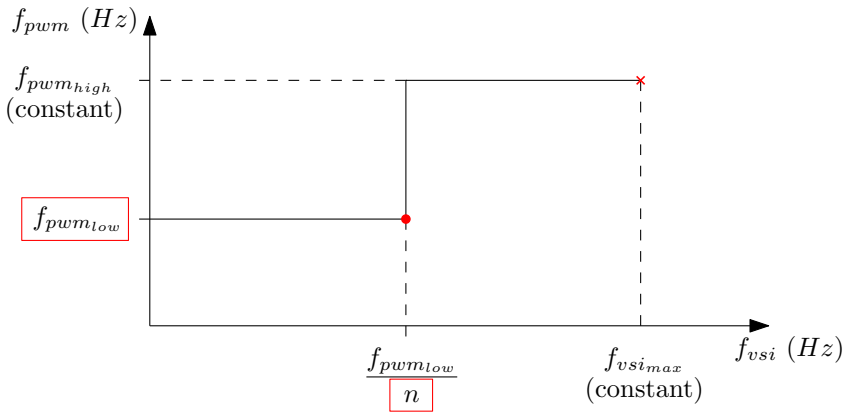


Figure 5.4: PWM scheme

These schemes allow the VSI to operate with a low switching frequency in low output frequency zones (low tram speeds) where the train driver always needs a high torque to accelerate the tram, hence high current, high IGBT losses and a high IGBT temperature. The minimum ratio between f_{pwm} and f_{vsi} is determined to respect some design rules for this application.

The last two design variables are the maximum torque limit and the maximum power limit of the motor. These limitations affect the IGBT temperature in 2 conflicting approaches:

(i) The low torque requires a low current, hence a low IGBT temperature.

(ii) The train driver can reach the maximum speed easier and faster with a high torque and a high power, but the high torque requirement applies only for a short period. As

the thermal response time is slow, it is possible that the IGBT temperature would not increase significantly.

Due to the component specifications and performance requirements, 11 inequality constraints given in Table 5.3 are required.

The summary of the optimization problem can be represented graphically as shown in Figure 5.5. The left side shows 4 design variables. The objective function and constraints are on the right side.

Table 5.3: List of constraints

Number	Inequality constraint	Unit	Limit
1	Inductor current–traction mode	A	≤ 675
2	Inductor current–regenerative braking mode	A	≤ 675
3	Inductor thermal current–braking resistor mode	A	≤ 300
4	Inductor thermal current–regenerative braking mode	A	≤ 300
5	Capacitor thermal current–braking resistor mode	A	≤ 240
6	Capacitor thermal current–regenerative braking mode	A	≤ 240
7	Peak phase current	A	≤ 600
8	Motor rotor temperature	°C	≤ 300
9	Motor winding temperature	°C	≤ 250
10	Motor iron temperature	°C	≤ 250
11	Round-trip time	s	≤ 5500

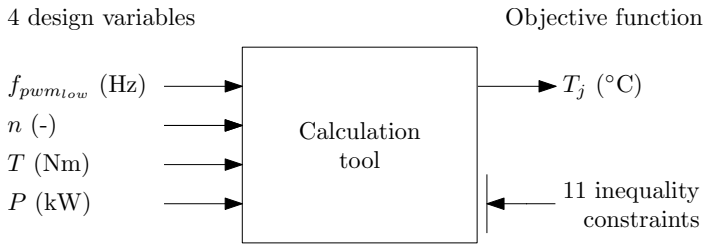


Figure 5.5: Optimization problem

The Kriging surrogate modelling approach presented in Section 2.2.3.3 is applied. Sample points are firstly selected. In this application, nonlinear relationships between design variables, constraints and the objective can be observed. At least three level-factorial design seem to be necessary in order to capture these nonlinear relationships. A grid design with 4 levels for $f_{pwm_{low}}$ and 3 levels for other design variables is used. The response of the objective and 11 constraints are obtained by evaluating CITHEL at 108 sample points allowing to construct 12 Kriging surrogate models (eleven constraints and one objective function). Figure 5.6 shows the surrogate model of the objective function. Finally, the optimizations are carried out with these 12 surrogate models.

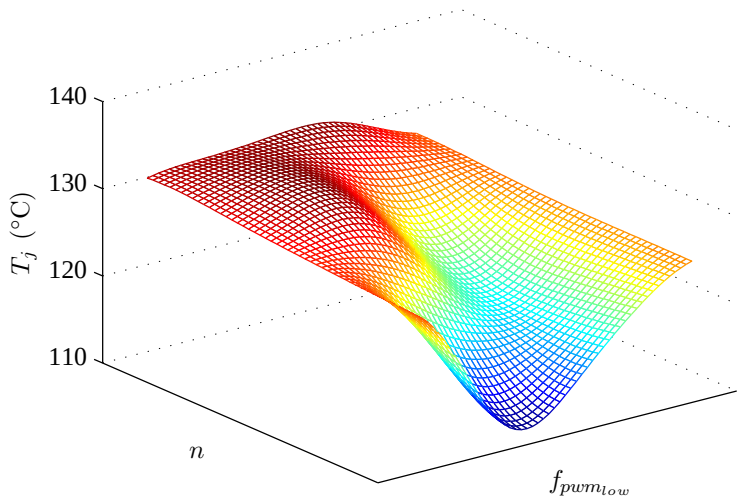


Figure 5.6: Surrogate model of the objective function

5.1.4 Comparative results

Three single-objective optimization algorithms are tested: SQP, GA and hybrid GA and SQP. The initial point set (SQP) and the initial population (GA, Hybrid) can influence the behaviour of algorithms. Three initial point sampling methods are studied: Monte Carlo Sampling, Grid Sampling and Latin Hypercube Sampling. In order to compare the results given by different studies, the optimal results found by a reference optimization using SQP with 1000 initial point multistart is chosen as a reference global optimum ($T_j = 114.25$°C).

5.1.4.1 SQP optimization results

A main drawback of the SQP is that it can be trapped in local optima if the initial point is far from the global solution. Therefore, a multistart strategy is used. The SQP is restarted with a set of different initial points in order to increase the probability to find the global optimum. Three sets of 108 initial points are chosen by using (i) Monte Carlo Sampling (MCS), (ii) grid sampling (Grid) and (iii) Latin Hypercube Sampling (LHS).

Table 5.4 compares the Percentage of Convergence (PC) towards the reference optimum point with a an Euclidean distance lower than 10^{-3} and the average number of evaluations for the three SQP approaches starting from 108 initial points. The computational time is given in terms of average number of evaluations. For example, the computation time of the SQP with a LHS approach is equal to $121 \times 0.08\text{s} = 9.68\text{s}$, where 0.08s is the average time needed to evaluate the objective and constraint functions and 121 is the number of SQP's evaluations. In this comparison, all design variables are scaled to [0, 1]. The LHS performs slightly better than other sampling approaches. As shown in Table 5.4, the SQP converged at an average of one hundred evaluations of the objective function. However, whatever the choice of the initial point set is, with any ten points, the SQP can converge at least once (PC=15.74%) or at most twice (PC=23.15%) towards the global solution because the percentage of convergence

is between 15.74% and 23.15%.

Table 5.4: SQP optimization results

	MCS	Grid	LHS
PC precision 10^{-3} (%)	20.37	15.74	23.15
Average number of evaluation	116	113	121

As shown in Figure 5.7, the studied problem is a hard optimization problem with several local optima. Using 108 initial points chosen randomly. Only 20% of the 108 initial points chosen randomly allow the SQP method to converge towards the reference optimum point (first bar with temperature=114.2514°C). It means that approximately 80% of the initial points lead to the local optima. Therefore, GA algorithms are performed in order to overcome this disadvantage.

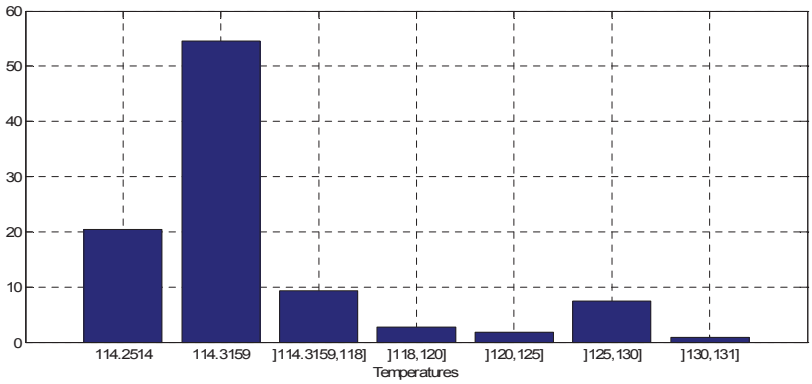


Figure 5.7: Multimodal problem

5.1.4.2 GA optimization results

Like the SQP, the GA is an initial population dependent algorithm. To compare rigorously the SQP results with those of GA, the three sets of SQP initial points, presented above, are used as the GA initial population containing 108 individuals each. The GA parameters have the following values: population size N=108, maximum number of generations T=100, crossover probability 0.8, rank fitness scaling, scattered crossover, and stochastic uniform selection operators. If the objective value is not improved over 10 generations or if the objective function tolerance is of less than 10^{-6}, the algorithm stops.

In Table 5.5, it can be observed that the GA results are less accurate than the SQP results. The GA found solutions, which are very close to the global optimum (with an accuracy of 10^{-1}). However, the percentages of convergence for a precision of 10^{-3} are of zero for all initial populations. To improve the results, the hybrid method has been used.

Table 5.5: GA optimization results

	MCS	Grid	LHS
PC precision 10^{-1} (%)	1	19	16
Average number of evaluations	5987	5025	9906

5.1.4.3 Hybrid method (GA & SQP) results

The GA parameters are tuned like previously. However, in order to reduce the number of evaluations, the GA algorithm is stopped prematurely. For this purpose, the number of generations has been reduced to 5 generations. It is the sole stopping criterion. Then the solutions found by the GA are used as initial points for the SQP algorithm. Table 5.6 shows the performance of the hybrid method. The results are more accurate than the SQP and GA alone, while fewer evaluations are required. With the GA and the grid sampling approach, the solutions are very close to the global one. Each solution is used by the SQP as an initial point in order to overcome the local optima. This explains the highest percentage of convergence of the hybrid method with the grid sampling approach.

Table 5.6: Hybrid method optimization results

	MCS	Grid	LHS
PC precision 10^{-3} (%)	11	70	32
Average number of evaluation	3380	3380	3323

5.1.4.4 Summary

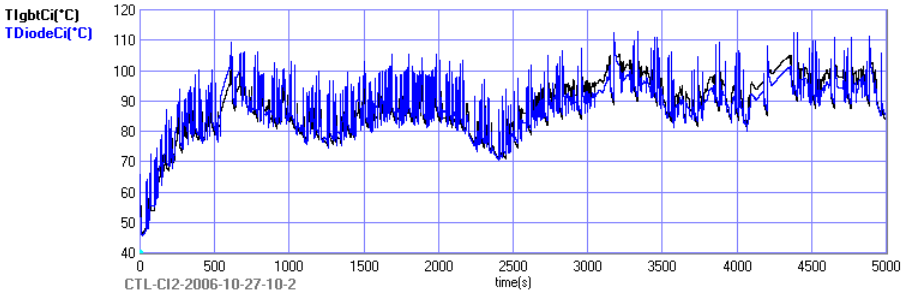
Three optimization algorithms have been tested. The SQP with multistart and the GA can be considered as effective. Due to the lack of precision of the GA and the risk to be trapped in the local optima of the SQP, a hybrid method combining the SQP and the GA with three different initial populations has been applied. The hybrid method using the grid sampling initial population gives a greater percentage of convergence and a lower computational time than the Monte Carlo and Latin hypercube sampling approaches.

5.1.5 Traction system design results

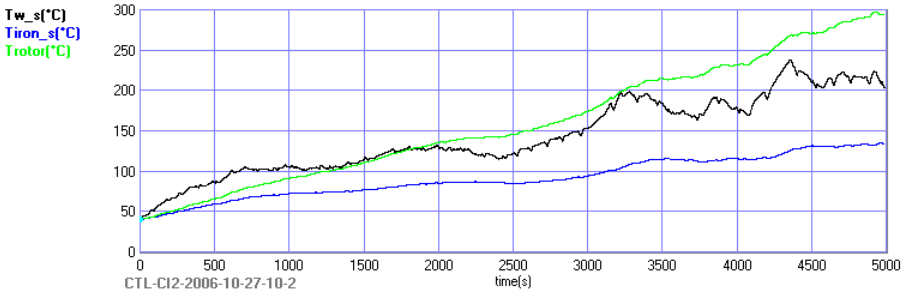
The final results are presented in this section. The minimum IGBT temperature is found at 114.25°C on the basis of the Kriging surrogate model. This design point is verified against CITHEL, which gives 113.05°C. The relative error of the surrogate model at the optimal point is of 1%. At this point, all the constraints are satisfied. The rotor temperature constraint is active at 300°C.

Figures 5.8a and 5.8b show the evolution of the IGBT and motor temperatures. These results are obtained from CITHEL. It can be observed that the IGBT temperature is decreased significantly compared to the initial case (170°C in Figure 5.1). In order to decrease

inverter losses (and the IGBT temperature), the PWM switching frequency is reduced. As a result, the motor harmonic currents are increased. This causes high losses in the motor and hence a high motor temperature.



(a) IGBT temperature



(b) Motor temperature

Figure 5.8: Simulation results at optimal solution

The Design of Experiments methodology (DOE) [29] is used to perform the sensitivity analysis. The objective function is observed when one or more input values are changed. A variation of 5% is applied to the optimal solution as shown in Table 5.7. The result of the sensitivity analysis is shown in Figure 5.9. Its intepretation is that the motor torque (T) is the most significant design variable to the objective function. The increase in the motor torque leads to the increase in the IGBT temperature. Other design variables are not considered as significant factors because their effects are lower than 95% of the confidential level.

Table 5.7: Variation of design variables for sensitivity analysis

Number	Symbol	Lower level (-5%)	Optimal value	Upper level (+5%)
1	f_{pwm_low}	437.03	460.03	483.03
2	n	9.5	10	10.5
3	T	791.13	832.77	874.40
4	P	77.49	81.57	85.65

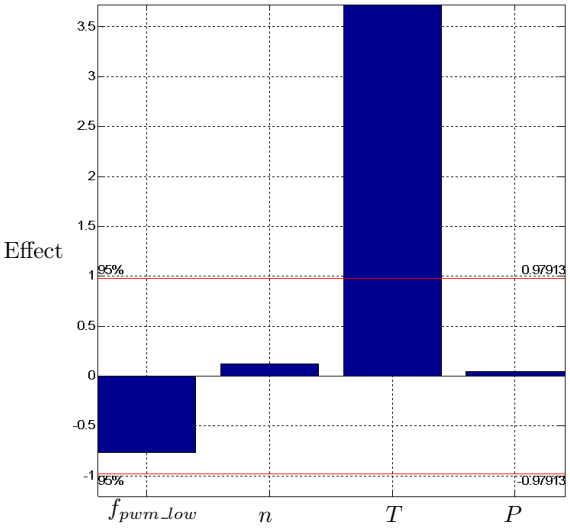


Figure 5.9: Sensitivity analysis

5.1.6 *Conclusion on the re-design problem*

This application shows the advantages of the optimal design using the surrogate model. The computation time is significantly decreased with this approach. Three optimization algorithms and initial point sampling methods are also studied. The GA and SQP hybrid algorithm offers a compromise between the global search capability and the optimization speed.

The optimal result of the traction design problem allows minimising the IGBT temperature while satisfying a number of constraints such as component specifications and tram performances. According to the simulation using Alstom's traction design tools, the tram can be perfectly operated in both normal and faulty modes.

In this application, the optimal design is applied in the late design phase. Therefore, only a few design variables can be taken into consideration. In the following section, the optimal design and the multilevel approach are used in the design of a standard tram traction system and its components. The problem is in the early design phase during which the degree of freedom is higher and many design variables are accessible.

5.2 Multilevel tram traction system design

This section demonstrates an application of the decomposition approach solved with the Target Cascading (TC) method described in Section 3.2.2. A standard tram traction system design problem consists of a traction system and a component design. According to the decomposition approach presented in Section 1.2.2, a train design problem is decoupled into several linked design problems as shown in Figure 5.10. These problems form a hierarchical design problem. The Target Cascading method allows coordinating and solving such optimal design problems. The tram traction system and its components are designed simultaneously.

