Integration of expert knowledge into radial basis function surrogate models

Zuzana Nedělková · Peter Lindroth · Ann-Brith Strömberg · Michael Patriksson

Abstract A current application in a collaboration between Chalmers University of Technology and Volvo Group Trucks Technology concerns the global optimization of a complex simulation-based function describing the rolling resistance coefficient of a truck tyre. This function is crucial for the optimization of truck tyres selection considered. The need to explicitly describe and optimize this function provided the main motivation for the research presented in this article. Many optimization algorithms for simulation-based optimization problems use sample points to create a computationally simple surrogate model of the objective function. Typically, not all important characteristics of the complex function (as, e.g., non-negativity)—here referred to as expert knowledge—are automatically inherited by the surrogate model. We demonstrate the integration of several types of expert knowledge into a radial basis function interpolation. The methodology is first illustrated on a simple example function and then applied to a function describing the rolling resistance coefficient of truck tyres. Our numerical results indicate that expert knowledge can be advantageously incorporated and utilized when creating global approximations of unknown functions from sample points.

Keywords Radial basis functions · Interpolation · Approximation · Expert knowledge · Optimization · Rolling resistance coefficient

Mathematics Subject Classification 90C20 · 90C34 · 41A29 · 41A30

Zuzana Nedělková, Ann-Brith Strömberg, Michael Patriksson
Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, SE-412 96 Gothenburg, Sweden
E-mail: zuzana@chalmers.se, anstr@chalmers.se, mipat@chalmers.se

Peter Lindroth
Volvo Group Trucks Technology, Chassis & Vehicle Dynamics, Chassis Strategies & Vehicle Analysis, SE-405 08 Gothenburg, Sweden
E-mail: peter.lindroth@volvo.com

1 Introduction

This article proposes a means to integrating certain types of expert knowledge about the reality behind a simulation-based function into a radial basis function (RBF) interpolation, based on sample points. We wish to create an explicit model of an unknown function that realistically relates its design to expert knowledge. The model is then used to represent and/or optimize the function. The research is motivated by a project conducted by Volvo Group Trucks Technology and with the purpose to find an optimal tyre configuration for each truck combination with respect to the rolling resistance, i.e., energy losses caused by the tyre. The function describing the rolling resistance coefficient (RRC) is computationally expensive: each simulation performed by a finite element analysis used in this article requires four hours of computing time ([2]). Many algorithms for solving simulation-based optimization problems utilize a simple surrogate model of the unknown function, which is based on sample points and uses a particular class of interpolating functions. In [21,32,36] a set of sample points is generated in the variable space, the simulation is run for each element of this set, and the surrogate model is based on the resulting values. The surrogate model is then optimized, yielding an approximation of the optimal solution to the unknown function. In order to improve the approximation the surrogate model can be updated with new sample points during the optimization process.

So far, we have considered six variables influencing the RRC: the speed of the truck, the inflation pressure, the width and diameter of the tyre, the spindle load, and the groove depth. We consider the interpolation based on RBF introduced by Wendland [45]. The resulting surrogate model—for which no further sample points can be evaluated—contains inaccuracies and even physically absurd values (e.g., negative values of the RRC, and the model being non-smooth), leading to a poor correspondence with the experts' expectations. This article describes how expert knowledge can be explicitly incorporated into the surrogate model. The surrogate model and its optimal solution are then more easily accepted by the expert community and improved in accordance with physical laws and/or the desired behaviour.

1.1 Previous work

Certain phenomena, such as the rolling resistance, cannot be described by explicit functions, but rather by solutions to, e.g., partial differential equations. In [34, Ch. 2] the rolling tyre is described by so-called fundamental differential equations for a rolling body. The evaluation of the underlying functions is typically then computationally expensive and can only be performed using simulations; derivatives of these functions are hence seldom available. Consequently, optimization problems including such simulation-based functions can in practice not be solved by algorithms requiring many function evaluations, such as, e.g., direct search methods (reviewed in [25]). E.g., the MADS al-

gorithm (described in [3]) finds local optima of the function (the function is assumed to meet certain additional criteria ([1])) which are not guaranteed to be global optima. However, many optimization problems of practical relevance are non-convex and exhibit multiple local optima, thus demanding the use of global optimization techniques for their solution. Many global optimization algorithms developed for solving simulation-based optimization problems provide a surrogate model that mimic the behaviour of the expensive function as closely as possible, while being computationally cheap to evaluate; this surrogate model is then optimized (see the review in [22]). Algorithm 1 is a simple description of a general response surface method.

Algorithm 1 General response surface optimization method

0: Create and evaluate an initial set of sample points.
1: Construct a surrogate model of the simulation-based function using the points evaluated.
2: Refine the surrogate model by selecting and evaluating a new sample point.
3: Go to step 1 unless a stopping criterion is met.
4: Compute an optimal solution to the surrogate model constructed.

The initial set of sample points is created by a design of experiments technique, such as the latin hypercube ([29]). The strategy to select a point to evaluate in step 2 is algorithm specific; it must balance local and global searches so that the information in the surrogate model is utilized, but also so that no part of the feasible set is left unexplored. The stopping criterion in step 3 is problem specific (e.g., a maximum number of function evaluations being calculated, or the attainment of a certain quality measure of the model with respect to some model validation technique, such as the cross-validation studied in [30]). In step 4 the resulting surrogate model is optimized by a global optimization solver; the resulting solution approximates the optimal solution of the underlying simulation-based function. If the function simulation is computationally more demanding than the expert knowledge incorporation then the expert knowledge may be utilized in step 1 else the expert knowledge is used in the last iteration of Algorithm 1.

The expert knowledge may also be utilized when constructing surrogate models needed for local optimization within a trust region algorithm for simulation-based optimization ([46]), or when no optimization is required, but the primary aim is to construct a physically meaningful surrogate of the true function.

The surrogate model can be a linear or quadratic approximation ([9]), a general regression function ([5]), or an interpolation ([33]). Interpolations are usually constructed as linear combinations of basis functions being, e.g., linear functions, thin plate splines, or multi-quadratics. A surrogate model from sample points is typically constructed to have a closed form expression of a number of coefficients and the unknown function's variables; an overview of optimization algorithms utilizing a surrogate model is found in [9]. The coefficients are found as the solution either to a system of equations (e.g., a system

of linear equations for approximation using linear regression; see [11]) or to an optimization problem. Both the approximation by a given function and the linear multiple regression can typically be reformulated as linear programming problems, the objective function and the constraints of which are formed by the approximation or regression requirements (see [11,44]).

The regression and approximation using polynomials are computationally more demanding and the accuracy of fit decreases with an increasing number of variables and/or sample points, as is shown in Section 2.1. Promising alternatives to effectively model the multivariate functions are RBF interpolation ([45]) or Kriging approximation ([26]). While the regression and polynomial approximations usually fit the unknown function only locally, the RBF interpolation and Kriging approximation often yield good global representations of the unknown function (see [7]) and are hence frequently used in algorithms for the global optimization of simulation-based functions (e.g., [16,21,40]). The latter treats more properly the covariance of individual variables ([10]) and is more suitable for noisy simulations ([13]). Kriging approximations are constructed iteratively, thus demanding more computational effort, and the closed form expression of the model of the unknown function is nonlinear with respect to its coefficients; therefore, any expert knowledge will result in nonlinear constraints with respect to the Kriging coefficients. RBF interpolations, on the contrary, are suitable for reformulations as optimization problems, since the corresponding closed form expression of the function model is linear with respect to (wrt.) its coefficients, see Section 3. By relaxing the interpolation requirement—that the value of the surrogate function has to meet the simulated value at each sample point—using a least squares approach a flexibility is released which can be utilized to introduce expert knowledge in terms of linear constraints added to the optimization model. In the case that the sample points are given with an error this relaxation will not result in a worse fit of the surrogate model to the true function. On the other hand, the convergence of the RBF interpolation to the unknown function, which we wish to explicitly approximate, is then no longer guaranteed. The expert knowledge may also give rise to infinitely many constraints (if the surrogate function must obey certain constraints over a whole domain); the resulting problem is then a semi-infinite program, which calls for specialized algorithms ([37, Ch. 3]).

1.2 Motivation

The main contribution of this article is the incorporation of explicit models of expert knowledge in an RBF interpolation. There are always two criteria when creating the surrogate model—the wish to fit the unknown function better over the entire design space and to obey the added expert knowledge. We assume that a decision-maker determines the relative importance of each of these criteria. Here, we consider the special case of the expert knowledge being enforced through explicit constraints on the interpolation coefficients. This case allows for the preservation of some of the properties of the origi-

nal model (especially in domains that are separated from those affected by the description of the expert knowledge, due to the strong locality of RBF interpolation; [12]). If the model would be constructed to fully comply with the expert knowledge—omitting the RBF interpolation—then the quality of the model would be constrained by that of the available expert knowledge. Moreover, then the framework of RBF interpolation could not be utilized to construct the surrogate model.

To the authors' knowledge, there are hitherto no approaches available in the open literature connecting RBF interpolation with expert knowledge aiming at producing surrogate models of unknown simulation-based functions.

1.3 Outline

This article is organized as follows. The components used to construct the surrogate model of the simulation-based function are introduced in Section 2; the RBF interpolation is described and a sample of expert knowledge is listed. In Section 3 we reformulate the RBF interpolation as an unconstrained optimization problem and incorporate expert knowledge into it. Section 4 describes the implementation of each of the types of expert knowledge considered. The methodology is applied to a one-dimensional test function (Section 5), as well as to the RRC function (Section 6). Finally, Section 7 provides conclusions and topics for future research and development.

2 Components of the surrogate model

A number of sample points with corresponding function values are to be interpolated. This can be achieved through the utilization of an RBF interpolation, as described in Section 2.1. To guarantee the convergence of the RBF interpolation to the unknown function, the latter has to be smooth which is a natural assumption for the RRC function. The interpolating function possesses theoretical properties, of which we list those being relevant for this study.

- As the set of data points grows dense, the surrogate models converge locally uniformly to the unknown function on a bounded domain (see [20]).
- The interpolation function is the simplest possible (in terms of its norm) function that interpolates data using the chosen RBF (see [21]).

We wish to include properties of the true function stemming from expert knowledge, which can contribute to reducing the distance between the unknown function and the surrogate model. Our surrogate model is constructed by standard RBF methods ([45, Sec. 8.5]). The theory of RBF interpolation and the corresponding error estimates are described in [45, Ch. 11 and 13], and are briefly summarized next.

