Manifold Mapping for Multilevel Optimization*

Pieter W. Hemker and David Echeverría

Centrum voor Wiskunde en Informatica
Kruislaan 413, NL 1098 SJ Amsterdam, The Netherlands
{P.W.Hemker,D.Echeverria}@cwi.nl

Summary. We first show the idea behind a space-mapping iteration technique for the efficient solution of optimization problems. Then we show how space-mapping optimization can be understood in the framework of defect correction. We observe a difference between the solution of the optimization problem and the computed space-mapping solutions. We repair this discrepancy by exploiting the correspondence with defect correction iteration and we construct the manifold-mapping algorithm, which is as efficient as the space-mapping algorithm but converges to the accurate solution.

1 Introduction

Space mapping (Bandler et al. [1, 2]) is a technique to reduce the computing time in demanding optimization procedures by means of simple surrogate models. Space mapping makes use of both accurate (and time-consuming) models and less accurate (but cheaper) ones.

The original space-mapping procedure corresponds with right-preconditioning the coarse (inaccurate) model in order to accelerate the iterative procedure for the optimization of the fine (accurate) one. The iterative procedure used in space mapping for optimization can be understood as a defect correction iteration [3] and the convergence can be analyzed accordingly. We show that, right-preconditioning is generally insufficient and (also) left-preconditioning is needed. This leads to the improved space-mapping or 'manifold-mapping' procedure. This manifold mapping is shown in some detail in Section 5

2 Fine and coarse models in optimization

The optimization problem.

The specifications of an optimization problem are denoted by $(\mathbf{t}, \mathbf{y}) \equiv (t_i, y_i)_{i=1,\ldots,m}$. The independent variable is $\mathbf{t} \in \mathbb{R}^m$. The dependent variable $\mathbf{y} \in Y \subset \mathbb{R}^m$ represents

*Invited Paper at SCEE-2006

the quantities that describe the behavior of the phenomena under study. The set Y is the *set of possible aims*.

The variable $\mathbf{y}$ does not only depend on $\mathbf{t}$ but also on control/design variables, $\mathbf{x}$. The difference between the measured data y_i and the values $y(t_i, \mathbf{x})$ may be the result of, e.g., measurement errors or the imperfection of the mathematical description.

Models that describe reality appear in several degrees of sophistication. Space mapping exploits the combination of the simplicity of the less sophisticated methods with the accuracy of the more complex ones. Therefore we distinguish the fine and the coarse model.

The fine model.

The *fine model* response is denoted by $\mathbf{f}(\mathbf{x}) \in \mathbb{R}^m$, with $\mathbf{x} \in X \subset \mathbb{R}^n$ the *fine model control variable*. The set $\mathbf{f}(X) \subset \mathbb{R}^m$ represents the fine model reachable aims. Notice that, with $n < m$, $\mathbf{f}(X)$ is an n-dimensional manifold in $Y \subset \mathbb{R}^m$. The fine model is assumed to be *accurate* but *expensive* to evaluate. For the optimization problem a *fine model cost function* $\| \mathbf{f}(\mathbf{x}) - \mathbf{y} \|$ should be minimized. So we look for

$$\mathbf{x}^* = \underset{\mathbf{x} \in X}{\operatorname{argmin}} \| \mathbf{f}(\mathbf{x}) - \mathbf{y} \| . \tag{1}$$

A design problem, characterized by the model $\mathbf{f}(\mathbf{x})$, the aim $\mathbf{y} \in Y$, and the space of possible controls $X \subset \mathbb{R}^n$, is a *reachable design* if the equality $\mathbf{f}(\mathbf{x}^*) = \mathbf{y}$ can be achieved for some $\mathbf{x}^* \in X$.

The coarse model.

The *coarse model* is denoted by $\mathbf{c}(\mathbf{z}) \in \mathbb{R}^m$, with $\mathbf{z} \in Z \subset \mathbb{R}^n$ the *coarse model control variable*. This model is assumed to be *cheap* to evaluate but *less accurate* than the fine model. The set $\mathbf{c}(Z) \subset \mathbb{R}^m$ is the set of *coarse model reachable aims* . For the coarse model we have the *coarse model cost function* $\| \mathbf{c}(\mathbf{z}) - \mathbf{y} \|$ and we denote its minimizer by $\mathbf{z}^*$,

$$\mathbf{z}^* = \underset{\mathbf{z} \in Z}{\operatorname{argmin}} \| \mathbf{c}(\mathbf{z}) - \mathbf{y} \| . \tag{2}$$

The space-mapping function.