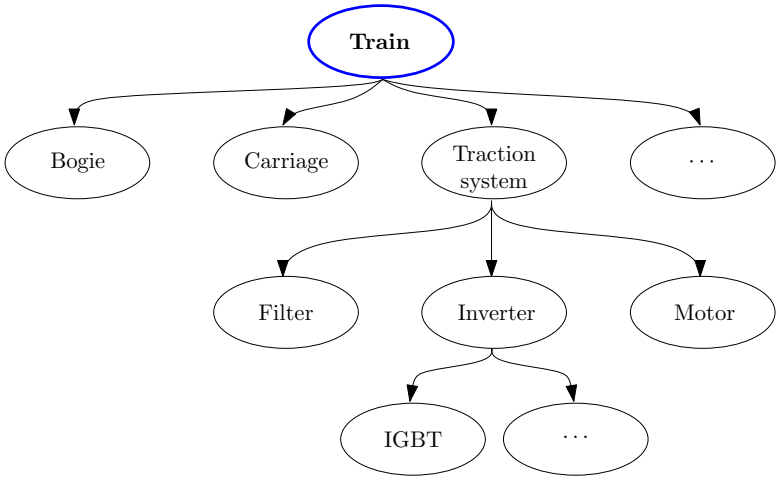


Figure 5.10: Object-based train decomposition

5.2.1 *Multilevel optimal design problem*

The tram traction system design problem is a hierarchical design problem. In this research, a simplified traction system design problem is investigated. The system level describes the whole traction system. Only two subsystems are considered: (i) the heat sink of traction inverter and (ii) the Permanent Magnet Motor (PMM) (Note that the modelling of each system is presented in the following section.). Therefore, three subproblems are defined: one system level problem and 2 subsystem level problems as shown in Figure 5.11. The system level design criteria are to achieve the target of total energy consumption (E) and total mass (M_t). Each subproblem has its own design variables and constraints. The system level is linked to the subsystem level via subtargets and subresponses. The system level problem (P_{sys}) consists of 9 design variables in total:

- 2 system level local variables: pulse width modulation (PWM) frequencies (F_{pwm_1} and F_{pwm_2}),
- 3 sublevel targets for the subsystem heat sink: the thermal parameters (the R^U_{th} thermal resistance, the C^U_{th} thermal capacitance) and the M^U_{HS} heat sink mass,
- 4 sublevel targets for the subsystem PMM: the electric circuit parameters (the R^U_s phase resistance, the L^U_s inductance, the Φ^U_s flux) and the M^U_{mot} motor mass.

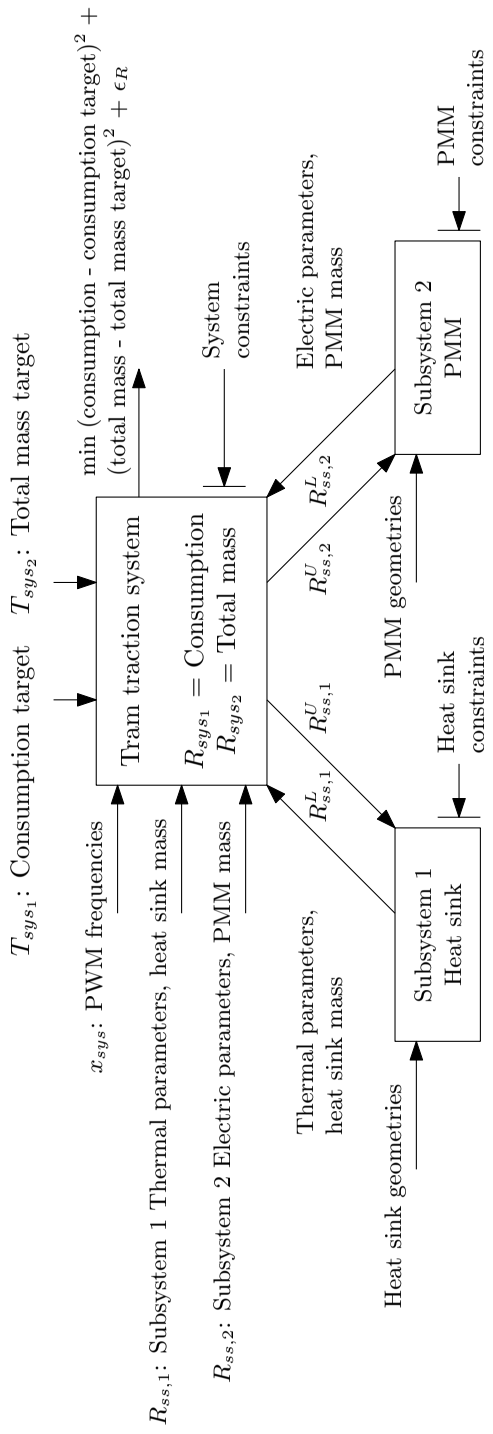


Figure 5.11: Tram traction system optimization problem

The system level optimization is a target setting problem according to the TC method. The above mentioned design variable are sought while minimising the discrepancy between targets and computed responses. The first system level optimization allows finding the optimal design vector. This design vector includes the "optimal" value of the sublevel targets for the subsystem heat sink and the PMM, which are then cascaded to subsystem optimization problems.

The subsystem heat sink problem ($P_{sub_{HS}}$) has 6 design variables (heat sink geometries) and the subsystem PMM ($P_{sub_{PMM}}$) has 8 design variables (motor geometries). They try to match the cascaded targets with the computed sublevel responses. As each subproblem is a constrained problem, the subsystem optimization may not be successful to match these targets and return the best achieved sublevel responses (e.g. R^L_{th}, C^L_{th} and M^U_{HS} for the subsystem heat sink problem) to the system level. At this stage, the TC has finished the first iteration. Starting the second iteration, the TC alternates with the system optimization problem, which, again, solves the target setting problem as well as doing its best to minimise discrepancy between sublevel targets and sublevel responses. The process continues until the convergence criteria are met i.e. the sublevel target at the system level is equal to the sublevel response computed at the subsystem level. Table 5.8 summarises the three optimization subproblems. For each subproblem, the design variables, targets, responses and objective function to be minimised are described.

Table 5.8: Summary of 3 optimal design subproblems

Problem	Find	Response	Target	Minimisation
System P_{sys}	**Local design variable:** F_{pwm_1} and F_{pwm_2} **Heat sink subproblem target:** R^U_{th}, C^U_{th} and M^U_{HS} **PMM subproblem target:** R^U_s, L^U_s, Φ^U_s and M^U_{mot}	E^R, M^R	E^T, M^T	$\left(E^R - E^T\right)^2 +$ $\left(M^R - M^T\right)^2 + \epsilon_R$
Heat sink $P_{sub_{HS}}$	**Local design variable:** 6 Geometries	R^L_{th}, C^L_{th}, M^L_{HS}	R^U_{th}, C^U_{th}, M^U_{HS}	$\left(R^L_{th} - R^U_{th}\right)^2 +$ $\left(C^L_{th} - C^U_{th}\right)^2 +$ $\left(M^L_{HS} - M^U_{HS}\right)^2$
PMM $P_{sub_{PMM}}$	**Local design variable:** 8 Geometries	R^L_s, L^L_s, Φ^L_s, M^L_{mot}	R^U_s, L^U_s, Φ^U_s, M^U_{mot}	$\left(R^L_s - R^U_s\right)^2 +$ $\left(L^L_s - L^U_s\right)^2 +$ $\left(\Phi^L_s - \Phi^U_s\right)^2 +$ $\left(M^L_{mot} - M^U_{mot}\right)^2$

5.2.2 *Tram traction system modelling*

For a hierarchical design problem, several models are needed. The main difficulty is to choose the models representing the system and components with suitable details. The system model can be less detailed i.e. presented in a more global point of view, the component models at the subsystem level should typically represent the components in detail and, therefore, provide a higher degree of freedom to the designer. These models are used in a complimentary way to design the traction system.

5.2.2.1 Modelling of traction systems

The traction system model is presented in Section 5.1.2. In this design problem, a standard tram traction system is considered. In order to ensure the performances of the system, two tracks are chosen as references. The traction system must be optimised for these scenarios:

- 2 reference tracks.
- 3 operation modes: normal operation, 75% motorization and 50% motorization faulty modes.
- 2 driving strategies: with and without regenerative braking.

This yields 12 different simulations in total for each design evaluation. These computations take more than one hour. Among 12 simulations, only the extreme value is used as a constraint or objective. For example, the IGBT temperature is considered as a constraint in the optimization problem. The calculation results obtained from CITHEL are a profile of the IGBT temperature as function of time. Firstly, the maximal temperature is selected. For one design vector, 12 values from 12 CITHEL simulation cases are obtained. Then, the maximum among these 12 values is considered as the output provided to the optimization process.

Similary to the previous example, the Kriging surrogate modelling approach is used to reduce the computation time. The Kriging surrogate model is constructed independently for each output. According to the design problem, there are 5 surrogate models of 7 inputs. Figure 5.12 depicts the system level model. It contains an analytic equation for total mass computation.

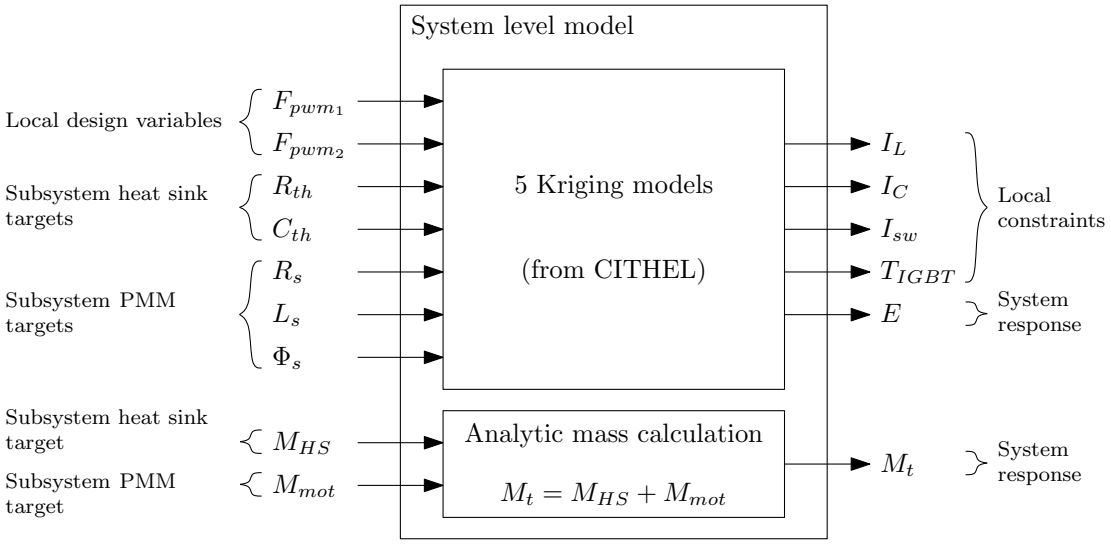


Figure 5.12: System level model

5.2.2.2 Modelling of heat sinks

The inverter is equipped with a cooling system to dissipate the heat losses generated by IGBTs. In tram application, the inverter cooling system is typically of air type. An air cooling system is more robust, cheaper and simpler than a water cooling system. Even with a fan forced-air convection cooling, its performance is not as good as the water cooling type but it is usually sufficient for this application.

The heat sink model aims at computing thermal resistance, thermal capacitance and mass from the geometry of the heat sink shown in Figure 5.13. These parameters are used in CITHEL to simulate the thermal behaviour of an inverter cooling system. The model is based on the theory of a thin plate heat conduction and forced-air convection through fins due to the fan equipped in the inverter cooling system [46, 111]. This model includes 72 equations in total with several if-else conditions according to experimental coefficients.

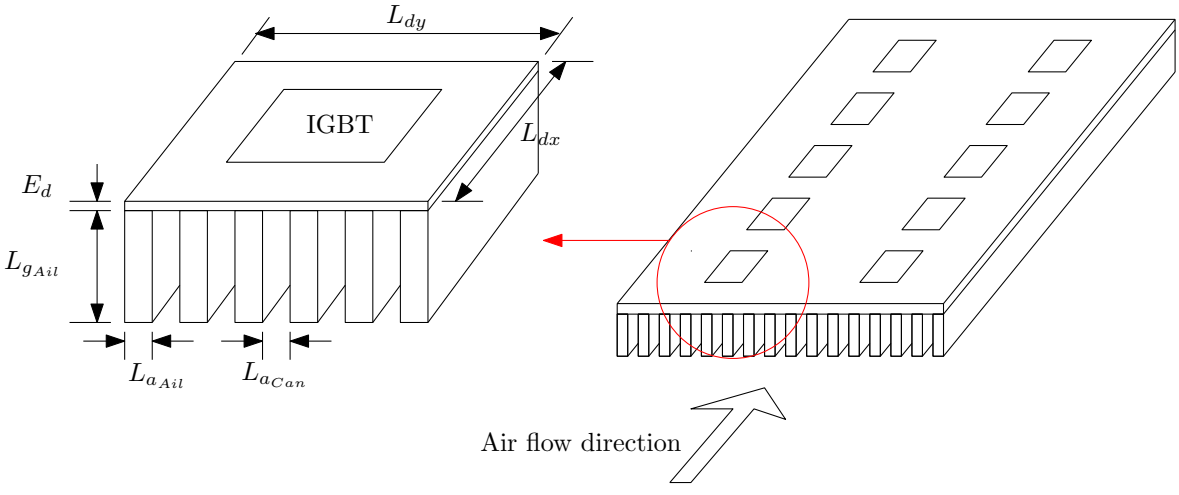


Figure 5.13: heat sink geometries

The thermal resistance, thermal capacitance and mass are used in the subproblem objective function, as described in Table 5.8. Three constraints are defined: pressure loss, air speed between two fins and minimum number of fins. Inputs and outputs of the heat sink model are shown in Figure 5.14.

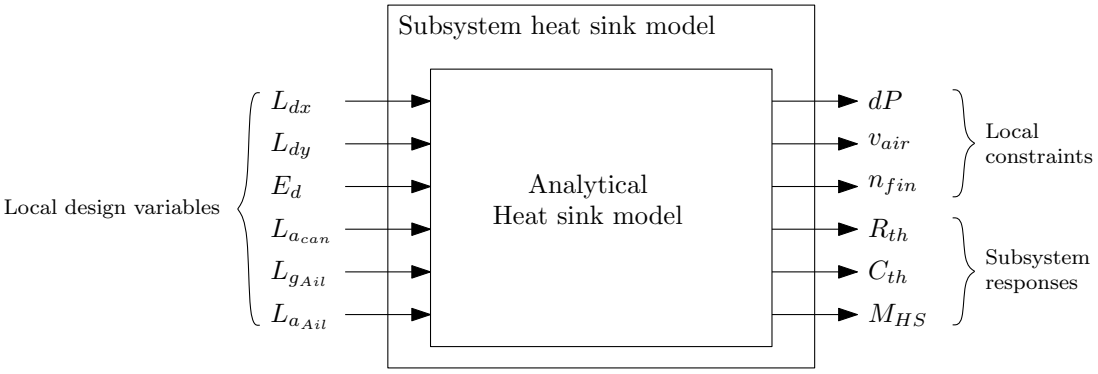


Figure 5.14: Subsystem heat sink model

5.2.2.3 Modelling of traction motors

The traction motor model is given in Appendix B. The rated torque and base speed are defined as performance specifications (see Section 4.2). The rated torque is set to 75% of the maximal torque computed in the system level by the traction system design tool. According to the PMM model, the steady-state thermal module is used i.e. the motor must be able to operate at this base point permanently without any excessive temperature.

Figure 5.15 shows inputs and outputs of the traction motor model. Local design variables are the dimensions of the motor. Constraints are defined in order to satisfy the following requirements:

- Model hypothesis: maximum flux density (linear B-H characteristic)
- Material properties: maximum winding and PM temperatures
- Failure protection: demagnetization of PM
- Mechanical interface: external diameter and length
- Inverter interface: maximum current and voltage

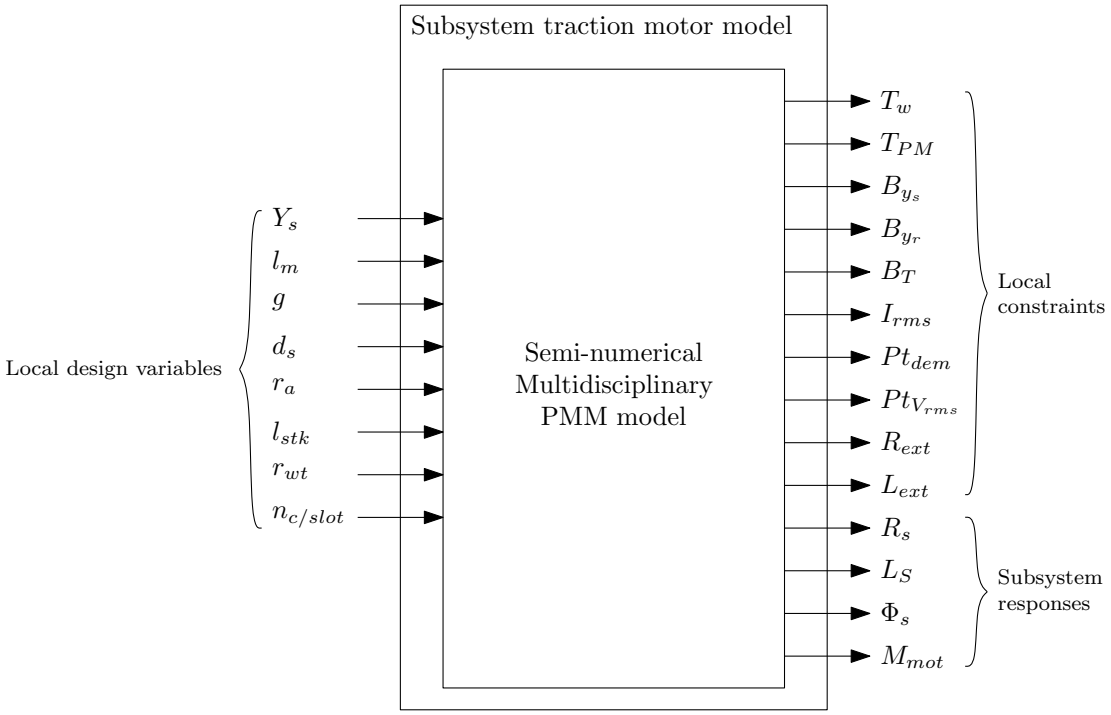


Figure 5.15: Subsystem traction motor model

5.2.3 Results

In this section, the TC convergence property is discussed firstly. After that, the optimal solution is presented.

