

# Towards a Rigorous Formulation of the Space Mapping Technique for Engineering Design

Slawomir Koziel, John W. Bandler

Simulation Optimization Systems Research Laboratory  
ECE Department, McMaster University  
Hamilton, L8S 4K1, Canada  
koziels@mcmaster.ca, bandler@mcmaster.ca

Kaj Madsen

Informatics and Mathematical Modelling  
Technical University of Denmark  
Lyngby, DK-2800, Denmark  
km@imm.dtu.dk

**Abstract**—This paper deals with the Space Mapping (SM) approach to engineering design optimization. We attempt here a theoretical justification of methods that have already proven efficient in solving practical problems, especially in the RF and microwave area. A formal definition of optimization algorithms using surrogate models based on SM technology is given. Convergence conditions for the chosen subclass of algorithms are discussed and explained using a synthetic example, the so-called generalized cheese-cutting problem. An illustrative, circuit-theory based example is also considered.

## I. INTRODUCTION

Mapping (SM) technology involves well-established and efficient optimization methods [1-4]. The main idea behind SM is that direct optimization of a so-called (accurate but computationally expensive) fine model of interest is replaced by the iterative optimization and updating of a corresponding so-called coarse model (less accurate but cheap to evaluate). Provided their misalignment is not significant, SM based algorithms are able to converge after only a few evaluations of the fine model. SM was originally demonstrated on microwave circuit optimization [1], where fine models may be based on electromagnetic simulators, while coarse models are physics-based circuit models.

SM techniques have recently solved modeling and optimization problems in a growing number of areas, not only RF and microwave design [1-4] but also structural design [5], vehicle crashworthiness design [6], magnetic systems [7], and others. For a review see [4].

Although SM algorithms have been developed that solve practical optimization problems, the unified formulation and theory of SM, including convergence proofs (with the exception of a subclass of algorithms based on trust-region methods [8]) is lacking. Besides the theoretical importance of a unified formulation and classification of SM methods, there are two important questions an SM theory should be able to answer: (i) how good the coarse model should be (compared with the fine model) in order to make an SM algorithm converge, (ii) how to design better (more efficient) algorithms. We try to answer these questions for algorithms based on the so-called Output Space Mapping (OSM).

## II. OPTIMIZATION USING SURROGATE MODELS

We are concerned with a class of optimization algorithms that use surrogate models. Let  $\mathbf{R}_f: X_f \rightarrow R^m$  denote the response vector of the so-called fine model of a given object, where  $X_f \subseteq R^n$ . 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 suitable objective function and  $\mathbf{R}_{spec} \in R^m$  is a given specification vector. In many engineering problems, we are concerned with so-called one-sided specifications: if  $\mathbf{R}_f = (R_{f,1}, \dots, R_{f,m})$ ,  $\mathbf{R}_{spec} = (R_{sp,1}, \dots, R_{sp,m})$ , and  $I_l, I_u \subset \{1, 2, \dots, m\}$  are such that  $I_l \cap I_u = \emptyset$ , then we require that  $R_{f,i} \leq R_{sp,i}$  for  $i \in I_u$  and  $R_{f,i} \geq R_{sp,i}$  for  $i \in I_l$ . Typically,  $U$  is defined as

$$U(\mathbf{R}_f) = \max \left\{ \max_{i \in I_u} (R_{f,i} - R_{sp,i}), \max_{i \in I_l} (R_{sp,i} - R_{f,i}) \right\} \quad (2)$$

In some problems,  $U$  can be defined by a norm, i.e.,

$$U(\mathbf{R}_f) = \|\mathbf{R}_f - \mathbf{R}_{spec}\| \quad (3)$$

We shall denote by  $X_f^*$  the set of all  $\mathbf{x} \in X_f$  satisfying (1) and call it the set of fine model minimizers.

