Advanced Optimization Techniques for Modern Filter Design—From Newton to Space Mapping

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

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

Bandler Corporation, www.bandler.com, john@bandler.com

Engineering Optimization and Modeling Center
Reykjavik University, www.ru.is, koziel@ru.is

Technical University of Denmark, www.dtu.dk, km@imm.dtu.dk

presented at

WSN: The State of Art of Microwave Filter Synthesis, Optimization and Realization
2010 IEEE MTT-S International Microwave Symposium, Anaheim, CA, May 23, 2010
Traditional Gradient-based Optimization

minimize w.r.t. $x$ a general, real-valued, non-linear function $F(x)$ in $n$ variables

traditional optimization algorithms are based on local information and Taylor’s formula

eyearly milestones in filter design by modern optimization methods
(Temes and Calahan, 1967, the state of the art)
(Lasdon et al., 1966, 1967, linear arrays and filters)
(Bandler, 1969, the state of the art)
(Director and Rohrer, 1969, adjoint sensitivity evaluation)
Variable Metric Methods (Quasi-Newton Methods)

local approximation at $\hat{x}$

$$q(x) = F(\hat{x}) + (x - \hat{x})^T F'(\hat{x}) + \frac{1}{2}(x - \hat{x})^T B(x - \hat{x})$$

where $B$ is a positive definite approximation to the Hessian of $F$ at $\hat{x}$

minimize $q$ and find the next iterate by a line search

(Davidon, 1959, Fletcher and Powell, 1963),
(Broyden, Fletcher, Goldfarb and Shanno (BFGS), independently around 1970)

trust regions were introduced by several authors in the early 1970s
Sequential Quadratic Programming

minimize w.r.t. $x$ a general, real-valued, non-linear function $F(x)$ in $n$ variables subject to a finite set of non-linear constraints

Han and Powell (1970s) developed a method similar to the variable metric method, with

- local quadratic approximation to the function
- constraints approximated by linear terms using first-order Taylor expansions
- the local subproblems solved by quadratic programming
- line search applied
Type of Approximation/Optimization Problem Considered

minimize w.r.t. $x$ the absolute values of the deviations between response $r(x, t_i)$ and specifications $y_i$

$$f_i(x) = r(x, t_i) - y_i, \ i = 1, \ldots, m$$

traditional methods are based on local information and Taylor’s formula, including

- least-squares formulation
- minimax formulation
- $L_1$ formulation
- general formulation
Traditional Least-Squares Formulation
(Levenberg, 1944, Marquardt, 1963)

$$F(x) = \sum_{i=1}^{m} f_i^2(x)$$

local approximation at $\hat{x}$:

$$\hat{L}(x) = \sum_{i=1}^{m} \hat{l}_i^2(x)$$

$$\hat{l}_i(x) = f_i(\hat{x}) + f_i'(\hat{x})^T (x - \hat{x})$$

minimize a damped version of $\hat{L}$

minimize $\hat{L}$ subject to some trust region (Moré, 1983)
Traditional Minimax Formulation

(Madsen, 1975)

\[ F(x) = \max_i \left| f_i(x) \right| \]

local approximation at \( \hat{x} \):

\[ \hat{L}(x) = \max_i \left| \hat{l}_i(x) \right| \]

\[ \hat{l}_i(x) = f_i(\hat{x}) + f'_i(\hat{x})^T (x - \hat{x}) \]

minimize \( \hat{L} \) subject to some trust region
Traditional $L_1$ Formulation  
(Hald and Madsen, 1985)

\[
F(x) = \sum_{i=1}^{m} |f_i(x)|
\]

local approximation at $\hat{x}$:

\[
\hat{L}(x) = \sum_{i=1}^{m} |\hat{l}_i(x)|
\]

\[
\hat{l}_i(x) = f_i(\hat{x}) + f_i'(\hat{x})^T (x - \hat{x})
\]

minimize $\hat{L}$ subject to some trust region
General Formulation
(Madsen, 1986)

\[
\text{minimize } F(x) = H(f(x))
\]

at the iteration \( \hat{x} \):

\[
\hat{L}(x) = H(\hat{l}(x))
\]

\[
\hat{l}_i(x) = f_i(\hat{x}) + f_i'(\hat{x})^T (x - \hat{x})
\]

minimize \( \hat{L} \) subject to some trust region
The **Space Mapping** Concept

*(Bandler et al., 1994-)*

- **Validation Space**
- **Optimization Space**
- **Mapping**
- **Surrogate Update**
- **Prediction**

