

A Space-Mapping Framework for Engineering Optimization—Theory and Implementation

Slawomir Koziel, *Member, IEEE*, John W. Bandler, *Fellow, IEEE*, and Kaj Madsen

Abstract—This paper presents a comprehensive approach to engineering design optimization exploiting space mapping (SM). The algorithms employ input SM and a new generalization of implicit SM to minimize the misalignment between the coarse and fine models of the optimized object over a region of interest. Output SM ensures the matching of responses and first-order derivatives between the mapped coarse model and the fine model at the current iteration point in the optimization process. We provide theoretical results that show the importance of the explicit use of sensitivity information to the convergence properties of our family of algorithms. Our algorithm is demonstrated on the optimization of a microstrip bandpass filter, a bandpass filter with double-coupled resonators, and a seven-section impedance transformer. We describe the novel user-oriented software package SMF that implements the new family of SM optimization algorithms.

Index Terms—Computer-aided design (CAD), design automation, engineering optimization, optimization, space mapping (SM).

I. INTRODUCTION

SPACE-MAPPING (SM) technology is a recognized engineering optimization paradigm consisting of a number of efficient optimization approaches [1]–[5]. The iterative optimization and updating of a so-called “coarse” or surrogate model (cheap to evaluate) replaces the direct optimization of an accurate, but computationally expensive high fidelity or “fine” model of interest. If the misalignment between the fine and coarse models is not significant, SM-based optimization algorithms typically provide excellent results after only a few evaluations of the fine model.

In the microwave arena [1], fine (validation) models are often based on time-consuming full-wave electromagnetic (EM) solvers, while coarse models may be physically based equivalent-circuit models. The advantages of SM techniques have been widely demonstrated in a growing number of engineering disciplines (see, e.g., [6]–[8]). Bandler *et al.* [5] recently reviewed these advances.

Manuscript received February 22, 2006; revised July 11, 2006. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada under Grant OGP0007239 and Grant STGP269760 and by Bandler Corporation.

S. Koziel is with the Simulation Optimization Systems Research Laboratory, Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada L8S 4K1 (e-mail: koziels@mcmaster.ca).

J. W. Bandler is with the Simulation Optimization Systems Research Laboratory, Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada L8S 4K1, and also with Bandler Corporation, Dundas, ON, Canada L9H 5E7 (e-mail: bandler@mcmaster.ca).

K. Madsen is with Informatics and Mathematical Modelling, Technical University of Denmark, DK-2800 Lyngby, Denmark (e-mail: km@imm.dtu.dk).

Color versions of Figs. 5, 8, and 10–13 are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMTT.2006.882894

Recent efforts have focused on efficient optimization algorithms involving implicit space mapping (ISM) [3], [9] and output space mapping (OSM) [4]. Space mapping interpolating surrogates (SMISs) [10] utilize surrogate models based on SM and OSM and has proven successful for difficult optimization problems. Its performance is obtained, however, at the expense of high complexity of the underlying surrogate model. Simpler, yet powerful algorithms, namely, generalized space mapping (GSM) and its frequency-dependent version frequency-dependent generalized space mapping (FDGSM) are presented in [11].

Recent relevant work covers other important aspects of SM such as the development of new SM-based models [12], [13], the theoretical justification of SM, and a convergence theory for SM optimization algorithms [14], as well as neuro-SM [15]–[19], and microwave applications of SM (e.g., [20] and [21]).

This paper describes a family of algorithms that exploit surrogate models based on the OSM concept that forces exact matching of responses and Jacobians between the surrogate and fine model. We present theoretical results that show the influence of Jacobian matching on the convergence of the optimization algorithm. We introduce design-variable-dependent ISM to increase the flexibility of the surrogate model in a consistent way. Finally, we describe the SMF system, a novel prototype user-oriented software package that implements our SM optimization algorithms.¹ Our system is supposed to make SM accessible to engineers inexperienced in SM. It provides sockets to popular simulators (e.g., Sonnet’s *em*, Agilent’s ADS, and FEKO) that allow automatic fine/coarse model data acquisition and, consequently, fully automatic SM optimization.

II. SM-BASED SURROGATE OPTIMIZATION

The optimization problem can be stated as follows. Let $\mathbf{R}_f : X_f \rightarrow R^m$, $X_f \subseteq R^n$ denote the response vector of the fine model of a given object. In the microwave area, components of \mathbf{R}_f may be the model evaluations (e.g., scattering parameter $|S_{21}|$) at m different frequency points. Our goal is to solve

$$\mathbf{x}_f^* = \arg \min_{\mathbf{x} \in X_f} U(\mathbf{R}_f(\mathbf{x})) \quad (1)$$

where U is a given objective function. We assume that the fine model is computationally expensive and solving (1) by means of direct optimization is impractical. Instead, we exploit inexpensive surrogates, i.e., models that are not as accurate as the fine model, but are computationally cheap, which allows us to use them in the optimization process. We consider an optimization algorithm that generates a sequence of points $\mathbf{x}^{(i)} \in X_f$,

¹SMF, Bandler Corporation, Dundas, ON, Canada, 2006.

$i = 1, 2, \dots$, and a family of surrogate models $\mathbf{R}_s^{(i)} : X_s^{(i)} \rightarrow R^m$, $i = 0, 1, \dots$ so that

$$\mathbf{x}^{(i+1)} = \arg \min_{\mathbf{x} \in X_s^{(i)} \cap X_f} U(\mathbf{R}_s^{(i)}(\mathbf{x})) \quad (2)$$

and $\mathbf{R}_s^{(i+1)}$ is constructed using suitable matching conditions with the fine model at previous points $\mathbf{x}^{(k)}$, $k = 1, \dots, i$.

SM assumes the existence of a coarse model $\mathbf{R}_c : X_c \rightarrow R^m$, $X_c \subseteq R^n$ that describes the same object as the fine model. \mathbf{R}_c is less accurate, but much faster to evaluate than \mathbf{R}_f . In the context of SM, a family of surrogate models is constructed from the coarse model in such a way that each $\mathbf{R}_s^{(i)}$ is a suitable distortion of \mathbf{R}_c such that given matching conditions are satisfied. In Section III, we describe a family of SM-based models that fit this scheme.

III. GENERALIZED IMPLICIT SPACE-MAPPING (GISM) FRAMEWORK

Our proposed GISM framework is a generalization of the GSM concept introduced in [11]. It is enhanced by design-variable-dependent ISM.

Let us recall that ISM [3] assumes that the coarse model depends on additional (preassigned) parameters, i.e., we have $\mathbf{R}_c : X_c \times X_p \rightarrow R^m$, where $X_p \subseteq R^q$ is the domain of such preassigned parameters.

An ISM optimization algorithm aims at predistortion of the coarse model by adjustment of its preassigned parameters \mathbf{x}_p so that, at the current point $\mathbf{x}^{(i)}$, the fine and coarse model response vectors coincide. The predistorted model becomes a surrogate, which, in turn, is optimized in order to obtain the next point $\mathbf{x}^{(i+1)}$. Thus, the surrogate model defined by ISM is

$$\mathbf{R}_s^{(i)}(\mathbf{x}) = \mathbf{R}_c(\mathbf{x}, \mathbf{x}_p^{(i)}) \quad (3)$$

where $\mathbf{x}_p^{(i)}$ is determined by solving a parameter-extraction problem of the form

$$\mathbf{x}_p^{(i)} = \arg \min_{\mathbf{x}} \left\| \mathbf{R}_f(\mathbf{x}^{(i)}) - \mathbf{R}_c(\mathbf{x}^{(i)}, \mathbf{x}) \right\|. \quad (4)$$

One of the ways of increasing the number of degrees of freedom of the surrogate model (3) is to make preassigned parameters dependent on the design variables. In particular, we can define

$$\mathbf{R}_s^{(i)}(\mathbf{x}) = \mathbf{R}_c(\mathbf{x}, \mathbf{G}^{(i)}\mathbf{x} + \mathbf{x}_p^{(i)}) \quad (5)$$

where $\mathbf{G}^{(i)} \in M_{q \times n}$ and $\mathbf{x}_p^{(i)}$ are determined by solving the parameter-extraction problem

$$\left(\mathbf{G}^{(i)}, \mathbf{x}_p^{(i)} \right) = \arg \min_{(\mathbf{G}, \mathbf{x})} \left\| \mathbf{R}_f(\mathbf{x}^{(i)}) - \mathbf{R}_f(\mathbf{x}^{(i)}, \mathbf{G} \cdot \mathbf{x}^{(i)} + \mathbf{x}) \right\|. \quad (6)$$