We consider the fine model to be expensive to compute and solving (1) by direct optimization to be impractical. Instead, we use surrogate models, i.e., models that are not as accurate as the fine model but are computationally cheap, hence suitable for iterative optimization. We consider a general optimization algorithm that generates a sequence of points  $\mathbf{x}^{(i)} \in X_f$ ,  $i=1, 2, \dots$ , and a family of surrogate models  $\mathbf{R}_s^{(i)}: X_f \rightarrow R^m$ ,  $i=0, 1, \dots$ , so that

$$\mathbf{x}^{(i+1)} = \arg \min_{\mathbf{x} \in X_f} U(\mathbf{R}_s^{(i)}(\mathbf{x})) \quad (4)$$

and  $\mathbf{R}_s^{(i+1)}$  is constructed using suitable matching conditions with the fine model at  $\mathbf{x}^{(i+1)}$  (and, perhaps, some of  $\mathbf{x}^{(k)}$ ,  $k=1, \dots, i$ ). If the solution to (4) is non-unique we may impose regularization. We may match responses, i.e.,

$$\mathbf{R}_s^{(i)}(\mathbf{x}^{(i)}) = \mathbf{R}_f(\mathbf{x}^{(i)}) \quad (5)$$

and/or match first order derivatives

$$\mathbf{J}_{\mathbf{R}_s^{(i)}}(\mathbf{x}^{(i)}) = \mathbf{J}_{\mathbf{R}_f}(\mathbf{x}^{(i)}) \quad (6)$$

where  $\mathbf{J}_{\mathbf{R}_s^{(i)}}$  and  $\mathbf{J}_{\mathbf{R}_f}$  denote jacobians of the surrogate and fine models, respectively. More precisely, we try to define models so that conditions such as (5) and (6) are satisfied.

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.

### III. SM BASED SURROGATE MODELS

The family of surrogate models  $\{\mathbf{R}_s^{(i)}\}$  can be implemented in various ways. SM assumes the existence of a so-called coarse model that describes the same object as the fine model: less accurate but much faster to evaluate. Let  $\mathbf{R}_c: X_c \rightarrow R^m$  denote the response vectors of the coarse model, where  $X_c \subseteq R^n$ . In the sequel, we assume for simplicity that  $X_c = X_f$ . By  $\mathbf{x}_c^*$  we denote the optimal solution of the coarse model, i.e.,

$$\mathbf{x}_c^* = \arg \min_{\mathbf{x} \in X_c} U(\mathbf{R}_c(\mathbf{x})) \quad (7)$$

We denote by  $X_c^*$  the set of all  $\mathbf{x} \in X_c$  satisfying (7) and call it the set of coarse model minimizers. In the SM framework, the 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.

#### A. Original SM Based Surrogate Model

The original SM approach assumes the existence of the mapping  $\mathbf{P}: X_f \rightarrow X_c$  such that  $\mathbf{R}_c(\mathbf{P}(\mathbf{x}_f)) \approx \mathbf{R}_f(\mathbf{x}_f)$  on  $X_f$  at least on some subset of  $X_f$  which is of interest. The proximity of  $\mathbf{R}_c$  and  $\mathbf{R}_f$  is measured using a suitable metric; in the ideal case we have  $\mathbf{R}_c(\mathbf{P}(\mathbf{x}_f)) = \mathbf{R}_f(\mathbf{x}_f)$ . For any given  $\mathbf{x}_f \in X_f$ ,  $\mathbf{P}(\mathbf{x}_f)$  is defined using parameter extraction:

$$\mathbf{P}(\mathbf{x}_f) = \arg \min_{\mathbf{x}} \|\mathbf{R}_c(\mathbf{x}) - \mathbf{R}_f(\mathbf{x}_f)\| \quad (8)$$

The surrogate model  $\mathbf{R}_s^{(i)}$  can be defined as

$$\mathbf{R}_s^{(i)}(\mathbf{x}) = \mathbf{R}_c(\mathbf{P}(\mathbf{x}^{(i)})) + \mathbf{B}^{(i)} \cdot (\mathbf{x} - \mathbf{x}^{(i)}) \quad (9)$$

for  $i=0,1,\dots$ , where  $\mathbf{P}$  is defined by (8) and  $\mathbf{B}^{(i)}$  is an approximation of  $\mathbf{J}_P(\mathbf{x}^{(i)})$ , the jacobian of  $\mathbf{P}$  at  $\mathbf{x}^{(i)}$ , obtained, e.g., by solving a Parameter Extraction (PE) problem of the form  $\mathbf{B}^{(i)} = \arg \min_{\mathbf{B}} \|\mathbf{J}_{R_f}(\mathbf{x}^{(i)}) - \mathbf{J}_{R_c}(\mathbf{P}(\mathbf{x}^{(i)})) + \mathbf{B} \cdot (\mathbf{x} - \mathbf{x}^{(i)})\|$ .

