On the Convergence of Space Mapping Optimization Algorithms

S. Koziel, J.W. Bandler and K. Madsen

Simulation Optimization Systems Research Laboratory
McMaster University, www.sos.mcmaster.ca, koziels@mcmaster.ca

Bandler Corporation, www.bandler.com, john@bandler.com
Technical University of Denmark, www.dtu.dk, km@imm.dtu.dk

presented at

THE EIGHTH SIAM CONFERENCE ON OPTIMIZATION
Stockholm, Sweden, May 15-19, 2005

Outline

space mapping (SM) concept

rigorous formulation of SM

SM optimization algorithms

convergence results: original SM

convergence results: output SM

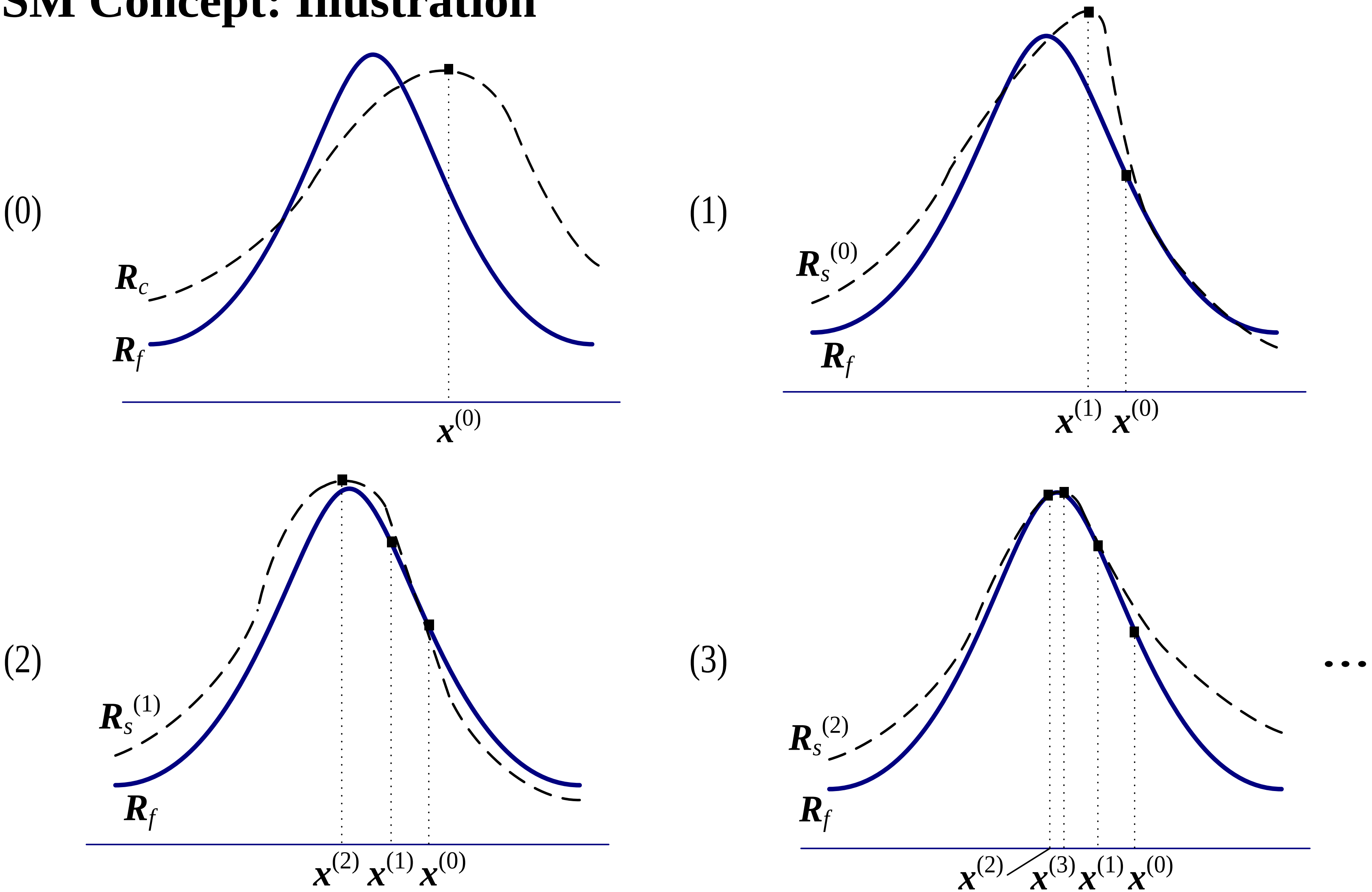
conclusions

Space Mapping Concept (*Bandler et al., 1994*)