Now, we can integrate the above concept into GSM [11] and define the GISM framework as the family of surrogate models $\mathbf{R}_s^{(i)}$ defined as

$$\mathbf{R}_s^{(i)}(\mathbf{x}) = \mathbf{A}^{(i)} \cdot \mathbf{R}_c(\mathbf{B}^{(i)} \cdot \mathbf{x} + \mathbf{c}^{(i)}, \mathbf{G}^{(i)} \cdot \mathbf{x} + \mathbf{x}_p^{(i)}) + \mathbf{d}^{(i)} + \mathbf{E}^{(i)} \cdot (\mathbf{x} - \mathbf{x}^{(i)}) \quad (7)$$

where

$$\begin{aligned} & \left(\mathbf{A}^{(i)}, \mathbf{B}^{(i)}, \mathbf{c}^{(i)}, \mathbf{G}^{(i)}, \mathbf{x}_p^{(i)} \right) \\ & = \arg \min_{(\mathbf{A}, \mathbf{B}, \mathbf{c}, \mathbf{G}, \mathbf{x}_p)} \varepsilon^{(i)}(\mathbf{A}, \mathbf{B}, \mathbf{c}, \mathbf{G}, \mathbf{x}_p) \end{aligned} \quad (8)$$

$$\begin{aligned} \mathbf{d}^{(i)} & = \mathbf{R}_f(\mathbf{x}^{(i)}) - \mathbf{A}^{(i)} \mathbf{R}_c(\mathbf{B}^{(i)} \cdot \mathbf{x}^{(i)} + \mathbf{c}^{(i)}, \mathbf{G}^{(i)} \cdot \mathbf{x}^{(i)} + \mathbf{x}_p^{(i)}) \end{aligned} \quad (9)$$

$$\begin{aligned} \mathbf{E}^{(i)} & = \mathbf{J}_{\mathbf{R}_f}(\mathbf{x}^{(i)}) \\ & - \mathbf{A}^{(i)} \cdot \mathbf{J}_{\mathbf{R}_c, \mathbf{x}}(\mathbf{B}^{(i)} \cdot \mathbf{x}^{(i)} + \mathbf{c}^{(i)}, \mathbf{G}^{(i)} \cdot \mathbf{x}^{(i)} + \mathbf{x}_p^{(i)}) \cdot \mathbf{B}^{(i)} \\ & - \mathbf{A}^{(i)} \cdot \mathbf{J}_{\mathbf{R}_c, \mathbf{x}_p}(\mathbf{B}^{(i)} \cdot \mathbf{x}^{(i)} + \mathbf{c}^{(i)}, \mathbf{G}^{(i)} \cdot \mathbf{x}^{(i)} + \mathbf{x}_p^{(i)}) \cdot \mathbf{G}^{(i)}. \end{aligned} \quad (10)$$

Matrices $\mathbf{A}^{(i)} = \text{diag}\{a_1^{(i)}, \dots, a_m^{(i)}\}$, $\mathbf{B}^{(i)} \in M_{n \times n}$, $\mathbf{c}^{(i)} \in M_{n \times 1}$, and $\mathbf{G}^{(i)} \in M_{q \times n}$ and vector $\mathbf{x}_p^{(i)}$ are obtained using parameter extraction applied to the matching condition $\varepsilon^{(i)}$. Matrices $\mathbf{d}^{(i)} \in M_{m \times 1}$ and $\mathbf{E}^{(i)} \in M_{m \times n}$ are calculated using (9) and (10) after having determined $\mathbf{A}^{(i)}$, $\mathbf{B}^{(i)}$, $\mathbf{c}^{(i)}$, $\mathbf{G}^{(i)}$, and $\mathbf{x}_p^{(i)}$. $\mathbf{J}_{\mathbf{R}_c, \mathbf{x}}$ and $\mathbf{J}_{\mathbf{R}_c, \mathbf{x}_p}$ denote the Jacobian of the coarse model with respect to \mathbf{x} and \mathbf{x}_p , respectively. If derivative information is not available, matrix $\mathbf{E}^{(i)}$ can be estimated, for example, using the Broyden update.

A general form of the matching condition is

$$\begin{aligned} \varepsilon^{(i)}(\mathbf{A}, \mathbf{B}, \mathbf{c}, \mathbf{G}, \mathbf{x}_p) & = \sum_{k=0}^i w_k \left\| \mathbf{R}_f(\mathbf{x}^{(k)}) - \mathbf{A} \cdot \mathbf{R}_c(\mathbf{B} \cdot \mathbf{x}^{(k)} + \mathbf{c}, \mathbf{G} \cdot \mathbf{x}^{(k)} + \mathbf{x}_p) \right\| \\ & + \sum_{k=0}^i v_k \left\| \mathbf{J}_{\mathbf{R}_f}(\mathbf{x}^{(k)}) \right. \\ & \quad \left. - \mathbf{A} \cdot \mathbf{J}_{\mathbf{R}_c, \mathbf{x}}(\mathbf{B} \cdot \mathbf{x}^{(k)} + \mathbf{c}, \mathbf{G} \cdot \mathbf{x}^{(k)} + \mathbf{x}_p) \cdot \mathbf{B} \right. \\ & \quad \left. - \mathbf{A} \cdot \mathbf{J}_{\mathbf{R}_c, \mathbf{x}_p}(\mathbf{B} \cdot \mathbf{x}^{(k)} + \mathbf{c}, \mathbf{G} \cdot \mathbf{x}^{(k)} + \mathbf{x}_p) \cdot \mathbf{G} \right\|. \end{aligned} \quad (11)$$

We assume that the coefficients w_k and v_k are either 0 or 1 (although more general situations are conceivable). For example, setting $w_k = 1$, $k = 0, \dots, i$ and $v_k = 0$, $k = 0, \dots, i$ means that the surrogate tries to match the fine model response at all previous points $\mathbf{x}^{(k)}$ (including the current point), but Jacobian matching is not exploited. We will use this setting in our numerical experiments of Section V.

Input SM determined by matrices $\mathbf{B}^{(i)}$ and $\mathbf{c}^{(i)}$, the multiplication matrix $\mathbf{A}^{(i)}$, as well as ISM parameters $\mathbf{x}_p^{(i)}$ and $\mathbf{G}^{(i)}$, can be considered as preconditioning of the coarse model that reduces the initial misalignment between the coarse and fine models over a neighborhood of the current point $\mathbf{x}^{(i)}$. Term $\mathbf{d}^{(i)}$ ensures perfect matching of responses at $\mathbf{x}^{(i)}$, while term $\mathbf{E}^{(i)}$ gives perfect matching of first-order derivatives at $\mathbf{x}^{(i)}$.

Apart from model (7)–(11), there is an optional frequency scaling that works in such a way that the coarse model is evaluated at a different frequency than the fine model using the transformation $\omega \rightarrow f_0^{(i)} + f_1^{(i)}\omega$, where $\mathbf{F} = [f_0^{(i)} \ f_1^{(i)}] \in R^2$ is obtained together with other parameters using a parameter-extraction process similar to (11).

Note that the flexibility of the surrogate model, i.e., the number of model parameters, may affect the uniqueness of the parameter-extraction problem, as well as the extrapolation properties of the model. If the model is too flexible, it may provide good matching at the points $\mathbf{x}^{(k)}$, but provide poor matching elsewhere and, consequently, slow down (or even prevent) convergence of the SM algorithm. The issue of proper choice of flexibility of the surrogate model is an important problem and needs to be carefully studied.

Flexibility of the model can be adjusted by constraining some of the SM parameters to their initial values (identity for \mathbf{A} and \mathbf{B} , zero for \mathbf{c} , \mathbf{d} , \mathbf{E} , and \mathbf{G} , $[0 \ 1]$ for \mathbf{F} , or initial values of the preassigned parameters for \mathbf{x}_p).

In general, a suitable choice of SM requires both knowledge of the problem and engineering experience. In practice, a proper SM combination can be chosen by performing parameter extraction at a small number of test points (preferably at least two points, but it is possible to get some information using even one point, namely, the starting point, at which one needs to evaluate the fine model anyway) for different combinations of SM. We recommend using the simplest combination (i.e., the one containing the least number of parameters) for which it is possible to achieve a reasonably good match between the fine model and the SM surrogate. If one can get a good match for single-point parameter extraction at different points using a simple mapping such as \mathbf{c} or \mathbf{x}_p , but the matching deteriorates while extracting parameters for several points at the same time (multipoint parameter extraction), SM should be enhanced by \mathbf{B} or \mathbf{G} , respectively. In general, ISM is preferred over input SM whenever the user is able to wisely choose preassigned parameters. Frequency mapping (\mathbf{F} term) is recommended if the response of the coarse model seems shifted in frequency with respect to the response of the fine model. Multiplicative OSM (\mathbf{A} term) should be used with care because it introduces a large number of extractable parameters. Additive OSM (\mathbf{d} and \mathbf{E} terms) can be safely used all the time.

Having defined the family of surrogate models, we can define an optimization algorithm [which is, in fact, an implementation of the generic surrogate-model-based optimization algorithm (2)]

- Step 1 Set $\mathbf{x}^{(0)} = \arg \min_{\mathbf{x} \in X_f \cap X_c} U(\mathbf{R}_c(\mathbf{x}))$; Set $i = 0$;
- Step 2 Evaluate $\mathbf{R}_f(\mathbf{x}^{(i)})$ and $\mathbf{J}_{\mathbf{R}_f}(\mathbf{x}^{(i)})$;
- Step 3 Obtain $\mathbf{R}_s^{(i)}$ using (7)–(11);
- Step 4 Find $\mathbf{x}^{(i+1)} = \arg \min_{\mathbf{x} \in X_s^{(i)} \cap X_f} U(\mathbf{R}_s^{(i)}(\mathbf{x}))$;
- Step 5 If termination condition go to Step 7;
- Step 6 Set $i := i + 1$; go to Step 2;
- Step 7 END

In numerical experiments, we use a termination condition of the form $\|\mathbf{x}^{(i+1)} - \mathbf{x}^{(i)}\| < \delta$, where δ is a small constant.

IV. SENSITIVITY INFORMATION VERSUS CONVERGENCE RATE OF THE SM OPTIMIZATION ALGORITHM

The GISM framework explicitly uses sensitivity information, i.e., the correction term \mathbf{E} , which allows the satisfaction of a first-order consistency condition, i.e., the alignment of first-order derivatives of the fine and surrogate models at the current iteration. As we show below, this term is significant to the convergence rate of the algorithm.

For the purpose of theoretical considerations, we shall consider a simplified version of GISM that only uses OSM in the form of the \mathbf{d} and \mathbf{E} terms, and no input mapping. In particular, we first consider the generic algorithm (2) that uses surrogate models with only the \mathbf{d} term. Let us define function $\Delta\mathbf{R} : X_f \cap X_c \rightarrow R^m$ as

$$\Delta\mathbf{R}(\mathbf{x}) = \mathbf{R}_f(\mathbf{x}) - \mathbf{R}_c(\mathbf{x}). \quad (12)$$

We build surrogates that use zeroth-order models of $\Delta\mathbf{R}$, i.e.,

$$\mathbf{R}_s^{(i)}(\mathbf{x}) = \mathbf{R}_c(\mathbf{x}) + \mathbf{d}^{(i)} \quad (13)$$

where $\mathbf{d}^{(i)} = \Delta\mathbf{R}(\mathbf{x}^{(i)})$.