#### B. Output SM Based Surrogate Modeling

The Output Space Mapping (OSM) aims at reducing misalignment between the coarse and fine models by adding a difference (residual) between those two to  $\mathbf{R}_c$ . Let us define function  $\Delta \mathbf{R}: X_f \cap X_c \rightarrow R^m$  as

$$\Delta \mathbf{R}(\mathbf{x}) = \mathbf{R}_f(\mathbf{x}) - \mathbf{R}_c(\mathbf{x}) \quad (10)$$

We construct surrogates that use (local) models of  $\Delta \mathbf{R}$ , denoted as  $\Delta \mathbf{R}_m$ . A generic surrogate model defined by OSM is

$$\mathbf{R}_s^{(i)}(\mathbf{x}) = \mathbf{R}_c(\mathbf{x}) + \Delta \mathbf{R}_m(\mathbf{x}, \mathbf{x}^{(i)}) \quad (11)$$

We consider the zero-order model  $\Delta \mathbf{R}_m(\mathbf{x}, \mathbf{x}^{(i)}) = \Delta \mathbf{R}(\mathbf{x}^{(i)})$  which leads to the surrogate

$$\mathbf{R}_s^{(i)}(\mathbf{x}) = \mathbf{R}_c(\mathbf{x}) + \Delta \mathbf{R}(\mathbf{x}^{(i)}) \quad (12)$$

Model (12) is based on the matching condition (5).

The second model is a first-order approximation of  $\Delta \mathbf{R}$  of the form  $\Delta \mathbf{R}_m(\mathbf{x}, \mathbf{x}^{(i)}) = \Delta \mathbf{R}(\mathbf{x}^{(i)}) + \mathbf{J}_{\Delta \mathbf{R}}(\mathbf{x}^{(i)}) \cdot (\mathbf{x} - \mathbf{x}^{(i)})$ , where  $\mathbf{J}_{\Delta \mathbf{R}}(\mathbf{x}^{(i)})$  denotes the jacobian of  $\Delta \mathbf{R}$  at  $\mathbf{x}^{(i)}$ . This leads to

$$\mathbf{R}_s^{(i)}(\mathbf{x}) = \mathbf{R}_c(\mathbf{x}) + \Delta \mathbf{R}(\mathbf{x}^{(i)}) + \mathbf{J}_{\Delta \mathbf{R}}(\mathbf{x}^{(i)}) \cdot (\mathbf{x} - \mathbf{x}^{(i)}) \quad (13)$$

Instead of the exact jacobian (usually unavailable) we can use its approximation produced by the Broyden update. Model (13) is based on matching conditions (5) and (6).

#### C. Implicit Space Mapping Based Surrogate Modeling

Implicit Space Mapping (ISM) assumes that the coarse model depends on additional parameters, i.e., we have  $\mathbf{R}_c: X_c \times X_p \rightarrow R^m$  where  $X_p \subseteq R^p$  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

$$\mathbf{R}_s^{(i)}(\mathbf{x}) = \mathbf{R}_c(\mathbf{x}, \mathbf{x}_p^{(i)}) \quad (14)$$

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

$$\mathbf{x}_p^{(i)} = \arg \min_{\mathbf{x}_p} \|\mathbf{R}_f(\mathbf{x}^{(i)}) - \mathbf{R}_c(\mathbf{x}^{(i)}, \mathbf{x}_p)\| \quad (15)$$

The model (14), (15) is based on the matching condition (5).