problem: optimize an expensive (high fidelity or fine) model $\boldsymbol{R}_f$ with respect to given specifications

methodology: take advantage of a cheap but less accurate (low fidelity or coarse) model $\boldsymbol{R}_c$ so that the main optimization effort focuses on a (cheap) surrogate model $\boldsymbol{R}_s$ built through $\boldsymbol{R}_c$, while $\boldsymbol{R}_f$ is rarely referenced

SM Concept: Illustration



Space Mapping Algorithm Flow

Step 1 Set $i = 0$. Choose an initial solution $\boldsymbol{x}^{(0)}$

Step 2 Using data from $\boldsymbol{R}_c$ and $\boldsymbol{R}_f$ at $\boldsymbol{x}^{(k)}$, $k = 0, 1, \ldots, i$, determine the surrogate model $\boldsymbol{R}_s^{(i)}$

Step 3 Optimize $\boldsymbol{R}_s^{(i)}$ to obtain $\boldsymbol{x}^{(i+1)}$

Step 4 Evaluate $\boldsymbol{R}_f$ at $\boldsymbol{x}^{(i+1)}$

Step 5 Set $i = i + 1$

Step 6 **if** not termination condition go to 2; **else** go to 7

Step 7 END: return $\boldsymbol{x}^{(i)}$ as the final solution

Space Mapping: Rigorous Formulation (*Koziel et al., 2005*)

notation

$\boldsymbol{R}_f : X_f \rightarrow R^m$ fine model ($X_f \subseteq R^n$)

$\boldsymbol{R}_c : X_c \rightarrow R^m$ coarse model ($X_c \subseteq R^n$)

$U : R^m \rightarrow R$ objective function

optimization problem $\boldsymbol{x}_f^* = \arg\min_{\boldsymbol{x} \in X_f} U\left(\boldsymbol{R}_f(\boldsymbol{x})\right)$

generic SM optimization algorithm

$$\boldsymbol{x}^{(0)} \in X_f \cap X_c$$

$$\boldsymbol{x}^{(i+1)} = \arg\min_{\boldsymbol{x} \in X_s^{(i)}} U\left(\boldsymbol{R}_s^{(i)}(\boldsymbol{x})\right), \quad i = 0,1,\ldots$$

where $\boldsymbol{R}_s^{(i)} : X_s^{(i)} \rightarrow R^m$, i = 0, 1, 2, … is a family of surrogate models ($X_s^{(i)} \cap X_f \neq \emptyset$) based on $\boldsymbol{R}_c$ and determined using $\boldsymbol{R}_f$ data at the previous points $\boldsymbol{x}^{(k)}$, k = 0, 1, …, i.

SM Algorithms: Original SM (*Bandler et al., 1994*)

let $\boldsymbol{P} : X_f \longrightarrow X_c$ be defined as

$$\boldsymbol{P}(\boldsymbol{x}_f) = \arg\min_{\boldsymbol{x}\in X_c} \| \boldsymbol{R}_c(\boldsymbol{x}) - \boldsymbol{R}_f(\boldsymbol{x}_f) \|$$

define the *i*th surrogate model $\boldsymbol{R}_s^{(i)}$ as

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

where $\boldsymbol{B}^{(i)}$ is an approximation to the Jacobian of $\boldsymbol{P}$ at point $\boldsymbol{x}^{(i)}$