Next, we will consider the algorithm that utilizes surrogate models based on both the \mathbf{d} and \mathbf{E} terms. In particular, we construct surrogates that use first-order models of $\Delta\mathbf{R}$, i.e.,

$$\mathbf{R}_s^{(i)}(\mathbf{x}) = \mathbf{R}_c(\mathbf{x}) + \mathbf{d}^{(i)} + \mathbf{E}^{(i)} \cdot (\mathbf{x} - \mathbf{x}^{(i)}) \quad (14)$$

where $\mathbf{d}^{(i)} = \Delta\mathbf{R}(\mathbf{x}^{(i)})$, and $\mathbf{E}^{(i)} = \mathbf{J}_{\Delta\mathbf{R}}(\mathbf{x}^{(i)})$ with $\mathbf{J}_{\Delta\mathbf{R}}(\mathbf{x}^{(i)})$ being the Jacobian of $\Delta\mathbf{R}$ at $\mathbf{x}^{(i)}$.

We are going to formulate convergence results for algorithm (2) with surrogate models (13) and for algorithm (2) with surrogate models (14) in order to see the impact of the \mathbf{E} term on the convergence properties of the SM algorithm.

Let us assume, for simplicity, that $X_f = X_c = X \subseteq R^n$ (this implies, of course, that $X_s^{(i)} = X$ for $i = 0, 1, 2, \dots$). We also assume that X is a closed subset of R^n . We denote by Y_R the range of $\Delta\mathbf{R} = \mathbf{R}_f - \mathbf{R}_c$, i.e., $Y_R = \Delta\mathbf{R}(X) \subseteq R^m$.

Definition 1: Let $X_c^*(\mathbf{R})$ denote a set of solutions to the problem $\mathbf{x}_c^* = \arg \min_{\mathbf{x} \in X} U(\mathbf{R}_c(\mathbf{x}) + \mathbf{R})$.

Definition 2: We shall denote by X_f^* the set of solutions to (1) and call it the set of fine model minimizers.

Theorem 1: Suppose that X is a closed subset of R^n and

- (i) $X_c^*(\mathbf{R})$ is not empty for any $\mathbf{R} \in Y_R$, and the following condition is satisfied:

$$\sup_{\mathbf{z} \in X_c^*(\mathbf{R}_2)} \sup_{\mathbf{v} \in X_c^*(\mathbf{R}_1)} \|\mathbf{z} - \mathbf{v}\| \leq L_C \|\mathbf{R}_2 - \mathbf{R}_1\| \quad (15)$$

for any $\mathbf{R}_1, \mathbf{R}_2 \in Y_R$, where $L_C \in R_+$ (here, $R_+ = (0, \infty)$),

- (ii) the function $\Delta\mathbf{R}$ is Lipschitz continuous, i.e.,

$$\|\Delta\mathbf{R}(\mathbf{y}) - \Delta\mathbf{R}(\mathbf{x})\| \leq L_R \|\mathbf{y} - \mathbf{x}\| \quad (16)$$

for any $\mathbf{x}, \mathbf{y} \in X$, where $L_R \in R_+$,

- (iii) L_R and L_C are such that $L_R L_C < 1$.

For any $\mathbf{x}^{(0)} \in X$, the sequence $\{\mathbf{x}^{(i)}\}$ defined by (2) and (13) is then convergent, i.e., there is an $\mathbf{x}^* \in X$ such that $\mathbf{x}^{(i)} \rightarrow \mathbf{x}^*$ for $i \rightarrow \infty$. Moreover, for any $\bar{\mathbf{x}} \in X_f^*$, we have the estimate

$$U(\mathbf{R}_f(\bar{\mathbf{x}})) \leq U(\mathbf{R}_f(\mathbf{x}^*)) = U(\mathbf{R}_c(\mathbf{x}^*) + \Delta\mathbf{R}(\mathbf{x}^*)). \quad (17)$$

A proof of Theorem 1 is given in the Appendix.

Let U_{\min} be defined as $U_{\min} = \min_{\mathbf{x} \in X_f} U(\mathbf{R}_f(\mathbf{x}))$. We have the following corollary.

Corollary 1: Suppose that for any $\mathbf{R} \in Y_R$ there is an $\mathbf{x}_R \in X$ such that $U(\mathbf{R}_c(\mathbf{x}_R) + \mathbf{R}) \leq U_{\min}$ and the assumptions of Theorem 1 are satisfied. Then $\mathbf{x}^* \in X_f^*$, where $\mathbf{x}^* = \lim_{i \rightarrow \infty} \mathbf{x}^{(i)}$ is the limit of the sequence $\{\mathbf{x}^{(i)}\}$ defined by algorithm (2), (13).

Proof: The convergence of the sequence $\{\mathbf{x}^{(i)}\}$ follows from Theorem 1. The assumption of the corollary implies that $U(\mathbf{R}_c(\mathbf{x}^{(i+1)}) + \Delta\mathbf{R}(\mathbf{x}^{(i)})) \leq U_{\min}$ for any $i = 1, 2, \dots$. In the limit ($i \rightarrow \infty$), we have $\mathbf{R}_c(\mathbf{x}^{(i+1)}) + \Delta\mathbf{R}(\mathbf{x}^{(i)}) \rightarrow \mathbf{R}_c(\mathbf{x}^*) + \Delta\mathbf{R}(\mathbf{x}^*) = \mathbf{R}_f(\mathbf{x}^*)$ so $U(\mathbf{R}_f(\mathbf{x}^*)) = U_{\min}$ and $\mathbf{x}^* \in X_f^*$.

Let us now formulate the convergence result for algorithm (2) with surrogate models (14). Again, we assume that $X_f = X_c = X \subseteq R^n$, X is a closed subset of R^n . We also admit the following notation: $Y = \mathbf{R}_c(X) \subseteq R^m$ (the range of \mathbf{R}_c) and $Y_R = \Delta\mathbf{R}(X) \subseteq R^m$ (the range of $\Delta\mathbf{R} = \mathbf{R}_f - \mathbf{R}_c$).

Theorem 2: Suppose that X is a closed subset of R^n . Let $\{\mathbf{x}^{(i)}\}$ be a sequence defined by algorithm (2) and (14) with arbitrary $\mathbf{x}^{(0)} \in X$ and

- (i) assumption (i) of Theorem 1 holds,
- (ii) function $\Delta\mathbf{R}$ is differentiable on X and $\mathbf{J}_{\Delta\mathbf{R}}$, the Jacobian of $\Delta\mathbf{R}$, is Lipschitz continuous on X , i.e.,

$$\|\mathbf{J}_{\Delta\mathbf{R}}(\mathbf{y}) - \mathbf{J}_{\Delta\mathbf{R}}(\mathbf{x})\| \leq L_J(\mathbf{x})\|\mathbf{y} - \mathbf{x}\| \quad (18)$$

for any $\mathbf{x}, \mathbf{y} \in X$, where $L_J : X \rightarrow R_+$ is a bounded function on X ,

- (iii) there is an $M_J \in R_+$ such that $\|\mathbf{J}_{\Delta\mathbf{R}}(\mathbf{x})\| \leq M_J$ on X ,
- (iv) for $i = 1, 2, \dots$, we have $L_C M_J < 1 - \delta$, where $\delta > 0$ is a given constant,
- (v) there is an i such that the following estimate holds:

$$q_k = \frac{L_C L_J(\mathbf{x}^{(k)})}{2(1 - L_C M_J)} \cdot \|\mathbf{x}^{(i+1)} - \mathbf{x}^{(i)}\| < 1 - \varepsilon \quad (19)$$

for all $k > i$, where $\varepsilon > 0$ is a given constant.

The sequence $\{\mathbf{x}^{(i)}\}$ is then convergent, i.e., there is an $\mathbf{x}^* \in X$ such that $\mathbf{x}^{(i)} \rightarrow \mathbf{x}^*$ for $i \rightarrow \infty$.

A proof of Theorem 2 is given in the Appendix.

Corollary 2: Suppose that for each $\bar{\mathbf{x}} \in X$ there is an $\mathbf{x} \in X$ such that $U(\mathbf{R}_c(\mathbf{x}) + \Delta\mathbf{R}(\bar{\mathbf{x}}) + \mathbf{J}_{\Delta\mathbf{R}}(\bar{\mathbf{x}}) \cdot (\mathbf{x} - \bar{\mathbf{x}})) \leq U_{\min}$, that the assumptions of Theorem 2 are satisfied, and that \mathbf{R}_c and U

are continuous. Then $\mathbf{x}^* \in X_f^*$, where $\mathbf{x}^* = \lim_{i \rightarrow \infty} \mathbf{x}^{(i)}$ is the limit of the sequence $\{\mathbf{x}^{(i)}\}$ defined by algorithm (2) and (14).

We omit the proof.

Remark 1: Theorems 1 and 2 can be formulated in such a way that constants L_C and L_R in assumption (i) of Theorem 1 are local with respect to $\mathbf{R} \in Y_R$ and $\mathbf{x} \in X$, respectively. This would relax the assumptions, but at the same time, make the notation more complicated.

Theorems 1 and 2 show that there are basically two fundamental (and natural) requirements for convergence of algorithms using \mathbf{d} and \mathbf{E} terms: (i) regularity of the perturbed coarse model optimal solution with respect to the perturbation vector and (ii) similarity between the fine and coarse models in terms of the difference between their first-order derivatives.