(high-fidelity physics model)

(surrogate optimization)

(low-fidelity physics model)
Original **Space Mapping Optimization**

*(Bandler et al., 1994-)*

find mapping $P(x)$ through parameter extraction

\[
\begin{align*}
\text{coarse model} & \quad \text{mapping} \quad \text{fine model} \\
\mathbf{z} & = P(x) \equiv \arg\min_{\mathbf{z}} \left\{ \| f(x) - c(z) \| \right\} 
\end{align*}
\]
Aggressive Space Mapping Optimization
(Bandler et al., 1995)

estimate mapping $P$ at the $k$th iteration

assume $P$ has been computed at $x_0, x_1, \ldots, x_k$

$$P(x) \approx P(x_k) + P'(x_k)(x - x_k)$$
$$\approx P(x_k) + B_k(x - x_k)$$
$$\equiv P_k(x)$$

where $B_k \approx P'(x_k)$ is, e.g., a Broyden (1970) update

approximate aim: $P_k(x) = z^* \rightarrow x_{k+1}$
Aggressive Space Mapping Optimization
(Bandler et al., 1995)

first iteration

\[ B_0 = I \]

let

\[ P_0(x) \equiv P(x_0) + B_0(x - x_0) \]

\[ \approx P(x) \]

solve

\[ P_0(x) = z^* \rightarrow x_1 \]
Space Mapping Optimization Methodologies

Taylor approximation

local information

error

space mapping

error

\( f(x) \)

\( c(P_k(x)) \)

\( f(x) \)
Space Mapping Approximation Errors (Bakr et al., 2001)

local Taylor approximation error at $x_k$

space mapping surrogate approximation error for $k$th mapping at $x_k$
Space Mapping vs. Taylor Approximation

use of a suitable coarse (surrogate) model may provide large iteration steps

space mapping may provide a good approximate solution in a few iteration steps

large iteration steps: space mapping is best
small iteration steps: Taylor is best?

beyond aggressive space mapping: to enhance space mapping for all size steps
Linking Companion Coarse (Empirical) and Fine (EM) Models Via Space Mapping (Bandler et al., 1994-)

\[ \nabla \times H = j \omega D + J \\
D = \varepsilon E \\
\nabla \times E = -j \omega B \\
B = \mu H \\
\n\varepsilon = \frac{\rho}{\omega} \\
\mu = \frac{\rho}{\omega} \\
\]

Obtain a mapping to match the models (parameter extraction).
Implicit, Input and Output Space Mappings
(Bandler et al., 2003-)

expert engineering knowledge helpful (few designable variables)
expertise helpful in “tuning the surrogate” (many possibilities, e.g., dielectric constant)
engineering expertise perhaps less necessary (many output variables)
The Novice-Expert Continuum

**output space mapping**: a “band-aid” solution for engineers and non-engineers; the parameter extraction step does not require coarse model re-analysis; good for final touch-ups

**input space mapping**: an engineering approach to find and cure the root-cause of a defect; but the parameter extraction step can be a difficult inverse optimization problem to solve w.r.t. the coarse model

**tuning space mapping** (new): simulator-based expert approach

but all types of **space mapping** can be viewed as special cases of **implicit space mapping**
Aggressive Space Mapping Design of Dielectric Resonator Multiplexers
(Ismail et al., 2003, Com Dev, Canada)

10-channel output multiplexer, 140 variables
For the Expert: **Tuning Space Mapping (TSM)**

*Bandler et al., 2006-

design parameters

---

space mapping

tuning parameters

---

fine model

---

responses

---

internal tuning port

---

surrogate based on the fine model with internal tuning ports
Tuning Space Mapping (TSM) Flowchart

Classical Space Mapping
*(Bandler et al., 2004)*

- **Start**
- Select models and mapping framework
- Find initial guess
- Simulate fine model
- Criterion satisfied?
  - Yes: End
  - No: Update surrogate model
- Optimize surrogate model

Tuning Space Mapping
*(Koziel et al., 2008)*

- **Start**
- Select models
- Find initial guess
- Simulate fine model
- Criterion satisfied?
  - Yes: End
  - No: Update tuning model
- Optimize tuning model
- Translate tuning parameters to design parameters
Tuning Space Mapping (TSM): Type 1 and Type 0 Embedding