SM Algorithms: Input SM (*Bandler et al., 1994*)

define the ith surrogate model $\boldsymbol{R}_s^{(i)}$ as

$$\boldsymbol{R}_s^{(i)}(\boldsymbol{x}) = \boldsymbol{R}_c(\boldsymbol{B}^{(i)} \cdot \boldsymbol{x} + \boldsymbol{c}^{(i)})$$

where $\boldsymbol{B}^{(i)}$ and $\boldsymbol{c}^{(i)}$ are determined using the parameter extraction procedure

$$(\boldsymbol{B}^{(i)}, \boldsymbol{c}^{(i)}) = \arg\min_{(\boldsymbol{B},\boldsymbol{c})} \Big\{ \sum_{k=0}^{i} w_k \, \| \boldsymbol{R}_f(\boldsymbol{x}^{(k)}) - \boldsymbol{R}_c(\boldsymbol{B} \cdot \boldsymbol{x}^{(k)} + \boldsymbol{c}) \| + \\ + \sum_{k=0}^{i} v_k \, \| \boldsymbol{J}_{\boldsymbol{R}_f}(\boldsymbol{x}^{(k)}) - \boldsymbol{J}_{\boldsymbol{R}_c}(\boldsymbol{B} \cdot \boldsymbol{x}^{(k)} + \boldsymbol{c}) \cdot \boldsymbol{B} \| \Big\}$$

where w_k, v_k are weighting coefficients

SM Algorithms: Output SM

(*Bandler et al., 2004, Koziel et al., 2005*)

define the *i*th surrogate $\boldsymbol{R}_s^{(i)}$ as

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

where $\Delta\boldsymbol{R}_m(\cdot, \boldsymbol{x}^{(i)})$ is a local model of $\Delta\boldsymbol{R}(\boldsymbol{x}) = \boldsymbol{R}_f(\boldsymbol{x}) - \boldsymbol{R}_c(\boldsymbol{x})$ defined at point $\boldsymbol{x}^{(i)}$

first-order model: $\boldsymbol{R}_s^{(i)}(\boldsymbol{x}) = \boldsymbol{R}_c(\boldsymbol{x}) + \Delta\boldsymbol{R}(\boldsymbol{x}^{(i)})$

second-order model: $\boldsymbol{R}_s^{(i)}(\boldsymbol{x}) = \boldsymbol{R}_c(\boldsymbol{x}) + \Delta\boldsymbol{R}(\boldsymbol{x}^{(i)}) + \boldsymbol{J}_{\Delta\boldsymbol{R}}(\boldsymbol{x}^{(i)}) \cdot (\boldsymbol{x} - \boldsymbol{x}^{(i)})$

SM Algorithms: Implicit SM (*Bandler et al., 2004*)

let $\boldsymbol{R}_c : X_c{\times}X_p \rightarrow R^m$, i.e., $\boldsymbol{R}_c$ depend on additional (preassigned) parameters ($X_p \subseteq R^p$)

define the *i*th surrogate model as

$$\boldsymbol{R}_s^{(i)}(\boldsymbol{x}) = \boldsymbol{R}_c(\boldsymbol{x}, \boldsymbol{x}_p^{(i)})$$

where

$$\boldsymbol{x}_p^{(i)} = \arg\min_{\boldsymbol{x}\in X_p} \| \boldsymbol{R}_f(\boldsymbol{x}^{(i)}) - \boldsymbol{R}_c(\boldsymbol{x}^{(i)}, \boldsymbol{x}) \|$$

SM Algorithms: Combined Methods (*Koziel et al., 2005*)

it is possible (and utilized in practice) to combine the concepts discussed so far

for example, define the *i*th surrogate $\boldsymbol{R}_s^{(i)}$ as

$$\boldsymbol{R}_s^{(i)}(\boldsymbol{x}) = \boldsymbol{R}_c(\boldsymbol{B}^{(i)} \cdot \boldsymbol{x} + \boldsymbol{c}^{(i)}) + \Delta\boldsymbol{R}(\boldsymbol{x}^{(i)}) + \boldsymbol{J}_{\Delta R}(\boldsymbol{x}^{(i)}) \cdot (\boldsymbol{x} - \boldsymbol{x}^{(i)})$$

where $\boldsymbol{B}^{(i)}$ and $\boldsymbol{c}^{(i)}$ are determined using parameter extraction, whereas

$$\Delta\boldsymbol{R}(\boldsymbol{x}) = \boldsymbol{R}_f(\boldsymbol{x}) - \boldsymbol{R}_c(\boldsymbol{x})$$

Characterization of Space Mapping

shifting the optimization burden to an inexpensive coarse model

convergence to a reasonable fine model solution after a small number of fine model evaluations

significant speed-up of the optimization process in comparison with direct optimization

the possibility of solving problems that cannot be dealt with by direct optimization

the necessity of providing a computationally cheap and reasonably good coarse model