2.1 RBF interpolation

The interpolation is the task of determining a continuous function $S_{\boldsymbol{\alpha}} : \mathbb{R}^D \mapsto \mathbb{R}$, where $\boldsymbol{\alpha} \in \mathbb{R}^N$ is a vector of interpolation coefficients, such that at each point in the set $X^N := \{\bar{\mathbf{x}}^1, \ldots, \bar{\mathbf{x}}^N\} \subset \mathbb{R}^D$ of sample points, the unknown function $f : \mathbb{R}^D \mapsto \mathbb{R}$ satisfies

$$S_{\boldsymbol{\alpha}}(\bar{\mathbf{x}}^n) = f(\bar{\mathbf{x}}^n), \quad n = 1, \ldots, N, \tag{1}$$

where $f(\bar{\mathbf{x}}^n)$ is the value of the unknown function computed by a simulation at the sample point $\bar{\mathbf{x}}^n$. Multivariate functions can be interpolated by introducing RBF, as described below.

Definition 1 (radial basis function, RBF). Let $\|\cdot\|$ denote the Euclidean norm. A function $g : \mathbb{R}^D \to \mathbb{R}$ is called a *radial basis function* if there exists a univariate function $\phi : [0, \infty) \to \mathbb{R}$ such that

$$g(\mathbf{x}) = \phi(\|\mathbf{x}\|), \quad \mathbf{x} \in \mathbb{R}^D.$$

The interpolation of multivariate functions using the RBF $g(\cdot) = \phi(\|\cdot\|)$ and a space π_{s-1} of $(s-1)$-degree polynomials on $\mathbb{R}^D$, with dimension $Q := \dim \pi_{s-1}(\mathbb{R}^D)$ and basis $\{p_q\}_{q=1}^Q$, is defined as follows ([45]). Assume that the set X^N of points and the corresponding vector $\mathbf{f} := (f(\bar{\mathbf{x}}^1), \ldots, f(\bar{\mathbf{x}}^N))^{\mathrm{T}} \in \mathbb{R}^N$ of function values are given. Let $\mathbf{x} = (x_1, \ldots, x_D)^{\mathrm{T}} \in \mathbb{R}^D$ denote the vector of variables, and define the surrogate function $S_{\boldsymbol{\alpha}}$ as

$$S_{\boldsymbol{\alpha}}(\mathbf{x}) := \sum_{n=1}^{N} \alpha_n \phi(\|\mathbf{x} - \bar{\mathbf{x}}^n\|) + \sum_{q=1}^{Q} \alpha_{N+q}\, p_q(\mathbf{x}). \tag{2}$$

The interpolation problem is then defined as that to find a vector $\boldsymbol{\alpha} \in \mathbb{R}^{N+Q}$ such that the equations

$$\sum_{n=1}^{N} \alpha_n \phi(\|\bar{\mathbf{x}}^i - \bar{\mathbf{x}}^n\|) + \sum_{q=1}^{Q} \alpha_{N+q}\, p_q(\bar{\mathbf{x}}^i) = f(\bar{\mathbf{x}}^i), \quad i = 1, \ldots, N, \tag{3a}$$

$$\sum_{n=1}^{N} \alpha_n p_q(\bar{\mathbf{x}}^n) = 0, \qquad q = 1, \ldots, Q, \tag{3b}$$

hold. Since p_q, $q = 1, \ldots, Q$, is the basis for π_{s-1}, the equations (3b) are equivalent to the requirement that the vector $(\alpha_1, \ldots, \alpha_N)$ belongs to the space π_{s-1}.

To include only the most common RBFs, such as linear or cubic, we require that not all points in the set X^N belong to a common hyperplane and choose to use the 1st degree polynomials space, π_1. Then, the system (3) of linear equations is non-singular (see [21, p. 33]), and hence possessing a unique solution.

Considering the 1st degree polynomials space π_1 with dimension $Q = D+1$ and basis $(1, x_1, \ldots, x_D)$, the interpolation problem can be reformulated as

follows (see also [21]). Consider a set X^N of $N \geq D+1$ points with a corresponding vector $\mathbf{f} \in \mathbb{R}^N$ of function values. A unique interpolation can be found for an arbitrary $\mathbf{f}$ if at most $N-1$ of the points in X^N belong to a common hyperplane. The vector $\boldsymbol{\alpha} \in \mathbb{R}^{N+1+D}$ determining the interpolation function $S_{\boldsymbol{\alpha}}$ (2) is defined by the following system

$$\begin{pmatrix} \mathbf{A} & \mathbf{P} \\ \mathbf{P}^{\mathrm{T}} & \mathbf{0} \end{pmatrix} \boldsymbol{\alpha} = \begin{pmatrix} \mathbf{f} \\ \mathbf{0} \end{pmatrix} \tag{4}$$

of linear equations, where $\mathbf{A}_{ij} := \phi(\|\bar{\mathbf{x}}^i - \bar{\mathbf{x}}^j\|)$ and $\mathbf{P}_{i\cdot} := (1, (\bar{\mathbf{x}}^i)^{\mathrm{T}})$, $i, j = 1, \ldots, N$. Due to the use of a polynomial space, the matrix of coefficients in (4) is non-singular (see [21]); therefore the system (4) (which is equivalent to the system (3) for our choice of the polynomial space) has a unique solution. Defining, for the sake of simplicity,

$$\tilde{\mathbf{A}} := \begin{pmatrix} \mathbf{A} & \mathbf{P} \\ \mathbf{P}^{\mathrm{T}} & \mathbf{0} \end{pmatrix} \quad \text{and} \quad \tilde{\mathbf{f}} := \begin{pmatrix} \mathbf{f} \\ \mathbf{0} \end{pmatrix},$$

the system (4) of $N+1+D$ equations is equivalently expressed as

$$\tilde{\mathbf{A}}\boldsymbol{\alpha} = \tilde{\mathbf{f}}. \tag{5}$$

Hence, the assignment (2) defines the unique RBF interpolation $S_{\boldsymbol{\alpha}}$ of the unknown function f on the set X^N of sample points, where the vector $\boldsymbol{\alpha}$ of interpolation coefficients is uniquely determined by the system (5) of linear equations. Such models are called RBF interpolations, since $S_{\boldsymbol{\alpha}}(\mathbf{x}) - \sum_{q=1}^{Q} \alpha_{N+q}\, p_q(\mathbf{x})$ is a linear combination of the function $\phi(\|\cdot\|)$ values, which are constant on spheres in $\mathbb{R}^D$ ([46]). When deciding which RBF to use in the surrogate model a strategy for measuring the error (e.g., cross-validation; see [23]) must also be chosen.

2.2 RBF approximation

Constructing an approximation of the unknown function f corresponds to finding a continuous function $S_{\boldsymbol{\alpha}}$ such that, for the given points $\bar{\mathbf{x}}^1, \ldots, \bar{\mathbf{x}}^N \subset \mathbb{R}^D$ and the vector of function values $\mathbf{f} := (f(\bar{\mathbf{x}}^1), \ldots, f(\bar{\mathbf{x}}^N))^{\mathrm{T}}$, it holds that

$$S_{\boldsymbol{\alpha}}(\bar{\mathbf{x}}^n) = f(\bar{\mathbf{x}}^n) + e_n, \quad n = 1, \ldots, N, \tag{6}$$

where $\mathbf{e} := (e_1, \ldots, e_N)^{\mathrm{T}}$ denotes possible perturbations of the vector $\mathbf{f}$ and $\boldsymbol{\alpha} \in \mathbb{R}^{N+1+D}$ denotes the vector of approximation coefficients. The approximation $S_{\boldsymbol{\alpha}}$ should be as simple as possible in terms of its L^2 norm (since physical relations are often simple), and hence the values of $\mathbf{e}$ should be chosen accordingly.

Solving a perturbed (with a vector $\mathbf{e}$) version of the approximation problem (6) is equivalent to solving the system

$$\tilde{\mathbf{A}}\boldsymbol{\alpha} := \begin{pmatrix} \mathbf{A} & \mathbf{P} \\ \mathbf{P}^{\mathrm{T}} & \mathbf{0} \end{pmatrix} \boldsymbol{\alpha} = \begin{pmatrix} \mathbf{f} + \mathbf{e} \\ \mathbf{0} \end{pmatrix} \tag{7}$$

of linear equations. In order to find the surrogate model with the least L^2 norm using the chosen RBF, Jakobsson et al. ([21]) derive a convex quadratic program for determining the perturbation vector $\mathbf{e}$. This program involves a parameter $\eta \in (0, 1)$ controlling the balance between reducing the interpolation error and decreasing the norm of $S_{\boldsymbol{\alpha}}$; its optimality conditions (see [4, Ch. 4]) are given by

$$\left(\eta \mathbf{I}^N - \mathbf{I}^N - \eta \mathbf{B}^{\mathrm{T}} \mathbf{A} \mathbf{B}\right) \mathbf{e} = \eta \mathbf{B}^{\mathrm{T}} \mathbf{A} \mathbf{B} \mathbf{f}, \tag{8}$$

where $\mathbf{B}_{ij} := (\tilde{\mathbf{A}}^{-1})_{ij}$ for $i, j = 1, \ldots, N$. A value of η close to 1 (respectively, 0) means that decreasing the norm of the surrogate model is prioritized much higher (respectively, lower) than minimizing the interpolation error. An appropriate value of η (based on expert knowledge) yields a balance such that some error is allowed, but the characteristics of the sample points are still present.

The assignment (7) then defines the RBF approximation $S_{\boldsymbol{\alpha}}$ of the unknown function f on the set X^N of sample points where the vector $\mathbf{e}$ fulfills (8).

2.3 Expert knowledge

We assume that—in addition to the sample points—expert knowledge about additional properties of the unknown function is available. The information obtained from experts is modeled by requirements on the unknown function and often comes with some uncertainty which has to be considered and processed (e.g., [43]). We next classify—in a mathematical context—the types of specifications of the expert knowledge considered in this article.

Property 1 (types of expert knowledge) *Expert knowledge is classified as specifications of*

a) *function values at certain points or in subdomains,*
b) *derivative values at certain points or in subdomains, and*
c) *functions wrt. variables* $\mathbf{x}$ *(e.g., linear) on subdomains.*

The requirements on function values may stem from physical laws (e.g., non-negativity) or mathematical relations (e.g., the value at a specific point should be close to a convex combination of the closest sample points' values). The values of the function and/or its derivatives may be required to lie in specific ranges (e.g., limit the oscillations of the function).