- Fine model
- Internal ports
- Design parameters
- Responses
- Tuning model
- Tuning element (Type 1 embedding)
- Tuning element (Type 0 embedding)
Open-loop Ring Resonator Bandpass Filter \textit{(Koziel et al., 2008)}

\begin{align*}
\text{design parameters} \\
x &= [L_1 \ L_2 \ L_3 \ L_4 \ S_1 \ S_2 \ g]^T \text{ mm} \\
\text{specifications} \\
|S_{21}| &\geq -3 \text{ dB for } 2.8 \text{ GHz} \leq \omega \leq 3.2 \text{ GHz} \\
|S_{21}| &\leq -20 \text{ dB for } 1.5 \text{ GHz} \leq \omega \leq 2.5 \text{ GHz} \\
|S_{21}| &\leq -20 \text{ dB for } 3.5 \text{ GHz} \leq \omega \leq 4.5 \text{ GHz}
\end{align*}
Open-loop Ring Resonator Bandpass Filter (Type 1 and Type 0)

Sonnet *em* model with internal (co-calibrated) ports

![Diagram of the open-loop ring resonator bandpass filter](image)
Open-loop Ring Resonator Bandpass Filter (Type 1 and Type 0)

Sonnet *em* model with internal (co-calibrated) ports

![Diagram of open-loop ring resonator bandpass filter](image)
Open-loop Ring Resonator Bandpass Filter (Type 1 and Type 0)

initial responses: tuning model (—), fine model (○),
fine model with co-calibrated ports (---)
Open-loop Ring Resonator Bandpass Filter (Type 1 and Type 0)

responses after two iterations: the tuning model (—), corresponding fine model (○)
Tuning (Implicit) Space Mapping Algorithm (Cheng et al., 2010)

original problem

\[ x^*_f = \arg\min_x U \left( R_f(x) \right) \]

align the surrogate to match fine model

\[ x^{(i)}_p = \arg\min_{x_p} \left\| R_f(x^{(i)}) - R^{(i)}_s(x^{(i)}, x_p) \right\| \]

design parameter value prediction

\[ x^{(i+1)} = \arg\min_x U \left( R^{(i)}_s(x, x^{(i)}_p) \right) \]
Response-Corrected **Tuning Space Mapping Algorithm**

*(Cheng et al., 2010)*

the response-corrected tuning model at optimum \( x^* \)

\[
\bar{R}_s(x) = R_s(x, x_p^*) + R_f(s(x^*)) - R_s(s(x^*), x_p^*)
\]

\( s \) is a function that snaps a point to the nearest fine model grid point
Yield Analysis and Yield Optimization (Cheng et al., 2010)

**Step 1** Use tuning space mapping to obtain a nominal optimal design. A tuning model or surrogate is also obtained.

**Step 2** Snap the optimal design to the nearest on-grid fine model point.

**Step 3** Simulate the snapped design (EM fine model).

**Step 4** Calculate the response difference between the fine model and the surrogate at the nearest on-grid point.

**Step 5** Add the response difference to the surrogate to form a new surrogate: the response corrected surrogate.

**Step 6** Perform yield analysis and yield optimization on the response-corrected surrogate.

**Step 7** Compare this response to that of the fine model.
Second-order Tapped-line Microstrip Filter (Type 1)

Term 1
\[ Z = 50 \text{ Ohm} \]

Term 2
\[ Z = 50 \text{ Ohm} \]

MLIN TL1
\[ W = 1 \text{ mm} \]
\[ L = (L_1 - 1) \text{ mm} \]

MLIN TL2
\[ W = 1 \text{ mm} \]
\[ L = (L_2 - 1) \text{ mm} \]

MCLIN TL1
\[ W = 1 \text{ mm} \]
\[ L = L_3 \text{ mm} \]
\[ S = g \text{ mm} \]

MLIN TL3
\[ W = 1 \text{ mm} \]
\[ L = (L_2 - 1) \text{ mm} \]
Second-order Tapped-line Microstrip Filter (Type 1)

tuning model (―), fine model (○),
response corrected surrogate (―)

yield analysis (500 trials)
Space Mapping with Constrained Parameter Extraction: Concept (Koziel et al., 2010)

selection of surrogate is critical for space mapping performance

a novel technique replaces “manual” adjustment of the type and number of space mapping parameters, based on an adaptively constrained parameter extraction process

• construct initial, over-flexible surrogate with excellent approximation capability (so that $\varepsilon^{(i)} = ||R_f(x^{(i)}) - R_c(x^{(i)}, x_p^{(i)})||$ can be brought to a very small value)
• adjust its generalization capability by constraining the parameter space
Wideband Bandstop Microstrip Filter

fine model (FEKO)

coarse model (Agilent ADS)