Convergence Theory: Original SM Algorithm (*Koziel et al., 2005*)

consider the idealized original SM algorithm

$$\boldsymbol{x}^{(0)} \in X_f$$

$$\boldsymbol{x}^{(i+1)} = \arg\min_{\boldsymbol{x} \in X_s^{(i)}} U(\boldsymbol{R}_s^{(i)}(\boldsymbol{x})), \quad i = 0, 1, 2, \ldots$$

where

$$\boldsymbol{R}_s^{(i)}(\boldsymbol{x}) = \boldsymbol{R}_c(\boldsymbol{P}(\boldsymbol{x}^{(i)}) + \boldsymbol{J}_{\boldsymbol{P}}(\boldsymbol{x}^{(i)}) \cdot (\boldsymbol{x} - \boldsymbol{x}^{(i)}))$$

$$\boldsymbol{P}(\boldsymbol{x}^{(i)}) = \arg\min_{\boldsymbol{x} \in X_c} \| \boldsymbol{R}_c(\boldsymbol{x}) - \boldsymbol{R}_f(\boldsymbol{x}^{(i)}) \|$$

$$X_s^{(i)} = \{\boldsymbol{x} \in X_f : \boldsymbol{P}(\boldsymbol{x}_f^{(i)}) + \boldsymbol{J}_{\boldsymbol{P}}(\boldsymbol{x}_f^{(i)}) \cdot (\boldsymbol{x} - \boldsymbol{x}_f^{(i)}) \in X_c\}$$

Convergence Theory: Original SM Algorithm

Assumption 1.1. Let X_f be a closed subset of R^n, and
(i) the space mapping $\boldsymbol{P}$ exists, it is differentiable on X_f and its Jacobian $\boldsymbol{J_P}$ is non-singular on X_f,
(ii) $\boldsymbol{J_P}^{-1}$ satisfies the Lipschitz condition with constant L_1, i.e.,

$$\| \boldsymbol{J}_{\boldsymbol{P}}^{-1}(\boldsymbol{x}) - \boldsymbol{J}_{\boldsymbol{P}}^{-1}(\boldsymbol{y}) \| < L_1 \| \boldsymbol{x} - \boldsymbol{y} \| \text{ on } X_f,$$

(iii) let $\boldsymbol{I_P}(\boldsymbol{x},\boldsymbol{y}) = \left[I_{P.ij}(\boldsymbol{x},\boldsymbol{y}) \right]_{i,j=1}^{n} = \boldsymbol{I} - \boldsymbol{J}_{\boldsymbol{P}}^{-1}(\boldsymbol{x}) \cdot \boldsymbol{J_P}(\boldsymbol{y})$. There is a matrix $\boldsymbol{J} = \left[J_{ij} \right]_{i,j=1}^{n}$ such that $|I_{P.ij}(\boldsymbol{x},\boldsymbol{y})| \le J_{ij}$ on X_f for i , j = 1, …, n.
(iv) the optimal solution of the coarse model exists and is unique (we shall denote it by $\boldsymbol{x}_c^*$),
(v) for any i = 0, 1, …, the optimal solution of $\boldsymbol{R}_s^{(i)}$ is in X_f,
(vi) there exists $\delta>0$ and $k>0$ such that $\boldsymbol{P}(\boldsymbol{x}^{(i)}) \in B(\boldsymbol{x}_c^*, \delta)$ for $i \ge k$,
(vii) L_1, δ and $\|\boldsymbol{J}\|$ are such that $L = L_1\delta + \|\boldsymbol{J}\| < 1$.

Convergence Theory: Original SM Algorithm

Theorem 1.1. Suppose that Assumption 1.1 is satisfied. Then, the sequence $\{\boldsymbol{x}^{(i)}\}$ is well defined and it is convergent to $\boldsymbol{x}^* \in X_f$.

Assumption 1.2. Suppose that the mapping $\boldsymbol{P}$ is exact, i.e., $\boldsymbol{R}_c(\boldsymbol{P}(\boldsymbol{x}_f)) = \boldsymbol{R}_f(\boldsymbol{x}_f)$ for any $\boldsymbol{x}_f \in X_f$, and the set of fine model minimizers is not empty.