The TC solves the target setting problem. The system targets and the optimal responses are given in Table 5.9. It can be observed that there are small discrepancies between the targets and the responses (approximately 1%). In this work, attainable targets are defined. Therefore, the original version of the TC (described in Section 3.2.2) works quite well. These targets are given empirically after encountering several non-convergence problems due to the fact that the original TC might suffer from the convergence difficulty when solving the same design problem with non-attainable system target (or minimisation problem). We propose an improved version of TC described in [85]. The similar traction design problem is solved with this improved algorithm in the case of a non-attainable target. The reader should refer to [5] for details of proposed algorithm.

Table 5.9: Traction system level targets and responses

Quantity	Unit	Target	Final results
Total mass	kg	420	420.21
Consumption	kWh	190	191.93

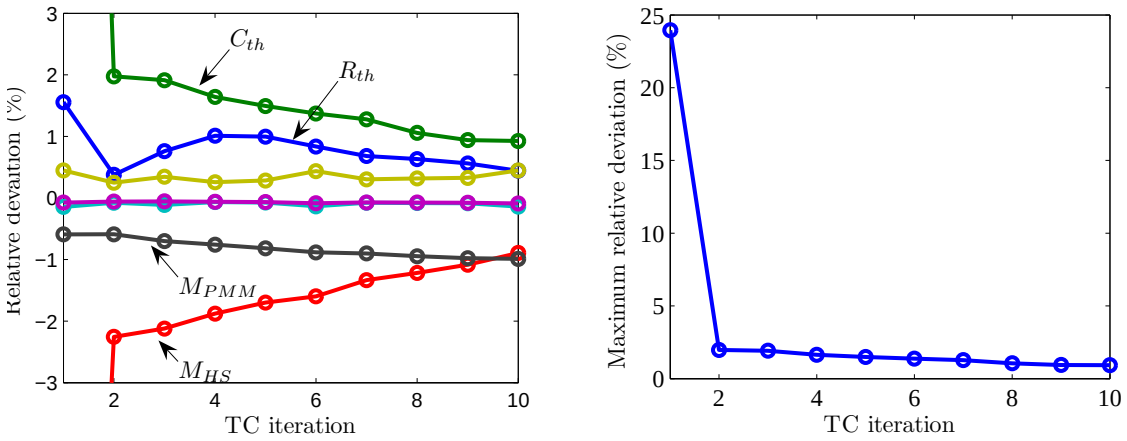
At each TC iteration (alternate between system and subsystem levels), relative deviation (*Devia*) of each sublevel target is computed. TC algorithm stops when the maximum absolute

value of relative deviation is less than 1%:

$$\max(Devia) = \max\left(\frac{\left|R^U_{ss,i} - R^L_{ss,i}\right|}{R^U_{ss,i}} \cdot 100\%\right) \qquad \text{for } i = 1, \ldots, n_T \tag{5.1}$$

where n_T is number of sublevel target/response (7 for this problem i.e. 3 from the subproblem heat sink and 4 from the subproblem PMM).

The relative deviation of all sublevel targets are shown in Figure 5.16a and the maximum relative deviation is depicted in Figure 5.16b. Ten iterations have been performed by TC. Note that *Devia* for the first iteration is not plotted in Figure 5.16a (24% and -19% for C_{th} and M_{HS}, respectively). At the final iteration, all relative deviations are lower than 1%. The numerical values are given in Table 5.10. Note that the responses are the value computed by the subsystem models and the targets are found by system level optimization subproblems and used in the calculation at the system level. They are defined as design variables in the system level problem. Their upper and lower limits are also given in the table.



(a) Deviation for all responses

(b) Maximum deviation

Figure 5.16: Deviation of subsystem targets and subsystem responses

The traction system design problem involves 3 subproblems. The optimal value of local design variables for each subproblem are given in Table 5.11. All constraints are satisfied. Figure 5.17 shows some simulation results obtained from CITHEL (system level). The simulation is made for 50% motorization in faulty mode, which is the most difficult scenario. It can be observed that the maximum IGBT temperature slightly exceeds the constraint limit defined at 120°C (see Table 5.12). This is caused by the surrogate model of this constraint (less than 2% error). Five surrogate models are used in the system level problem, maximum error at the optimal point is only 3.86% for the energy consumption model. The energy consumption computed with CITHEL is 199.63 kWh compared with 191.93 kWh obtained from the surrogate model.

For the heat sink subproblem, three design variables are located at their boundary limit. However, there is not any active constraint. Table 5.12 lists constraints of system and subsystem problems.

Table 5.10: Sublevel target achievement

Problem	Quantity	Symbol	Unit	Limit	Target	Response	*Devia* (%)
Heat sink	Thermal resistance	R_{th}	K/W	[0.08, 0.24]	0.1678	0.1670	0.44
	Thermal capacitance	C_{th}	J/K	[597.5, 1792.5]	1543.10	1531.78	0.93
	Heat sink mass	M_{HS}	kg	[10, 50]	13.78	13.91	-0.89
PMM	Phase resistance	R_s	mΩ	[25.61, 76.83]	30.9	31.00	-0.14
	Inductance	L_s	mH	[0.3934, 1.1802]	0.9626	0.9635	-0.09
	Flux	Φ_s	Wb	[0.1781, 0.5343]	0.5106	0.5083	0.45
	PMM mass	M_{mot}	kg	[200, 1000]	406.43	410.46	-0.99

Table 5.11: Optimization results – Local optimal design variables

Problem	Quantity	Symbol	Unit	Limit	Optimal results
System	PWM frequency 1	F_{pwm_1}	Hz	[300, 1500]	796
	PWM frequency 2	F_{pwm_2}	Hz	[1500, 2000]	1693
Heat sink	Width of heat sink	L_{dx}	mm	[150, 500]	150.19
	Length of heat sink	L_{dy}	mm	[140, 500]	140
	Diffuser thickness	E_d	mm	[10, 50]	27.93
	Heat sink fin gap	$L_{a_{Can}}$	mm	[5, 20]	20
	Fin height	$L_{g_{Ail}}$	mm	[10, 100]	59
	Fin thickness	$L_{a_{Ail}}$	mm	[3, 20]	3
PMM	Stator yoke height	Y_s	mm	[10, 50]	39.56
	PM height	l_m	mm	[1, 15]	15
	Air gap	g	mm	[2, 10]	4.33
	Slot height	d_s	mm	[20, 40]	37.31
	Armature radius	r_a	mm	[100, 500]	125.28
	Stack length	l_{stk}	mm	[100, 700]	367
	Tooth width ratio	r_{wt}	-	[0.3, 0.7]	0.6
	Conductor per slot	$n_{c/slot}$	-	[1, 10]	5

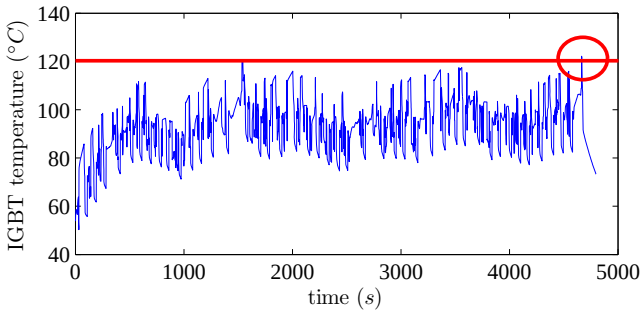


Figure 5.17: System level simulation results at optimal design solution – IGBT temperature during round-trip of tram operation

For the PMM subproblem, the PM temperature constraint is active at 150°C. The flux density in the stator yoke, rotor yoke and teeth constraints are active. According to the required torque and speed and sublevel targets, the PMM control strategy does not use maximum inverter voltage and current.

Table 5.12: Constraint values

Problem	Symbol	Quantity	Unit	Limit	Optimal result
System	I_L	Inductor thermal current	A	≤ 230	188.89
	I_C	Capacitor thermal current	A	≤ 140	121.18
	I_{sw}	Phase switching current	A	≤ 1000	558.71
	T_{IGBT}	IGBT temperature	°C	≤ 120	<u>120</u>
Heat sink	dP	Pressure loss	Pa	≤ 200	176
	v_{air}	Air speed	m/s	≤ 20	19.63
	n_{fin}	Number of fin	-	≥ 5	38
PMM	T_w	Winding temp.	°C	≤ 200	184.25
	T_{PM}	PM temp.	°C	≤ 150	<u>150.00</u>
	B_{Y_s}	Stator yoke flux density	T	≤ 1.6	<u>1.60</u>
	B_{Y_r}	Rotor yoke flux density	T	≤ 1.6	<u>1.60</u>
	B_T	Tooth flux density	T	≤ 1.6	<u>1.60</u>
	I_{rms}	RMS current	A	≤ 300	171.64
	Pt_{dem}	PM demag. flag	-	≤ 0	-0.1
	$Pt_{V_{rms}}$	RMS voltage flag	-	≤ 0	-23.98
	R_{ext}	External radius	mm	≤ 700	212.10
	L_{ext}	External length	mm	≤ 1000	564.80

Note that the optimal results for the system level problem are obtained from surrogate model.

5.2.4 Conclusion on the multilevel problem

An application of the Target Cascading method to the design of tram traction system was presented. The tram traction system design problem has been decomposed into several subsystem and component subproblems. To perform an optimal design, such decomposed system can be coordinated via TC algorithm. TC gives a new approach to formulate and to solve the complex system optimal design problem. The problem formulation is more suitable to this kind of problem than the conventional method, where the models are merged and used in the optimal design problem.

A tram traction system and 2 components: a heat sink for an inverter and a traction motor, have been designed simultaneously and consistently using the TC method. The coordination process via sublevel targets and responses has been performed successfully with a relative error of 1%. The component designs (sublevel problem) are performed in parallel. This allows limiting the computation time, which is significantly increased as the number of subproblems increases. Some convergence problems must be stated in the case of a non-attainable target. An improved TC algorithm is described and used to solve the same problem in [82, 85].

5.3 Conclusion

In this chapter, two typical design problems in the railway industry were presented. The techniques given in Part I have been applied to these problems.

The first application was the tram traction re-design problem. This problem appeared in the late design phase. Only a few control parameters at the system level can be considered as design variables. Even with a limited number of degrees of freedom, the surrogate modelling technique and several optimization techniques were successfully used to solve this problem. Combining these techniques, the solution can be found with high precision and acceptable computation time.

The second application was a more general design problem. The problem is to design a standard tram traction system and its components. As the problem is in the early design phase, it involves many components and there is also a great number of degrees of freedom as well as interactions. For these reasons, the design problem has become very complex. In the design of a complex system such as the traction system, the design problem is traditionally decomposed into several subproblems. TC was used to formulate such optimal design subproblems coordinated together. It can be considered as the optimization of the optimization process or metaoptimization. With TC, large-scale optimal design problems can be performed and integrated into the existing structure of the company. The optimal design problem is mapped directly to each engineering team/unit. This makes it easier for a company to accept and use the optimal design in their everyday design task.

Conclusion

This thesis presents an optimal design approach applied to the railway traction system design. The approach is general and can be used in the complex system design problem. The methodology is demonstrated via real application cases: a railway traction motor and hierarchical traction system design problems.

A complex design methodology is presented in Chapter 1. The conventional V-cycle is firstly described. It is composed of 2 main phases: design phase (Top-down) and validation phase (Bottom-up). A design problem is always an inverse problem i.e. the performances of the system are given and the component characteristics are sought. The behaviour model referred to as the model that simulates the behaviour of the system, is generally used in the computer aided engineering. Therefore, an iterative method must be used to inverse the behaviour model. To solve the design problem, an optimal design methodology can be applied. It is considered as a tool that helps the designer to find the solution more rapidly.

In the last section of Chapter 1, the railway traction system is introduced. Reflections on the optimal design applied to Alstom's design process are discussed. The optimal design should be applied in the early phase in which many design variables are accessible. Regarding the design criteria, mass, cost and reliability seem to be very important in the railway application.

Chapter 2 presents several modelling techniques, which can be classified into 2 groups: physical models and surrogate models. The physical modelling approach is traditional. It requires a lot of experience and time to develop. Depending on the technique, the model can be high fidelity and very expensive in terms of computation time. Integration of such time-consuming model into the optimal design process is sometimes prohibited. Therefore, an approach using a surrogate model tends to solve this problem. A surrogate model is constructed from a low number of high fidelity model evaluations. Three surrogate modelling techniques are presented: Polynomial, Radial Basis Function and Kriging. Assessment results show that Kriging model is more accurate than the other techniques.

The design problem can be formulated as an optimization problem. It can use the global system approach in which the design problem is formulated as an optimization problem. The problem can be a single-objective or multi-objective optimization problem. For a single-objective optimization problem, SQP, GA and Hybrid algorithm are described. The hybrid algorithm profits from the advantage of both SQP and GA. It gives a compromise between the number of function calls, global search capability and accuracy of the final results. For a multi-objective optimization problem, a set of single-objective optimizations is launched in order to generate a group of optimal solutions, a Pareto front. Such strategy is referred

to as Transformation strategy. The weighted-sum and the epsilon-constraint are presented. The epsilon-constraint strategy is more robust and can generate a nonconvex Pareto front. A multi-objective optimization algorithm such as NSGA-II searches the Pareto front based on sets of individuals. This stochastic method provides an accurate and well-distributed Pareto front. However, it requires a very high number of function evaluations.

The surrogate-assisted algorithm is a promising approach. It uses the surrogate model and modelling error associated with the model to guide and iteratively refine the surrogate model. This allows performing both local and global search (exploitation and exploration). The GEI infill criteria is used in the single-objective optimization. We proposed a novel infill criteria called Pseudo distance that extends the algorithm to the multi-objective optimization case. We show that these algorithms cope with noisy function and they are able to find the global solutions by using a very low number of function evaluations. These properties are very useful when using FEA in the optimal design problem.

The complex system design problem is typically decomposed into several small problems. It also involves many models representing components or disciplines. Multidisciplinary Design Optimization allows solving the interaction problem between different models. The implicit variables as well as interaction variables are treated by the optimization algorithm. The optimization problem is unavoidably more complex because of the additional design variables and constraints. However, since the models are completely decoupled, they can be evaluated in parallel. This reduces significantly the total optimization time.

In multilevel hierarchical design problem, Target Cascading allows coordinating the decoupled subproblems. It can be considered as the optimization of optimization process or metaoptimization. With TC, the optimization problem formulation is more likely to the existing management structure of company. It is mapped directly to each engineering team/unit. This makes it easier for the company to accept and use optimal design in their everyday design task.

In Part II, the applications of the optimal design are presented. Chapter 4 focuses on the traction motor design. The comparative study emphasizes that the transient thermal behaviour of a motor must be taken into account in case that the load cycle is known. This allows downsizing the motor and ensures the correct operation in the flux-weakening mode.

The model of a surface-mounted permanent magnet motor, used in the comparative study, is a semi-nemerical multidisciplinary model. It is composed of several interacting modules. MDO formulations are applied to the optimal design using this model. In the IDF formulation, interaction variables e.g. the winding temperature, the PM temperature, the phase resistance and the flux are treated by the optimization algorithm. This removes any direct link between modules, therefore, they can be launched in parallel. In AAO formulation, implicit variables in the thermal module are also added to the optimization problem. Fix point iteration in the thermal module has been removed. As a result, AAO formulation leads to the most complex optimization problem but it requires the lowest optimization time i.e. 4 times lower than in the MDF case.

In this research, a commercial FEA code, Vector Filed Opera2D is also directly integrated into the optimal design process. Both single and multi-objective optimal design problems are solved using a surrogate-assisted algorithm. A very low number of FEA evaluations is needed

to locate the Pareto optimal solutions. The proposed algorithm outperformes NSGA-II for this problem.