### IV. CONVERGENCE PROPERTIES OF OSM ALGORITHM

In this section we examine the convergence properties of the algorithm (4) using the OSM based surrogate model (12). We give sufficient conditions for convergence. We assume that  $X_f = X_c = X \subseteq R^n$ . 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  $\mathbf{R}_c^*: Y_R \rightarrow R^n$  be the function defined as

$$\mathbf{R}_c^*(\mathbf{R}) = \arg \min_{\mathbf{x} \in X} U(\mathbf{R}_c(\mathbf{x}) + \mathbf{R}) \quad (16)$$

**Theorem 1.** Suppose that  $X$  is a closed subset of  $R^n$  and (i) the function  $\mathbf{R}_c^*$  is Lipschitz continuous on  $Y_R$ , i.e.,

$$\|\mathbf{R}_c^*(\mathbf{R}_2) - \mathbf{R}_c^*(\mathbf{R}_1)\| \leq L_C \|\mathbf{R}_2 - \mathbf{R}_1\| \quad (17)$$

for any  $\mathbf{R}_1, \mathbf{R}_2 \in Y_R$ , where  $\mathbf{R}_c^*(\mathbf{R}_2)$  is the solution to (16) which is closest in norm to  $\mathbf{R}_c^*(\mathbf{R}_1)$ , whereas  $L_C \in R_+$ ,

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

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

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

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

Then, for any  $\mathbf{x}^{(0)} \in X$  the sequence  $\{\mathbf{x}^{(i)}\}$  defined by (4), (12) is convergent, i.e., there is  $\mathbf{x}^* \in X$  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

$$U(\mathbf{R}_f(\bar{\mathbf{x}})) \leq U(\mathbf{R}_f(\mathbf{x}^*)) = U(\mathbf{R}_c(\mathbf{x}^*) + \Delta \mathbf{R}(\mathbf{x}^*)) \quad (19)$$

A proof of Theorem 1 is given in the Appendix.

This result shows that there are basically two fundamental (and natural) requirements for convergence of OSM based algorithms: (i) regularity of the perturbed coarse model optimal solution with respect to the perturbation vector, and (ii) similarity between the fine and coarse model in terms of the difference between their first-order derivatives.

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

**Corollary 1.** Suppose that for any  $\mathbf{R} \in Y_R$  there is  $\mathbf{x}_R \in X$  such that  $U(\mathbf{R}_c(\mathbf{x}_R) + \mathbf{R}) \leq U_{\min}$  and 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 (4), (12).

*Proof.* Convergence of the sequence  $\{\mathbf{x}^{(i)}\}$  follows from Theorem 1. 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  $\mathbf{x}^* \in X_f^*$ .

Under similar assumptions (and the requirement that the jacobian of function  $\Delta\mathbf{R}$  is bounded and Lipschitz continuous on  $X$ ) one can show convergence of the algorithm (4) using the OSM based surrogate model (13). It can also be shown that the convergence rate of algorithm (4), (13) is much better (we have  $\|\mathbf{x}^{(i+2)} - \mathbf{x}^{(i+1)}\| \leq C_2 \cdot \|\mathbf{x}^{(i+1)} - \mathbf{x}^{(i)}\|^2$ ) than algorithm (4), (12) (we have  $\|\mathbf{x}^{(i+2)} - \mathbf{x}^{(i+1)}\| \leq C_1 \cdot \|\mathbf{x}^{(i+1)} - \mathbf{x}^{(i)}\|$  in this case).

We consider an illustration of the convergence conditions. We analyze a generalized “cheese-cutting” problem [4]. The fine model is the irregular two-dimensional object of Fig. 1a. Its upper edge is defined by a positive function  $f$ . The corresponding coarse model is a rectangle of height  $H$  (Fig. 1b). Both models use the design parameter  $x$  (length).

We aim at finding  $x^*$  so that the area of the irregular object in Fig. 1a equals  $A_{opt}$ . The fine model  $\mathbf{R}_f: X \rightarrow R$  is given by

$$\mathbf{R}_f(x) = \int_0^x f(t) dt \quad (20)$$

The coarse model  $\mathbf{R}_c: X \rightarrow R$  is given by

$$\mathbf{R}_c(x) = Hx \quad (21)$$

In the sequel, we check the assumptions of Theorem 1 for  $\mathbf{R}_f$  and  $\mathbf{R}_c$  defined above assuming  $f(x) = H + \sin(x)\exp(-x/5)$ ,  $H=2$ ,  $X=[0,10]$ . Objective function  $U$  is defined by a norm (cf. (3)). In particular