The similarity or discrepancy between the responses of two models is expressed by the *misalignment function* $r(\mathbf{z}, \mathbf{x}) = \| \mathbf{c}(\mathbf{z}) - \mathbf{f}(\mathbf{x}) \|$. For a given $\mathbf{x} \in X$ it is useful to know which $\mathbf{z} \in Z$ yields the smallest discrepancy. Therefore, the *space-mapping function* $\mathbf{p} : X \subset \mathbb{R}^n \to Z \subset \mathbb{R}^n$ is introduced,

$$\mathbf{p}(\mathbf{x}) = \underset{\mathbf{z} \in Z}{\operatorname{argmin}}\, r(\mathbf{z}, \mathbf{x}) = \underset{\mathbf{z} \in Z}{\operatorname{argmin}} \| \mathbf{c}(\mathbf{z}) - \mathbf{f}(\mathbf{x}) \| . \tag{3}$$

Perfect mapping.

To identify the cases where the accurate solution $\mathbf{x}^*$ is related with the less accurate solution $\mathbf{z}^*$ by the space mapping function, a space-mapping function $\mathbf{p}$ is called a *perfect mapping* iff $\mathbf{z}^* = \mathbf{p}(\mathbf{x}^*)$.

We notice that *perfection* is not a property of the space-mapping function alone, but it also depends on the data $\mathbf{y}$ considered. A space-mapping function can be perfect for one data set but imperfect for a different data set, and if a design is reachable a space mapping is always perfect irrespective of the coarse model used.

3 Primal and dual space-mapping solutions

In literature many space mapping based algorithms can be found [1, 2], where two types can be distinguished: the primal and the dual.

The *primal* space-mapping approach seeks for a solution of the minimization problem

$$\mathbf{x}_p^* = \underset{\mathbf{x}\in X}{\operatorname{argmin}} \|\mathbf{p}(\mathbf{x}) - \mathbf{z}^*\| \,. \tag{4}$$

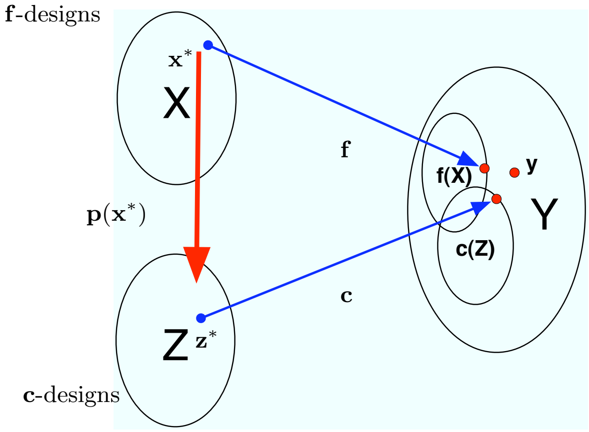
The dual determines

$$\mathbf{x}_d^* = \underset{\mathbf{x}\in X}{\operatorname{argmin}} \| \, \mathbf{c}(\mathbf{p}(\mathbf{x})) - \mathbf{y} \| \;, \tag{5}$$

where we can recognize $\mathbf{c}(\mathbf{p}(\mathbf{x}))$ as a *surrogate model.*

Both approaches coincide when $\mathbf{z}^* \in \mathbf{p}(X)$ and $\mathbf{p}$ is injective. If, in addition, the mapping is perfect both $\mathbf{x}_p^*$ and $\mathbf{x}_d^*$ are equal to $\mathbf{x}^*$. However, in general the space-mapping function $\mathbf{p}$ will not be perfect, and hence, a space-mapping based algorithm may *not* yield the solution of the fine model optimization. The principle of the approach is summarized in Figure 1.