Theorem 1.2. Suppose that Assumptions 1.1 and 1.2 hold. Then, the sequence $\{\boldsymbol{x}^{(i)}\}$ is well defined on X_f and it is convergent to $\boldsymbol{x}_f^* \in X_f$, the minimizer of the fine model $\boldsymbol{R}_f$.

Convergence Theory: Original SM Algorithm

Theorem 1.3. Suppose that Assumption 1.1 is satisfied and X_f^* is not empty. Suppose further that the objective function U is Lipschitz continuous with constant L_U on R^m, and $\boldsymbol{R}_c(\boldsymbol{x}_c^*) = \boldsymbol{R}_f(\boldsymbol{x}_f^*)$, where $\boldsymbol{x}_c^*$ and $\boldsymbol{x}_f^*$ are the coarse and fine model optimal solutions, respectively. Let $\boldsymbol{R}_\varepsilon : X_f \rightarrow R^m$ be a function defined, for any $\boldsymbol{x}_f \in X_f$, as

$$\boldsymbol{R}_\varepsilon(\boldsymbol{x}_f) = \boldsymbol{R}_c(\boldsymbol{P}(\boldsymbol{x}_f)) - \boldsymbol{R}_f(\boldsymbol{x}_f)$$

Then, the sequence $\{\boldsymbol{x}_f^{(i)}\}$ is well defined and convergent on X_f and there is $\boldsymbol{x}^* = \lim_{i\rightarrow\infty} \boldsymbol{x}_f^{(i)}$ satisfying $U(\boldsymbol{R}_f(\boldsymbol{x}^*)) \leq U_{\min} + L_U \|\boldsymbol{R}_\varepsilon(\boldsymbol{x}^*)\|$, where $U_{\min} = \min_{\boldsymbol{x}\in X_f} U(\boldsymbol{R}_f(\boldsymbol{x}))$.

Convergence Theory: Output SM Algorithm (*Koziel et al., 2005*)

consider an OSM algorithm of the form

$$\boldsymbol{x}^{(0)} \in X$$

$$\boldsymbol{x}^{(i+1)} = \arg\min_{\boldsymbol{x} \in X} U(\boldsymbol{R}_s^{(i)}(\boldsymbol{x})), \quad i = 0, 1, 2, \ldots$$

where

$$\boldsymbol{R}_s^{(i)}(\boldsymbol{x}) = \boldsymbol{R}_c(\boldsymbol{x}) + \Delta\boldsymbol{R}(\boldsymbol{x}^{(i)})$$

$$\Delta\boldsymbol{R}(\boldsymbol{x}) = \boldsymbol{R}_f(\boldsymbol{x}) - \boldsymbol{R}_c(\boldsymbol{x})$$

(we assume here that $X_f = X_c = X$)

Convergence Theory: Output SM Algorithm

Assumption 2.1. Let $\boldsymbol{x}_c^* : \Delta\boldsymbol{R}(X) \rightarrow R^n$ be a function defined as $\boldsymbol{x}_c^*(\boldsymbol{R}) = \arg\min_{\boldsymbol{x}\in X} U(\boldsymbol{R}_c(\boldsymbol{x}) + \boldsymbol{R})$. Suppose that the function $\boldsymbol{x}_c^*$ is Lipschitz continuous on $\Delta\boldsymbol{R}(X)$ with constant L.

Assumption 2.2. Let the function $\Delta\boldsymbol{R}$ be Lipschitz continuous on X with constant L_R.

Theorem 2.1. Suppose that X is a closed subset of R^n, Assumptions 2.1 and 2.2 hold, and L_R and L are such that $q = L_R L < 1$. Then, for any $\boldsymbol{x}^{(0)} \in X$ the sequence $\{\boldsymbol{x}^{(i)}\}$ is convergent to $\boldsymbol{x}^* \in X$.

Convergence Theory: Output SM Algorithm

Assumption 2.3. Suppose that the coarse model has the following property: for each $\boldsymbol{R} \in \Delta\boldsymbol{R}(X)$ there is $\boldsymbol{x}_R \in X$ such that $U(\boldsymbol{R}_c(\boldsymbol{x}_R)+\boldsymbol{R}) \leq U_{\min}$, where $U_{\min} = \min_{\boldsymbol{x} \in X_f} U(\boldsymbol{R}_f(\boldsymbol{x}))$.