(i) we have  $Y_R = \{y \in R: y = \int_0^x (f(t) - H) dt, x \in X\}$  and

$\mathbf{R}_c^*(\mathbf{R}) = \arg \min_x \|\mathbf{R}_c(x) + \mathbf{R} - A_{opt}\| = (A_{opt} - \mathbf{R}) / H$ . Thus,  $\|\mathbf{R}_c^*(\mathbf{R}_2) - \mathbf{R}_c^*(\mathbf{R}_1)\| \leq \|(A_{opt} - \mathbf{R}_2) / H - (A_{opt} - \mathbf{R}_1) / H\| \leq H^{-1} \|\mathbf{R}_2 - \mathbf{R}_1\|$ , i.e., assumption (i) is satisfied with  $L_C = H^{-1}$ .

(ii) since  $\Delta\mathbf{R}(x) = \int_0^x (f(t) - H) dt$ , we have

$$\|\Delta\mathbf{R}(y) - \Delta\mathbf{R}(x)\| \leq \left| \int_y^x (f(t) - H) dt \right| \leq \sup_{t \in X} |f(t) - H| \cdot \|y - x\|$$

i.e., assumption (ii) is satisfied with  $L_R(x) = \sup\{|f(t) - H|: t \in X\}$ . This estimate is pessimistic, because the local Lipschitz constant (i.e., the constant valid within some neighborhood of  $x$ ) is usually much lower,

(iii) we have  $L_R L_C = \sup\{|f(t) / H - 1|: t \in X\}$ . In the worst case we have  $L_R L_C < 1$  if  $f(x) / H < 2$  on  $X$ . For the assumed data we have  $L_R L_C < 0.5$ , enough to ensure convergence. Note that in our example, the assumptions of Corollary 1 are satisfied. Indeed, let  $\Delta\mathbf{R}$  be arbitrary except it is not larger than  $A_{opt}$ . Then  $\|\mathbf{R}_c(\mathbf{x}_{\Delta\mathbf{R}}) + \Delta\mathbf{R} - A_{opt}\| = 0 = U_{\min}$  is satisfied for  $\mathbf{x}_{\Delta\mathbf{R}} = (A_{opt} - \Delta\mathbf{R}) / H$ . Thus, we have ensured convergence to the optimal solution of the fine model.

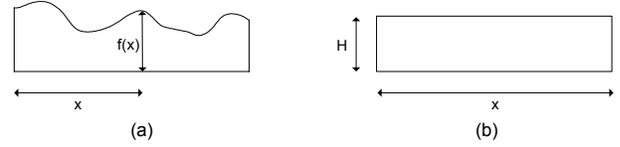


Fig. 1. The fine (a) and coarse (b) model in a two-dimensional generalized cheese-cutting problem.

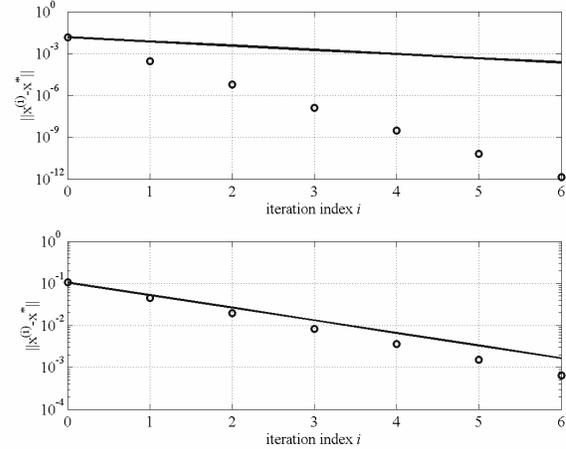


Fig. 2. Generalized cheese-cutting example: lower limit for the convergence rate (solid line), and actual convergence (circles) for  $A_{opt}=10$  (upper graph), and  $A_{opt}=2$  (lower graph).