Fig. 1. The space mapping function $\mathbf{p}(\mathbf{x}) = \underset{\mathbf{z}\in Z}{\operatorname{argmin}} \| \, \mathbf{c}(\mathbf{z}) - \mathbf{f}(\mathbf{x}) \| \,.$



The surrogate model: $\mathbf{c}(\mathbf{p}(\mathbf{x})) \approx \mathbf{f}(\mathbf{x})$.

4 Defect correction iteration

The efficient solution of a complex problem by the iterative use of a simpler one, is known since long in computational mathematics as defect correction iteration [3].

To solve a nonlinear operator equation

$$\mathcal{F}\,\mathbf{x} = \mathbf{y}, \tag{6}$$

where $\mathcal{F} : D \subset E \to \widehat{D} \subset \widehat{E}$ is a continuous, generally nonlinear operator and E and $\widehat{E}$ are Banach spaces, defect correction iteration reads

$$\begin{cases} \mathbf{x}_0 \quad = \widetilde{\mathcal{G}}_0\,\mathbf{y}\,, \\ \mathbf{x}_{k+1} = \widetilde{\mathcal{G}}_{k+1}\left(\widetilde{\mathcal{F}}_k\,\mathbf{x}_k - \mathcal{F}\,\mathbf{x}_k + \mathbf{y}\right)\,, \end{cases} \tag{7}$$

where $\widetilde{\mathcal{F}}_k$ is a simpler version of $\mathcal{F}$ and $\widetilde{\mathcal{G}}_k$ is the (simple-to-evaluate) left-inverse of $\widetilde{\mathcal{F}}_k$.

For our optimization problems, where the design may be not reachable, $\mathbf{y} \in \widehat{D}$ but $\mathbf{y} \notin \mathcal{F}(D)$, so that no solution for (6) exists. We want to find the solution of (1), or

$$\mathbf{x}^* = \operatorname{argmin}_{\mathbf{x}\in D}\, \|\mathcal{F}\mathbf{x} - \mathbf{y}\|_{\widehat{E}}\,,$$

which we associate with it a defect correction process for iterative optimization by taking $E = \mathbb{R}^n$, $\widehat{E} = \mathbb{R}^m$, $D = X$, $\widehat{D} = Y$ and by substitution of the operators:

$$\begin{array}{rcl} \mathcal{F}\mathbf{x} = \mathbf{y} & \Leftrightarrow & \mathbf{f}(\mathbf{x}) = \mathbf{y}\,, \\ \mathbf{x} = \mathcal{G}\mathbf{y} & \Leftrightarrow & \mathbf{x} = \underset{\xi\in E}{\operatorname{argmin}}\, \|\mathbf{f}(\xi) - \mathbf{y}\|_{\widehat{E}}\,, \\ \widetilde{\mathcal{F}}_k\mathbf{x} = \mathbf{y} & \Leftrightarrow & \mathbf{c}(\overline{\mathbf{p}}_k(\mathbf{x})) = \mathbf{y}\,, \\ \mathbf{x} = \widetilde{\mathcal{G}}_k\mathbf{y} & \Leftrightarrow & \mathbf{x} = \underset{\xi\in E}{\operatorname{argmin}}\, \|\mathbf{c}(\overline{\mathbf{p}}_k(\xi)) - \mathbf{y}\|_{\widehat{E}}\,. \end{array} \tag{8}$$

Here $\overline{\mathbf{p}}_k$ is *not* the space-mapping function but an arbitrary (easy to compute) bijection, e.g., the identity if $X = Z$. Notice that, in principle, also $\mathbf{c} = \mathbf{c}_k$ might be adapted during the iteration.

With (8) we derive from (7) the defect-correction iteration scheme for optimization:

$$\mathbf{x}_0 = \operatorname{argmin}_{\mathbf{x}\in X}\, \|\mathbf{c}(\overline{\mathbf{p}}_0(\mathbf{x})) - \mathbf{y}\|\,, \tag{9}$$

$$\mathbf{x}_{k+1} = \operatorname{argmin}_{\mathbf{x}\in X}\, \|\mathbf{c}(\overline{\mathbf{p}}_{k+1}(\mathbf{x})) - \mathbf{c}(\overline{\mathbf{p}}_k(\mathbf{x}_k)) + \mathbf{f}(\mathbf{x}_k) - \mathbf{y}\|\,. \tag{10}$$

In this iteration every minimization involves the surrogate model, $\mathbf{c} \circ \overline{\mathbf{p}}_k$.