The tram traction system design problem is discussed in Chapter 5. A comparative study of single-objective algorithm is presented. Multistart SQP, GA and hybrid algorithms are used to solve a tram traction system re-design problem. CITHEL, Alstom's traction system design tool is used as the high fidelity tool in the Kriging surrogate model construction. The results show that the hybrid algorithm is better than GA in terms of accuracy and number of function evaluations. The percentage of convergence (to the global optimum) is the best among the three algorithms. From an engineering aspect, the optimal design using a surrogate modelling approach allows improving the tram traction system while respecting 11 constraints (component specifications, tram performances). The Kriging models are also used in the sensitivity analysis. It shows that the final result is robust.

The decomposition approach is demonstrated via a multilevel tram traction system design problem. The global design problem is decoupled (object-based decomposition) into 3 subproblems: the traction system design, heat sink design and PMM design problems. The Target Cascading method is used to coordinate and solve these subproblems. Two system targets are defined: total mass and energy consumption. The results show that the system targets as well as subsystem targets can be achieved. TC requires 10 TC iterations i.e. alternation between the system and subsystem levels. The traction system and two components have been designed concurrently. The complex system decomposition and multilevel hierarchical optimization methods are very promising. They allow solving large-scale and complex optimal design problems, which usually have a strong link to the company management structure.

This work has opened up new research opportunities. The surrogate modelling technique and a surrogate-assisted optimization algorithm are very promising approaches. The applications to traction system design have shown a great potential of this method. Several enhancements may be considered in further works. The surrogate modelling technique must be able to deal with integer and non-classifiable variables. The problem involving mixed design variables is very common in the electrical engineering field e.g. type of component chosen from catalogue, number of conductors. It is interesting to develop an infill criterion that can handle the design tool with different levels of fidelity. It seems very useful in our community as the designer typically uses low fidelity analytical model and high fidelity FEA model.

The complex system design methodology as well as the multilevel optimization algorithm (MDO and TC approaches) must be studied in detail. Such methods are suitable for companies with a large management structure. In order to gain more acceptance from the railway industry, the methods must be tested in more complex problems taking into account many components. Also, more Alstom's existing design tools must be integrated in the optimal design process. In order to ensure the company's achievement, non-technical criteria such as marketing and environmental issues should be considered as design criteria.

Appendices

Appendix A

Surrogate modelling assessment

This appendix gives the description of test functions and the full result of the surrogate modelling assessment presented in Section 2.2.4.

A.1 Test function

1. Branin function (Br): 2 variables,

$$\begin{aligned} f_{Br}(x_1, x_2) = & \left(x_2 - \frac{5.1}{4\pi^2}x_1^2 + \frac{5}{\pi}x_1 - 6\right)^2 + 10\left(1 - \frac{1}{8\pi}\right)\cos x_1 + 10 \\ \text{where} \quad & x_1 \in [-5,\ 10],\ x_2 \in [0,\ 15] \end{aligned} \tag{A.1}$$

2. Peaks function (Pk): 2 variables,

$$\begin{aligned} f_{Pk}(x_1, x_2) = & \ 3(1 - x_1)^2 \exp\left(-x_1^2 - (x_2 + 1)^2\right) \\ & -10\left(\frac{x_1}{5} - x_1^3 - x_2^5\right)\exp\left(-x_1^2 - x_2^2\right) - \frac{1}{3}\exp\left(-(x_1 + 1)^2 - x_2^2\right) \\ \text{where} \quad & x_1, x_2 \in [-3,\ 3] \end{aligned} \tag{A.2}$$

3. Goldstein and Price function (GP): 2 variables,

$$\begin{aligned} f_{GP}(x_1, x_2) = & \left(1 + (x_1 + x_2 + 1)^2\left(19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1x_2 + 3x_2^2\right)\right) \cdot \\ & \left(30 + (2x_1 - 3x_2)^2\left(18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1x_2 + 27x_2^2\right)\right) \\ \text{where} \quad & x_1, x_2 \in [-2,\ 2] \end{aligned} \tag{A.3}$$

4. Hartman function (H): 6 variables,

$$\begin{aligned} f_H(x) = & -\sum_{i=1}^{4} c_i \exp\left(-\sum_{j=1}^{6}\alpha_{ij}(x_j - p_{ij})^2\right) \\ \text{where} \quad & x_1, \ldots, x_6 \in [0,\ 1] \end{aligned} \tag{A.4}$$

α_{ij}, c_i and p_{ij} are defined as in Table A.1 and A.2.

Table A.1: Hartman function parameter – α_{ij} and c_i

i	α_{ij}, $j = 1, \ldots, 6$						c_i
	1	2	3	4	5	6	
1	10	3	17	3.5	1.7	8	1
2	0.05	10	17	0.1	8	14	1.2
3	3	3.5	1.7	10	17	8	3
4	17	8	0.05	10	0.1	14	3.2

Table A.2: Hartman function parameter – p_{ij}

i	p_{ij}, $j = 1, \ldots, 6$					
	1	2	3	4	5	6
1	0.1312	0.1696	0.5569	0.0124	0.8283	0.5886
2	0.2329	0.4135	0.8307	0.3736	0.1004	0.9991
3	0.2348	0.1451	0.3522	0.2883	0.3047	0.6650
4	0.4047	0.8828	0.8732	0.5743	0.1091	0.0381

5. Six-hump camel-back function (SC): 2 variables,

$$\begin{aligned} f_{SC}(x_1, x_2) = \quad & 4x_1^2 - 2.1x_1^4 + \frac{1}{3}x_1^6 + x_1x_2 - 4x_2^2 + 4x_2^4 \\ \text{where} \quad & x_1, x_2 \in [-2,\ 2] \end{aligned} \tag{A.5}$$

6. Shekel 10 function (Sh): 4 variables,

$$\begin{aligned} f_{Sh}(x_1, x_2) = \quad & \sum_{i=1}^{10} \frac{1}{c_i + \sum_{j=1}^{4} (x_j - a_{ji})^2} \\ \text{where} \quad & x_1, \ldots, x_4 \in [0,\ 10] \end{aligned} \tag{A.6}$$

a_{ji} and c_i are given in Table A.3.

Table A.3: Shekel 10 function parameter – a_{ji} and c_i

i	a_{ji}, $j = 1, \ldots, 4$				c_i
	1	2	3	4	
1	4	4	4	4	0.1
2	1	1	1	1	0.2
3	8	8	8	8	0.2
4	6	6	6	6	0.4
5	3	7	3	7	0.4
6	2	9	2	9	0.6
7	5	5	3	3	0.3
8	8	1	8	1	0.7
9	6	2	6	2	0.5
10	7	3.6	7	3.6	0.5

A.2 Results

Test criteria are:

- Sampling technique: LHS, HSS and MCS
- Size of a sample set: The sample sets are scaled by the number of design variables. They are set to 5, 10 and 20 times number of design variables for small, medium, and large sample set, respectively.
- Surrogate techniques: RBF and Kriging.

Table A.4: Kriging – Normalized maximum absolute error (NEMAX)

Test function	LHS			HSS			MCS		
	small	medium	large	small	medium	large	small	medium	large
1	0.33	0.18	0.04	0.25	0.03	0.02	0.42	0.16	0.01
2	0.92	2.18	0.25	0.94	1.20	1.01	2.02	2.17	0.32
3	1.25	1.24	0.59	1.68	1.64	1.25	0.73	0.60	1.48
4	2.20	1.73	3.36	1.19	2.07	1.95	0.84	0.79	3.29
5	1.85	0.53	0.11	0.66	0.20	0.29	2.00	0.68	0.10
6	1.58	0.75	2.62	0.65	0.46	0.74	4.07	1.02	1.57
Average	1.35	1.10	1.16	0.90	0.94	0.88	1.68	0.90	1.13

Table A.5: Kriging – Normalized root mean squared error (NRMSE)

Test function	LHS			HSS			MCS		
	small	medium	large	small	medium	large	small	medium	large
1	0.16	0.11	0.01	0.39	0.04	3e-4	0.44	0.10	0.04
2	1.72	0.72	2.02	0.34	0.15	0.06	1.94	1.54	1.99
3	1.88	0.04	0.16	0.56	0.11	0.22	1.33	0.45	0.01
4	1.59	0.80	0.97	0.61	0.35	0.07	0.30	1.47	1.26
5	0.07	0.33	0.03	0.11	0.03	0.04	0.03	0.36	0.04
6	0.41	0.07	0.97	0.63	0.52	0.13	1.98	0.16	0.35
Average	0.97	0.34	0.69	0.44	0.20	0.09	1.00	0.68	0.62

Table A.6: RBF – Normalized maximum absolute error (NEMAX)

Test function	LHS			HSS			MCS		
	small	medium	large	small	medium	large	small	medium	large
1	1.70	0.21	0.60	2.09	0.38	0.12	1.23	1.37	0.11
2	0.73	0.57	0.48	0.81	1.31	1.18	5.66	4.95	1.54
3	1.03	1.18	0.91	1.66	0.87	2.25	7.48	3.06	2.29
4	3.63	2.52	3.80	1.53	1.81	1.56	1.88	1.06	2.93
5	2.52	0.75	1.44	2.77	1.02	1.64	1.63	2.03	1.35
6	0.98	1.42	2.36	0.90	0.73	1.30	3.10	1.46	2.53
Average	1.76	1.11	1.60	1.63	1.02	1.34	3.50	2.32	1.79

Table A.7: RBF – Normalized root mean squared error (NRMSE)

Test function	LHS			HSS			MCS		
	small	medium	large	small	medium	large	small	medium	large
1	0.02	0.49	0.16	0.37	0.07	0.17	0.72	0.08	0.61
2	1.22	1.77	2.22	0.36	0.78	0.17	2.84	4.00	1.09
3	1.79	9e-3	0.01	0.086	0.10	0.25	6.20	0.05	0.19
4	2.74	0.15	0.62	0.37	0.85	0.17	0.27	1.30	1.20
5	0.23	0.74	4e-3	0.28	0.25	0.23	0.67	1.08	0.31
6	0.43	0.29	0.70	1.32	1.01	0.18	1.00	0.21	0.26
Average	1.07	0.57	0.62	0.59	0.51	0.20	1.95	1.11	0.61

Appendix B

Surface-mounted permanent magnet synchronous motor modelling

This appendix intends to detail the semi-numerical model of surface-mounted permanent magnet motor.

B.1 Introduction

A design of a surface-mounted permanent magnet (SMPM) motor implies a multidisciplinary design: magnetic, electric, heat transfer, mechanic, cost, etc. The modelling of SMPM is based on a modular approach. Input and output data of each module are defined in the first place. These input and output describe interactions between modules. Once the input and the output are identified, the model used in each module can be selected based on the need of fidelity. For example, the analytical model or Finite Element Analysis (FEA) can be used in the magnetic or thermal module. In the same manner, the flux-weakening control or the maximal torque per ampere strategy can be used in the control module. The structure of SMPM motor model is shown in Figure B.1. It consists of eight modules. Each module represents a discipline; magnetic, electric, control, thermal etc. Regarding the interaction, there are two temperature feedbacks:

(i) The magneto-thermal loop allows taking into account permanent magnet properties change due to the PM temperature change;

(ii) The electro-thermal loop modifies phase resistance value according to the winding temperature.

There are two types of input and output: time dependent and time independent variable. Time dependent inputs consist of torque and speed (as a function of time). These can be obtained using vehicle dynamic equations. Time independent inputs are voltage limit, material property and motor geometry. Time dependent outputs e.g. flux density, phase

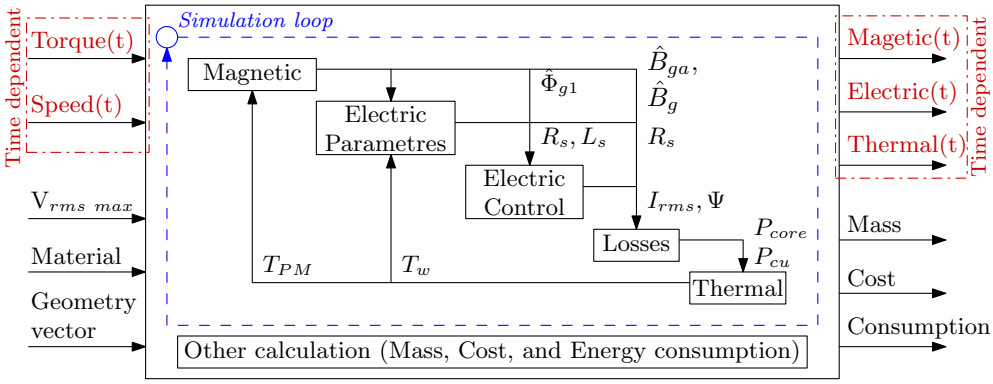


Figure B.1: Structure of SMPM motor model

current, temperature (as a function of time) and time independent outputs; mass, cost, and consumption are computed by the SMPM model. When using the model in optimization, each input variable can be fixed as a design parameter or assigned as a design variable depending on how an optimal design problem is formulated. Outputs have to be conditioned before used in the optimization e.g. taking the maximal value of time dependent output. Table B.1 - Table B.11 provide an exhaustive list of model's inputs and outputs.

In this research, an approach using semi-numerical model is used. Physical phenomena are described by analytical equations. Thermal interdisciplinary interactions and temperature-material property interactions are solved by an iterative method. This allows fast computation and acceptable accuracy since it will be used in preliminary design phase. Model accuracy and complexity are sufficient to capture interactions between disciplines. The following sections describe each module in detail and a simulation example will be presented.

B.2 Magnetic module

A simplified model was built. Several assumptions were made:

(i) Squarewave PM flux density distribution is considered.

(ii) Flux crossing PM and air gap is in radial direction.

(iii) First harmonic of flux density is extracted from square-wave and used for further calculation.

(iv) Steel permeability is assumed infinity.

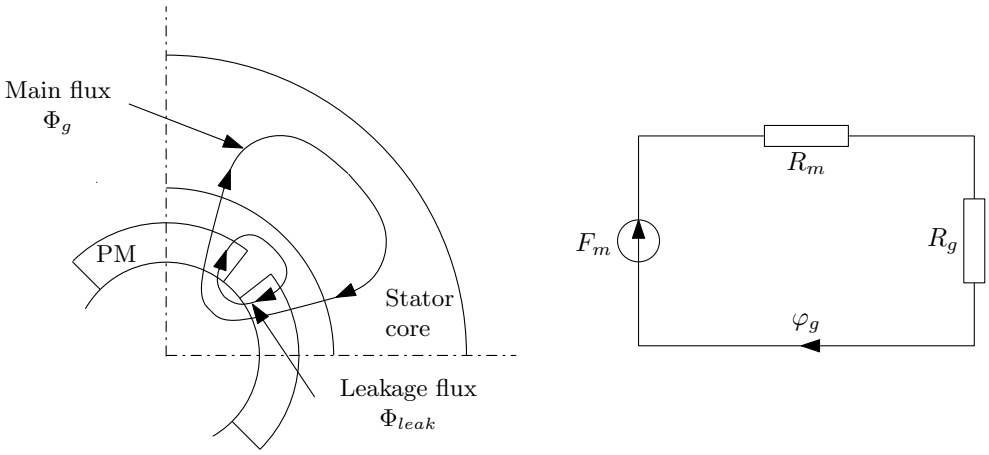
The main flux path of the SMPM motor is shown in Figure B.2a. Magnet flux is the actual flux that passes the magnet. It is composed of two main flux paths: air gap flux and leakage flux. The air gap flux is effective flux crossing the air gap. The leakage flux can be taking into account by leakage coefficient k_{leak}, which is a ratio of air gap flux and magnet flux.

Figure B.2b shows magnetic equivalent circuit of one pole. The air gap reluctance R_g and the magnet reluctance R_m are expressed as:

$$R_g = \frac{1}{\frac{\pi}{p} \cdot k_p \cdot \mu_0 \cdot \mu_r \cdot l_{stk}} \cdot \ln\left(\frac{r_a}{r_a - g'}\right) \tag{B.1}$$

$$R_m = \frac{1}{\frac{\pi}{p} \cdot k_p \cdot \mu_0 \cdot l_{stk}} \cdot \ln\left(\frac{r_a - g' - l_m}{r_a - g'}\right) \tag{B.2}$$

where $g' = k_c \cdot g$, k_c Carter coefficient given in [27], l_{stk} stack length, k_p magnet span coefficient. Other dimensions can be found in Figure B.3.