Input $W_c$  

Output

$G_c$

$L_c$

$L_r$

$W_r$

design parameters: $x = [L_r \ W_r \ L_c \ W_c \ G_c]^T$

initial surrogate model has 10 parameters
## Wideband Bandstop Microstrip Filter: Optimization Results

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Specification Error</th>
<th>Number of Fine Model Evaluations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Best Found</td>
<td>Final</td>
</tr>
<tr>
<td>Standard SM</td>
<td>–1.8 dB</td>
<td>–1.7 dB</td>
</tr>
<tr>
<td>Constrained SM</td>
<td>–2.0 dB</td>
<td>–2.0 dB</td>
</tr>
</tbody>
</table>

*algorithm terminated after 20 iterations without convergence*
Wideband Bandstop Microstrip Filter: Responses

fine model at initial (dashed line) and final (solid line) designs obtained using the constrained space mapping algorithm
Wideband Bandstop Microstrip Filter: Convergence

constrained SM algorithm (o) versus standard SM algorithm (×):

convergence plot

specification error evolution
Conclusions

filters have been designed by modern optimization techniques for some 45 years

traditional Newton-based methods employ Taylor approximations
  ● required for "coarse" or "surrogate" optimizations
  ● required for model alignment (parameter extraction)

space mapping harnesses physics-based surrogates to remove expensive "fine" models from traditional optimization loops

space mapping facilitates full-wave EM-based, as well as multidisciplinary design optimization
References 1


References 2


References 3


References 4


References 5


Space Mapping with Constrained Parameter Extraction: Details

constrained parameter extraction process:

\[ x_p^{(i)} = \arg \min_{l^{(i)} \leq x_p \leq u^{(i)}} \| R_f(x^{(i)}) - R_c(x^{(i)}, x_p) \| \]  

\( (*) \)

where \( l^{(i)} = x_p^{(i-1)} - \delta^{(i)} \) and \( u^{(i)} = x_p^{(i-1)} + \delta^{(i)} \)

\( x_p^{(i-1)} \) model parameters at iteration \( i - 1 \)

\( \delta^{(i)} \) surrogate model parameter space size at iteration \( i \)

Updating \( l^{(i)} \) and \( u^{(i)} \) (\( \delta^{(i)} \), \( x_p^{(i-1)} \) and \( \varepsilon_{\text{max}} \) are input arguments):

1. Calculate \( l^{(i)} = x_p^{(i-1)} - \delta^{(i)} \) and \( u^{(i)} = x_p^{(i-1)} + \delta^{(i)} \); 
2. Find \( x_p^{(i)} \) using \( (*) \); 
3. If \( \varepsilon^{(i)} \leq \alpha_{\text{decr}} \cdot \varepsilon_{\text{max}} \) then \( \delta^{(i+1)} = \delta^{(i)} / \beta_{\text{decr}} \); Go to 5; 
4. If \( \varepsilon^{(i)} > \alpha_{\text{incr}} \cdot \varepsilon_{\text{max}} \) then \( \delta^{(i+1)} = \delta^{(i)} \cdot \beta_{\text{incr}} \); Go to 5; 
5. END; 

Typically: \( \alpha_{\text{decr}} = 1, \alpha_{\text{incr}} = 2, \beta_{\text{decr}} = 5, \beta_{\text{incr}} = 2 \)
Space Mapping with Constrained Parameter Extraction: Interpretation

our algorithm tightens the constraints if the approximation error is sufficiently small, loosens them otherwise

constraint tightening improves the generalization capability of the surrogate: low error \( \varepsilon^{(i-1)} = \| R_f(x^{(i-1)}) - R_c(x^{(i-1)}, x_p^{(i-1)}) \|_p \)
and \( \varepsilon^{(i)} = \| R_f(x^{(i)}) - R_c(x^{(i)}, x_p^{(i)}) \|_p \) makes it more likely to have \( \| R_f(x^{(i-1)}) - R_c(x^{(i-1)}, x_p^{(i)}) \|_p \) small if \( \delta^{(i)} \) is reduced because a small \( \| x_p^{(i)} - x_p^{(i-1)} \|_\infty \leq \| \delta^{(i)} \|_\infty \) implies similarity of the subsequent surrogate models

improved performance follows from rigorous convergence results for space mapping algorithms