Orthogonality and the need for left-preconditioning.

For the stationary points of the above process, $\lim_{k\to\infty} \mathbf{x}_k = \overline{\mathbf{x}}$, we can derive [4]:

$$\mathbf{f}(\overline{\mathbf{x}}) - \mathbf{y} \in \mathbf{c}(Z)^{\perp}(\mathbf{z}^*)\,. \tag{11}$$

Like the space-mapping methods, the above iteration has the disadvantage that, in general, the fixed point does not coincide with the solution of the fine model minimization problem. This is due to the fact that the approximate solution $\overline{\mathbf{x}}$ satisfies (11), whereas the (local) minimum $\mathbf{x}^*$ satisfies

$$\mathbf{f}(\mathbf{x}^*) - \mathbf{y} \in \mathbf{f}(X)^{\perp}(\mathbf{x}^*)\,.$$

Hence, differences between $\overline{\mathbf{x}}$ and $\mathbf{x}^*$ will be larger for larger distances between $\mathbf{y}$ and the sets $\mathbf{f}(X)$ and $\mathbf{c}(Z)$ and for larger angles between the linear manifolds tangential at $\mathbf{c}(Z)$ and $\mathbf{f}(X)$ near the optima.

By these orthogonality relations we see that it is advantageous, both for the conditioning of the problem and for the minimization of the residual, if the manifolds $\mathbf{f}(X)$ and $\mathbf{c}(Z)$ are found parallel in the neighborhood of the solution. However, by space mapping or by right-preconditioning, the relation between $\mathbf{f}(X)$ and $\mathbf{c}(Z)$ remains unchanged. This causes that the fixed point of traditional space mapping does, generally, not correspond with $\mathbf{x}^*$. This, however, can be improved by the introduction of an additional left-preconditioner. Therefore we consider such a preconditioner $\mathbf{S}$ so that near $\mathbf{c}(\mathbf{z}^*) \in Y$ the manifold $\mathbf{c}(Z) \subset Y$ is mapped onto $\mathbf{f}(X) \subset Y$:

$$\mathbf{f}(\mathbf{x}) \approx \mathbf{S}(\mathbf{c}(\overline{\mathbf{p}}(\mathbf{x}))) .$$

In the next section we propose our *manifold-mapping* algorithm, where an affine operator maps $\mathbf{c}(Z)$ onto $\mathbf{f}(X)$ in the neighborhood of the solution. More precisely: it maps $\mathbf{c}(\overline{\mathbf{p}}(\mathbf{x}_k))$ to $\mathbf{f}(\mathbf{x}_k)$ and it approximately maps one tangential linear manifold onto the other. This restores the orthogonality relation $\mathbf{f}(\overline{\mathbf{x}}) - \mathbf{y} \in \mathbf{f}(X)^{\perp}(\overline{\mathbf{x}})$. Thus it improves significantly the traditional space-mapping approach and makes the solution $\mathbf{x}^*$ a stationary point of the iteration.

5 Manifold Mapping, the improved space mapping algorithm

We introduce the affine mapping $\mathbf{S} : Y \to Y$ such that $\mathbf{S}\,\mathbf{c}(\overline{\mathbf{z}}) = \mathbf{f}(\mathbf{x}^*)$ for a proper $\overline{\mathbf{z}} \in Z$, and the linear manifold tangential to $\mathbf{c}(Z)$ in $\mathbf{c}(\overline{\mathbf{z}})$ maps onto the one tangential to $\mathbf{f}(X)$ in $\mathbf{f}(\mathbf{x}^*)$. Because both $\mathbf{f}(X)$ and $\mathbf{c}(Z)$ are n-dimensional manifolds in $\mathbb{R}^m$, the mapping $\mathbf{S}$ can be described by

$$\mathbf{S}\,\mathbf{v} = \mathbf{f}(\mathbf{x}^*) + S\ \left(\mathbf{v} - \mathbf{c}(\overline{\mathbf{z}})\right)\ ,$$

where S is an $m \times m$-matrix of rank n. A full rank $m \times m$-matrix S can be constructed, which has a well-determined part of rank n, while a remaining part of rank $m - n$ is free to choose. Because of the supposed similarity between the models $\mathbf{f}$ and $\mathbf{c}$ we keep the latter part close to the identity. The meaning of the mapping $\mathbf{S}$ is illustrated in Figure 2.