3 Integration of expert knowledge into RBF interpolation

We wish to take into account the expert knowledge when the interpolation coefficients $\boldsymbol{\alpha}$ of the RBF interpolation are computed. As described in Section 2.1, the RBF interpolation of the sample points is uniquely determined. Therefore we have to introduce a certain degree of freedom—utilizing an appropriate reformulation—in order to allow the surrogate model to take the given expert knowledge into account. The strict interpolation condition (5) then has

to be relaxed. We base this relaxation on least squares, i.e., we search for a model that fits the sample points as good as possible wrt. the sum of squared residuals criterion; see [38, Ch. 2] for details. While the least squares approach utilizes the l^2 norm, the l^1 and l^∞ norms also yield tractable optimization problems.

Since the system (5) comprises $N+1+D$ equations as well as unknowns, additional equations stemming from the expert knowledge lead to an overdetermined system of equations. The vector $\tilde{\mathbf{f}}$ contains the observed function values while the surrogate model predicted values are determined by $\tilde{\mathbf{A}}\boldsymbol{\alpha}$, in the terminology of least squares. Since the RBF interpolation given by the solution to the system (5) does not necessarily conform to the expert knowledge, we solve instead the optimization problem to minimize the squared Euclidean distance between the surrogate model and the sample points, according to

$$\underset{\boldsymbol{\alpha}\in\mathbb{R}^{N+1+D}}{\text{minimize}} \left\| \tilde{\mathbf{A}}\boldsymbol{\alpha} - \tilde{\mathbf{f}} \right\|^2 . \tag{9a}$$

According to Property 1, the expert knowledge can be expressed by inequality and/or equality constraints, as

$$g_t(\boldsymbol{\alpha}) \geq 0, \quad t \in \mathcal{T}, \tag{9b}$$
$$h_r(\boldsymbol{\alpha}) = 0, \quad r \in \mathcal{R}. \tag{9c}$$

The objective function in (9a) is quadratic with a positive definite Hessian (see [6]). Hence, if the functions g_t and h_r are linear and the sets $\mathcal{T}$ and $\mathcal{R}$ are finite, then the optimization problem (9) can be solved in polynomial time (see [24]). Note that each type of expert knowledge listed in Property 1 results in linear constraints wrt. $\boldsymbol{\alpha}$ (i.e., g_t, $t \in \mathcal{T}$, and h_r, $r \in \mathcal{R}$, are linear functions). The surrogate model given by the solution to the optimization problem (9)—approximating the original RBF interpolation—will typically not coincide exactly with the evaluated values of the function f at the sample points. However, the values will be as close as possible into the sense of least squares, taking in account the additional constraints (9b)–(9c).

3.1 On the differences between the interpolation and its approximation

The original RBF interpolation is approximated as a result of the incorporation of the expert knowledge into the model. We next investigate how and to what extent the expert knowledge influences the magnitude of the difference between the original interpolation and its approximation.

The constraints (9b)–(9c) shift the values of the original interpolation $S_{\boldsymbol{\alpha}}$ at the sample points. Due to the uniqueness of the RBF interpolation for the given sample points, adding the expert knowledge to the interpolation problem corresponds to a perturbation of the vector $\mathbf{f}$ in (3a) to $\mathbf{f}+\mathbf{e}$, according to (6). At each sample point $\bar{\mathbf{x}}^n$, the value of e_n is proportional to how much the surrogate model is enforced—by the expert knowledge—to behave differently

(i.e., if $f_{\boldsymbol{\alpha}}(\bar{\mathbf{x}}^n) < 0$, then the requirement that $S_{\boldsymbol{\alpha}}(\bar{\mathbf{x}}^n) \geq 1$ yields a larger value of e_n than does $S_{\boldsymbol{\alpha}}(\bar{\mathbf{x}}^n) \geq 0$). For each specific type of expert knowledge the value of $\mathbf{e}$ is found by solving the corresponding optimization problem (9). Its solution yields an approximation with a minimal residual wrt. the difference between the original and the perturbed vector of function values at the sample points, i.e., between $\mathbf{f}$ and $\mathbf{f} + \mathbf{e}$.

Let $\boldsymbol{\alpha} = \boldsymbol{\alpha}^1$ denote the solution to the system (5). Further, let $\boldsymbol{\alpha} = \boldsymbol{\alpha}^2$ denote the solution to the system (7) of linear equations which is found by solving (9). Defining $\tilde{\mathbf{e}}^{\mathrm{T}} := \left(\mathbf{e}^{\mathrm{T}}, \mathbf{0}^{\mathrm{T}}\right)$ yields the equivalence expression

$$\boldsymbol{\alpha}^2 = \boldsymbol{\alpha}^1 + \tilde{\mathbf{A}}^{-1}\tilde{\mathbf{e}}, \tag{10}$$

implying that

$$\begin{aligned} S_{\boldsymbol{\alpha}^2}(\mathbf{x}) - S_{\boldsymbol{\alpha}^1}(\mathbf{x}) &= \sum_{n=1}^{N} (\tilde{\mathbf{A}}^{-1}\tilde{\mathbf{e}})_n \phi(\|\mathbf{x} - \bar{\mathbf{x}}^n\|) + (\tilde{\mathbf{A}}^{-1}\tilde{\mathbf{e}})_{N+1} + \sum_{d=1}^{D} (\tilde{\mathbf{A}}^{-1}\tilde{\mathbf{e}})_{N+1+d}\, x_d \\ &= \tilde{\mathbf{e}}^{\mathrm{T}} (\tilde{\mathbf{A}}^{-1})^{\mathrm{T}} \left(\phi(\|\mathbf{x} - \bar{\mathbf{x}}^1\|), \ldots, \phi(\|\mathbf{x} - \bar{\mathbf{x}}^N\|), 1, \mathbf{x}^{\mathrm{T}}\right)^{\mathrm{T}}. \end{aligned} \tag{11}$$

Hence, the difference between the interpolation and its approximation depends linearly on the elements of $\tilde{\mathbf{e}}$, i.e, on the perturbations of the function values at the sample points. The expert knowledge corresponds to such a perturbation.

In Section 4 we describe how different specifications of the expert knowledge lead to different types of optimization problems, accordingly motivating designated optimization algorithms. As demonstrated in (11), the inclusions of constraints representing expert knowledge will result in adjustments of the approximation that vary linearly with the enforced difference between the interpolation and its approximation.

4 Implementation of expert knowledge

The specifications listed in Property 1 do not guide towards a specific RBF, or a specific polynomial space. Each specification has to be appropriately modeled, resulting in optimization models with different characteristics. We next develop optimization models for incorporating each of the specifications into the RBF interpolation. We demonstrate that even complex expert knowledge can be modeled using linear constraints. We start with the non-negativity requirement, which motivated this work and is a special case of Property 1.a).

4.1 Non-negativity requirement

Assume that the expert knowledge expresses that the unknown function must be non-negative everywhere. This is a special case of Property 1.a) described

in Section 4.2 with $\varepsilon_m = 0$, $m = 1, \ldots, M$ which results in minimizing the function (9a) subject to the non-negativity constraints according to

$$\underset{\boldsymbol{\alpha}}{\text{minimize}} \ \|\tilde{\mathbf{A}}\boldsymbol{\alpha} - \tilde{\mathbf{f}}\|^2, \tag{12a}$$

$$\text{subject to } S_{\boldsymbol{\alpha}}(\mathbf{x}) \geq 0, \quad \mathbf{x} \in X, \tag{12b}$$

where the surrogate function $S_{\boldsymbol{\alpha}}$ is defined by (2), and the set $X \subseteq \mathbb{R}^D$ is a box, in which the surrogate model $S_{\boldsymbol{\alpha}}$ is assumed to be valid. To enforce non-negativity, the function values $f(\bar{\mathbf{x}}^n)$ at the sample points are assumed to be non-negative. Since the objective function is quadratic and the constraints linear wrt. $\boldsymbol{\alpha}$, the main complexity of the problem (12) stems from the presence of infinitely many constraints: the problem forms a quadratic semi-infinite program. General semi-infinite optimization problems cannot be solved without a discretization, such that the feasible set is defined by a finite number of constraints. Algorithms for solving semi-infinite programs are found in [15,37], covering also theoretical analysis of the algorithms, and in [18], focusing on implementation. We propose Algorithm 2 for solving the optimization problem (12). The algorithm is based on a discretization of the feasible region, which utilizes the minimization of $S_{\boldsymbol{\alpha}}$ over X in order to generate new candidate points in which the non-negativity is required.

Algorithm 2 Calculate $\boldsymbol{\alpha}^P$ defining $S_{\boldsymbol{\alpha}^P}(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in X$

1: Choose the RBF ϕ to be used and the maximum allowed number, Z, of iterations, set $P := 0$, and compute $\boldsymbol{\alpha}^0 = \tilde{\mathbf{A}}^{-1}\tilde{\mathbf{f}}$.

2: Assemble the surrogate model $S_{\boldsymbol{\alpha}^P}$, as defined in (2). Compute

$$\mathbf{x}^P_{\text{opt}} \in \underset{\mathbf{x} \in X}{\text{argmin}} \ S_{\boldsymbol{\alpha}^P}(\mathbf{x}). \tag{13}$$

3: If $S_{\boldsymbol{\alpha}^P}(\mathbf{x}^P_{\text{opt}}) \geq 0$, then terminate: the vector $\boldsymbol{\alpha}^P$ defines the required surrogate model.

4: Compute $\boldsymbol{\alpha}^{P+1}$ as solution to the problem to

$$\begin{aligned} &\underset{\boldsymbol{\alpha}}{\text{minimize}} \ \|\tilde{\mathbf{A}}\boldsymbol{\alpha} - \tilde{\mathbf{f}}\|^2, \\ &\text{subject to } S_{\boldsymbol{\alpha}}(\mathbf{x}^p_{\text{opt}}) \geq 0, \quad p = 0, \ldots, P. \end{aligned} \tag{14}$$

If $P > Z$, then terminate. Otherwise, let $P := P + 1$, and go to step 2.

Summarizing Algorithm 2, in the 1st step the RBF interpolation of the unknown function is computed. In the 2nd step the surrogate model is assembled and minimized. The point corresponding to the most negative function value defines a constraint to be added to the optimization problem, which is solved in the 4th step. The algorithm is terminated if the required surrogate model is found or if the maximum allowed number Z of iterations is attained. The latter termination criteria implies that an additional simulation of the computationally expensive function should be performed and can be replaced by prescribing the minimum allowed difference ε between the optimal values of two consecutive surrogate models, $|S_{\boldsymbol{\alpha}^{P-1}}(\mathbf{x}^{P-1}_{\text{opt}}) - S_{\boldsymbol{\alpha}^P}(\mathbf{x}^P_{\text{opt}})| \leq \varepsilon$.