Theorems 1 and 2 indicate that the convergence rate of algorithm (2) and (14) is much better than the convergence of algorithm (2) and (13). In particular, we have $\|\mathbf{x}^{(i+2)} - \mathbf{x}^{(i+1)}\| \leq C_1 \cdot \|\mathbf{x}^{(i+1)} - \mathbf{x}^{(i)}\|$ [cf. (A2)] for the algorithm using only the \mathbf{d} term, and $\|\mathbf{x}^{(i+2)} - \mathbf{x}^{(i+1)}\| \leq C_2 \cdot \|\mathbf{x}^{(i+1)} - \mathbf{x}^{(i)}\|^2$ [cf. (A8)] for the algorithm using the \mathbf{d} and \mathbf{E} terms (in fact, C_2 is not a constant, but it does not depend significantly on $\mathbf{x}^{(i)}$). Thus, the explicit use of sensitivity information in the SM optimization algorithm has a significant impact on the convergence rate.

Theorems 1 and 2 provide convergence results for the SM algorithm using the \mathbf{d} and \mathbf{E} terms working without convergence safeguards (in other words, it may happen that the algorithms are not convergent if the coarse model is not good enough in terms of satisfying the assumptions of the theorems). It should be noted that, in practice, SM algorithms are often used together with trust region methods [23] so that the optimization of the current surrogate model $\mathbf{R}_s^{(i)}$ is constrained to a neighborhood of the previous iteration point $\mathbf{x}^{(i)}$, and the size of this neighborhood is adjusted based on the actual improvement of the fine model solution. It can be shown using classical methods (see, e.g., [24]) that our SM algorithm working with the \mathbf{d} and \mathbf{E} terms and trust region methods guarantees convergence to the local fine model optimum (under classical conditions such as smoothness of the fine/coarse models and convexity of the objective function) [23]. This is probably the most important strength of explicit use of Jacobian information in SM optimization algorithms. In the numerical experiments of Section V, we use the SM optimization algorithm with trust region methods.

Let us consider a third-order elliptic OTA-C low-pass filter example [22] shown in Fig. 1. A coarse model of the filter is the transfer function formula assuming ideal transconductors and no parasitic elements, shown in (20), at bottom of the page.

A fine model is the transfer function formula that takes into consideration parasitic conductances g_p and capacitors C_p

$$H_f(s) = g_b (sC_f(sC_2 + g_p) + g^2) / D(s) \quad (21)$$

$$H_c(s) = \frac{g_b(s^2 C_2 C_f + g^2)}{s^3 C_2 (C_1 C_3 + C_1 C_f + C_3 C_f) + s^2 C_2 (C_1 + C_f) g + s(C_1 + C_3) g^2 + g^3} \quad (20)$$

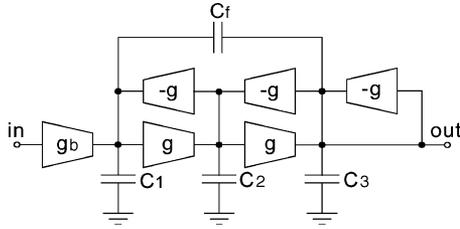
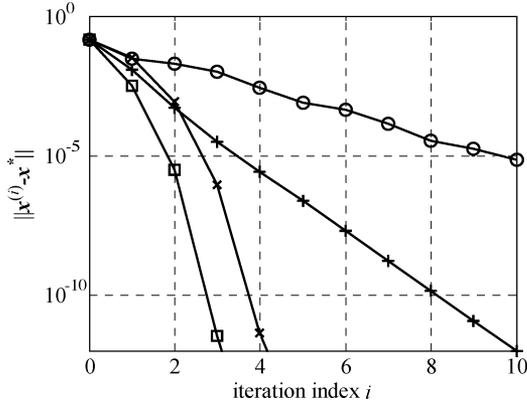


Fig. 1. Diagram of the second-order OTA-C low-pass filter [22].


 Fig. 2. Convergence of the SM algorithms for the elliptic OTA-C design problem: algorithm (2) and (13) without (o) and with (+) the coarse model improvement by input SM (\mathbf{B} and \mathbf{c} terms), and algorithm (2) and (14) without (x) and with (□) the input SM improvement.

where

$$D(s) = (s(\overline{C}_1 + C_f) + g_p)(s(\overline{C}_2 + g_p)(s(\overline{C}_3 + C_f) + g + g_p) - 2sC_f g^2 - s^2 C_f (s\overline{C}_2 + g_p) + (s(\overline{C}_1 + C_f) + g_p) g^2 + (s(\overline{C}_3 + C_f) + g + g_p) g^2) \quad (22)$$

and $\overline{C}_i = C_i + C_p$, $i = 1, 2, 3$. We use normalized elements with fixed $g = 1$, $g_p = 0.03$, and $C_p = 0.05$. Optimization variables are $\mathbf{x} = [g_b \ C_1 \ C_2 \ C_3 \ C_f]^T$. The optimization problem is to find \mathbf{x} such that the following specifications are satisfied:

$$\begin{aligned} 0.9 \leq |H_f(j\omega)| \leq 1.0, & \quad \text{for } 0 \leq \omega \leq 1 \text{ GHz} \\ |H_f(j\omega)| \leq 1.0, & \quad \text{for } 1.0 \text{ GHz} \leq \omega \leq 1.3 \text{ GHz} \\ |H_f(j\omega)| \leq 0.12, & \quad \text{for } 1.3 \text{ GHz} \leq \omega \leq 3 \text{ GHz}. \end{aligned} \quad (23)$$

We have applied algorithms (2) and (13) and (2) and (14) to find the fine model solution assuming $\mathbf{x}^{(0)} = [0.9824 \ 1.2249 \ 0.8310 \ 0.5205 \ 0.5628]^T$ (the coarse model optimal solution). We consider: (i) that the surrogate is constructed directly using \mathbf{R}_c , as in (13) or (14) and (ii) that the coarse model is improved using the input SM, i.e., the GISM framework with the \mathbf{B} , \mathbf{c} and \mathbf{d} terms, and with the \mathbf{B} , \mathbf{c} , \mathbf{d} , and \mathbf{E} terms, respectively.

Fig. 2 shows the convergence properties of our algorithms (w.r.t. the l_2 norm of the difference between current iteration point $\mathbf{x}^{(i)}$ and \mathbf{x}^*). The optimal fine model solution is $\mathbf{x}^* = [1.0424 \ 1.1131 \ 0.7967 \ 0.5594 \ 0.5530]^T$.

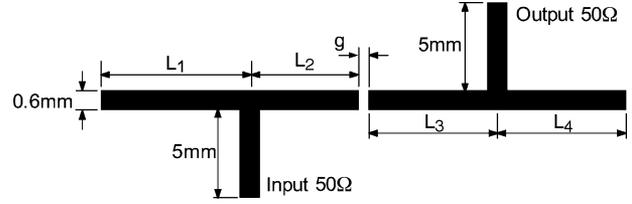


Fig. 3. Geometry of the microstrip bandpass filter [25].

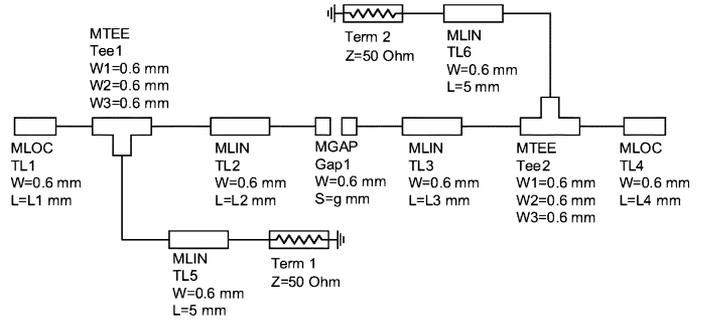


Fig. 4. Coarse model of microstrip bandpass filter (Agilent's ADS).

The results clearly show that using the \mathbf{E} term plays a crucial role in improving the convergence rate of the optimization process. Using the input SM also speeds up convergence of the algorithm, however, the impact of the \mathbf{E} term is significantly larger, as shown in Fig. 2. This complies with the theory presented here.

V. EXAMPLES

A. Bandpass Filter

The GISM framework was applied to the optimization of the microstrip bandpass filter [25] shown in Fig. 3. The design parameters are $\mathbf{x} = [L_1 \ L_2 \ L_3 \ L_4 \ g]^T$. The fine model is simulated in FEKO [26], and the coarse model is the circuit model implemented in Agilent's ADS [27] (Fig. 4). For this problem, we use ISM with preassigned parameters being electrical permittivities (initial value: 9) and heights (initial value: 0.66 mm) of microstrip elements MLIN and MLOC, shown in Fig. 4, frequency SM (\mathbf{F} term), as well as OSM (\mathbf{d} and \mathbf{E} terms). The Jacobian of the fine/coarse model was estimated using the forward finite-difference method. The design specifications are

$$\begin{aligned} |S_{21}| \leq -20 \text{ dB}, & \quad \text{for } 4.5 \text{ GHz} \leq \omega \leq 4.7 \text{ GHz} \\ |S_{21}| \geq -3 \text{ dB}, & \quad \text{for } 4.9 \text{ GHz} \leq \omega \leq 5.1 \text{ GHz} \\ |S_{21}| \leq -20 \text{ dB}, & \quad \text{for } 5.3 \text{ GHz} \leq \omega \leq 5.5 \text{ GHz}. \end{aligned}$$

The initial design is the coarse model optimal solution $\mathbf{x}^{(0)} = [6.784 \ 4.890 \ 6.256 \ 5.28 \ 0.0956]^T$ mm. The fine model response at $\mathbf{x}^{(0)}$, as well as the response at the solution obtained using GISM after four iterations ($\mathbf{x}^{(4)} = [6.431 \ 4.760 \ 6.175 \ 4.886 \ 0.0604]^T$ mm) are shown in Fig. 5. The response at $\mathbf{x}^{(4)}$ satisfies the design specifications (specification error -1.4 dB). The optimization time on an Intel P4 3.4-GHz machine was 2 h and 32 min. For the sake of comparison, we also performed a direct optimization for this

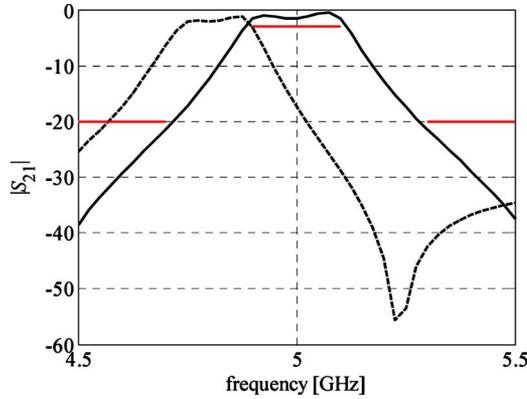


Fig. 5. Initial (dashed line) and optimized (solid line) $|S_{21}|$ versus frequency for the microstrip bandpass filter.