Theorem 2.2. Suppose that Assumptions 2.1, 2.2 and 2.3 hold, a fine model minimizer exists, and $\boldsymbol{R}_c$ and U are continuous. Then, the sequence $\{\boldsymbol{x}^{(i)}\}$ is well defined on X and it is convergent to $\boldsymbol{x}_f^* \in X$, the minimizer of the fine model $\boldsymbol{R}_f$.

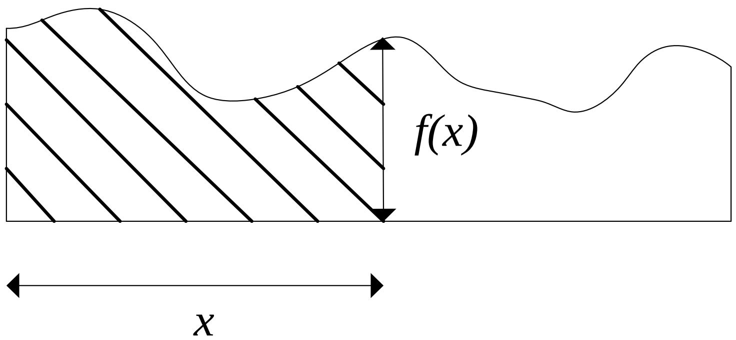
Convergence Theory: Output SM Algorithm

Remark 2.1. Under similar assumptions (and the requirement that the Jacobian of function $\boldsymbol{\Delta R}$ is bounded and Lipschitz continuous on X) one can show convergence of the OSM algorithm using the surrogate model $\boldsymbol{R}_s^{(i)}(\boldsymbol{x}) = \boldsymbol{R}_c(\boldsymbol{x}) + \boldsymbol{\Delta R}(\boldsymbol{x}^{(i)}) + \boldsymbol{J}_{\boldsymbol{\Delta R}}(\boldsymbol{x}^{(i)}) \cdot (\boldsymbol{x} - \boldsymbol{x}^{(i)})$.
It can also be shown that the convergence rate of this algorithm is much better (we have $\| \boldsymbol{x}^{(i+2)} - \boldsymbol{x}^{(i+1)} \| \leq C_2 \cdot \| \boldsymbol{x}^{(i+1)} - \boldsymbol{x}^{(i)} \|^2$) than the algorithm that uses the surrogate model $\boldsymbol{R}_s^{(i)}(\boldsymbol{x}) = \boldsymbol{R}_c(\boldsymbol{x}) + \boldsymbol{\Delta R}(\boldsymbol{x}^{(i)})$ (we have $\| \boldsymbol{x}^{(i+2)} - \boldsymbol{x}^{(i+1)} \| \leq C_1 \cdot \| \boldsymbol{x}^{(i+1)} - \boldsymbol{x}^{(i)} \|$ in this case).

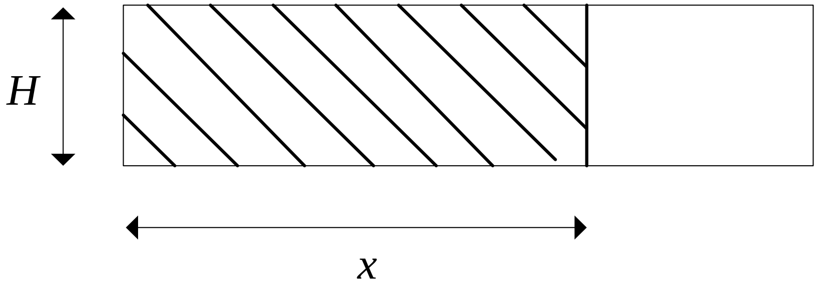
Example: Generalized "Cheese Cutting Problem"

(*Koziel et al., 2005*)

fine model: $\boldsymbol{R}_f(x) = \int_0^x f(t)\,dt$



coarse model: $\boldsymbol{R}_c(x) = Hx$



problem: find x^* such that $\boldsymbol{R}_f(x^*) = A_{opt}$ using the OSM algorithm

Example: Generalized "Cheese Cutting Problem"

assumptions: (i) $f(x) = H + \sin(x)\exp(-x/5)$

(ii) $H = 2$

(iii) fine/coarse model domain $X = [0,10]$

it follows that assumptions of Theorem 2.1 are satisfied with $L = H^{-1}$ and $L_R(x) = \sup\{|f(t) - H| : t \in X\}$. For our data, we have $L_R L = \sup\{|f(t)/H - 1| : t \in X\} < 0.5$