The optimization problem (14) is a convex quadratic program with finitely many linear constraints and can be solved, e.g., by a modification of the simplex method ([14]) or an interior point method ([28]). The problem (13) is generally a non-convex optimization problem with box constraints. An external global solver (e.g., BARON [41] based on a branch-and-cut approach; see [42]) can be used to find its global minimum. Alternatively, explicit formulations of the first and second partial derivatives of the surrogate function $S_{\boldsymbol{\alpha}^P}$ can be used to find its global minimum using an interval analysis and convex underestimates (see [17]). Convergence of the algorithm used in BARON to a global minimum is guaranteed for the problem (13), because the variables are both upper and lower bounded ([41]).

Algorithm 2, which is a special case of the modified Remez algorithm ([39]), converges to an optimal solution to the original optimization problem (12), provided that the global minimum is computed in the 3$^{\text{rd}}$ step of each iteration; this holds since X is a compact set and since $S_{\boldsymbol{\alpha}}$ is continuous on X by the definition (2). However, Algorithm 2 becomes numerically inefficient for higher dimensional problems, if either the number of iterations P grows very large, or if the simulation-based function f possesses a zero value at some of the sample points. Numerical tests using Algorithm 2 are presented in Section 5.1.

4.2 Prescribed function values

A typical representative of Property 1.a) is the requirement that the surrogate function meets (with a proper tolerance) a specific value at each of a number of specified points. Three approaches to meet this goal are discussed next.

Utilizing expert knowledge as constraints of the optimization problem

The optimization problem corresponding to the requirement to (approximately) meet prescribed function values is to

$$\underset{\boldsymbol{\alpha}}{\text{minimize}} \ \|\tilde{\mathbf{A}}\boldsymbol{\alpha} - \tilde{\mathbf{f}}\|^2, \tag{15a}$$

$$\text{subject to} \ \left|S_{\boldsymbol{\alpha}}(\hat{\mathbf{x}}^{N+m}) - \hat{f}^{N+m}\right| \leq \varepsilon_m, \quad m = 1, ..., M, \tag{15b}$$

where $\hat{f}^{N+m}$ denotes the function value at the point $\hat{\mathbf{x}}^{N+m}$, as given by an expert, and $\varepsilon_m > 0$ denotes the tolerance. The triples $(\hat{\mathbf{x}}^{N+m}, \hat{f}^{N+m}, \varepsilon_m)$ might stem from a one-dimensional approximation of the D-dimensional function f, or represent points of the unknown function with well-known values.

The quadratic program, obtained when each of the M constraints in (15b) is expressed by two affine inequalities, can be solved with a standard quadratic programming solver (see [14,28]); its optimal solution $\boldsymbol{\alpha}^{\text{opt}}$ determines a new surrogate model $S_{\boldsymbol{\alpha}^{\text{opt}}}$ complying with the expert knowledge.

The resulting system of $2M$ affine constraints in (15b) may be inconsistent, indicating that a human error is involved, or that the model is not properly

constructed. In such cases, the model needs to be adjusted to resolve the inconsistency. Lagrange multiplier values (see [4, Section 6.1]) for the $2M$ affine constraints indicate which of them should be weakened. The Lagrange function for the problem (15) is defined as

$$L(\boldsymbol{\alpha}, \boldsymbol{\mu}^1, \boldsymbol{\mu}^2) := \|\tilde{\mathbf{A}}\boldsymbol{\alpha} - \tilde{\mathbf{f}}\|^2 + \sum_{m=1}^{M} \Big[(\mu_m^1 - \mu_m^2) \Big(S_{\boldsymbol{\alpha}}(\hat{\mathbf{x}}^{N+m}) - \hat{f}^{N+m} \Big) - (\mu_m^1 + \mu_m^2)\varepsilon_m \Big]$$

and the corresponding Lagrangian dual problem is given by

$$(\boldsymbol{\mu}^{1*}, \boldsymbol{\mu}^{2*}) \in \operatorname{argmax}_{\boldsymbol{\mu}^1, \boldsymbol{\mu}^2 \in \mathbb{R}_+^M} \{ \operatorname{minimum}_{\boldsymbol{\alpha} \in \mathbb{R}^{N+1+D}} L(\boldsymbol{\alpha}, \boldsymbol{\mu}^1, \boldsymbol{\mu}^2) \}.$$

A large value of any of the Lagrange multipliers, μ_m^{i*}, $i = 1, 2$, indicates that the corresponding expert knowledge is not well represented by the triple $(\hat{\mathbf{x}}^{N+m}, \hat{f}^{N+m}, \varepsilon_m)$, $m = 1, \ldots, M$. If no feasible solution to the constraints (15b) exists, then the tolerances ε_m corresponding to large Lagrange multiplier values, μ_m^{1*} or μ_m^{2*}, should be increased or an alternative surrogate model $S_{\boldsymbol{\alpha}}$ should be chosen. Hence, the values of the tolerances ε_m, required for the system of constraints (15b) to be consistent, reflect the goodness of the chosen RBF for the problem at hand; see Section 5.2 for a specific example.

Utilizing expert knowledge directly

The expert knowledge can be used directly, i.e., the pairs $(\hat{\mathbf{x}}^{N+m}, \hat{f}^{N+m})$ can be included in the set of sample points. Then the updated set of sample points $\hat{X}^{N+M} := \{\bar{\mathbf{x}}^1, \ldots, \bar{\mathbf{x}}^N, \hat{\mathbf{x}}^{N+1}, \ldots, \hat{\mathbf{x}}^{N+M}\}$ and the corresponding vector $\hat{\mathbf{f}} := (f(\bar{\mathbf{x}}^1), \ldots, f(\bar{\mathbf{x}}^N), \hat{f}^{N+1}, \ldots, \hat{f}^{N+M})^{\mathrm{T}}$ of function and expert knowledge values are used in (2)–(3), yielding an updated interpolation, $S_{\boldsymbol{\alpha}^{\mathrm{upd}}}$, of the unknown function. The function $S_{\boldsymbol{\alpha}^{\mathrm{upd}}}$ uses more points than the original interpolation $S_{\boldsymbol{\alpha}}$. Therefore, $S_{\boldsymbol{\alpha}^{\mathrm{upd}}}$ is expected to be a more accurate model of the unknown function, provided that the values given by the expert possesses the same level of accuracy as do the values of the original sample points.

Interpolating sample points and approximating expert points

Even if the values provided by experts are less accurate than the values of the original sample points, they may still be utilized; rather than being interpolated, these points are approximated (see Section 2.2).

Assume that the set $\hat{X}^{N+M}$ of sample and expert points is given, with the corresponding vector $\hat{\mathbf{f}}$ of function values. The corresponding RBF model of the unknown function is expressed as

$$S_{\boldsymbol{\alpha}}(\mathbf{x}) := \sum_{n=1}^{N} \alpha_n \phi(\|\mathbf{x} - \bar{\mathbf{x}}^n\|) + \sum_{m=1}^{M} \alpha_{N+m} \phi(\|\mathbf{x} - \hat{\mathbf{x}}^{N+m}\|) + \alpha_{N+M+1} + \sum_{d=1}^{D} \alpha_{N+M+1+d} x_d,$$

where the vector $\boldsymbol{\alpha} \in \mathbb{R}^{N+M+1+D}$ of coefficients is found by solving a system of $N + M + 1 + D$ linear equations analogical to (7), which contains the unknown

perturbation vector $\mathbf{e}$. We assume here that the perturbations at the original sample points $(\bar{\mathbf{x}}^n, f(\bar{\mathbf{x}}^n))$ are zero, so we require that

$$S_{\boldsymbol{\alpha}}(\bar{\mathbf{x}}^n) = f(\bar{\mathbf{x}}^n), \quad n = 1, \ldots, N,$$

hold, i.e., $e_n = 0$ for $n = 1, \ldots, N$, while some perturbations are possible for the expert points, which is expressed as

$$S_{\boldsymbol{\alpha}}(\hat{\mathbf{x}}^{N+m}) = \hat{f}^{N+m} + e_{N+m}, \quad m = 1, \ldots, M.$$

The corresponding expression of the equations (8), for computing the perturbation vector $\mathbf{e}$, is given by

$$\left[\begin{pmatrix} (\eta_1 - 1)\eta_1^{-1}\mathbf{I}^N & \mathbf{0}^{N\times M} \\ \mathbf{0}^{M\times N} & (\eta_2 - 1)\eta_2^{-1}\mathbf{I}^M \end{pmatrix} - \mathbf{B}^{\mathrm{T}}\mathbf{A}\mathbf{B} \right] \mathbf{e} = \mathbf{B}^{\mathrm{T}}\mathbf{A}\mathbf{B}\mathbf{f}, \tag{16}$$

where the value of $\eta_1 \in (0, 1)$ is chosen to be close to 0 in order to near-interpolate the real function at the N sample points, while $\eta_2 \in (0, 1)$ is assigned an appropriate value to approximate the M points given by the expert knowledge; a value of η_2 near 0 (respectively, 1) designates a high (respectively, low) level of confidence of the expert knowledge. The roles of η_1 and η_2 can be interchanged if, e.g., the expert sample points are entirely trusted and the original sample points are assumed to be less accurate. The expression in the left-hand side of (16) can be extended to more than two levels of accuracy among the given points.

The perturbation vector $\mathbf{e}$ computed in (16) is then used in the analogy of (7) with $N + M + 1 + D$ equations to find $\boldsymbol{\alpha}^{\mathrm{approx}}$ and the corresponding surrogate function $S_{\boldsymbol{\alpha}^{\mathrm{approx}}}$, which interpolates the N sample points and approximates the M expert points.

Numerical tests of the three approaches described in this subsection, i.e., [i] utilizing the expert knowledge as constraints of the optimization problem, [ii] interpolating the points given by the expert knowledge, and [iii] interpolating the sample points while approximating the expert points, are presented in Section 5.2.

4.3 Boundedness requirement on function derivatives

Expert knowledge may imply that the derivatives of the surrogate function $S_{\boldsymbol{\alpha}}$ are bounded [Property 1.b)]. The partial derivatives wrt. x_d, $d = 1, \ldots, D$, of the function $S_{\boldsymbol{\alpha}}$ defined in (2) are given by

$$\frac{\partial}{\partial x_d} S_{\boldsymbol{\alpha}}(\mathbf{x}) = \sum_{n=1}^{N} \alpha_n \frac{\partial \phi\left(\|\mathbf{x} - \bar{\mathbf{x}}^n\|\right)}{\partial x_d} + \alpha_{N+1+d}, \quad d = 1, \ldots, D. \tag{17}$$

Note that, for any choice of an explicit radial basis function, there exists an explicit expression of the derivative of the univariate function ϕ.