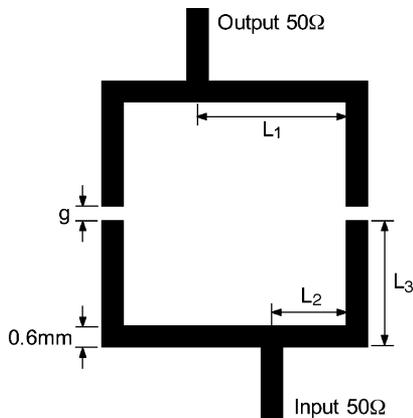


Fig. 6. Geometry of the bandpass filter with double-coupled resonators [25].

problem, which ended up in a local minimum that does not satisfy the design specifications (specification error +1.4 dB). Direct optimization took 32 h and 36 min using the same machine.

B. Bandpass Filter With Double-Coupled Resonators

Our next example is the microstrip bandpass filter with double-coupled resonators [25] shown in Fig. 6. The design parameters are $\mathbf{x} = [L_1 \ L_2 \ L_3 \ g]^T$. The fine model is simulated in FEKO [26], the coarse model is the circuit model implemented in Agilent's ADS [27] (Fig. 7). For this problem, we use input SM (\mathbf{c} term), ISM with preassigned parameters being electrical permittivities (initial value: 9) and heights (initial value: 0.66 mm) of microstrip elements MLIN, MCORN, and MTEE, shown in Fig. 7 (grouped into five groups in order to reduce the number of parameters), design-variable-dependent ISM (\mathbf{G} term), frequency SM (\mathbf{F} term), as well as OSM (\mathbf{d} and \mathbf{E} terms). The Jacobian of the fine/coarse model was estimated using the forward finite-difference method.

The design specifications are

$$\begin{aligned} |S_{21}| &\leq -20 \text{ dB}, & \text{for } 4.5 \text{ GHz} \leq \omega \leq 4.85 \text{ GHz} \\ |S_{21}| &\geq -3 \text{ dB}, & \text{for } 4.95 \text{ GHz} \leq \omega \leq 5.05 \text{ GHz} \\ |S_{21}| &\leq -20 \text{ dB}, & \text{for } 5.15 \text{ GHz} \leq \omega \leq 5.5 \text{ GHz}. \end{aligned}$$

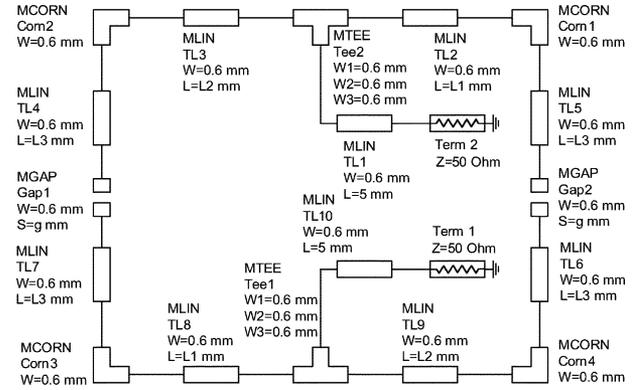


Fig. 7. Coarse model of bandpass filter with double-coupled resonators (Agilent's ADS).

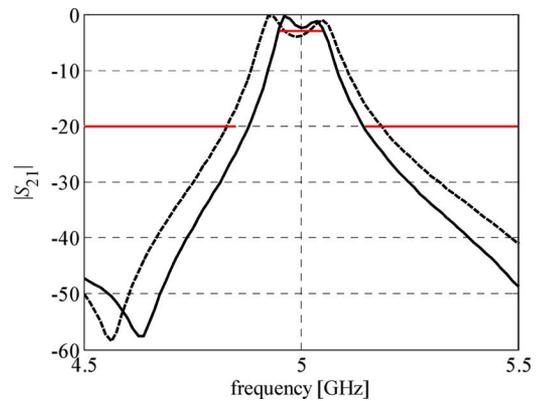


Fig. 8. Initial (dashed line) and optimized (solid line) $|S_{21}|$ versus frequency for the bandpass filter with double-coupled resonators.

The initial design is $\mathbf{x}^{(0)} = [3.0 \ 1.5 \ 3.0 \ 0.25]^T$ mm. The fine model response at $\mathbf{x}^{(0)}$, as well as the response at the solution obtained using GISM after three iterations ($\mathbf{x}^{(3)} = [3.125 \ 1.730 \ 2.796 \ 0.354]^T$ mm) are shown in Fig. 8. The response at $\mathbf{x}^{(3)}$ satisfies the design specifications (specification error -0.7 dB). The optimization time on an Intel P4 3.4-GHz machine was 2 h and 56 min. Direct optimization for this problem took 18 h and 20 min on the same machine. The specification error for the solution obtained by direct optimization is -0.8 dB.

C. Seven-Section Impedance Transformer

We also apply the proposed GISM framework to the seven-section capacitively loaded impedance transformer [28]. We consider a "coarse" model as an ideal seven-section transmission line (TL), while the "fine" model is a capacitively loaded TL with capacitors $C_{1,\dots,8} = 0.025$ pF.

The characteristic impedances in the fine model are kept fixed at the values $[Z_1 \ Z_2 \ Z_3 \ Z_4 \ Z_5 \ Z_6 \ Z_7] = [91.9445 \ 85.5239 \ 78.1526 \ 70.7101 \ 63.9774 \ 58.4632 \ 54.3806]$. The models are shown in Fig. 9. Both models are implemented in MATLAB [29].

Our design parameters are taken as the normalized lengths $\mathbf{x} = [L_1 \ L_2 \ L_3 \ L_4 \ L_5 \ L_6 \ L_7]^T$ with respect to the quarter-wave length L_q at the center frequency of 4.35 GHz. Design specifications are $|S_{11}| \leq 0.07$ for $1 \text{ GHz} \leq \omega \leq 7.7 \text{ GHz}$ with 68 points

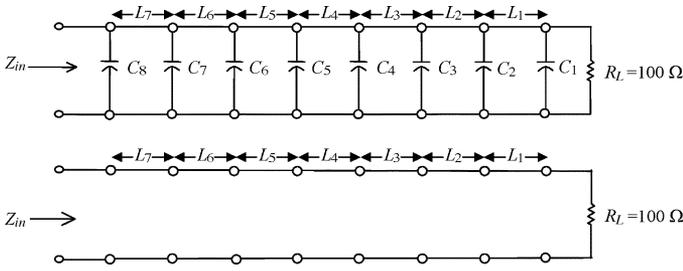


Fig. 9. Seven-section capacitively loaded impedance transformer [28]: “fine” model (upper graph) and “coarse” model (lower graph).

TABLE I
COMPARISON OF THE OPTIMAL SOLUTIONS OF
THE THREE SURROGATE MODELS

Model index	Terms Used	$U(\mathbf{R}_i^*)$	$\frac{U(\mathbf{R}_f^*) - U(\mathbf{R}_i^*)}{U(\mathbf{R}_f^*)}$	$\ \mathbf{x}_i^* - \mathbf{x}_f^*\ $	$\ \mathbf{R}_i^* - \mathbf{R}_f^*\ $
1	$\mathbf{A}\mathbf{B}\mathbf{c}$	$-8.55 \cdot 10^{-3}$	0.134	0.042	0.014
2	$\mathbf{A}\mathbf{B}\mathbf{c}\mathbf{x}_p$	$-9.30 \cdot 10^{-3}$	0.058	0.029	0.013
3	$\mathbf{A}\mathbf{B}\mathbf{c}\mathbf{x}_p\mathbf{G}$	$-9.71 \cdot 10^{-3}$	0.016	0.005	0.003

per frequency sweep. The initial design is the coarse model optimal solution $\mathbf{x}^{(0)} = [1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1]^T$. In order to use ISM, we introduce preassigned parameters to the coarse model, which are the characteristic impedances $\mathbf{x}_p = [Z_1 \ Z_2 \ Z_3 \ Z_4 \ Z_5 \ Z_6 \ Z_7]^T$.

We use the following optimization technique. First, we create an SM surrogate model based on a number of *a priori* fine model data, and then we optimize the surrogate. The whole optimization process takes a single parameter extraction and only a single surrogate model optimization, however, at the expense of the initial effort of acquiring necessary fine model information.

In the first stage, we set up our surrogate model as in (7)–(11) using the set of evaluation points (also called the *base set*) denoted as $\mathbf{x}^{(k)}$, $k = 0, 1, \dots, 14$ (i.e., $2n + 1$ points, where $n = \dim(X_f)$). We have $\mathbf{x}^{(0)} = [1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1]^T$ (coarse model optimal solution) and $\mathbf{x}^{(j)} = \mathbf{x}^{(0)} + (-1)^j \delta_{\lceil j/2 \rceil} \cdot \mathbf{u}_{\lceil j/2 \rceil}$, $j = 1, \dots, 14$, where $\mathbf{u}_k = [0 \ \dots \ 0 \ 1 \ 0 \ \dots \ 0]^T$ is a unit vector with 1 at the k th position; δ_k is the size of the region X_R along the k th axis. This distribution of points is called the star distribution [30].