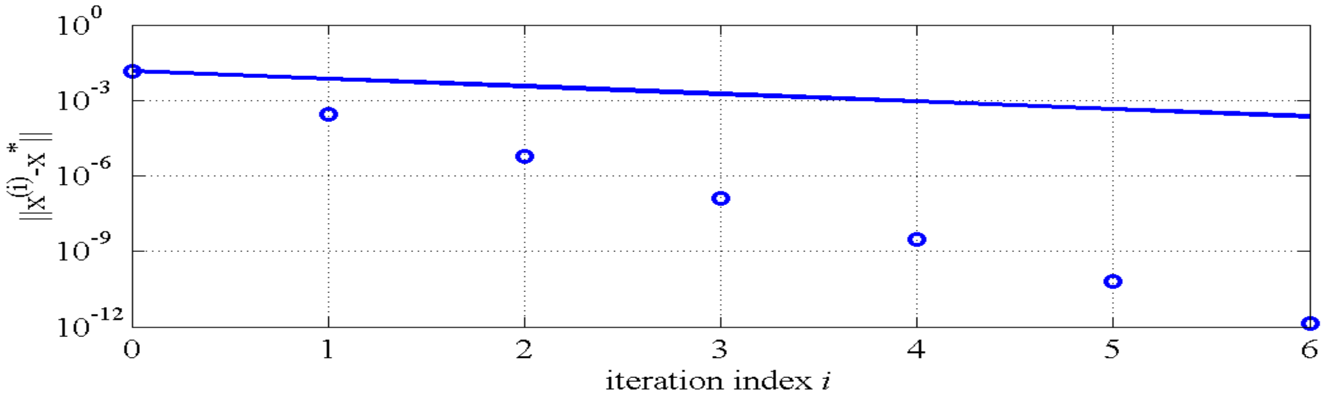
this assures global convergence of the OSM algorithm

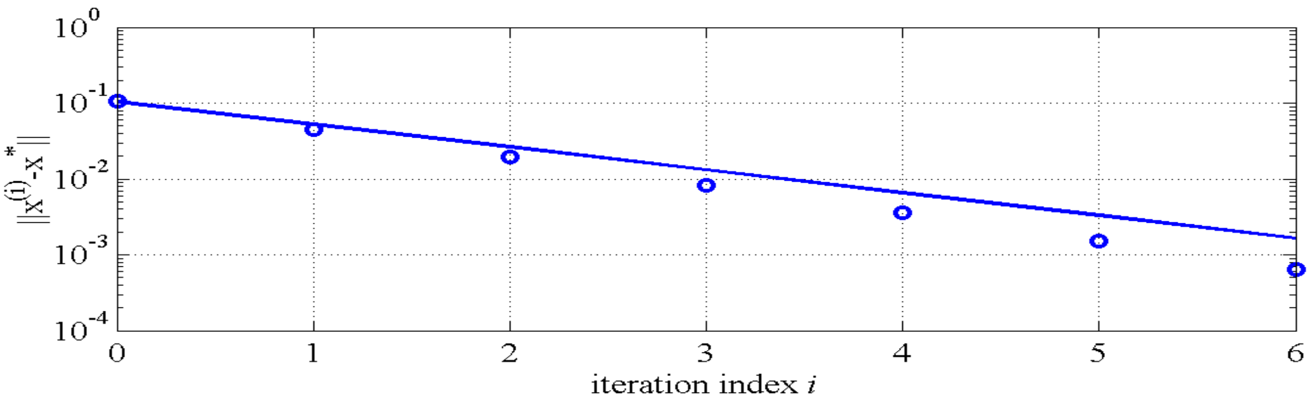
note: (i) the actual convergence rate depends on local Lipschitz constants around optimal solution.

(ii) it is easy to find f for which the OSM algorithm is not convergent for the considered problem.

Example: Generalized “Cheese Cutting Problem”

Lower limit for the convergence rate (line), and actual convergence (circles) for A_{opt} = 10 (upper graph), and A_{opt} = 2 (lower graph).