From (17) we conclude that the gradient (wrt. $\mathbf{x}$) of the function $S_{\boldsymbol{\alpha}}$ is linear wrt. the interpolation coefficients $\boldsymbol{\alpha}$. Hence, the optimization problem including requirements on the derivatives of the surrogate model can be formulated as to

$$\begin{aligned} &\underset{\boldsymbol{\alpha}}{\text{minimize}} \ \|\tilde{\mathbf{A}}\boldsymbol{\alpha} - \tilde{\mathbf{f}}\|^2, \\ &\text{subject to} \ \left| \frac{\partial}{\partial x_d} S_{\boldsymbol{\alpha}}(\hat{\mathbf{x}}^m) - \hat{b}^m \right| \leq \varepsilon_{md}, \quad m = 1, ..., M, \ d = 1, \ldots, D, \end{aligned} \tag{18}$$

where $\hat{b}^m$ is the desired value of the derivative at the point $\hat{\mathbf{x}}^m$ and $\varepsilon_{md} > 0$ represent the tolerance levels. Note that the optimization problem (18) possesses the same mathematical properties as (15); it may thus be solved by a standard quadratic programming solver. The optimal solution $\boldsymbol{\alpha}^{\text{opt}}$ to (18) determines a new surrogate model $S_{\boldsymbol{\alpha}^{\text{opt}}}$ fulfilling the expert knowledge.

The expert knowledge may imply that also higher-order derivatives should be constrained. E.g., obtaining a smooth enough surrogate function requires upper and lower bounds on its second derivatives, which are linear wrt. the coefficients $\boldsymbol{\alpha}$ [cf. (17)]. Numerical tests on the constrained second derivative are presented in Section 5.3.

4.4 Specific functions prescribed on subdomains

The type of expert knowledge listed as Property 1.c) concerns prescribed values of the unknown function on lower-dimensional subdomains. Such function values may stem from an accurate lower-dimensional approximation of the unknown function or from other physical models.

The search for a surrogate function that matches a given function on a given subdomain leads to the optimization problem to

$$\begin{aligned} &\underset{\boldsymbol{\alpha}}{\text{minimize}} \ \|\tilde{\mathbf{A}}\boldsymbol{\alpha} - \tilde{\mathbf{f}}\|^2, \\ &\text{subject to} \ S_{\boldsymbol{\alpha}}(\mathbf{x}) = h_r(\mathbf{x}), \quad \mathbf{x} \in X_r, \ r = 1, ..., R, \end{aligned} \tag{19}$$

where $h_r : \mathbb{R}^D \mapsto \mathbb{R}$ is a given function with which the surrogate model must coincide on the subdomain $X_r \subseteq X$. The program (19) is semi-infinite with a quadratic objective function and an infinite number of linear constraints.

A finite-dimensional approximation of (19) is created by discretizations of the sets X_r, $r = 1, \ldots, R$. The finite optimization problem then is to

$$\begin{aligned} &\underset{\boldsymbol{\alpha}}{\text{minimize}} \ \|\tilde{\mathbf{A}}\boldsymbol{\alpha} - \tilde{\mathbf{f}}\|^2, \\ &\text{subject to} \ S_{\boldsymbol{\alpha}}(\hat{\mathbf{x}}^{rn}) = h_r(\hat{\mathbf{x}}^{rn}), \quad n = 1, \ldots, N_r, \ r = 1, \ldots, R, \end{aligned} \tag{20}$$

where $\{\hat{\mathbf{x}}^{r1}, \ldots, \hat{\mathbf{x}}^{rN_r}\} \subset X_r$ represents the discretization of X_r. The optimal solution $\boldsymbol{\alpha}^{\text{opt}}$ of (20) tends to the optimal solution of (19) as the set $\{\hat{\mathbf{x}}^{r1}, \ldots, \hat{\mathbf{x}}^{rN_r}\}$ grows dense in X_r (i.e., as $N_r \to \infty$); see [39].

Alternatively, an algorithm for solving semi-infinite programs can be applied. Such algorithms include interior point methods ([35]), in which the convergence properties are dependent on the properties of the matrix $\tilde{\mathbf{A}}$, and global optimization algorithms ([31]). A simple example (with an affine function h_r) of this kind of expert knowledge is presented in Section 5.4.

4.5 Other kinds of expert knowledge

One may of course come across other kinds of expert knowledge, such as, e.g., negativity requirements, or relations to other functions on given domains. These can be implemented by means similar to those presented above, as long as they allow for constraint descriptions that are simple enough wrt. the interpolation/approximation coefficients $\boldsymbol{\alpha}$.

5 Numerical examples

We next describe and perform numerical experiments testing the performance of the methodology proposed in Section 4 by using a known, simple one-dimensional test function as the "unknown" function. The accuracy of the approximation can easily be measured.

All numerical studies were performed in MATLAB R2012b ([27]). The finite quadratic programming problems (14), (15), (18), and (20), were solved by the trust region implementation `quadprog.m` from the MATLAB Optimization Toolbox ([8]). The tolerance $1e^{-12}$ wrt. the function values was used as termination criterion. All the algorithms were initiated with $\boldsymbol{\alpha}^0 = (1, \dots, 1)^{\mathrm{T}}$. To find global minima of the non-convex programs (13) the global solver BARON was used. The cubic radial function was used for all the RBF interpolations.

5.1 Non-negativity requirement

Consider the simple one-dimensional function f defined by

$$f(x) := c_1 - c_2 \cos x + \frac{c_3}{x}, \quad x \in \mathbb{R}, \tag{21}$$

where $c_1 := 4.75$, $c_2 := 5$, and $c_3 := 0.04$. Since this particular function is one-dimensional, a spline interpolation would be more suitable. This example is, however, used to demonstrate the methodology presented, and which is designed primarily for higher dimensions. The function (21) and the nine sample points $\{\bar{x}^1, \dots, \bar{x}^9\} := \{0.01, 0.10, 0.28, 0.58, 0.66, 0.72, 0.76, 0.98, 1\}$ are plotted in the left part of Figure 1, which also illustrates the RBF interpolation $S_{\boldsymbol{\alpha}^0}$ of these sample points. We assume, according to the expert knowledge, that the surrogate model has to be non-negative.

The original RBF interpolation forms the surrogate model $S_{\boldsymbol{\alpha}^0}$ [see (14)] which obviously does not fulfill the expert's non-negativity requirement. The

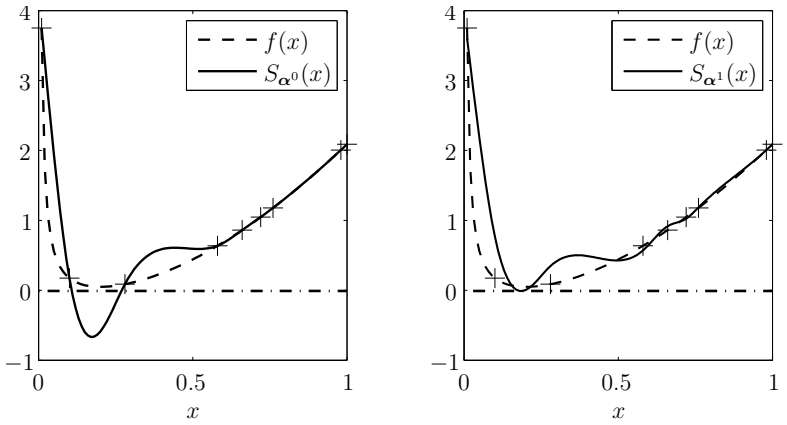

Fig. 1 (Left) The test function f with sample points (+) and its RBF interpolation $S_{\boldsymbol{\alpha}^0}$. (Right) The surrogate model $S_{\boldsymbol{\alpha}^1}$ which meets the non-negativity requirement, as computed by Algorithm 2. The function f is plotted for comparison.

desired non-negative model $S_{\boldsymbol{\alpha}^1}$ was found at the 2^{nd} iteration of Algorithm 2; see the rightmost plot of Figure 1.

Since our test function is known and the RBF interpolation yields an explicit function $S_{\boldsymbol{\alpha}}$, a one-dimensional measure of the error of the P^{th} model (P denoting the number of iterations in Algorithm 2) is given by their L^i distance for $i = 1, 2$, i.e.,

$$v_{Pi} := \left(\int_0^1 |S_{\boldsymbol{\alpha}^P}(x) - f(x)|^i \, \mathrm{d}x \right)^{1/i}. \tag{22}$$

From Table 1 we deduce that the model $S_{\boldsymbol{\alpha}^1}$, which in addition to the sample points uses the expert knowledge about the non-negativity of the unknown function at $x^0_{\text{opt}} = 0.2$ found when minimizing the original surrogate model, is more accurate than the original RBF interpolation $S_{\boldsymbol{\alpha}^0}$ wrt. the L^1 distance (22) and its correspondence with the expert knowledge is increased. The L^2 distance is, however, higher for the model $S_{\boldsymbol{\alpha}^1}$. Note also that a transformation of function values prior constructing the surrogate model can be advantageously used to fulfill the expert expectations in this case.[1]

Table 1 Errors of different surrogate models

	v_{P1}	v_{P2}
Model $S_{\boldsymbol{\alpha}^0}$	0.29	1.00
Model $S_{\boldsymbol{\alpha}^1}$	0.26	1.01

[1] A logarithmic transformation possesses a non-negative and smooth surrogate model of the unknown function in this case. On the other hand, the resulting surrogate function then becomes exponential, which may contradict the expert knowledge.

5.2 Prescribed function values

The expert may state that the function must not oscillate. This requirement may be expressed, e.g., such that the surrogate function should interpolate the centre of a line between two specific sample points. In Figure 2, an undesired oscillation and two sample points (+) are circled (left) and the expert point (•), denoted $(\hat{x}^1, f(\hat{x}^1))$, is marked (right). The implementations of the three approaches for the requirements described in Section 4.2 are described below.

Utilizing expert knowledge as constraints of the optimization problem

This approach amounts to solving the optimization problem (15), yielding $S_{\boldsymbol{\alpha}^{\text{opt}}}$. Here, $M := 1$, representing the point $(\hat{x}^1, f(\hat{x}^1)) = (0.19, 0.13)$; we set the tolerance $\varepsilon_1 := 0.2$, reflecting the requirement that $S_{\boldsymbol{\alpha}^{\text{opt}}}(\hat{x}^1) \in [\, f(\hat{x}^1) - \varepsilon_1, f(\hat{x}^1) + \varepsilon_1]$. The resulting model is shown in Figure 2 (right).