For the sake of comparison, we set up three models, which are special cases of our GISM-based model (7)–(11). The first one, i.e., $\mathbf{R}_{s,1}^{(14)}$, uses only matrices \mathbf{A} , \mathbf{B} , and \mathbf{c} ; the second one, i.e., $\mathbf{R}_{s,2}^{(14)}$, uses matrices \mathbf{A} , \mathbf{B} , \mathbf{c} and preassigned parameters \mathbf{x}_p ; and the third model, i.e., $\mathbf{R}_{s,3}^{(14)}$, uses \mathbf{A} , \mathbf{B} , \mathbf{c} , \mathbf{x}_p , and \mathbf{G} (we call this model the full GISM). The preassigned parameters are kept fixed at their initial values in model $\mathbf{R}_{s,1}^{(14)}$. We do not use the \mathbf{d} and \mathbf{E} terms in models $\mathbf{R}_{s,i}^{(14)}$, $i = 1, 2, 3$ because they are suitable for local modeling, which is not the case in this example.

Having the models, we performed their optimization and compared the results with the optimal solution \mathbf{x}_f^* of the fine model (obtained by direct optimization). Table I shows the results. We use the following notation: \mathbf{x}_i^* , $i = 1, 2, 3$, the optimal solution of the surrogate model $\mathbf{R}_{s,i}^{(14)}$; $\mathbf{R}_i^* = \mathbf{R}_f(\mathbf{x}_i^*)$, the response of the fine model at the optimal solution of the i th surrogate; $\mathbf{R}_f^* = \mathbf{R}_f(\mathbf{x}_f^*)$, the response of the fine model at

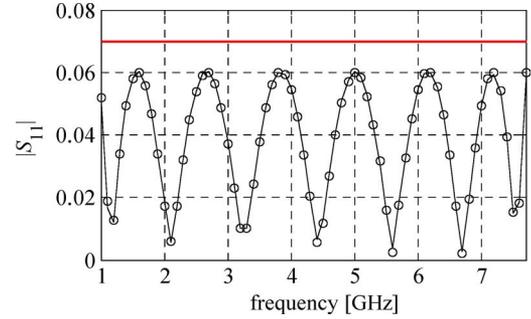


Fig. 10. Fine (solid line) and surrogate (circles) model responses at the surrogate model optimal solutions for model $\mathbf{R}_{s,3}^{(14)}$.

its final solution; and $U(\mathbf{R})$, the value of the minimax error of response \mathbf{R} .

The results in Table I show that increasing flexibility by means of preassigned parameters and our design-variable-dependent ISM allows us to obtain a better quality of the surrogate model (i.e., a better match between the fine and surrogate models), as well as better accuracy of the optimization outcome. In particular, the relative difference between the minimax error of the fine model at its optimal solution and the minimax error of the fine model at the surrogate model optimal solution is more than 13% for the model using only the \mathbf{A} , \mathbf{B} , and \mathbf{c} terms, and 1.6% for the model using the full GISM. Similarly, the distance between the fine model optimal solution and the surrogate model optimal solution is more than eight times smaller for the model using the full GISM than for the model using only the \mathbf{A} , \mathbf{B} , and \mathbf{c} terms.

Fig. 10 shows the fine (solid line) and surrogate (circles) model responses at the surrogate model optimal solutions for model $\mathbf{R}_{s,3}^{(14)}$.

VI. AUTOMATIC SM OPTIMIZATION: SMF SYSTEM

In order to make SM accessible to engineers not experienced in this technology, a prototype comprehensive user-oriented software package is under development. SMF is a graphical-user-interface-based MATLAB system that can perform SM-based constrained optimization, modeling, and statistical analysis. It implements existing SM approaches, including input SM, output SM, ISM, and frequency SM (in particular, the GISM framework). It contains drivers for simulators (Sonnet’s *em*, MEFiSTo, Agilent’s ADS, FEKO) that allow the linking of commercial fine/coarse models to the algorithm and make the optimization process fully automatic.

We only focus on one aspect of SMF: a module for automatic SM optimization. Fig. 11 shows a block diagram of the SM optimization module in SMF. Optimization is performed in several steps. First, the user enters problem arguments: starting point, frequency sweep, optimization type and design specifications. Next, the user sets up SM itself, i.e., the kind of SM to be used (e.g., input, output, implicit), specifies termination conditions, parameter-extraction options, and optional constraints.

The next step is to link the fine and coarse models to SMF by setting up the data that will be used to create the model drivers. Using user-provided data (e.g., simulator input files and design variable identification data), SMF creates drivers that are later

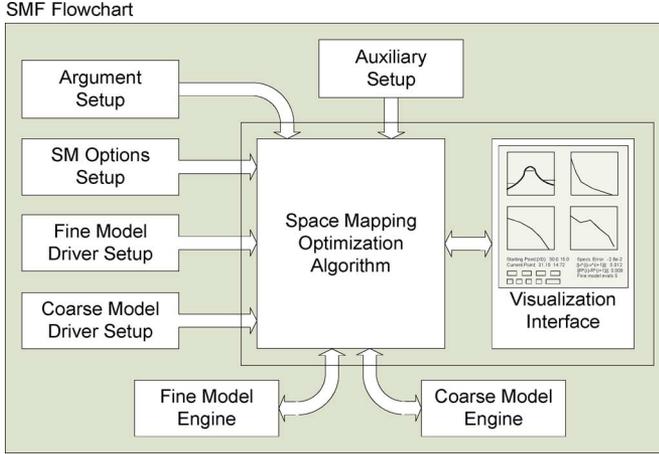


Fig. 11. Flowchart of the optimization module in the SMF system.

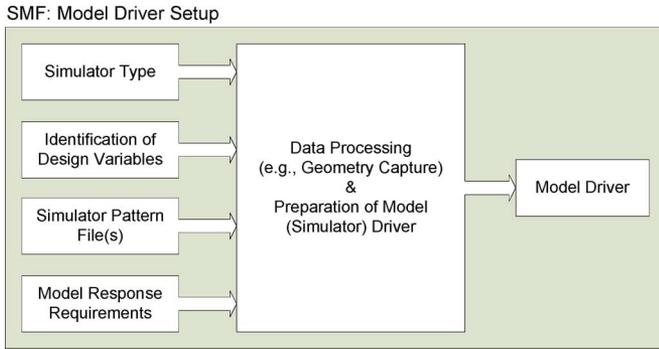


Fig. 12. Setup of model driver in SMF.

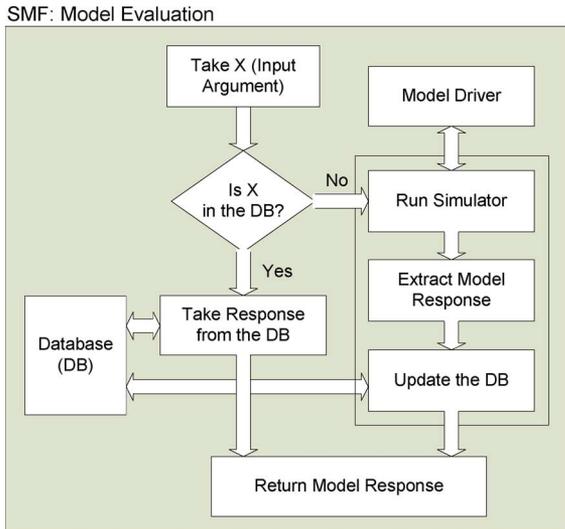


Fig. 13. Evaluation of simulator-based models in SMF.

used to evaluate fine/coarse models for any necessary design variable values. Model evaluation is accomplished by generating the simulator input file corresponding to the required design, running the simulator, and acquiring the results. Fig. 12 shows the flowchart for preparing the model driver in SMF. Fig. 13 shows the flowchart for evaluating the simulator-based

model in SMF. Note that all the model responses are stored in the database and retrieved if necessary.

Parameter extraction, surrogate model optimization, and optional trust-region specific options are set in the next step using auxiliary interfaces.

Having done the setup, the user runs the execution interface, which allows execution of the SM optimization algorithm and visualization of the results, including model responses, specification error plots, as well as convergence plots.

All the test problems presented in Section V have been solved using SMF. In order to do that, we only needed to prepare projects corresponding to fine and coarse models in FEKO and ADS, respectively (or in MATLAB in the case of the seven-section transformer example). The models are then linked to SMF by simply pointing into corresponding input files, as well as providing data concerning the design variable labels. The SM options are set up as described above. In particular, the selection of SM type is performed by checking check boxes corresponding to the $\mathbf{A}, \mathbf{B}, \dots, \mathbf{G}$ terms. All the other setup is done using separate graphical user interfaces.

VII. CONCLUSION

We have presented a novel GISM optimization framework. Our algorithms have exploited “traditional” input-SM-based preconditioning of the coarse model, design-variable-dependent ISM, and output-SM external terms to ensure perfect matching of the responses and first-order derivatives between the surrogate and the fine model. GISM performance is illustrated through a microstrip bandpass filter, a bandpass filter with double-coupled resonators, and a seven-section capacitively loaded impedance transformer.

We have also provided theoretical results that have demonstrated the importance to the convergence properties of the algorithm of the explicit use of sensitivity information.

Our SMF system, a novel prototype user-oriented software package that implements our SM optimization algorithms, has been described. It provides sockets to popular commercial simulators (e.g., Sonnet’s *em*, Agilent’s ADS, FEKO) that allow automatic fine/coarse model data acquisition and, consequently, fully automatic SM optimization and modeling.