(a) Main flux path (b) Magnetic equivalent circuit

Figure B.2: Magnetic circuit

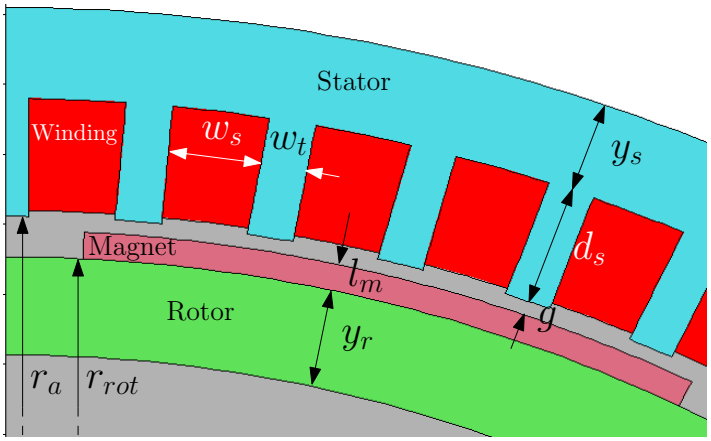


Figure B.3: Motor geometry

Magnetomotrice force (MMF) of magnet F_m and air gap flux φ_g are related by equation:

$$F_m = \frac{B_r}{\mu_R \mu_0} l_m = (R_g + R_m)\varphi_g = (R_g + R_m)\hat{B}_g A_m \tag{B.3}$$

with A_m magnet pole area. Therefore, flux density in air gap due to PM can be expressed as:

$$\hat{B}_g = \frac{k_{leak} \cdot B_r \cdot l_m}{r_g} \cdot \left(\ln\left(\frac{r_{rot} + l_m + g'}{r_{rot} + l_m}\right) + \frac{1}{\mu_R} \ln\left(\frac{r_{rot} + l_m}{r_{rot}}\right)\right)^{-1} \tag{B.4}$$

where B_r remanence flux density, μ_R relative permeability of PM, r_g mid-air gap radius ($r_g = r_{rot} + l_m + g/2$). The Carter coefficient has been verified with FEA and the relative error between flux for the case of a slotted stator and a smooth stator with a corrected air gap has been studied. In order to keep relative error below 5%, tooth width ratio $\left(r_{wt} = \frac{w_t}{w_s + w_t}\right)$ higher than 0.4 is preferred.

In sinewave back-EMF motor, fundamental air gap flux density ($\hat{B}_{g1}$) given in (B.5)is extracted from squarewave flux density distribution as shown in Figure B.4.

$$\hat{B}_{g1} = \frac{4}{\pi} \cdot \hat{B}_g \cdot \sin\left(\frac{\pi}{2} k_p\right) \tag{B.5}$$

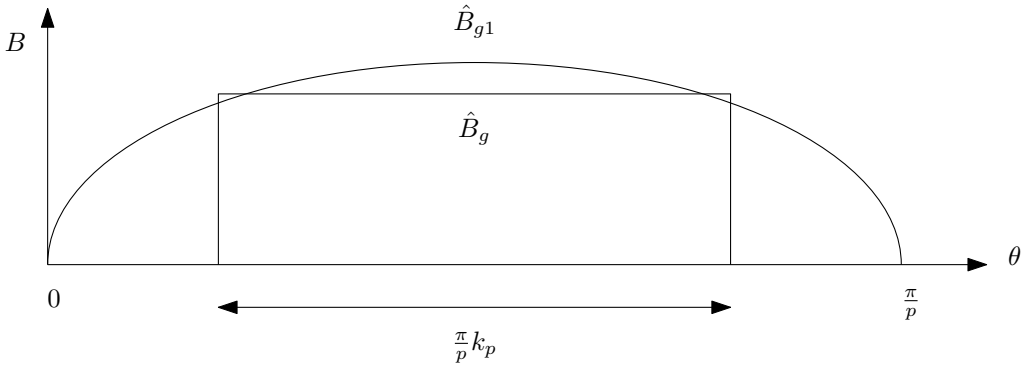


Figure B.4: Squarewave flux density distribution and fundamental flux density

Peak value of fundamental flux $\hat{\phi}_{g1}$ can be computed by integrating flux density over a pole area as in (B.6). Total flux per phase $\hat{\Phi}_{g1}$ is obtained by multiplying effective coil number per phase.

$$\hat{\phi}_{g1} = \int_0^{\frac{\pi}{p}} \hat{B}_{g1} l_{stk} r_g \cdot d\theta \tag{B.6}$$

$$\hat{\Phi}_{g1} = \hat{\phi}_{g1} \cdot N_{s,eff} = \frac{2}{p} \cdot \hat{B}_{g1} \cdot \left(r_{rot} + l_m + \frac{g}{2}\right) \cdot l_{stk} \cdot N_{s,eff} \tag{B.7}$$

where p poles pair, k_p PM span ratio, l_{stk} stack length, and $N_{s,eff}$ effective coil number per phase described as:

$$N_{s,eff} = N_s \cdot k_{pitch} \cdot k_w = \left(\frac{N_c \cdot N_{slot} \cdot p}{a}\right) \cdot k_{pitch} \cdot k_w \tag{B.8}$$

where N_s coil number per phase, N_c conductor number per slot, N_{slot} slot number per pole and per phase, a parallel path number, k_{pitch} winding pitch factor, and k_w winding distribution factor.

Flux equation (B.7) is validated against FEA. Figure B.5 shows flux lines computed by FEA. The relative error between (B.7) and FEA is less than 10%, which is acceptable for preliminary design. The main error is due to Carter coefficient, which is inaccurate for the structure with narrow teeth (r_{wt} is equal to 0.26 for this motor.). However, correction factors can be used to calibrate the model in case that the model error is known a priori.

The properties of the PM are critical in the design of a SMPM. The major factors are the protection against the demagnetization, the property change due to temperature changes

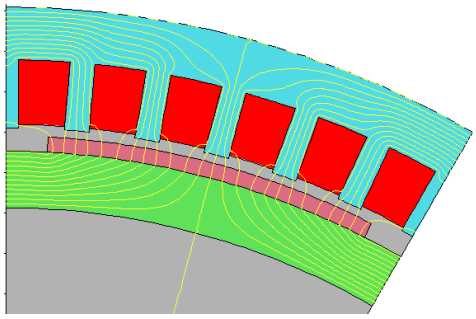


Figure B.5: Flux line from FEA

and the working temperature limit. The characteristic of permanent magnet such as B_r and the intrinsic coercive field strength (H_{ci}) change as the temperature changes. It can be described by a reversible temperature coefficient of B_r and H_{ci}, α_{B_r} and $\beta_{H_{ci}}$, respectively [106]. For NdFeB, B_r decreases as the temperature increases. However, reversible losses can be recovered when the temperature returns to the initial point. In order to take into account the PM temperature effects, at each simulation step, B_r will be updated with the actual permanent magnet temperature T_{PM} using the remanence flux density at 20°C (B_{r20}) and remanence temperature coefficient α_{B_r}.

$$B_r = B_{r20}\left(1 + \frac{\alpha_{B_r}}{100} \cdot (T_{PM} - 20)\right) \tag{B.9}$$

Note that T_{PM} is in °C.

The change in "knee point" of normal demagnetization curve is very important (red points in Figure B.6a). In order to avoid irreversible loss, the operating point cannot be lower than this point [34] i.e. the allowable motor current is limited by the knee point. Unfortunately, the relationship between the knee point and the temperature is generally not provided by the manufacturer. However, one can always approximate this relationship from B-H curve at various temperatures given in the data sheet. It will be used as a constraint in the optimization problem. For NdFeB 39H, Br_{min} can be given as:

$$B_{r_{min}} = -5.58 \cdot 10^{-5} \cdot T_{PM}^2 + 0.022 \cdot T_{PM} - 1.445 \tag{B.10}$$

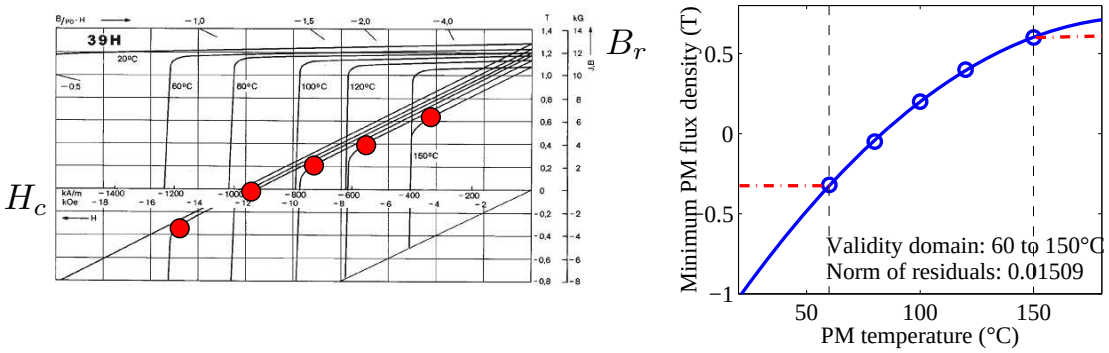
with T_{PM} in °C. This relationship is valid from 60°C to 150°C. For T_{PM} lower than 60°C or higher than 150°C, constant values are given as shown in Figure B.6b.

B.3 Electric module

This module computes circuit parameters (resistance R_s and inductance L_s) from motor geometry. The phase resistance at 20°C is defined as:

$$R_{s20} = \frac{\rho_{cu} \cdot l}{k_r \cdot w_s \cdot d_s} \cdot 2N_s \tag{B.11}$$

where l one conductor length (including end winding), k_r slot-fill factor and ρ_{cu} copper conductivity at 20°C.



(a) B-H characteristic from the PM data sheet

(b) Approximate model

Figure B.6: Minimal value of the remanence flux density as a function of the temperature can be estimated from the PM B-H curve

During the simulation phase, the resistance is updated with the actual winding temperature (T_w, in °C) from the heat transfer module.

$$R_s = R_{s20}\left(1 + \alpha_{cu} \cdot (T_w - 20)\right) \tag{B.12}$$

where α_{cu} is the conductivity temperature coefficient of copper.

The phase inductance is composed of 4 components, (i) air gap self inductance L_g, (ii) air gap mutual inductance M_g, (iii) slot leakage inductance L_{slot}, and (iv) end winding inductance L_{end}.

$$L_s = L_g - M_g + (L_{slot} - L_{end}) \tag{B.13}$$

L_g is computed by simply dividing the peak value of fundamental air gap flux due to armature current ($\hat{\Phi}_{ga1}$) by current I. For sinusoidal distributing winding, M_g is obtained using (B.17).

$$L_g = \frac{\hat{\Phi}_{ga1}}{I} \tag{B.14}$$

$$\hat{\Phi}_{ga1} = \frac{2 \cdot l_{stk} \cdot r_a}{p} \cdot \hat{B}_{ga1} \cdot N_{s,eff} \tag{B.15}$$

$$\hat{B}_{ga1} = \frac{4}{\pi} \cdot \frac{\mu_0}{2 \cdot p \cdot (g' + l_m/\mu_R)} \cdot N_{s,eff} \cdot \tag{B.16}$$

$$M_g = -\frac{1}{2} L_g \tag{B.17}$$

End winding leakage inductance L_{end} formula are given in [8]:

$$L_{end} = 3.15 \cdot 10^{-6} \left(2N_s\right)^2 \frac{3k_w^2}{4\pi p} \left(b_1 + \frac{(b_4 + b_3/2)}{2}\right) \tag{B.18}$$

where b_1, b_3, b_4 are geometries of end winding as shown in Figure B.7[1].

For open rectangular slot, L_{slot} can be computed as [34]:

$$L_{slot} = \mu_0 N_s^2 l_{stk} \left(\frac{h2}{3w_s} + \frac{h1}{w_s} + \frac{h2}{4w_s}\right) p \left(b_1 + \frac{(b_4 + b_3/2)}{2}\right) \tag{B.19}$$

h_1 and h_2 are slot and winding dimensions (see Figure B.8).

[1] Figure from [8]

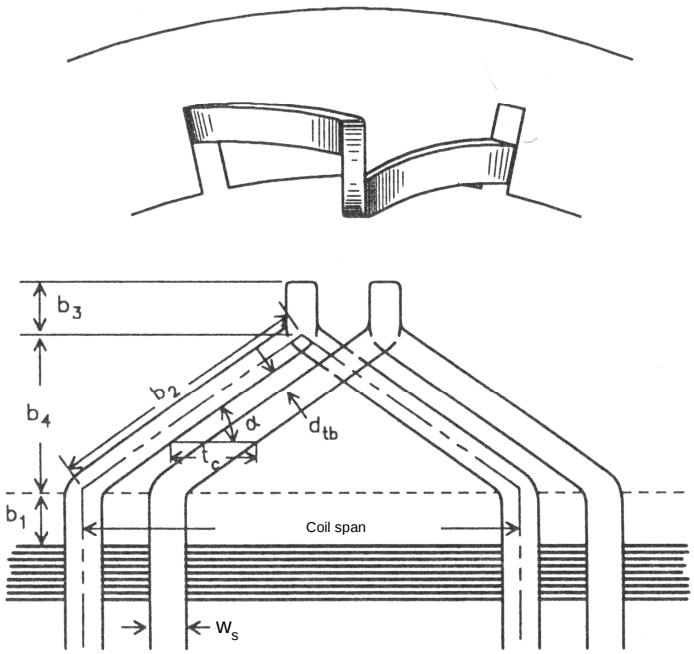


Figure B.7: Dimension of end winding

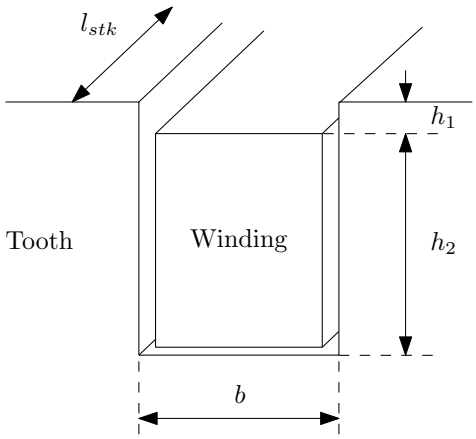


Figure B.8: Dimension of slot

B.4 Control strategy module

Since only electrical steady-state operation is considered, it is assumed that there is no error between the controllers' reference signal and real output. The model is an inverse steady-state single-phase equivalent model as shown in Figure B.9. It is then transformed in d-q reference frame in order to perform the flux-weakening strategy [81]. It computes necessary d-axis, q-axis, phase current and voltage to produce the required torque and speed at each time step.

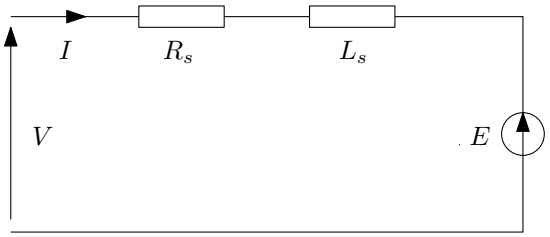


Figure B.9: Single-phase equivalent electric model

Relationships between RMS current (I_{rms}) and d-q axis current i_d and i_q are given as:

$$i_d = \sqrt{3} I_{rms} \sin \psi \tag{B.20}$$

$$i_q = \sqrt{3} I_{rms} \cos \psi \tag{B.21}$$

Therefore voltage equations in d-q coordinates (v_d and v_q) and RMS voltage V_{rms} can be described as follows:

$$v_d = R_s i_d - \omega_s L_q i_q \tag{B.22}$$

$$v_q = R_s i_q + \omega_s \left(L_d i_d + \phi_d \right) \tag{B.23}$$

$$V_{rms} = \sqrt{\frac{v_d^2 + v_q^2}{3}} \tag{B.24}$$

where ω_s electrical speed, L_d and L_q d- and q-axis inductance.

The electromagnetic torque T_{em} can be computed from electromagnetic power balance:

$$T_{em} = \frac{P_{em}}{\Omega} = p \left(L_d - L_q \right) i_d i_q + p \phi_d i_q \tag{B.25}$$

For SMPM motor, both L_d and L_q are equal to L_s (B.13). Equation B.25 becomes

$$T_{em} = p \phi_d i_q \tag{B.26}$$

where ϕ_d d-axis flux computed from fundamental air gap flux (B.7):

$$\phi_d = \sqrt{\frac{3}{2}} \cdot \hat{\Phi}_{g1} \tag{B.27}$$

Two operating modes are shown in Figure B.10. Zone 1: constant torque, in this zone, the torque is proportional to q-axis current i_q. d-axis current i_d is controlled to zero in order to maximise the torque per ampere. As the speed increases, the voltage increases.

Once the voltage reaches the maximum inverter voltage, the controller switches to flux-weakening mode (zone 2). The speed at this transition point is called base speed (Ω_b). In this zone, Ψ phase angle between back-EMF and phase current is changed by decreasing i_d. This results in negative value of i_d. It creates a flux in the opposite direction of the air gap flux. The result flux, hence back-EMF, then decreases. Therefore, the speed can be increased. As mentioned above, if the operating flux density is lower than the knee point at the corresponding temperature in the PM characteristic curve, it would cause an irreversible demagnetization of the magnet, which is not preferred.