Fig. 2 shows convergence of algorithm (4), (12) with the starting point being the coarse model optimal solution for  $A_{opt}=10$  (upper graph) and  $A_{opt}=2$  (lower graph). Solid lines denote a lower limit for the convergence rate (i.e., assuming  $L_R L_C = 0.5$ ). Circles denote the actual convergence rate, which is faster for  $A_{opt}=10$  because in this case, all iterations except  $x^{(0)}$  are located in the interval  $[4.65, 4.70]$  in which we have  $\sup\{|f(t) - H|: t \in [4.65, 4.7]\} < 0.05$ , which gives  $L_R L_C < 0.025$  (the starting point is  $x^{(0)}=5$ , and the final solution  $x^*=4.64663401$ ). For  $A_{opt}=2$  convergence is slower, due to the fact that the local Lipschitz constant around the optimal solution  $x^*=0.77715486$  is about  $L_R L_C \approx 0.5$ .

## V. EXAMPLE

We discuss the application of OSM based optimization algorithms to active filter design. This is a synthetic problem that aims at illustrating the performance of algorithm (4) with OSM based surrogate models (12) and (13). We consider a second-order OTA-C [9] low-pass filter (Fig. 3) implementing a Butterworth transfer function of the form

$$H_0(s) = (s^2 + \sqrt{2}s + 1)^{-1} \quad (22)$$

A coarse model of the filter is the transfer function formula assuming ideal transconductors and no parasitic elements:

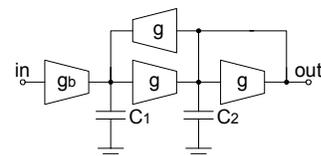


Fig. 3. Diagram of the second-order OTA-C low-pass filter.

$$H_c(s) = \frac{g_b g}{s^2 C_1 C_2 + s C_1 g + g^2} \quad (23)$$

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

$$H_f(s) = \frac{g_b g}{s^2 C_1^* C_2^* + s(C_1^*(2g_o + g) + 2C_2^* g_o) + g^2 + 2g_o(g + 2g_o)} \quad (24)$$

where  $C_i^* = C_i + C_p$ ,  $i=1,2$ . We use normalized elements with fixed  $g=1$ ,  $g_o=0.1$ ,  $C_p=0.1$ . Optimization variables are  $C_1$ ,  $C_2$  and  $g_b$ . The optimization problem is to find  $C_1$ ,  $C_2$  and  $g_b$  so that the difference between the fine model response vector  $\mathbf{R}_f$  and the target vector  $\mathbf{R}_{spec}$  is minimal w.r.t. the  $l_2$  norm, where  $\mathbf{R}_f = (|H_f(j\omega_1)|, \dots, |H_f(j\omega_m)|)$ , and  $\mathbf{R}_{spec} = (|H_0(j\omega_1)|, \dots, |H_0(j\omega_m)|)$ . We set  $m=21$  and  $\omega_1=0.0$ ,  $\omega_2=0.1$ , ...,  $\omega_{21}=2.0$ . Optimization variable values that correspond to the optimal solution of the coarse model are  $C_1=1.4142$ ,  $C_2=0.7071$ ,  $g_b=1.0$ .

We have applied algorithms (4), (12) and (4), (13) to find the fine model solution assuming  $\mathbf{x}^{(0)}=(1.4142, 0.7071, 1.0)$ . We consider (i) the surrogate is constructed directly using  $\mathbf{R}_c$  as in (12) or (13); (ii) the coarse model is improved using the so-called input SM [4]. In particular, at the  $i$ -th iteration we use  $\bar{\mathbf{R}}_c^{(i)}(\mathbf{x}) = \mathbf{R}_c(\mathbf{B}^{(i)} \cdot \mathbf{x} + \mathbf{c}^{(i)})$ , where  $\mathbf{B}^{(i)}$  and  $\mathbf{c}^{(i)}$  are  $n \times n$  and  $n \times 1$  matrices, respectively, obtained using multipoint PE of the form  $(\mathbf{B}^{(i)}, \mathbf{c}^{(i)}) = \arg \min_{\mathbf{B}, \mathbf{c}} \sum_{k=1}^i \|\mathbf{R}_f(\mathbf{x}^{(k)}) - \mathbf{R}_c(\mathbf{B} \cdot \mathbf{x}^{(k)} + \mathbf{c})\|$ .

This improvement reduces misalignment between the coarse and fine models so that the Lipschitz constants in convergence conditions are smaller. As a result we can expect a better convergence rate of the algorithm.