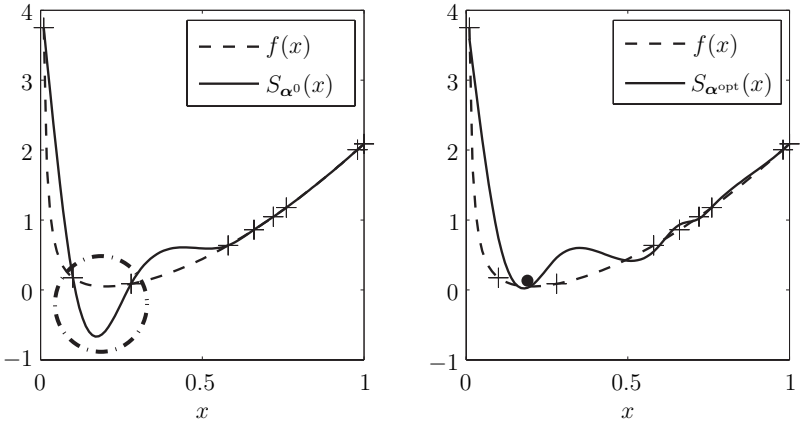

Fig. 2 Illustrations of the numerical test of prescribed function values utilized as constraints of the optimization problem. The true (i.e., "unknown") function f, the sample points (+), and the expert point (•) are plotted for comparison. (Left) The original RBF interpolation $S_{\boldsymbol{\alpha}^0}$ (with the undesired oscillation marked with a circle). (Right) The updated surrogate model $S_{\boldsymbol{\alpha}^{\text{opt}}}$ given by the optimal solution to (15).

The result from this approach, when adding just one expert point, does not fully accomplish our desires. While the oscillation is limited around the expert point, the function $S_{\boldsymbol{\alpha}^{\text{opt}}}$ oscillates in other subdomains and, hence, does not approximate the true function well.

We add the expert point $(\hat{x}^2, f(\hat{x}^2)) = (0.62, 0.75)$ in order to compare the values of the Lagrange multipliers for the constraints (15b) corresponding to the two expert points. We choose $\varepsilon_2 := \varepsilon_1 = 0.2$. The resulting Lagrange multiplier values $\mu_1^{1*} = 1.7e^{-17}$ and $\mu_1^{2*} = 1.48$ (corresponding to the triple $(\hat{x}^1, f(\hat{x}^1), \varepsilon_1)$), and $\mu_2^{1*} = 3.8e^{-17}$, and $\mu_2^{2*} = 3.2e^{-17}$ (corresponding to the

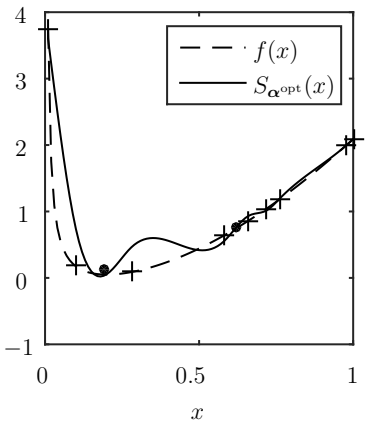

Fig. 3 Illustration of Lagrange multipliers comparison when the expert points (•) are utilized to define the constraints (15b).

triple $(\hat{x}^2, f(\hat{x}^2), \varepsilon_2))$ indicate that the chosen RBF surrogate model is more suitable for obeying the additional expert point $(\hat{x}^2, f(\hat{x}^2))$ than the initial one, $(\hat{x}^1, f(\hat{x}^1))$. Hence the tolerance $\varepsilon_2 = 0.2$ should be decreased. See Figure 3 for an illustration.

Utilizing expert knowledge directly

This approach amounts to use the point given by expert knowledge directly as a new sample point in the construction of $S_{\boldsymbol{\alpha}^{\mathrm{upd}}}$. The resulting model is illustrated in Figure 4 (left).

The result of this approach is a much less oscillating function, which also fits the unknown function better. Table 2 shows the errors, v_{Pi} defined in (22), for all the surrogate models utilizing the point given by expert knowledge. Nevertheless, this model still contains an undesirable dip, that can be restricted by posing constraints on the smoothness of the model

Table 2 Errors of surrogate models for three approaches for utilizing the prescribed function value and error of the original surrogate model

	v_{P1}	v_{P2}
Model $S_{\boldsymbol{\alpha}^0}$	0.29	1.00
Model $S_{\boldsymbol{\alpha}^{\mathrm{opt}}}$	0.28	1.01
Model $S_{\boldsymbol{\alpha}^{\mathrm{upd}}}$	0.18	0.96
Model $S_{\boldsymbol{\alpha}^{\mathrm{approx}}}$	0.16	0.96

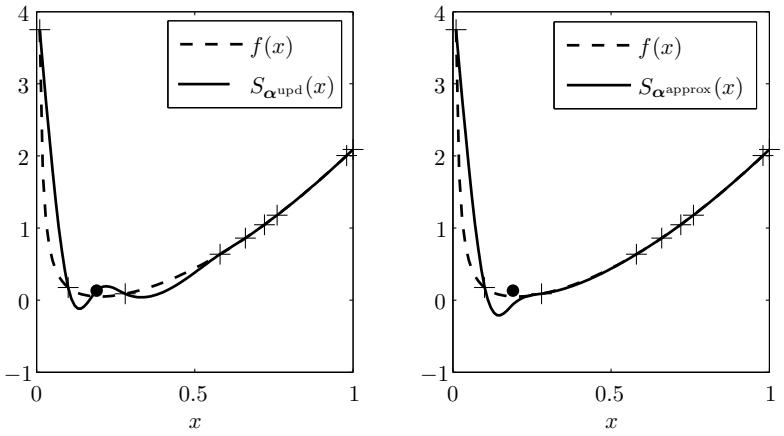

Fig. 4 Illustration of the numerical test of prescribed function values utilized directly with the resulting surrogate function $S_{\boldsymbol{\alpha}^{\mathrm{upd}}}$ (Left). The function $S_{\boldsymbol{\alpha}^{\mathrm{approx}}}$ interpolates the sample points and approximates the expert point (Right).

Interpolating sample points and approximating expert points

The third approach—constructing the surrogate model $S_{\boldsymbol{\alpha}^{\mathrm{approx}}}$ of the unknown function by interpolating the sample points and approximating the expert point—is suitable if we are doubtful about the expert knowledge. The resulting surrogate model depends on the tuning of the coefficients η_1 and η_2 in (16). Their values are given by expert knowledge. The resulting surrogate model $S_{\boldsymbol{\alpha}^{\mathrm{approx}}}$, with $\eta_1 = 1e^{-5}$ and $\eta_2 = 2.5e^{-4}$, is shown in Figure 4 (right).

We conclude that the third approach with well chosen values of η_1 and η_2 performs the best among the three approaches implemented for this simple case; see Table 2. As an alternative to adding a point, the oscillations can be avoided by requiring a positive curvature of the surrogate model on the interval $[0, 1]$, which can be implemented using the methodology described in Section 4.3.

All three approaches presented in this subsection provide a better goodness of fit than the original RBF interpolation, while using the same expert knowledge. Note that the goodness of fit always depends on the individual application as well as on the accuracy of the expert knowledge employed.

5.3 Boundedness requirement on function derivatives

In order to to obtain a surrogate model of the unknown function that is smoother than the original RBF interpolation, we have utilized the second derivative of the surrogate function over its domain; see [47]. Suppose that the expert knowledge prescribes the values of the second derivatives of the surrogate model to be in the interval $[-b, b]$, over the domain $X := (0, 1]$. This

is represented as constraints in the optimization problem (18), as

$$\underset{\boldsymbol{\alpha}}{\text{minimize}} \ \|\tilde{\mathbf{A}}\boldsymbol{\alpha} - \tilde{\mathbf{f}}\|^2, \tag{23a}$$

$$\text{subject to} \ \left| \frac{\mathrm{d}^2}{\mathrm{d}x^2} S_{\boldsymbol{\alpha}}(x) \right| \leq b, \quad x \in X. \tag{23b}$$

The optimization problem (23) is semi-infinite, since it includes infinitely many linear constraints. Its finite reformulation is obtained by representing X by the finite set of points $\{\hat{x}^1, \ldots, \hat{x}^M\} \subset X$, as that to

$$\begin{aligned} &\underset{\boldsymbol{\alpha}}{\text{minimize}} \ \|\tilde{\mathbf{A}}\boldsymbol{\alpha} - \tilde{\mathbf{f}}\|^2, \\ &\text{subject to} \ \left| \frac{\mathrm{d}^2}{\mathrm{d}x^2} S_{\boldsymbol{\alpha}}(\hat{x}^m) \right| \leq b, \quad m = 1, \ldots, M. \end{aligned} \tag{24}$$

The resulting surrogate model $S_{\boldsymbol{\alpha}^{\mathrm{opt}}}$ for $b = 12$ is plotted in Figure 5 together with the original RBF interpolation $S_{\boldsymbol{\alpha}^0}$. The maximum curvature of the re-

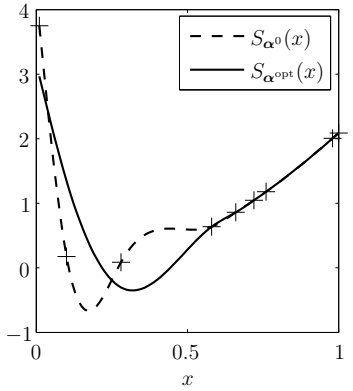

Fig. 5 Illustration of the numerical test of prescribed values of function derivatives. The second derivative of the surrogate model $S_{\boldsymbol{\alpha}^{\mathrm{opt}}}$ is constrained over its domain; the original interpolation $S_{\boldsymbol{\alpha}^0}$ is plotted for comparison.

sulting function is $\max_{x\in[0,1]} \left| \frac{\mathrm{d}^2}{\mathrm{d}x^2} S_{\boldsymbol{\alpha}_{\mathrm{opt}}}(x) \right| = 10.8$, while that of the original surrogate function is $\max_{x\in[0,1]} \left| \frac{\mathrm{d}^2}{\mathrm{d}x^2} S_{\boldsymbol{\alpha}_0}(x) \right| = 218$; we conclude that the resulting function is smoother than the original interpolation. The L^1 and L^2 distances (22), i.e., $v_{\mathrm{opt1}} = 0.35$ and $v_{\mathrm{opt2}} = 1.09$, respectively, of this model and of the function f are larger than that of the original interpolation. This indicates that expert knowledge about the second derivative should be combined with some other kind of expert knowledge (e.g., prescribed function values) in order to yield a satisfactory result.