APPENDIX

Proof of Theorem 1: Take any $\mathbf{x}^{(0)} \in X$. Define $\{\mathbf{x}^{(i)}\}$ according to (2) and (13), i.e., $\mathbf{x}^{(i+1)} = \arg \min_{\mathbf{x} \in X} U(\mathbf{R}_c(\mathbf{x}) + \Delta \mathbf{R}(\mathbf{x}^{(i)}))$ for $i = 1, 2, \dots$. From Definition 1, we have that

$$\mathbf{x}^{(i+1)} = \mathbf{R}_c^* \left(\Delta \mathbf{R} \left(\mathbf{x}^{(i)} \right) \right). \quad (\text{A1})$$

We would like to obtain an estimate for $\|\mathbf{x}^{(i+2)} - \mathbf{x}^{(i+1)}\|$. Bearing in mind that $\mathbf{x}^{(i+2)} \in X_c^*(\Delta \mathbf{R}(\mathbf{x}^{(i+1)}))$ and $\mathbf{x}^{(i+1)} \in X_c^*(\Delta \mathbf{R}(\mathbf{x}^{(i)}))$, and using assumptions (i)–(iii), we get

$$\begin{aligned} \|\mathbf{x}^{(i+2)} - \mathbf{x}^{(i+1)}\| &\leq L_C \|\Delta \mathbf{R}(\mathbf{x}^{(i+1)}) - \Delta \mathbf{R}(\mathbf{x}^{(i)})\| \\ &\leq q \|\mathbf{x}^{(i+1)} - \mathbf{x}^{(i)}\| \end{aligned} \quad (\text{A2})$$

where $q = L_C L_R < 1$. Now, for any $j > i$, we have

$$\begin{aligned} \|\mathbf{x}^{(j)} - \mathbf{x}^{(i)}\| &\leq \|\mathbf{x}^{(i+1)} - \mathbf{x}^{(i)}\| + \dots + \|\mathbf{x}^{(j)} - \mathbf{x}^{(j-1)}\| \\ &\leq (1 + q + q^2 + \dots + q^{j-i-1}) \\ &\quad \cdot \|\mathbf{x}^{(i+1)} - \mathbf{x}^{(i)}\| \\ &\leq \frac{1 - q^{j-i}}{1 - q} \|\mathbf{x}^{(i+1)} - \mathbf{x}^{(i)}\| \\ &\leq \frac{q^{i+1}}{1 - q} \|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}\| \end{aligned} \quad (\text{A3})$$

which is arbitrarily small for sufficiently large i , i.e., $\{\mathbf{x}^{(i)}\}$ is a Cauchy sequence. Thus, there is an $\mathbf{x}^* \in X$, $\mathbf{x}^* = \lim_{i \rightarrow \infty} \mathbf{x}^{(i)}$. Estimate (17) is obvious. This ends the proof of the theorem.

Proof of Theorem 2: Let us take any $\mathbf{x}^{(0)} \in X$. Define $\{\mathbf{x}^{(i)}\}$ as in (2) and (14), i.e., $\mathbf{x}^{(i+1)} = \arg \min_{\mathbf{x} \in X} U(\mathbf{R}_c(\mathbf{x}) + \mathbf{J}_{\Delta \mathbf{R}}(\mathbf{x}^{(i)}) \cdot (\mathbf{x} - \mathbf{x}^{(i)}))$ for $i = 1, 2, \dots$. From Definition 1, we have that $\mathbf{x}^{(i+1)} \in X_c^*(\Delta \mathbf{R}(\mathbf{x}^{(i)}) + \mathbf{J}_{\Delta \mathbf{R}}(\mathbf{x}^{(i)}) \cdot (\mathbf{x}^{(i+1)} - \mathbf{x}^{(i)}))$. We would like to obtain an estimate for $\|\mathbf{x}^{(i+2)} - \mathbf{x}^{(i+1)}\|$. From assumption (i), we get

$$\begin{aligned} &\|\mathbf{x}^{(i+2)} - \mathbf{x}^{(i+1)}\| \\ &\leq L_C \left\| \Delta \mathbf{R}(\mathbf{x}^{(i+1)}) - \Delta \mathbf{R}(\mathbf{x}^{(i)}) + \mathbf{J}_{\Delta \mathbf{R}}(\mathbf{x}^{(i+1)}) \right. \\ &\quad \cdot (\mathbf{x}^{(i+2)} - \mathbf{x}^{(i+1)}) - \mathbf{J}_{\Delta \mathbf{R}}(\mathbf{x}^{(i)}) \\ &\quad \left. \cdot (\mathbf{x}^{(i+1)} - \mathbf{x}^{(i)}) \right\|. \end{aligned} \quad (\text{A4})$$

It follows from assumption (ii) and [23, Th. 3.1.6] that

$$\begin{aligned} &\left\| \Delta \mathbf{R}(\mathbf{x}^{(i+1)}) - \Delta \mathbf{R}(\mathbf{x}^{(i)}) - \mathbf{J}_{\Delta \mathbf{R}}(\mathbf{x}^{(i)}) \cdot (\mathbf{x}^{(i+1)} - \mathbf{x}^{(i)}) \right\| \\ &\leq 0.5 L_J(\mathbf{x}^{(i)}) \|\mathbf{x}^{(i+1)} - \mathbf{x}^{(i)}\|^2. \end{aligned} \quad (\text{A5})$$

From (A4) and (A5), we have

$$\begin{aligned} \|\mathbf{x}^{(i+2)} - \mathbf{x}^{(i+1)}\| &= 0.5 L_C L_J(\mathbf{x}^{(i)}) \|\mathbf{x}^{(i+1)} - \mathbf{x}^{(i)}\|^2 \\ &\quad + L_C M_J \|\mathbf{x}^{(i+2)} - \mathbf{x}^{(i+1)}\| \end{aligned} \quad (\text{A6})$$

which gives

$$\|\mathbf{x}^{(i+2)} - \mathbf{x}^{(i+1)}\| \leq \frac{0.5 L_C}{1 - L_C M_J} \cdot L_J(\mathbf{x}^{(i)}) \|\mathbf{x}^{(i+1)} - \mathbf{x}^{(i)}\|^2. \quad (\text{A7})$$

Let k be the smallest index for which assumption (v) is satisfied. For any $i > k$, we then have

$$\|\mathbf{x}^{(i+2)} - \mathbf{x}^{(i+1)}\| \leq q_i \|\mathbf{x}^{(i+1)} - \mathbf{x}^{(i)}\| \quad (\text{A8})$$

where

$$q_i = \frac{0.5 L_C L_J(\mathbf{x}^{(i)})}{1 - L_C M_J} \|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\| < 1 - \varepsilon. \quad (\text{A9})$$

Thus, $\{\mathbf{x}^{(i)}\}$ is a Cauchy sequence, and since X is closed, there is an $\mathbf{x}^* \in X$, $\mathbf{x}^* = \lim_{i \rightarrow \infty} \mathbf{x}^{(i)}$. This ends the proof of the theorem.

ACKNOWLEDGMENT

The authors thank Agilent Technologies, Santa Rosa, CA, for making ADS available. The authors also acknowledge discussions with N. K. Nikolova, McMaster University, Hamilton, ON, Canada, and thank J. Zhu, McMaster University, for suggesting and setting up the microstrip filter examples.

REFERENCES

- [1] J. W. Bandler, R. M. Biernacki, S. H. Chen, P. A. Grobelny, and R. H. Hemmers, "Space mapping technique for electromagnetic optimization," *IEEE Trans. Microw. Theory Tech.*, vol. 2, no. 12, pp. 2536–2544, Dec. 1994.
- [2] J. W. Bandler, R. M. Biernacki, S. H. Chen, R. H. Hemmers, and K. Madsen, "Electromagnetic optimization exploiting aggressive space mapping," *IEEE Trans. Microw. Theory Tech.*, vol. 43, no. 12, pp. 2874–2882, Dec. 1995.
- [3] J. W. Bandler, Q. S. Cheng, N. K. Nikolova, and M. A. Ismail, "Implicit space mapping optimization exploiting preassigned parameters," *IEEE Trans. Microw. Theory Tech.*, vol. 52, no. 1, pp. 378–385, Jan. 2004.
- [4] J. W. Bandler, Q. S. Cheng, D. H. Gebre-Mariam, K. Madsen, F. Pedersen, and J. Søndergaard, "EM-based surrogate modeling and design exploiting implicit, frequency and output space mappings," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Philadelphia, PA, Jun. 2003, pp. 1003–1006.
- [5] J. W. Bandler, Q. S. Cheng, S. A. Dakroury, A. S. Mohamed, M. H. Bakr, K. Madsen, and J. Søndergaard, "Space mapping: The state of the art," *IEEE Trans. Microw. Theory Tech.*, vol. 52, no. 1, pp. 337–361, Jan. 2004.
- [6] S. J. Leary, A. Bhaskar, and A. J. Keane, "A constraint mapping approach to the structural optimization of an expensive model using surrogates," *Optim. Eng.*, vol. 2, no. 4, pp. 385–398, Dec. 2001.
- [7] M. Redhe and L. Nilsson, "Using space mapping and surrogate models to optimize vehicle crashworthiness design," presented at the 9th AIAA/ISSMO Multidisciplinary Anal. Optim. Symp., Atlanta, GA, Sep. 2002, Paper AIAA-2002-5536.
- [8] H.-S. Choi, D. H. Kim, I. H. Park, and S. Y. Hahn, "A new design technique of magnetic systems using space mapping algorithm," *IEEE Trans. Magn.*, vol. 37, no. 5, pp. 3627–3630, Sep. 2001.
- [9] J. W. Bandler, M. A. Ismail, and J. E. Rayas-Sanchez, "Expanded space-mapping EM-based design framework exploiting preassigned parameters," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 49, no. 12, pp. 1833–1838, Dec. 2002.
- [10] J. W. Bandler, D. M. Hailu, K. Madsen, and F. Pedersen, "A space mapping interpolating surrogate algorithm for highly optimized EM-based design of microwave devices," *IEEE Trans. Microw. Theory Tech.*, vol. 52, no. 11, pp. 2593–2600, Nov. 2004.
- [11] S. Koziel, J. W. Bandler, and K. Madsen, "Space mapping optimization algorithms for engineering design," in *IEEE MTT-S Int. Microw. Symp. Dig.*, San Francisco, CA, Jun. 2006, [CD ROM].
- [12] S. Koziel, J. W. Bandler, A. S. Mohamed, and K. Madsen, "Enhanced surrogate models for statistical design exploiting space mapping technology," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Long Beach, CA, Jun. 2005, pp. 1609–1612.
- [13] J. W. Bandler, Q. S. Cheng, and S. Koziel, "Implementable space mapping approach to enhancement of microwave device models," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Long Beach, CA, Jun. 2005, pp. 1139–1142.
- [14] S. Koziel, J. W. Bandler, and K. Madsen, "Towards a rigorous formulation of the space mapping technique for engineering design," in *Proc. Int. Circuits Syst. Symp.*, Kobe, Japan, May 2005, pp. 5605–5608.
- [15] M. H. Bakr, J. W. Bandler, M. A. Ismail, J. E. Rayas-Sanchez, and Q.-J. Zhang, "Neural space-mapping optimization for EM-based design," *IEEE Trans. Microw. Theory Tech.*, vol. 48, no. 12, pp. 2307–2315, Dec. 2000.