Fig. 4 shows the convergence properties of our algorithms (w.r.t. the  $l_2$  norm of the difference between  $\mathbf{R}_{spec}$  and  $\mathbf{R}_f$ ). The optimal fine model solution is  $\mathbf{x}^*=(1.157, 1.006, 1.390)$ . The results show that preconditioning of the coarse model by input SM and using a higher-order approximation of  $\Delta \mathbf{R}$  (i.e., model (13) versus (14)) both improve the convergence rate of the algorithm. This complies with the theory of Section IV.

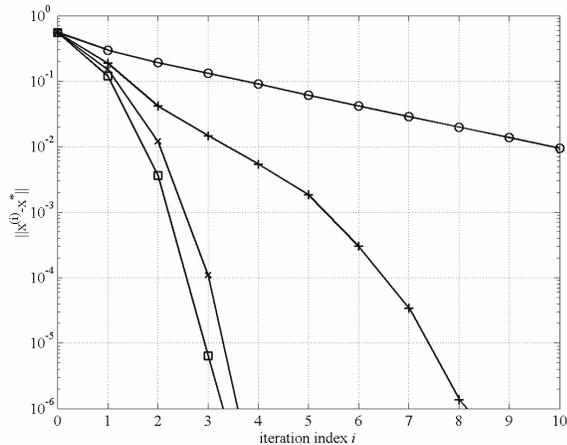


Fig. 4. Convergence of our OSM algorithms for the OTA-C design problem: algorithm (4), (12) without (o) and with (+) coarse model improvement, and algorithm (4), (13) without (x) and with (□) improvement.

## VI. CONCLUSIONS

A rigorous formulation of optimization algorithms using SM-based surrogate models is presented. Properties of a subclass, utilizing Output Space Mapping, are investigated. Theoretical results are followed by examples that explain the assumptions imposed on the coarse and fine model to ensure algorithm convergence. The paper is a first step in an SM theory that aims not only at a theoretical justification of SM methods but also at guidelines for designing more efficient algorithms.

### APPENDIX

*Proof of Theorem 1.* Take any  $\mathbf{x}^{(0)} \in X$ . Define  $\{\mathbf{x}^{(i)}\}$  according to (4), (12), 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

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

We would like to obtain an estimate for  $\|\mathbf{x}^{(i+2)} - \mathbf{x}^{(i+1)}\|$ .

Using (A1), and assumption (i)-(iii) we get

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

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

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

which is arbitrarily small for sufficiently large  $i$ , i.e.,  $\{\mathbf{x}^{(i)}\}$  is a Cauchy sequence. Thus, there is  $\mathbf{x}^* \in X$ ,  $\mathbf{x}^* = \lim_{i \rightarrow \infty} \mathbf{x}^{(i)}$ .

Estimate (19) is obvious. This ends the proof of the theorem.

### 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. Microwave Theory Tech.*, vol. 42, pp. 536-544, 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. Microwave Theory Tech.*, vol. 43, 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. Microwave Theory Tech.*, vol. 52, pp. 378-385, Jan. 2004.
- [4] J.W. Bandler, Q.S. Cheng, S.A. Dakroury, A.S. Mohamed, M.H. Bakr, K. Madsen and J. Sondergaard, "Space mapping: the state of the art," *IEEE Trans. Microwave Theory Tech.*, vol. 52, pp. 337-361, Jan. 2004.
- [5] S.J. Leary, A. Bhaskar and A. J. Keane, "A constraint mapping approach to the structural optimization of an expensive model using surrogates," *Optimization Eng.*, vol. 2, pp. 385-398, 2001.
- [6] M. Redhe and L. Nilsson, "Using space mapping and surrogate models to optimize vehicle crashworthiness design," 9th AIAA/ISSMO Multidisciplinary Analysis and Optimization Symp., Atlanta, GA, 2002, Paper AIAA-2002-5536.
- [7] 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, pp. 3627-3630, Sept. 2001.
- [8] K. Madsen and J. Søndergaard, Convergence of hybrid space mapping algorithms, *Optimization and Engineering*, vol. 5, pp. 145-156, Kluwer Academic Publishers, 2004.
- [9] Y. Sun (Editor), *Design of High Frequency Integrated Analogue Filters*, The Institution of Electrical Engineers, London, 2002.