5.4 Specific functions prescribed on subdomains

Suppose that expert knowledge provides the information that the unknown function is affine on the closed interval $X_1 = [0.28, 1]$ (see Figure 6). This is represented by the constraints of the optimization problem (19) to

$$\begin{aligned} &\underset{\boldsymbol{\alpha}}{\text{minimize}} \ \|\tilde{\mathbf{A}}\boldsymbol{\alpha} - \tilde{\mathbf{f}}\|^2, \\ &\text{subject to } \ S_{\boldsymbol{\alpha}}(x) = ax + b, \quad x \in X_1, \end{aligned} \tag{25}$$

where the values of a and b are given. A finite reformulation of the semi-infinite optimization problem (25) is obtained by representing X_1 by a finite set $\{\hat{x}^{11}, \ldots, \hat{x}^{1N_1}\} \subset X_1$ of points, according to

$$\begin{aligned} &\underset{\boldsymbol{\alpha}}{\text{minimize}} \ \|\tilde{\mathbf{A}}\boldsymbol{\alpha} - \tilde{\mathbf{f}}\|^2, \\ &\text{subject to } \ S_{\boldsymbol{\alpha}}(\hat{x}^{1n}) = a\hat{x}^{1n} + b, \quad n = 1, ..., N_1. \end{aligned} \tag{26}$$

The optimal solution to (26) for $N_1 = 9$ yields the surrogate model $S_{\boldsymbol{\alpha}^{\text{opt}}}$ in Figure 6 (right).

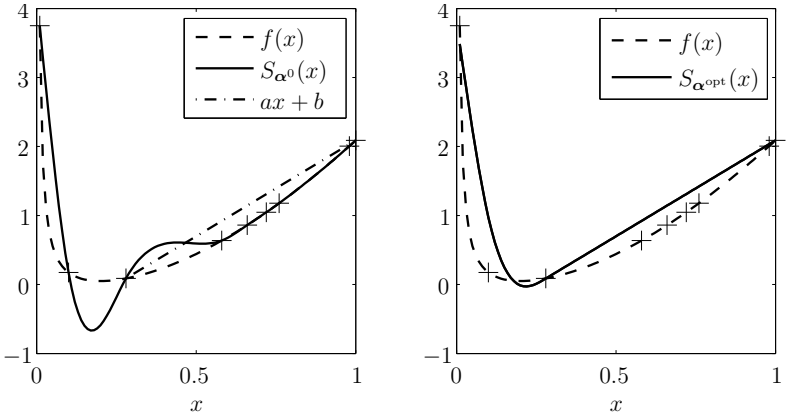

Fig. 6 Illustration of the numerical test of a prescribed function type. (Left) The original RBF interpolation $S_{\boldsymbol{\alpha}^0}$ (solid line) and the required function (dash-dotted line). (Right) the updated surrogate model $S_{\boldsymbol{\alpha}^{\text{opt}}}$ (solid line) given by the solution to (26). The true function f is plotted for comparison (dashed line).

The L^1 and L^2 distances of the new surrogate model $S_{\boldsymbol{\alpha}^{\text{opt}}}$ and the true function f are larger than that of the original RBF interpolation, because the linear part is forced to lie above the sample points. But the function $S_{\boldsymbol{\alpha}^{\text{opt}}}$ can be more suitable for some applications, e.g., when the minimum of the function f in (21) is to be localized.

The above illustrations indicate that expert knowledge of several different, rather complex, types can be expressed by a finite number of affine constraints and therefore also successfully incorporated in the surrogate model, in terms of solving tractable optimization problems.

6 The rolling resistance coefficient function

We next describe the application that motivated the research described in this article, namely the analysis of the RRC of truck tyres.

In order to improve the fuel efficiency of truck transportation various energy losses must be reduced. One of the most important losses is caused by the rolling resistance, which can be described as the effort required to keep a given tyre rolling. Rolling resistance includes energy losses due to aerodynamic drag associated with rolling, and within the structure of the tyre; it also includes friction between the tyre and the road and between the tyre and the rim.

The rolling resistance is determined by the RRC. Ali et al. [2] identified that the RRC is influenced mainly by the tyre inflation pressure, the tyre width, the tyre diameter, the groove depth, the vehicle speed, and the vertical load on the corresponding axle. The RRC can hence be represented by a six-dimensional function. In order to explicitly describe this function we have utilized the function values (all of which are positive) for a set X^N of $N = 252$ sample points from a finite element analysis (FEA) model of a representative truck tyre; see [2] for detailed descriptions of the tyre and the FEA model, for which each evaluation requires four hours of computation time. The RBF-based interpolation described in this article is next used to construct a surrogate model of the RRC function such that the most important characteristics of the FEA model are retained.

The surrogate model $S_{\boldsymbol{\alpha}}$ of the RRC as defined in (2) is found by solving the system (4) of linear equations. We utilized cross-validation (see [23]) to show that the linear RBF provides the best fit of the unknown function.

Incorporating the resulting model of the RRC into a complex joint model of the vehicle, the tyre, and the road revealed some contradictions with existing expert knowledge. Figure 7 shows a cut through the six-dimensional space parallel to the load axis and which illustrates the unphysical properties of the original model $S_{\boldsymbol{\alpha}^0}$. For this model the RRC possesses negative values in the interval $[6600, 8700]$ lbs (see Figure 7), which contradicts the physical meaning of the RRC. The function $S_{\boldsymbol{\alpha}^0}$ also possesses "kinks" at some of the sample points over the load axis (see the neighborhood of 6000 lbs in Figure 7) which is not desired from the experts' point of view. Below, we implement these two types of expert knowledge into the surrogate model of the RRC function.

To eliminate the negative values and the "kinks" we solve the optimization problem (12) with additional constraints (23b) to restrict the second partial derivative wrt. the load. In Figure 7 the same cut as for the model $S_{\boldsymbol{\alpha}^0}$ is plotted for the resulting model $S_{\boldsymbol{\alpha}^{\mathrm{opt}}}$.

The non-negativity of the updated surrogate model is established by the non-negative value of its global minimum, which could be found due to the explicit expression of the function and its first and second derivatives.

An approximate measure of the smoothness of the function $S_{\boldsymbol{\alpha}}$ can be provided by its coefficient of variation (see [19]), which can be computed through a discretization of the six-dimensional space and represented by W discretiza-

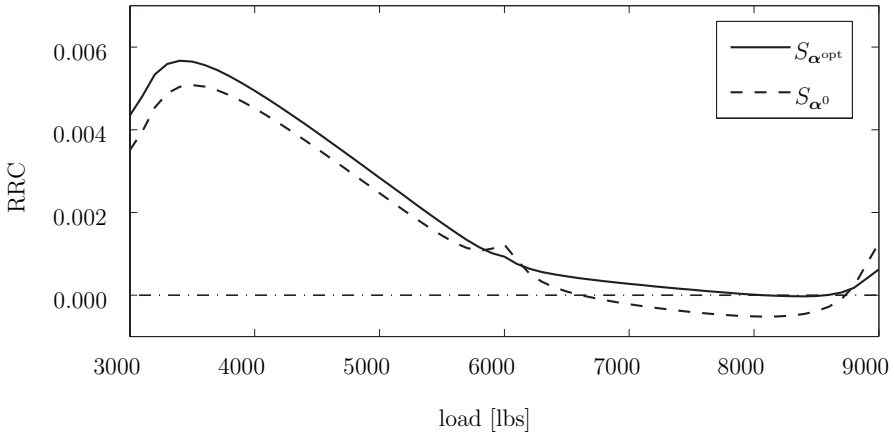

Fig. 7 Illustration of the undesired behaviour of the original model $S_{\boldsymbol{\alpha}^0}$ describing the RRC. In this cut, the negative values are attained in the interval $[6600, 8700]$ lbs and the "kink" is exhibited at the load = 6000 lbs. The undesired behaviour is limited for the updated model $S_{\boldsymbol{\alpha}^{\mathrm{opt}}}$, whose values are all non-negative while the "kink" is essentially removed.

tion points $\tilde{\mathbf{x}}^1, \ldots, \tilde{\mathbf{x}}^W$. Evaluations of the surrogate model of the RRC at these points then yield the values $S_{\boldsymbol{\alpha}}(\tilde{\mathbf{x}}^1), \ldots, S_{\boldsymbol{\alpha}}(\tilde{\mathbf{x}}^W)$. The coefficient of variation is defined as

$$c_{\mathrm{var}} := \frac{\sigma}{\mu},$$

where

$$\mu := \frac{1}{W}\sum_{w=1}^{W} f_{\mathrm{r}}(\tilde{\mathbf{x}}^w) \quad \text{and} \quad \sigma := \sqrt{\frac{1}{W}\sum_{w=1}^{W}[f_{\mathrm{r}}(\tilde{\mathbf{x}}^w) - \mu]^2}. \tag{27}$$

The smoothness $c_{\mathrm{var}} = 0.64$ of the model $S_{\boldsymbol{\alpha}_0}$ was slightly improved to $c_{\mathrm{var}} = 0.59$ when utilizing the expert knowledge (i.e., for the model $S_{\boldsymbol{\alpha}_{\mathrm{opt}}}$).

Since the updated surrogate model of the RRC is non-negative and "smoother" than the original one, we conclude that it is more easily acceptable by experts. However, the model still does not correspond fully to the expert expectations—at the end points of the load interval the function looks slightly unphysical wrt. smoothness. This behaviour is due to the limiting of the second derivative wrt. load and not wrt. other variables. The RRC attains the value zero (e.g., for load 8100 lbs), which neither seems to be natural (to eliminate this drawback the constraints (12b) can be replaced by $S_{\boldsymbol{\alpha}}(\mathbf{x}) \geq \varepsilon$, if an expert knowledge about the value of $\varepsilon > 0$ is available). The overall accuracy of the surrogate model of the true RRC function can be assessed, e.g., by cross-validation or bootstrap; see [23]. The leave-one-out cross-validation value equals 0.002 for both surrogate models because it assesses how well the surrogate model coincide with the simulations not with the expert knowledge.

The surrogate model $S_{\boldsymbol{\alpha}_{\mathrm{opt}}}$ of the RRC was incorporated into surrogate functions to define the objective and constraints of an optimization model for

finding an optimal set of tyres for each vehicle and operating environment specification. Available algorithms to solve this problem are currently investigated and will be the subject of a future publication.