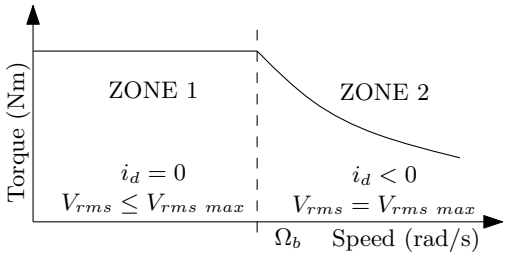


Figure B.10: Traction motor operating zone

A flux weakening control flowchart is shown in Figure B.11. By giving the required electromagnetic torque (T_{em}), i_q can be defined:

$$i_q = \frac{T_{em}}{p \cdot \phi_d} \tag{B.28}$$

V_{rms} (B.24) is then computed. If V_{rms} is higher than the maximal inverter output RMS voltage ($V_{rms\ max}$), this means that the inverter does not provide sufficient voltage and flux-weakening mode may be performed in order to achieve the required torque and speed.

In the flux weakening zone, i_d can be computed by solving the quadratic equation obtained with (B.22)-(B.24) substituting V_{rms} for $V_{rms\ max}$ and i_q from (B.28):

$$\begin{aligned} V_{rms\ max}^2 = \; & \left(R_s^2 + \omega_s^2 L_d^2\right) i_d^2 + \left(2\omega_s^2 L_d \Phi_d\right) i_d + \\ & \left(\omega_s^2 L_q^2 i_q^2 + \omega_s^2 \Phi_d^2 + 2\omega_s \Phi_d R_s i_q + R_s^2 i_q^2\right) \\ 0 = \; & A \cdot i_d^2 + B \cdot i_d + C \end{aligned} \tag{B.29}$$

The solution should be a real negative value. Otherwise, it can be concluded that the voltage is not sufficient to produce the required torque at the corresponding speed. In this case, the necessary voltage (which is higher than the maximum voltage) is given as an output and used as a constraint in optimal design process[2].

The RMS phase current (I_{rms}) can be computed:

$$I_{rms} = \sqrt{\frac{i_d^2 + i_q^2}{3}} \tag{B.30}$$

Ψ and φ, angle between current and voltage, can be written as:

$$\Psi = \arctan\left(\frac{i_d}{i_q}\right) \tag{B.31}$$

[2]This voltage output must be lower or equal to $V_{rms\ max}$ in order to ensure that motor can produce required torque at the corresponding speed.

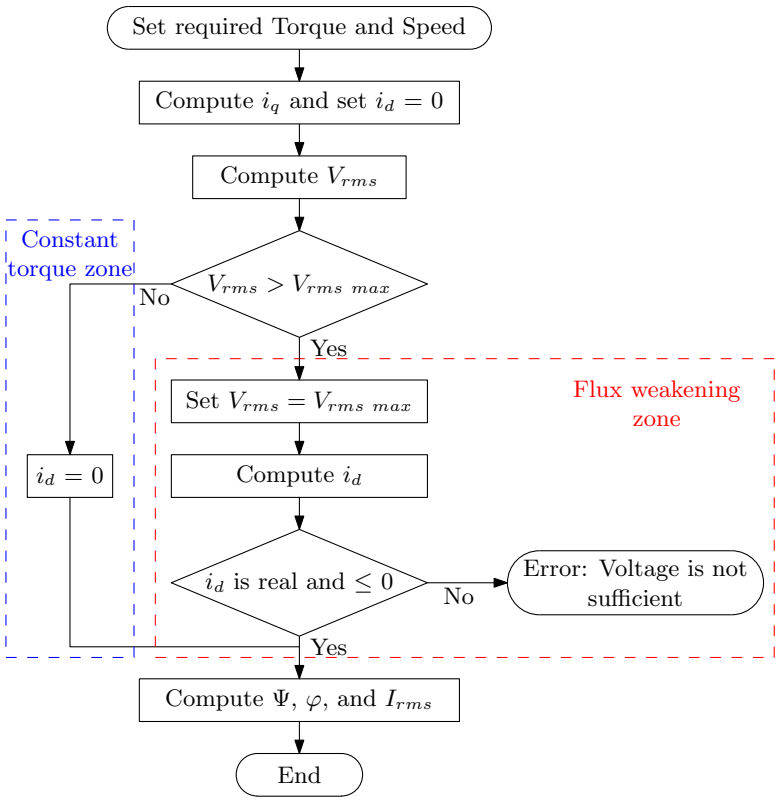


Figure B.11: Flux weakening control flowchart

$$\varphi = \arctan\left(\frac{v_d}{v_q}\right) - \Psi \tag{B.32}$$

B.5 Losses module

Copper losses can be easily obtained using conventional formula:

$$P_{cu} = 3I_{rms}^2 R_s \tag{B.33}$$

Core losses are computed using an approximate model obtained from a specific losses curve at 50Hz (q_t) given in the lamination datasheet. This model uses the peak value of the resulting flux density in tooth ($\hat{B}_t$) or stator yoke ($\hat{B}_{Y_s}$) as input. For M235-35A, q_t is given as in (B.34)

$$q_t = -1.2073\hat{B}^4 + 4.5706\hat{B}^3 - 4.6764\hat{B}^2 + 2.4067\hat{B} - 0.2025 \tag{B.34}$$

These specific losses are recalibrated to the corresponding electrical frequency (f_s). They are then multiplied by the mass of teeth and the stator yoke to achieve core losses in the teeth and stator yoke.

$$P_{core\ (tooth,yoke)} = mass_{(teeth,yoke)} \cdot q_t \cdot \left(\frac{f_s}{50}\right)^{1.5} \tag{B.35}$$

Figure B.12 shows the vector diagram of the resulting flux density $\hat{B}_{g\,res1}$, which is vector sum of the air gap flux density due to the PM, mutual, and armature current $\hat{B}_{g1}$, $\hat{B}_{Mg1}$,

and $\hat{B}_{ga1}$, respectively. The resulting flux density in the teeth, stator and rotor yoke can be expressed as:

$$\hat{B}_t = \hat{B}_{g\,res1} \cdot \left(\frac{\pi \cdot r_g}{3 \cdot N_{slot} \cdot w_t} \right) \tag{B.36}$$

$$\hat{B}_{Y_s} = \frac{1}{2} \hat{B}_{g\,res1} \cdot \left(\frac{\pi \cdot r_g}{p \cdot Y_s} \right) \tag{B.37}$$

$$\hat{B}_{Y_r} = \frac{1}{2} \hat{B}_{g\,res1} \cdot \left(\frac{\pi \cdot r_g}{p \cdot Y_r} \right) \tag{B.38}$$

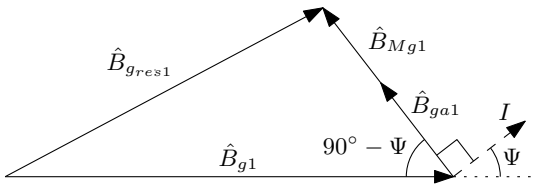


Figure B.12: Vector diagram of the resulting flux density

B.6 Heat transfer module

The transient thermal behaviour of motor is necessary to design a motor in variable speed applications, especially when the load profile is known such as in the traction applications [74]. Motor sizing is based on the steady-state temperature on the whole route profile, not on a rated point (base point).

A lumped parameter thermal network allows estimating the temperature in various parts of motor. The model is based on models developed for asynchronous motors [6, 79]. They are modified for Totally Enclosed, Fan-Cooled (TEFC) SMPM motors. Several hypotheses are made:

(i) Radial heat fluxes are considered, except shaft, heat transfer is axially.

(ii) Heat sources are concentrated at node.

(iii) Air flow at end winding is neglected.

(iv) Only tangential air flow is considered at air gap.

Model is composed of 8 nodes (see Figure B.13). The node temperature represents the average temperature of each element. For each node, the differential equation is:

$$C_{th_i} \frac{dT_i}{dt} = \sum_{i,j=1}^{n} \frac{(T_j - T_i)}{R_{th_{ij}}} + P_i \tag{B.39}$$

Thermal resistances R_{th} characterise conduction and convection heat transfers. Thermal capacitances C_{th} represent thermal inertia in case of a transient simulation. Thermal resistances and capacitors are determined from dimensions and motor materials. Losses generated by a motor act as heat sources P.

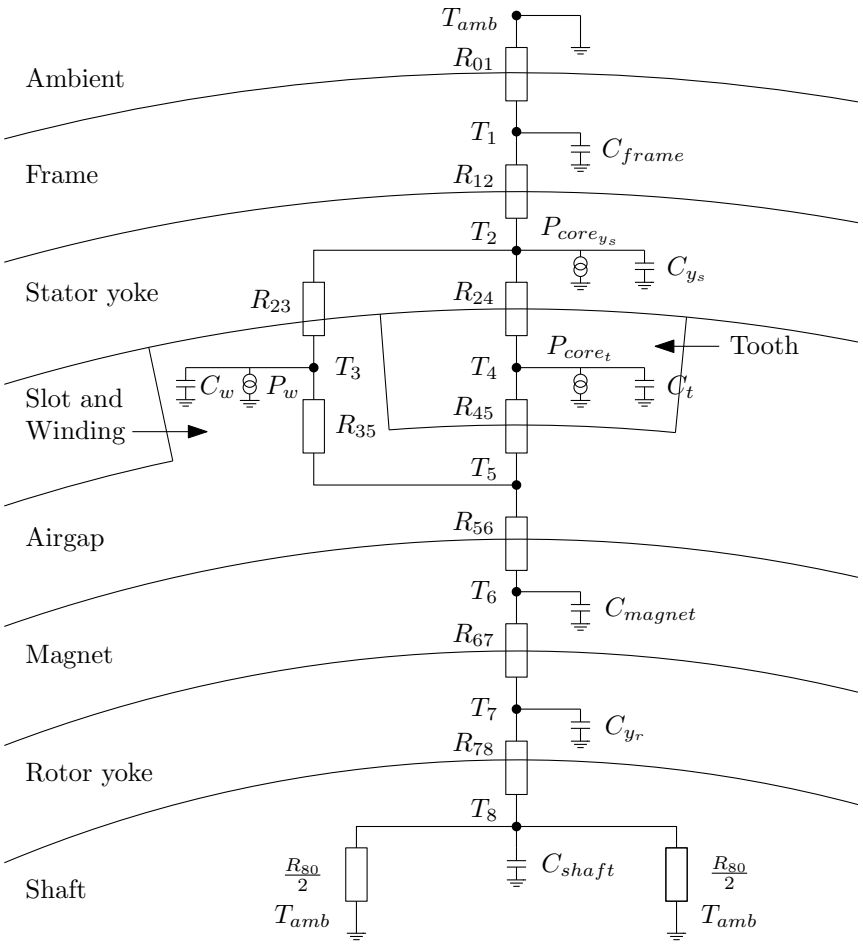


Figure B.13: Thermal model of SMPM motor

The $R_{th_{ij}}$ resistance between two nodes i and j can be defined as:

$$R_{th_{ij}} = \frac{R_{th_i}}{2} + \frac{R_{th_j}}{2} \tag{B.40}$$

Note that the temperature at node represents the average temperature of the element and the node position does not necessary have to be in the middle of elements.

For radial heat conduction, R_{th} of an element can be expressed as:

$$R_{th_i} = R_{cond_i} = \frac{1}{\theta k_i L} \ln\left(\frac{r_{ext_i}}{r_{int_i}}\right) \tag{B.41}$$

where k thermal conductivity coefficient (k is function of the temperature as given in [6]), r_{int} internal radius, r_{ext} external radius and L longueur and θ open angle comprising between 0–2π radian. (see Figure B.14).

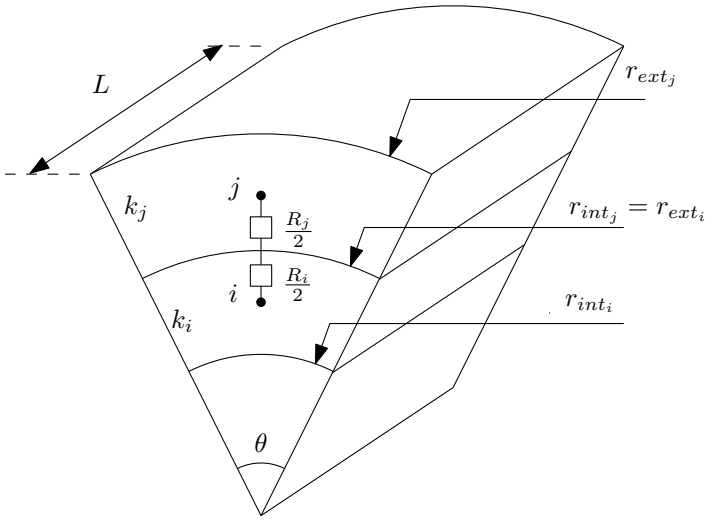


Figure B.14: Thermal resistance of radial convection

Heat convection is considered at the air gap and external frame. Thermal resistance due to heat convection is described as:

$$R_{th_i} = R_{conv_i} = \frac{1}{A \cdot h} \tag{B.42}$$

where A heat exchange surface and h heat convection coefficient.

Air gap heat convection coefficient: At air gap, tangential air speed due to rotation of rotor is:

$$v_{air_g} = (r_a - g)\,\Omega \tag{B.43}$$

Reynolds number and Prandtl number are expressed as:

$$Re_g = \frac{l_{stk} \cdot v_{air_g} \cdot \rho_{air_g}}{\mu_{air_g}} \tag{B.44}$$

$$Pr_g = \frac{C_{p_{air_g}} \cdot \mu_{air_g}}{k_{air_g}} \tag{B.45}$$

where ρ_{air_g} air mass density, μ_{air_g} air dynamic viscosity and $C_{p_{air_g}}$ specific heat capacity of air in the air gap.

h is expressed as:

$$h_g = \frac{Nu_g \cdot k_{air_g}}{g} \tag{B.46}$$

where Nu_g Nusselt number at air gap. Nu_g is function of Taylor number as shown in (B.47)[79].

$$\begin{aligned} &Ta \leq 41, && Nu_g = 2.2 \\ &41 \leq Ta \leq 100, && Nu_g = 0.23 Ta^{0.63} \cdot Pr_g^{0.27} \\ &Ta > 100, && Nu_g = 0.3855 Ta^{0.5} \cdot Pr_g^{0.27} \end{aligned} \tag{B.47}$$

where

$$Ta = Re_g \sqrt{\frac{g}{r_a - g}} \tag{B.48}$$

External frame heat convection coefficient: It depends on external air properties and external air speed in axial direction as depicted in Figure B.15

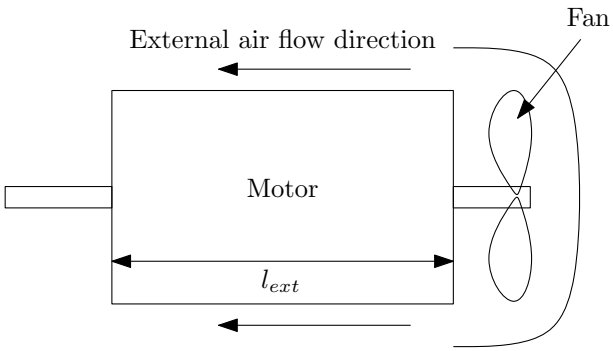


Figure B.15: External air flow

Reynolds number and Nusselt number are expressed as:

$$Re_{ext} = \frac{l_{ext} \cdot v_{air_{amb}} \cdot \rho_{air_{amb}}}{\mu_{air_{amb}}} \tag{B.49}$$

$$\begin{aligned} &Re_{air_{ext}} < 10000, && Nu_{ext} = 0.66 Re_{ext}^{0.5} \cdot Pr_{ext}^{0.33} \\ &Re_{air_{ext}} \geq 10000, && Nu_{ext} = 0.66 Re_{ext}^{0.75} \cdot Pr_{ext}^{0.33} \end{aligned} \tag{B.50}$$

Therefore, external air heat convection coefficient is defined as:

$$h_{ext} = \frac{Nu_{ext} \cdot k_{air_{amb}}}{l_{ext}} \tag{B.51}$$

The heat transfer module includes the numerical algorithm for solving the Ordinary Differential Equation (ODE). At the end of the load profile, steady state convergence over load cycle is checked. The temperatures at the end of the cycle are set as the new initial temperature and the simulation is rerun until the temperatures reach their steady state. This depends on the thermal time constant.

It also allows computing the steady-state temperature at any time step (do not take into account thermal capacitor). This can be useful when only the simple calculation is needed e.g. a design based on the torque as a function of speed characteristics or a base point design, in such cases the simulation is not needed since the time domain is not concerned.

Temperatures are implicit variables because material properties depend on temperatures as mentioned above. At any torque-speed point, the input temperatures are forced to be equal to the output temperatures through the fixed-point iteration method.