Conclusions

an exposition of recent convergence results for Space Mapping optimization algorithms has been presented

it follows that the fundamental (and natural) requirement for convergence of SM algorithms is similarity between the fine and coarse model (expressed by proper analytical conditions)

both convergence itself and the convergence rate depend on the quality of satisfying the above condition

future work will focus on obtaining convergence results for other types of SM optimization algorithms as well as unification of all SM algorithms in one theoretical framework

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. 2536–2544, Dec. 1994.

S. Koziel, J.W. Bandler and K. Madsen, “Towards a rigorous formulation of the space mapping technique for engineering design,” to appear, *Proc. Int. Symp. Circuits, Syst., ISCAS*, 2005.

J.W. Bandler, Q.S. Cheng, S.A. Dakroury, D.M. Hailu, K. Madsen, A.S. Mohamed and F. Pedersen, “Space mapping interpolating surrogates for highly optimized EM-based design of microwave devices,” *IEEE MTT-S Int. Microwave Symp. Digest* (Fort Worth, TX, 2004), pp. 1565–1568.

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.

S. Koziel, J.W. Bandler and K. Madsen, “An output space mapping framework for engineering optimization,” submitted, *Math. Programming*, 2005.

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. Microwave Theory and Tech.*, vol. 52, pp. 337–361, Jan. 2004.

References 2

M.H. Bakr, J.W. Bandler, N. Georgieva and K. Madsen, “A hybrid aggressive space mapping algorithm for EM optimization,” *IEEE Trans. Microwave Theory and Tech.*, vol. 47, pp. 2440–2449, Dec. 1999.

Q.S. Cheng, “Advances in space mapping technology exploiting implicit space mapping and output space mapping,” Ph.D. Thesis, Dept. Elect. Computer Eng., McMaster University, Hamilton, ON, Canada, 2004.

S. Koziel, J.W. Bandler and K. Madsen, “On the convergence of a space mapping optimization algorithm,” submitted, *SIAM Journal on Optimization*.

S. Koziel, J.W. Bandler, A.S. Mohamed and K. Madsen, “Enhanced surrogate models for statistical design exploiting space mapping technology,” to appear, *MTT-S Int. Microwave Symp. Dig.*, 2005.

J.W. Bandler, Q.S. Cheng and S. Koziel, “Implementable space mapping approach to enhancement of microwave device models,” to appear, *MTT-S Int. Microwave Symp. Dig.*, 2005.

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. Microwave Theory and Tech.*, vol. 49, pp. 67–79, Jan. 2001.

References 3

M. H. Bakr, J. W. Bandler, K. Madsen and J. Søndergaard, "An introduction to the space mapping technique," *Optimization Eng.*, vol. 2, pp. 369–384, 2001.

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.

M. H. Bakr, J.W. Bandler, R.M. Biernacki, S. H. Chen and K. Madsen, "A trust region aggressive space mapping algorithm for EM optimization," *IEEE Trans. Microwave Theory Tech.*, vol. 46, pp. 2412–2425, Dec. 1998.

Convergence Theory: Original SM Algorithm

Remark 1.1. Assumption 1.1 (v) is satisfied, particularly if $X_f = X_c = R^n$.

Remark 1.2. The global Lipschitz condition in Assumption 1.1 (ii) can be replaced by a (weaker) local one with slight changes of the proof.

Remark 1.3. Assumption 1.1 (vii) is motivated by the fact that in practice the coarse and fine models usually match each other closely.

Convergence Theory: Original SM Algorithm

Remark 1.4. When the function $\boldsymbol{R}_\varepsilon$ in Theorem 1.3 is not identically zero, the sequence $\{\boldsymbol{x}_f^{(i)}\}$ may not be convergent to the fine model optimal solution.

Remark 1.5. Theorem 1.3 says that the error of locating the fine model optimal solution is directly dependent on the error $\boldsymbol{R}_\varepsilon$ of the mapping $\boldsymbol{P}$. In particular, if $\boldsymbol{R}_\varepsilon \equiv 0$ then the limit point $\boldsymbol{x}^* = \lim_{i\to\infty} \boldsymbol{x}_f^{(i)}$ is the optimal solution of the fine model, i.e., $\boldsymbol{x}^* \in X_f^*$. However, Theorem 1.2 is more general than Theorem 1.3 with $\boldsymbol{R}_\varepsilon \equiv 0$, i.e., Theorem 1.2 cannot be obtained from Theorem 1.3 by letting $\boldsymbol{R}_\varepsilon \rightarrow 0$.