7 Conclusions

We have shown that utilizing expert knowledge in developing a surrogate function of an unknown function is possible and is often cheaper than performing additional costly simulations of the unknown function. We show that utilizing expert knowledge may lead to an increase of goodness of fit as compared to interpolating the sample points. We show that also the surrogate model can be enhanced by incorporating reliable expert knowledge. We have identified how the RBF interpolation can be reformulated as a tractable optimization problem allowing for the utilization of constraints stemming from expert knowledge.

The methodology of approximating an unknown function combining the RBF interpolation with expert knowledge was described and illustrated on a simple one-dimensional problem and on the six-dimensional case of the function describing the rolling resistance coefficient of a truck tyre. The user of the approach described always has to decide whether he or she prefers to fulfil the expert knowledge or the guaranteed accuracy of the RBF interpolation, even though these two criteria are not always conflicting. Applying the approach described results in an approximation, rather than an interpolation, at the sample points; the accuracy of the resulting model is thus not guaranteed, but the resulting function is more easily accepted by experts, since it also fulfils the physical laws invoked. Although surrogate models do not meet the simulation-based function values at the sample points, they might provide a higher level of accuracy over the entire design space in comparison with interpolation models.

Further efforts will be devoted to implementing the resulting model into existing global optimization algorithms. The implementation of the expert knowledge will typically be performed when the computation time required for each simulation of the expensive function is too long or when the final surrogate model, which by then cannot be further updated by simulating additional sample points, is constructed.

Acknowledgements The research leading to the results presented in this article was financially supported by the Swedish Energy Agency (project number P34882-1), Chalmers University of Technology, and Volvo Group Trucks Technology. We acknowledge four anonymous referees for their valuable comments.

References

1. Abramson, M.A., Audet, C.: Convergence of mesh adaptive direct search to second-order stationary points. SIAM Journal on Optimization **17**(2), 606–619 (2006)

2. Ali, R., Dhillon, R., El-Gindy, M., Öijer, F., Johansson, I., Trivedi, M.: Prediction of rolling resistance and steering characteristics using finite element analysis truck tyre model. International Journal of Vehicle Systems Modelling and Testing **8**(2), 179–201 (2013)
3. Audet, C., Dennis Jr, J.E.: Mesh adaptive direct search algorithms for constrained optimization. SIAM Journal on Optimization **17**(1), 188–217 (2006)
4. Bazaraa, M.S., Sherali, H.D., Shetty, C.M.: Nonlinear Programming: Theory and Algorithms, 3rd edition. John Wiley & Sons, Hoboken, NJ, USA (2006)
5. Billups, S., Larson, J., Graf, P.: Derivative-free optimization of expensive functions with computational error using weighted regression. SIAM Journal on Optimization **23**(1), 27–53 (2013)
6. Björck, Å.: Numerical Methods for Least Squares Problems. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA (1996)
7. Buhmann, M.D.: Radial Basis Functions: Theory and Implementations, *Cambridge Monographs on Applied and Computational Mathematics*, vol. 12. Cambridge University Press, Cambridge, UK (2003)
8. Coleman, T.F., Branch, M.A., Grace, A., MathWorks, I.: Optimization Toolbox: For Use with MATLAB®. The MathWorks, Inc., Natick, MA, USA (1999)
9. Conn, A.R., Scheinberg, K., Vicente, L.N.: Introduction to Derivative-Free Optimization, *MPS-SIAM Series on Optimization*, vol. 8. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA (2009)
10. Elsayed, K., Lacor, C.: Robust parameter design optimization using kriging, RBF and RBFNN with gradient-based and evolutionary optimization techniques. Applied Mathematics and Computation **236**, 325–344 (2014)
11. Fisher, W.D.: A note on curve fitting with minimum deviations by linear programming. Journal of the American Statistical Association **56**(294), 359–362 (1961)
12. Fornberg, B., Flyer, N., Hovde, S., Piret, C.: Locality properties of radial basis function expansion coefficients for equispaced interpolation. IMA Journal of Numerical Analysis **28**(1), 121–142 (2008)
13. Forrester, A.I.J., Keane, A.J., Bressloff, N.W.: Design and analysis of "noisy" computer experiments. AIAA Journal **44**(10), 2331–2339 (2006)
14. Gärtner, B., Schönherr, S.: An efficient, exact, and generic quadratic programming solver for geometric optimization. In: S.W. Cheng, O. Cheong, P. Argarwal, S. Fortune (eds.) Proceedings of the Sixteenth Annual Symposium on Computational Geometry, pp. 110–118. ACM, New York, NY, USA (2000)
15. Guerra-Vázquez, F., Rückmann, J.-J.: Semi-infinite programming. In: J.J. Cochran (ed.) Wiley Encyclopedia of Operations Research and Management Science. John Wiley & Sons, Inc. (2010). 10 pages, `dx.doi.org/10.1002/9780470400531.eorms1037`
16. Gutmann, H.-M.: A radial basis function method for global optimization. Journal of Global Optimization **19**(3), 201–227 (2001)
17. Hansen, E.: Global optimization using interval analysis—The multi-dimensional case. Numerische Mathematik **34**(3), 247–270 (1980)
18. Hettich, R.: An implementation of a discretization method for semi-infinite programming. Mathematical Programming **34**(3), 354–361 (1986)
19. Hürlimann, W.: A uniform approximation to the sampling distribution of the coefficient of variation. Statistics & Probability Letters **24**(3), 263–268 (1995)
20. Jackson, I.R.H.: Convergence properties of radial basis functions. Constructive Approximation **4**(1), 243–264 (1988)
21. Jakobsson, S., Patriksson, M., Rudholm, J., Wojciechowski, A.: A method for simulation based optimization using radial basis functions. Optimization & Engineering **11**(4), 501–532 (2010)
22. Jones, D.R.: A taxonomy of global optimization methods based on response surfaces. Journal of Global Optimization **21**(4), 345–383 (2001)
23. Kohavi, R.: A study of cross-validation and bootstrap for accuracy estimation and model selection. In: M. Kaufmann (ed.) Proceedings of the 14th International Joint Conference on Artificial Intelligence, vol. 2, pp. 1137–1143. Montréal, PQ, Canada (1995)
24. Kozlov, M.K., Tarasov, S.P., Khachiyan, L.G.: The polynomial solvability of convex quadratic programming. USSR Computational Mathematics and Mathematical Physics **20**(5), 223–228 (1980)

25. Lewis, R.M., Torczon, V., Trosset, M.W.: Direct search methods: Then and now. Journal of Computational and Applied Mathematics **124**(1–2), 191–207 (2000)
26. Lophaven, S.N., Nielsen, H.B., Sondergaard, J.: DACE–a MATLAB kriging toolbox, version 2.0. Tech. rep., Department of Informatics and Mathematical Modelling, Technical University of Denmark, Kongens Lyngby, Denmark (2002)
27. MATLAB, MATLAB-Version: 8.0. 0.783 (R2012b). Natick, Massachusetts: The MathWorks Inc **82** (2012)
28. Nesterov, Yu, Nemirovskii, A.S., Ye, Y.: Interior-Point Polynomial Algorithms in Convex Programming, *Studies in Applied Mathematics*, vol. 13. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA (1994)
29. McKay, M.D., Beckman, R.J., Conover, W.J.: Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics **21**(2), 239–245 (1979)
30. Meckesheimer, M., Booker, A.J., Barton, R.R., Simpson, T.W.: Computationally inexpensive metamodel assessment strategies. AIAA Journal **40**(10), 2053–2060 (2002)
31. Mitsos, A.: Global optimization of semi-infinite programs via restriction of the right-hand side. Optimization **60**(10–11), 1291–1308 (2011)
32. Müller, J., Shoemaker, C.A., Piché, R.: SO-MI: A surrogate model algorithm for computationally expensive nonlinear mixed-integer black-box global optimization problems. Computers & Operations Research **40**(5), 1383–1400 (2013)
33. Ni, Q., Hu, S.: A new derivative free optimization method based on conic interpolation model. Acta Mathematica Scientia Series B **24**(2), 281–290 (2004)
34. Pacejka, H.B., Besselink, I.: Tire and Vehicle Dynamics. Automotive engineering. Elsevier, Oxford, UK (2012)
35. Pereira, A.I., Costa, M.F.P., Fernandes, E.M.G.P.: Interior point filter method for semi-infinite programming problems. Optimization **60**(10–11), 1309–1338 (2011)
36. Peri, D., Tinti, F.: A multistart gradient-based algorithm with surrogate model for global optimization. Communications in Applied and Industrial Mathematics **3**(1) (2012)
37. Polak, E.: Optimization: Algorithms and Consistent Approximation, *Applied Mathematical Sciences Series*, vol. 124. Springer-Verlag, New York, NY, USA (1997)
38. Rao, C.R., Toutenburg, H., Shalabh, Heumann, C.: Linear Models and Generalizations: Least Squares and Alternatives. Springer Series in Statistics. Springer-Verlag, Berlin Heidelberg, Germany (2009). dx.doi.org/10.1007/978-3-540-74227-2
39. Reemtsen, R.: Modifications of the first Remez algorithm. SIAM Journal on Numerical Analysis **27**(2), 507–518 (1990)
40. Regis, R.G., Shoemaker, C.A.: Constrained global optimization of expensive black box functions using radial basis functions. Journal of Global Optimization **31**(1), 153–171 (2005)
41. Sahinidis, N.V.: BARON 12.1.0: Global optimization of mixed-integer nonlinear programs, *User's Manual* (2013). www.gams.com/dd/docs/solvers/baron.pdf
42. Tawarmalani, M., Sahinidis, N.V.: A polyhedral branch-and-cut approach to global optimization. Mathematical Programming **103**(2), 225–249 (2005)
43. Tsunoda, M., Monden, A., Keung, J., Matsumoto, K.: Incorporating expert judgment into regression models of software effort estimation. In: K.R.P.H. Leung, P. Muenchaisri (eds.) Proceedings of the 19th Asia-Pacific Software Engineering Conference, vol. 1, pp. 374–379. IEEE Computer Society, Hong Kong (2012)
44. Ward, L.E.: Linear programming and approximation problems. The American Mathematical Monthly **68**(1), 46–53 (1961)
45. Wendland, H.: Scattered Data Approximation. Cambridge Monographs on Applied and Computational Mathematics, Cambridge, UK (2005)
46. Wild, S.M., Shoemaker, C.: Global convergence of radial basis function trust-region algorithms for derivative-free optimization. SIAM Review **55**(2), 349–371 (2013)
47. Zhu, W., Wang, Y., Zhu, Q.F.: Second-order derivative-based smoothness measure for error concealment in DCT-based codecs. IEEE Transactions on Circuits and Systems for Video Technology **8**(6), 713–718 (1998)