B.7 Simulation example

The simulation can be done by giving the motor's geometry and material, inverter voltage limit. Motor load profile (torque and speed as a function of time) is given as shown in Figure B.16. The simulation takes 40 load cycles to achieve the steady-state temperatures.

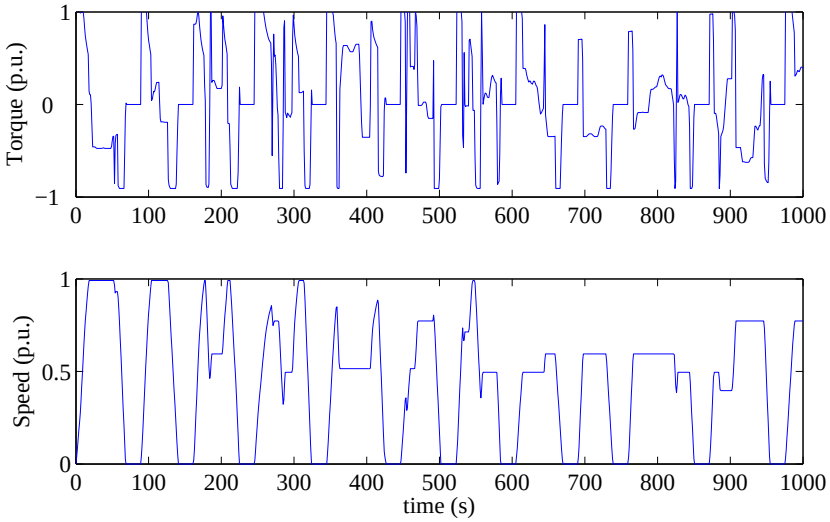


Figure B.16: Torque and speed requirement

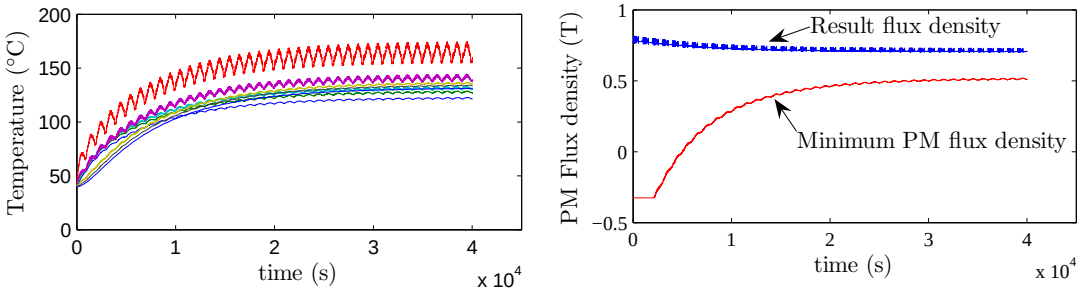
Figure B.17a shows the temperatures as a function of time. The temperature evolution starts from the ambient temperature at 40°C. The winding and the PM temperatures reach their steady-state after 40 load cycles. At the steady-state, the winding and the PM temperatures can be observed at 175°C and 145°C, respectively.

The resulting flux density and minimal PM flux density are shown in Figure B.17b. As the PM temperature increases, the minimal PM flux density increases. The PM demagnetization can be occurred if the resulting flux density is lower than the minimal PM flux density. In this example, it cannot be observed.

Figure B.18 show the RMS voltage and current. The maximum RMS voltage is given 290V in traction mode and 350V in braking mode. At the beginning of the simulation, the required RMS current is lower due to the fact that when the PM temperature is low, the PM can provide high remanence flux density. Therefore, motor requires less current to produce the same amount of torque.

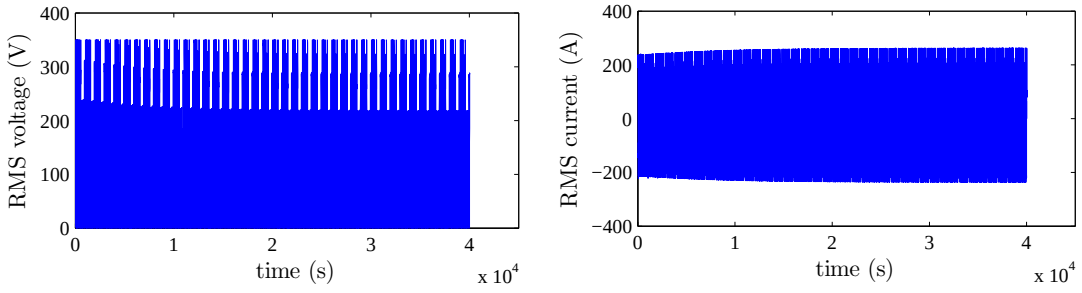
In d-q coordinate, currents are shown in Figure B.19. It can be observed that i_d is negative when the motor is operated in a flux weakening mode.

When using these simulation results in an optimal design process, the results must be reconditioned. The maximum value of the PM temperature is taken. Different value between



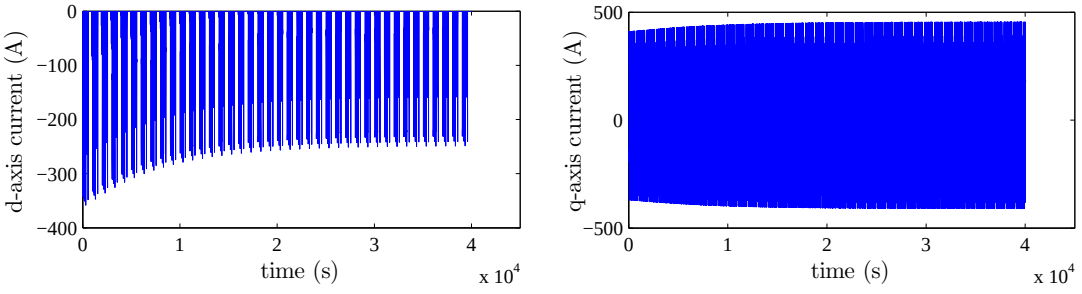
(a) (b)

Figure B.17: (b) PM flux density, (a) Temperatures



(a) RMS voltage (b) RMS current

Figure B.18: RMS voltage and current



(a) d-axis current (b) q-axis current

Figure B.19: Current in d-q coordinate

the resulting flux density and the minimal PM flux density ($B_{res} - B_{r_{min}}$) is used to indicate the demagnetization of the PM.

B.8 Conclusion

In this appendix, a model of a surface-mounted permanent magnet motor is presented. This model is built using a modular approach. Each module represents a discipline. They are described by analytical relationship. Implicit variables and interaction loops are solved by using the iterative method. This model is suitable for preliminary design and for using in an optimal design process. Applications of this model are presented in Chapter 4 and 5.

Table B.1: Input–Geometries

Symbol	Quantity	Unit
Y_s	Stator yoke height	m
Y_r	Rotor yoke height	m
l_m	PM height	m
g	Air gap	m
d_s	Slot height	m
r_a	Armature radius	m
l_{stk}	Stack length	m
d_{open}	Slot opening depth	m
t_{insu}	Insulation thickness	m
t_{frame}	Frame thickness	m
r_{wt}	Tooth width ratio	-
k_p	PM span coefficient	-
p	Pole pair	-
N_{slot}	Slot/pole/phase	-
k_{leak}	Leakage flux coefficient	-

Table B.2: Input–Winding

Symbol	Quantity	Unit
n_{ph}	Number of phases	-
N_c	Conductor/slot	-
A	Parallel path	-
$pitch$	Winding pitch	-
k_r	Slot filling coefficient	-

Table B.3: Input–Inverter

Symbol	Quantity	Unit
$V_{max_{Tr}}$	Maximal inverter voltage (traction)	V
$V_{max_{Br}}$	Maximal inverter voltage (braking)	V
$flag_{control}$	Flag control model: 1=with R_s, 2=without	-
$flag_{fw}$	Flag flux-weakening: true=with R_s, false=without	-

Table B.4: Input–Performances

Symbol	Quantity	Unit
T	Required torque vector	Nm
Ω	Required speed vector	rad/s
t	Time vector	s

Table B.5: Input–Simulation options

Symbol	Quantity	Unit
h	Time step	s
ode	1=transient, 0=steady-state	-
$cyclemax$	Maximal cycle to simulate	-

Table B.6: Input–Thermal

Symbol	Quantity	Unit
T_{amb}	Ambient temperature	K
T_{init}	Initial temperature vector	K

Table B.7: Input–Material properties

Symbol	Quantity	Unit
General		
ρ	Mass density	kg/m^3
cost	Material cost	Euro/kg
k	Thermal conductivity	W/(m·K)
c_p	Specific heat	J/kg/K
Copper		
ρ_{elec20}	Resistivity at 20°C	Ωm
α_{cu}	Conductivity temp. coefficient	°C^{-1}
PM		
B_{r20}	Remanence flux density at 20°C	T
μ_r	Relative permeability	-
α_{Br}	Temperature coefficient	% °C^{-1}
-	Demagnetization model	-
Steel sheet		
-	Core loss model	-

Table B.8: Output–Electric parameter module

Symbol	Quantity	Unit
R_{s20}	Phase resistance at °C	Ω
R_s	Phase resistance at actual temperature	Ω
L_s	Inductance	H

Table B.9: Output–Magnetic module

Symbol	Quantity	Unit
$\hat{B_{ga1}}$	Peak air gap flux density due to armature current	T
$\hat{B_{g1_{Br1}}}$	Peak air gap flux density due to PM for B_r=1	T
$\hat{\Phi_{g1_{Br1}}}$	Total flux per phase for B_r=1	Wb
B_r	PM remanence flux density	T
$\hat{B_{g1}}$	Peak air gap flux density due to PM	T
$\hat{\Phi_{g1_{Br1}}}$	Total flux per phase	Wb
$\hat{B_{gres1}}$	Result flux density in air gap	T
B_{Y_s}	Resulting flux density in stator yoke	T
B_{Y_r}	Resulting flux density in rotor yoke	T
B_t	Resulting flux density in tooth	T
B_{rmin}	Minimum allowable PM flux density	T

Table B.10: Output–Electric control module

Symbol	Quantity	Unit
I_{rms}	RMS current	A
i_d	d-axis current	A
i_q	q-axis current	A
V_{rms}	RMS voltage	V
V_d	d-axis voltage	V
V_q	q-axis voltage	V
Ψ	Angle between back-EMF and phase current	°
φ	Angle between current and voltage	°
T_{em}	Electromagnetic torque	Nm

Table B.11: Output–Other modules

Symbol	Quantity	Unit
Losses		
P_{cu}	Copper losses	W
P_{core_t}	Core losses in teeth	W
$P_{core_{Ys}}$	Core losses in stator yoke	W
η	Efficiency	%
E	Energy consumption	kWh
Thermal		
T_i	Temperature at node i	K
T_w	Winding temperature	K
T_{PM}	PM temperature	K
Mass		
M_j	Mass of element j	kg
M	Total mass	kg
Cost		
$Cost_j$	Cost of element j	euro
$Cost$	Total cost	euro

Bibliography

[1] N. M. Alexandrov and R. M. Lewis. Comparative properties of collaborative optimization and other approaches to mdo. In *proceedings of the First ASMO UK/ISSMO Conference on Engineering Design Optimization*, July 1999.

[2] J. T. Allison. Complex system optimization: A review of analytical target cascading, collaborative optimization, and other formulations. Master's thesis, University of Michigan, 2004.

[3] J. T. Allison, D. Walsh, M. Kokkolaras, P. Y. Papalambros, and M. Cartmell. Analytical target cascading in aircraft design. In *proceeding of 44th AIAA Aerospace Sciences Meeting and Exhibit*, 2006.

[4] N. E. Antoine. *Aircraft optimization for minimal environment impact*. PhD thesis, Stanford University, August 2004.

[5] N. Bianchi. Radially-magnetised interior-permanent-magnet synchronous motor for high-speed drive: an analysis and finite-element combined design procedure. *Electromotion*, 6:103–111, 1999.

[6] A. Boglietti, A. Cavagnino, M. Lazzari, and A. Pastorelli. A simplified thermal model for variable speed self cooled industrial induction motor. *IEEE Transactions on Industry Applications*, 39(4):945–952, July/August 2003.

[7] BOSCH. *Automotive handbook*. Bentley Publishers, 6th edition edition, 2004.

[8] R. Bouchard and G. Olivier. *Conception de moteurs asynchrones triphasés*. Editions de l'école polytechnique de montréal, 1997. In French.

[9] G. E. P. Box, Hunter W. G., and J. S. Hunter. *Statistics for Experimeters : design, discovery, and innovation*. Jonh Wiley & Sons, Inc., 2nd edition, 2005.

[10] R. D. Braun and I. M. Kroo. Development and application of the collaborative optimization architecture in a multidisciplinary design environment. In *Multidisciplinary Design Optimization: State of the Art*, pages 98–116. SIAM, 1997.

[11] S. Brisset. Démarches et outils pour la conception optimal des machines electriques. Habilitation á diriger des recherches, Université des Sciences et Techniques de Lille, France, December 2007. In French.

[12] S. Brisset and P. Brochet. Optimization of switched reluctance motors using deterministic methods with static and dynamic finite element simulations. *IEEE Transactions on Magnetics*, 34(5):2853–2856, September 1998.

[13] J.-P. Chiles and P. Delfiner. *Geostatistics: Modeling Spatial Uncertainty*. Wiley, New York, 1999.

[14] M. Cioffi, A. Formisano, and R. Martone. Statistical analysis in robust design of superconducting magnets. *COMPEL: The International Journal for Computation and Mathematics in Electrical and Electronic Engineering*, 24(2):427–435, 2005.

[15] Y. Collette and P. Siarry. *Optimisation multiobjectif*. Edition Eyrolles, 2002.

[16] R. D. Cook. *Finite Element Modeling for Stress Analysis*. John Wiley & Sons, Inc., New York, NY, USA, 1994.

[17] C. Coutel, F. Wurtz, J. Bigeon, and C. Chillet. Constrained optimisation of a linear actuator: comparison of two methods to deal with implicit parameters in the analytical model. In *proceedings of International conference Electric Machines and Drives (IEMD)*, pages 625–627, May 1999.

[18] E. Cramer, J. Dennis Jr., P. Frank, R. Lewis, and G. Shubin. Problem formulation for multidisciplinary optimization. *SIAM Journal on Optimization*, 4(4):754–776, 1994.

[19] F. de Toro, J. Ortega, J. Fernandez, and A. Diaz. Psfga: a parallel genetic algorithm for multiobjective optimization. In *proceedings of 10th Euromicro Workshop onParallel, Distributed and Network-based Processing*, pages 384–391, 2002.

[20] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, April 2002.

[21] M. Debruyne. High power igbt traction drives. presented at the 5th World Congress on Railway Research, Cologne, Germany, November 25–29 2001.

[22] D. Derks, F. Gillon, and P. Brochet. Modelling of an axial synchronous 3 phase motor. *Electromotion*, 10(4):359–364, 2003 2003.

[23] J. Dréo, A. Pé, P. Siarry, and E. Taillard. *Metaheuristics for Hard Optimization Methods and Case Studies*. Springer–Verlag, 2006.

[24] P. Eustache, G. Meunier, and J. L. Coulomb. Finite element toolbox for generic coupling (magnetic, thermal, etc.). *IEEE Transactions on Magnetics*, 32(3):1461–1464, May 1996.

[25] A. Fasquelle, D. Saury, S. Harmand, and A. Randria. Numerical study of fluid flow and heat transfer in an electrical motor. In *proceeding of ASME Joint U.S. – European Fluids Engineering Summer Meeting*, USA, July 2006.

[26] J. Figueroa, L. Radaorozandry, J. Cros, and P. Viarouge. Comparative analysis of surface mount pm motor structures for a traction application. In *proceeding of International conference on Electrical Machine (ICEM)*, Chania, Greece, September 2006.

[27] F. Gillon. *Modélisation et optimisation par plans d'expérience déun moteur à commutations électroniques*. PhD thesis, Université des sciences et technologies de lille, 1997. In French.

[28] F. Gillon and P. Brochet. Shape optimization of a permanent magnet motor using the experimental design method. *IEEE Transaction on Magnetics*, 35(3):1278–1281, 1999.

[29] F. Gillon and P. Brochet. Screening and response surface method applied to the numerical optimization of electromagnetic devices. *IEEE Transactions on Magnetics*, 36(4):1163–1167, 2000.

[30] A. Giunta. *Aircraft multidisciplinary design optimization using design of experiments theory and response surface modeling methods*. PhD thesis, Virginia polytechnic institute and state university, May 1997.

[31] A. Giunta, S. Wojtkiewicz, and M. Eldred. Overview of modern design of experiments methods for computational simulations. In *proceeding of the 41st Aerospace Sciences Meeting and Exhibit*, AIAA–2003–649, Reno, Nevada, USA, January 6–9 2003.

[32] S. Giurgea, H. S. Zire, and A. Miraoui. Two-stage surrogate model for finite-element-based optimization of permanent-magnet synchronous motor. *IEEE Transactions on Magnetics*, 43(9):3607–3613, September 2007.