- [16] V. K. Devabhaktuni, B. Chattaraj, M. C. E. Yagoub, and Q.-J. Zhang, "Advanced microwave modeling framework exploiting automatic model generation, knowledge neural networks, and space mapping," *IEEE Trans. Microw. Theory Tech.*, vol. 51, no. 7, pp. 1822–1833, Jul. 2003.
- [17] J. E. Rayas-Sanchez, "EM-based optimization of microwave circuits using artificial neural networks: The state-of-the-art," *IEEE Trans. Microw. Theory Tech.*, vol. 52, no. 1, pp. 420–435, Jan. 2004.
- [18] J. E. Rayas-Sanchez, F. Lara-Rojo, and E. Martinez-Guerrero, "A linear inverse space-mapping (LISM) algorithm to design linear and nonlinear RF and microwave circuits," *IEEE Trans. Microw. Theory Tech.*, vol. 53, no. 3, pp. 960–968, Mar. 2005.
- [19] L. Zhang, J. Xu, M. C. E. Yagoub, R. Ding, and Q.-J. Zhang, "Efficient analytical formulation and sensitivity analysis of neuro-space mapping for nonlinear microwave device modeling," *IEEE Trans. Microw. Theory Tech.*, vol. 53, no. 9, pp. 2752–2767, Sep. 2005.
- [20] M. A. Ismail, D. Smith, A. Panariello, Y. Wang, and M. Yu, "EM-based design of large-scale dielectric-resonator filters and multiplexers by space mapping," *IEEE Trans. Microw. Theory Tech.*, vol. 52, no. 1, pp. 386–392, Jan. 2004.
- [21] K.-L. Wu, Y.-J. Zhao, J. Wang, and M. K. K. Cheng, "An effective dynamic coarse model for optimization design of LTCC RF circuits with aggressive space mapping," *IEEE Trans. Microw. Theory Tech.*, vol. 52, no. 1, pp. 393–402, Jan. 2004.
- [22] S. Koziel, S. Szczepanski, and R. Schaumann, "Structure generation and performance comparison of elliptic $G_m - C$ filters," *Int. J. Circuit Theory Appl.*, vol. 32, no. 6, pp. 565–589, Nov. 2004.
- [23] A. R. Conn, N. I. M. Gould, and P. L. Toint, *Trust Region Methods*, ser. MPS-SIAM Optim.. Philadelphia, PA: MPS-SIAM, 2000.
- [24] N. M. Alexandrov, J. E. Dennis, R. M. Lewis, and V. Torczon, "A trust region framework for managing use of approximation models in optimization," *Struct. Multidisciplinary Optim.*, vol. 15, no. 1, pp. 16–23, 1998.
- [25] A. Hennings, E. Semouchkina, A. Baker, and G. Semouchkin, "Design optimization and implementation of bandpass filters with normally fed microstrip resonators loaded by high-permittivity dielectric," *IEEE Trans. Microw. Theory Tech.*, vol. 54, no. 3, pp. 1253–1261, Mar. 2006.
- [26] "FEKO User's Manual" EM Software & Systems-S.A. (Pty) Ltd., Stellenbosch, South Africa, Jun. 2004 [Online]. Available: <http://www.feko.info>
- [27] Agilent ADS. ver. 2003C, Agilent Technol., Santa Rosa, CA, 2003.
- [28] M. H. Bakr, J. W. Bandler, K. Madsen, and J. Søndergaard, "An introduction to the space mapping technique," *Optim. Eng.*, vol. 2, no. 4, pp. 369–384, Dec. 2001.
- [29] MATLAB. ver. 7.0, The MathWorks Inc., Natick, MA, 2005.
- [30] J. W. Bandler, N. Georgieva, M. A. Ismail, J. E. Rayas-Sánchez, and Q. J. Zhang, "A generalized space mapping tableau approach to device modeling," *IEEE Trans. Microw. Theory Tech.*, vol. 49, no. 1, pp. 67–79, Jan. 2001.



Slawomir Koziel (M'03) was born in Poland, in 1970. He received the M.Sc. and Ph.D. (with honors) degrees in electronic engineering from Gdansk University of Technology, Gdansk, Poland, in 1995 and 2000, respectively, and the M.Sc. degree in theoretical physics and mathematics and Ph.D. degree in mathematics (with honors) from the University of Gdansk, Gdansk, Poland, in 2000, 2002, and 2003, respectively.

He is currently a Research Associate with the Simulation Optimization Systems Research Laboratory, Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada. He has authored or coauthored over 90 papers. His research interests include SM-based modeling and optimization, circuit theory, analog signal processing, active filter design, evolutionary computation, and numerical analysis.



John W. Bandler (S'66–M'66–SM'74–F'78) was born in Jerusalem on November 9, 1941. He received the B.Sc.(Eng.), Ph.D., and D.Sc.(Eng.) degrees from the University of London, London, U.K., in 1963, 1967, and 1976, respectively.

He joined Mullard Research Laboratories, Redhill, Surrey, U.K., in 1966. From 1967 to 1969, he was a Postdoctorate Fellow and Sessional Lecturer at the University of Manitoba, Winnipeg, Canada. He joined McMaster University, Hamilton, ON, Canada, in 1969. He was Chairman of the Department of Electrical Engineering and Dean of the Faculty of Engineering. He is currently Professor Emeritus in Electrical and Computer Engineering, directing research in the Simulation Optimization Systems Research Laboratory. He has authored or coauthored more than 385 papers. He was a member of the Micronet Network of Centres of Excellence. He was President of Optimization Systems Associates Inc. (OSA), which he founded in 1983, until November 20, 1997, the date of acquisition of OSA by the Hewlett-Packard Company. OSA implemented a first-generation yield-driven microwave computer-aided design (CAD) capability for Raytheon in 1985, followed by further innovations in linear and nonlinear microwave CAD technology for the Raytheon/Texas Instruments Joint Venture MIMIC Program. OSA introduced the CAE systems RoMPE in 1988, HarPE in 1989, OSA90 and OSA90/hope in 1991, Empipe in 1992, and Empipe3D and EmpipeExpress in 1996. OSA created the product *empath* in 1996, which was marketed by Sonnet Software Inc. He is President of Bandler Corporation, Dundas, ON, Canada, which he founded in 1997. He joined the Editorial Boards of the *International Journal of Numerical Modelling* in 1987, the *International Journal of Microwave and Millimeterwave Computer-Aided Engineering* in 1989, and *Optimization and Engineering* in 1998. He was a Guest Editor of the *International Journal of Microwave and Millimeter-Wave Computer-Aided Engineering* Special Issue on "Optimization-Oriented Microwave CAD" (1997). He was Guest Coeditor of the *Optimization and Engineering* Special Issue on "Surrogate Modelling and Space Mapping for Engineering Optimization" (2001).

Dr. Bandler is a Fellow of the Canadian Academy of Engineering, the Royal Society of Canada, the Institution of Electrical Engineers (IEE), U.K., and the Engineering Institute of Canada. He is a member of the Association of Professional Engineers of the Province of Ontario, Canada, and the Massachusetts Institute of Technology (MIT) Electromagnetics Academy. He received the Automatic Radio Frequency Techniques Group Automated Measurements Career Award in 1994 and the IEEE Microwave Theory and Techniques Society (IEEE MTT-S) Microwave Application Award in 2004. He was an associate editor of the IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES (1969–1974) and has continued serving as a member of the Editorial Board. He was guest editor of the IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES Special Issue on "Computer-Oriented Microwave Practices" (1974) and "Automated Circuit Design Using Electromagnetic Simulators" (1997) and guest coeditor of the Special Issue on "Process-Oriented Microwave CAD and Modeling" (1992) and "Electromagnetics-Based Optimization of Microwave Components and Circuits" (2004). He was chair of the MTT-1 Technical Committee on Computer-Aided Design.



Kaj Madsen was born in Denmark in 1943. He received the cand.scient. degree in mathematics from the University of Aarhus, Aarhus, Denmark, in 1968 and the Dr.Techn. degree from the Technical University of Denmark (DTU), Lyngby, Denmark, in 1986.

From 1968 to 1988, he was a Lecturer in numerical analysis, apart from the 1973–1974 academic year, when he was with AERE Harwell, Didcot, U.K. Most of his career has been spent with the Department for Numerical Analysis, DTU. From 1981 to 1983, he was with the Computer Science Department, Copenhagen University, Copenhagen, Denmark. During Summer 1978, he visited McMaster University, Hamilton, ON, Canada. In 1988, he became a Full Professor. Since the 1990s, he has arranged several international workshops on linear programming, parallel algorithms, and surrogate modeling, and SM. In 1993, he joined the Department of Mathematical Modelling, DTU, and during 1995–2000 was Head of that department. In 2000, he took an active part in forming the new department Informatics and Mathematical Modelling, DTU, which includes computer science and applied mathematics. Since January 2001, he has been Head of that department. His primary fields of interest in teaching and research are nonlinear optimization, including SM techniques and global optimization, and validated computing using interval analysis.