[33] H.-M. Gutmann. A radial basis function method for global optimization. *Journal of Global Optimization*, 19(3):201–227, March 2001.

[34] J. R. Hendershot Jr. and T. J. E. Miller. *Design of brushless permanent-magnet motors*. Magna physics publishing and Clarendon press, Oxford, 1994.

[35] J. W. Herrmann. Decomposition in product development. Technical Report 2004-6, Institute for Systems Research, University of Maryland, College Park, MD 20742, USA, 2004.

[36] S. Hiwa, T. Hiroyasu, and M. Miki. Hybrid optimization using direct, ga, and sqp for global exploration. *IEEE Congress on Evolutionary Computation*, pages 1709–1716, Sept. 2007.

[37] J. H. Holland. *Adaptation in natural and artificial systems*. MIT Press, Cambridge, MA, USA, 1992.

[38] J. Holtz. Pulsewidth modulation - a survey. *IEEE Transactions on Industrial Electronics*, 39(5):410–420, October 1992.

[39] A. Hoppe, M. Kaufmann, and B. Lauber. Multidisciplinary optimization considering crash and nvh loadcases. In *proceeding of ATZ/MTZ Virtual Product Creation*, Stuttgart, 2005.

[40] B. Husslage, G. Rennen, E. R. V. Dam, and D. D. Hertog. Space-filling latin hypercube designs for computer experiments. Discussion Paper 18, Tilburg University, Center for Economic Research, 2006.

[41] C. H. Im, H. K. Kim, H.-K. Jung, and K. Choi. A novel algorithm for multimodal function optimization based on evolution strategy. *IEEE Transactions on Magnetics*, 40(2):1224–1227, March 2004.

[42] R. Jin, W. Chen, and T. W. Simpson. Comparative studies of metamodelling techniques under multiple modelling criteria. *Structural and Multidisciplinary Optimization*, 23(1):1–13, December 2001.

[43] D. R. Jones. A taxonomy of global optimization methods based on response surfaces. *Journal of Global Optimization*, 21(4):345–383, 2001.

[44] D. R. Jones, M. Schonlau, and W. J. Welch. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13(4):455–492, 1998.

[45] J. R. Kalagnanam and U. M. Diwekar. An efficient sampling technique for off-line quality control. *Technometrics*, 39(3):308–319, 1997.

[46] C.-K. Kim and K.-M. Nho. Heat sink design of high power converter. In *proceeding of 29th Annual IEEE Power Electronics Specialists Conference (PESC)*, volume 2, pages 2122–2130, 1998.

[47] H. M. Kim, N. F. Michelena, P. Y. Papalambros, and T. Jiang. Target cascading in optimal system design. *Transaction of ASME: Journal of Mechanical Design*, 125:481–489, 2003.

[48] H. M. Kim, D. G. Rideout, P.Y. Papalambros, and J. L. Stein. Analytical target cascading in automotive vehicle design. *Transaction of ASME: Journal of Mechanical Design*, 125:474–480, 2003.

[49] C. Kitts and S. Wood. *Engineering: Our digital future*. -, 2003.

[50] J. P. C. Kleijnen. Sensitivity analysis of simulation experiments: regression analysis and statistical design. *Mathematics and Computers in Simulation*, 34(3–4):297–315, September 1992.

[51] J. P. C. Kleijnen and W. Van Beers. Kriging for interpolation in random simulation. *Journal of the Operational Research Society*, 2002.

[52] D. Knight, D. G. Elliott, Y. Jaluria, and N. Langrana. Automated optimal design using concurrent integrated experiment and simulation. In *AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, USA, September 4–6 2002.

[53] P. N. Koch, T. W. Simpson, J. K. Allen, and F. Mistree. Statistical approximations for multidisciplinary optimization: The problem of size. *Special Multidisciplinary Design Optimization Issue of Journal of Aircraft*, 36(1):275–286, 1999.

[54] J. R. Koehler and A. B. Owen. Computer experiments. In S. Ghosh and C. R. Rao, editors, *Handbook of Statistics*, volume 13, pages 261–308. Elsevier Science, New York, 1996.

[55] S. Kreuawan, Gillon F., Brochet P., and Cypers D. Optimal design of damping circuit on dc link railway traction system. *International Journal of Electrical Engineering in Transportation (IJEET)*, 4(1):27–32, 2008.

[56] S. Kreuawan, F. Gillon, and P. Brochet. Efficient global optimization, an efficient tool for optimal design. In *proceeding of 16th International Conference on the Computation of Electromagnetic Fields (Compumag)*, Aachen, Germany, 2007.

[57] S. Kreuawan, F. Gillon, and P. Brochet. Comparative study of design approach for electric machine in traction application. *International Review of Electrical Engineering (IREE)*, 3:455–465, May–June 2008.

[58] S. Kreuawan, F. Gillon, and P. Brochet. Multidisciplinary design optimization formulation in electromagnetic device optimal design. In *proceeding of 13th Biennial IEEE Conference on Electromagnetic Field Computation (CEFC)*, Athens, Greece, May 2008.

[59] S. Kreuawan, F. Gillon, and P. Brochet. Surrogate-assisted multiobjective optimization for optimal design using finite element analysis. In *proceeding of 13th Biennial IEEE Conference on Electromagnetic Field Computation (CEFC)*, Athens, Greece, May 2008.

[60] S. Kreuawan, F. Gillon, F. Moussouni, S. Brisset, and P. Brochet. Optimal design of traction motor in railway propulsion system. In *proceeding of International Aegean Conference on Electric Machines, Power Electronics and Electromotion Joint Conference*, pages 343–348, September 2007.

[61] S. Kreuawan, F. Moussouni, F. Gillon, S. Brisset, P. Brochet, and F. Porcher. Optimal design process applied to electric railway traction system. *Transactions on Systems, Signals & Devices*, 2008. To appear in 2008.

[62] V. Lakshmanan. Using a genetic algorithm to tune a bounded weak echo region detection algorithm. *Journal of Applied Meteorology*, 39(2):222–230, 2000.

[63] L. Lebensztajn, C. A. R. Marretto, M. C. Costa, and J.-L. Coulomb. Kriging: a useful tool for electromagnetic device optimization. *IEEE Transactions on Magnetics*, 40(2):1196–1199, March 2004.

[64] M. Lebrun. Simulation et cao en automatique et mécatronique. *Technique de l'ingénieur*. In French.

[65] S. Lok Santosh Naidu. Neural network surrogate model for multidisciplinary design optimization. Master's thesis, Indian Instidute of Technology, Bombay, July 2004.

[66] P. Luniewski and U. Jansen. Benefits of system–oriented igbt module design for high power inverters. In *proceeding of European Conference on Power Electronics and Applications*, pages 1–10, September 2007.

[67] C. N. Madu and C. Kuei. Regression metamodeling in computer simulation - the state of the art. *Simulation Practice and Theory*, 2(1):27–41, January 1994.

[68] F. Marignetti. Coupled electro magnetic thermal and fluid dynamical simulation of axial flux pm synchronous machines. In *proceeding of the COMSOL Users Conference*, Grenoble, France, 2007.

[69] T. Maruyama and H. Igarashi. A reliability analysis of electromagnetic devices using surrogate model. In *proceeding of 13th Biennial IEEE Conference on Electromagnetic Field Computation (CEFC)*, Athens, Greece, May 2008.

[70] W. J. Marx, D. N. Mavris, and D. P. Schrage. Integrated product development for the wing structural design of the high speed civil transport. In *proceeding of 5th AIAA/USAF/NASA/OAI Symposium on Multidisciplinary Analysis and Optimization*, USA, September 1994.

[71] M. D. McKay, R. J. Beckman, and W. J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 42(1):55–61, February 2000.

[72] M. Meckesheimer, A. J. Booker, R. R. Barton, and T.W. Simpson. Computationally inexpensive metamodel assessment strategies. *AIAA Journal*, 40:2053–2060, 2002.

[73] V. Mester. *Conception optimal systémique des composants des chaînes de traction électrique*. PhD thesis, Ecole Centrale de Lille, France, 22 May 2007. In French.

[74] V. Mester, F. Gillon, S. Brisset, and P. Brochet. Global optimal design of a wheel traction motor by a systemic approach of the electric drive train. In *proceeding of IEEE Vehicle Power and Propulsion Conference (VPPC)*, Windsor, UK, September 2006.

[75] V. Mester, F. Gillon, and P. Brochet. Optimal design of electric vehicles drive train. *The International Journal for Computation and Mathematics in Electrical and Electronic Engineering (COMPEL)*, 24(3):967–983, 2005.

[76] V. Mester, F. Gillon, M. Hecquet, and P. Brochet. Multidisciplinary optimal design of electric machines using multimodeling approach. In *proceeding of 12th Biennial IEEE Conference on Electromagnetic Field Computation (CEFC)*, Miami, FL., USA, 2006.

[77] Z. Michalewicz. *Genetic algorithms + data structures = evolution programs*. Springer–Verlag, London, UK, 3rd edition, 1996.

[78] K. Miettinen. *Nonlinear multiobjective optimization*. Kluwer Academic Publishers, 1999.

[79] P. H. Millor, D. Roberts, and D. R. Turner. Lumped parameter thermal model for electrical machine of tefc design. *Electric Power Application, IEE Proceedings B*, 138(5):205–218, 1991.

[80] P. Monté. Smoothing noisy data by kriging with nugget effects. *Wavelet, Images and Surface Fitting*, pages 371–378, 1994.

[81] S. Morimoto, Y. Takeda, T. Hirasa, and K. Taniguchi. Expansion of operating limits for permanent magnet motor by current vector control considering inverter capacity. *IEEE Transactions on Industry Applications*, 26(5):866–871, September/October 1990.

[82] F. Moussouni. *Méthodologie et Outils adaptés à l'optimisation globale et multicritère de systèmes complexes*. PhD thesis, Ecole Centrale de Lille, France, 2008.

[83] F. Moussouni, S. Brisset, and P. Brochet. Some results on the design of brushless dc wheel motor using sqp and ga. *International Journal of Applied Electromagnetics and Mechanics (IJAEM)*, 26(3–4):233–241, January 2007.

[84] F. Moussouni, S. Brisset, and P. Brochet. Fast building of pareto optimal set: Application to a brushless dc wheel motor. In *proceeding of the 10th International Workshop on Optimization and Inverse Problems in Electromagnetism (OIPE)*, pages 48–49, Ilmenau, Germany, September 2008.

[85] F. Moussouni, S. Kreuawan, S. Brisset, F. Gillon, P. Brochet, and L. Nicod. Multilevel design optimization using target cascading. In *proceeding of the 10th International Workshop on Optimization and Inverse Problems in Electromagnetism (OIPE)*, pages 10–11, Ilmenau, Germany, September 2008.

[86] T. Ohnishi and N. Takahashi. Optimal design of efficient ipm motor using finite element method. *IEEE Transactions on Magnetics*, 36(5):3537–3539, September 2000.

[87] P. Y. Papalambros. Extending the optimization paradigm in engineering design. In *proceeding of the 3rd International Symposium on Tools and Methods of Competitive Engineering*, Delft, Netherlands, April 18–21 2002.

[88] D. Petrichenko. *Contribution à la modélisation et à la conception optimal des turbo-alternateurs de faible puissance*. PhD thesis, Ecole Centrale de Lille, France, July 2007.

[89] P. Rohl, D. N. Mavris, and D. P. Schrage. A multilevel decomposition procedure for the preliminary wing design of a high-speed civil transport aircraft. In *proceeding of 1st Industry/University Symposium on High Speed Civil Transport Vehicles*, December 1994.

[90] L. Rondot, V. Mazauric, Y. Delannoy, and Meunier G. Dedicating finite volume method (fvm) to electromagnetic plasma modeling: Circuit breaker application. In *proceeding of 13th Biennial IEEE Conference on Electromagnetic Field Computation (CEFC)*, Athens, Greece, May 2008.

[91] T. Ruimin and W. Erzhi. A new boundary element method for calculating the magnetic field of permanent magnetic machine. In *proceedings of 5th International Conference on Electrical Machines and Systems (ICEMS)*, volume 2, pages 1181–1183, August 2001.

[92] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn. Design and analysis of computer experiments. *Statistical Science*, 4:409–435, 1989.

[93] Magnet sales website. www.magnetsales.com.

[94] M. J. Sasena. *Flexibility qnd Efficiency Enhancements for Constrained Global Design Optimization with Kriging Approximations*. PhD thesis, University of Michigan, USA, 2002.

[95] M. J. Sasena, P. Y. Papalambros, and P. Goovaerts. Metamodeling sampling criteria in a global optimization framework. In *proceeding of 8th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, AIAA–2000–4921, Long Beach, CA, USA, September 6–8 2000.

[96] M. Schonlau. *Computer experiments and global optimization*. PhD thesis, University of Waterloo, Canada, 1997.

[97] M. Schonlau, W. Welch, and D. Jones. Global versus local search in constrained optimization of computer models. *New Developments and Applications in Experimental Design, Institute of Mathematical Statistics, Hayward, California*, 34:11–25, 1998.

[98] H. A. Simon. *The sciences of the artificial*. The MIT Press, 2nd edition, 1981.

[99] T. W. Simpson, J. D. Peplinski, P. N. Koch, and J. K. Allen. Metamodels for computer-based engineering design: Survey and recommendations. *Engineering with Computers*, 17:129–150, 2001.

[100] T. W. Simpson, Z Siddique, and J. Jiao, editors. *Product platform and product family design - Methods and applications*. Springer, 2005.

[101] J. Sobieszczanski-Sobieski, M. S. Emiley, J. S. Agte, and R. R. Sandusky. Advancement of bi-level integrated system synthesis (bliss). In *proceeding of 38th AIAA Aerospace Sciences Meeting and Exhibit, AIAA 2000–0421*, pages 28–30, 2000.

[102] J. Sobieszczanski-Sobieski and R. T. Haftka. Multidisciplinary aerospace design optimization: survey of recent developments. *Structural and Multidisciplinary Optimization*, 14(1):1–23, August 1997.

[103] N. P. Tedford and J. Martins. On the common structure of mdo problems: a comparison of architectures. In *proceedings of the 11th AIAA/ISSMO multidisciplinary analysis and optimization conference*, Portsmouth, VA, September 2006. AIAA 2006-7080.

[104] T. V. Tran, S. Brisset, D. Echeverria, D. Lahaye, and P. Brochet. Space-mapping techniques applied to the optimization of a safety isolating transformer. In *proceeding of the 13th International Symposium on the Electromagnetic Fields (ISEF)*, September 2007.

[105] T. V. Tran, F. Moussouni, S. Brisset, and P. Brochet. Combinatorial optimization of a safety isolating transformer using branch-and-bound method and genetic algorithm.

In *proceeding of 16th International Conference on the Computation of Electromagnetic Fields (Compumag)*, Aachen, Germany, 2007.

[106] S. R. Trout. Material selection of permanent magnets considering thermal properties correctly. In *proceeding of Electrical Insulation Conference and Electrical Manufacturing & Coil Winding Conference*, pages 365–370, USA, October 16–18 2001.

[107] A.M. Trzynadlowski, F. Blaabjerg, J.K. Pedersen, R.L. Kirlin, and S. Legowski. Random pulse width modulation techniques for converter-fed drive systems–a review. *IEEE Transactions on Industry Applications*, 30(5):1166–1175, September/October 1994.

[108] D. A. V. Veldhuizen and G. B. Lamont. Multiobjective evolutionary algorithm test suites. In *proceedings of the 1999 ACM symposium on Applied computing*, pages 351–357, New York, NY, USA, 1999. ACM.

[109] P. Venkataraman. *Applied Optimization with MATLAB Programming*. John Wiley and Son, 2002.

[110] H. K. Versteeg and W. Malalasakera. *An Introduction to Computational Fluid Dynamics – The Finite Volume Method*. Addison Wesley Longman Limited, UK, 1999.

[111] J. A. Visser, D. J. de Kock, and F. D. Conradie. Minimisation of heat sink mass using mathematical optimisation. In *proceeding of 16th Annual IEEE Semiconductor Thermal Measurement and Management Symposium*, pages 252–259, 2000.

[112] S. Vivier. *Stratégie d'optimisation par plans d'expériences et Application aux dispositifs électrotechniques modélisés par éléments finis*. PhD thesis, Université des Sciences et Techniques de Lille, France, July 2002. In French.

[113] S. Wang and J. Kang. Shape optimization of bldc motor using 3-d finite element method. *IEEE Transactions on Magnetics*, 36(4):1119–1123, July 2000.

[114] J. Y. Wong. *Theory of ground vehicles*. John Wiley & Sons, Inc., 3rd edition edition, 2001.

[115] R. Wrobel and P.H. Mellor. Design considerations of a direct drive brushless machine with concentrated windings. *IEEE Transaction on Energy Conversion*, 23(1):1–8, March 2008.