So we propose the following algorithm, where the optional right-preconditioner $\overline{\mathbf{p}} : X \to Z$ is still an arbitrary non-singular operator, which can be adapted to the problem. Often we will simply take $\overline{\mathbf{p}} = I$, the identity.

1. Set $k = 0$, set $S_0 = I$ the $m \times m$ identity matrix, and compute
$$\mathbf{x}_0 = \mathrm{argmin}_{\mathbf{x} \in X}\ \|\ \mathbf{c}(\overline{\mathbf{p}}(\mathbf{x})) - \mathbf{y}\|\ .$$
2. Compute $\mathbf{f}(\mathbf{x}_k)$ and $\mathbf{c}(\overline{\mathbf{p}}(\mathbf{x}_k))$.
3. If $k > 0$, with $\Delta\mathbf{c}_i = \mathbf{c}(\overline{\mathbf{p}}(\mathbf{x}_{k-i})) - \mathbf{c}(\overline{\mathbf{p}}(\mathbf{x}_k))$ and $\Delta\mathbf{f}_i = \mathbf{f}(\mathbf{x}_{k-i}) - \mathbf{f}(\mathbf{x}_k)$, $i = 1, \cdots, \min(n,k)$, we define ΔC and ΔF to be the rectangular $m \times \min(n,k)$-matrices with respectively $\Delta\mathbf{c}_i$ and $\Delta\mathbf{f}_i$ as columns.
The generalized singular value decomposition of these (rectangular) matrices is $\Delta C = U_c \Sigma_c V^T$ and $\Delta F = U_f \Sigma_f V^T$, with U_c, U_f orthonormal, Σ_c and Σ_f diagonal and V non-singular.

4. The next iterant is computed as

$$\mathbf{x}_{k+1} = \mathrm{argmin}_{\mathbf{x}\in X} \; \| \, \mathbf{c}(\overline{\mathbf{p}}(\mathbf{x})) - \mathbf{c}(\overline{\mathbf{p}}(\mathbf{x}_k)) + \left[U_c \Sigma_c \Sigma_f^{\dagger} U_f^T + (I - U_c U_c^T)(I - U_f U_f^T) \right] (\mathbf{f}(\mathbf{x}_k) - \mathbf{y}) \| \; . \tag{12}$$

5. Set $k := k+1$ and goto 2.

Here, $\Sigma_f^{\dagger}$ denotes the pseudo-inverse of Σ_f. It can be shown that (12) is asymtotically equivalent to

$$\mathbf{x}_{k+1} = \mathrm{argmin}_{\mathbf{x}\in X} \; \| \, \mathbf{S}_k(\mathbf{c}(\overline{\mathbf{p}}(\mathbf{x}))) - \mathbf{y} \| \; , \tag{13}$$

where the approximate affine mapping is

$$\mathbf{S}_k \, \mathbf{v} = \mathbf{f}(\mathbf{x}_k) + S_k(\mathbf{v} - \mathbf{c}(\overline{\mathbf{p}}(\mathbf{x}_k))) \, ,$$

with $S_k = U_f \Sigma_f \Sigma_c^{\dagger} U_c^T + (I - U_f \, U_f^T)\,(I - U_c \, U_c^T)$.

If the above iteration converges with fixed point $\overline{\mathbf{x}}$ and mapping $\overline{\mathbf{S}}$, we have

$$\mathbf{f}(\overline{\mathbf{x}}) - \mathbf{y} \in \overline{\mathbf{S}}(\mathbf{c}(\overline{\mathbf{p}}(X)))^{\perp}(\overline{\mathbf{x}}) = \mathbf{f}(X)^{\perp}(\overline{\mathbf{x}}) \; .$$

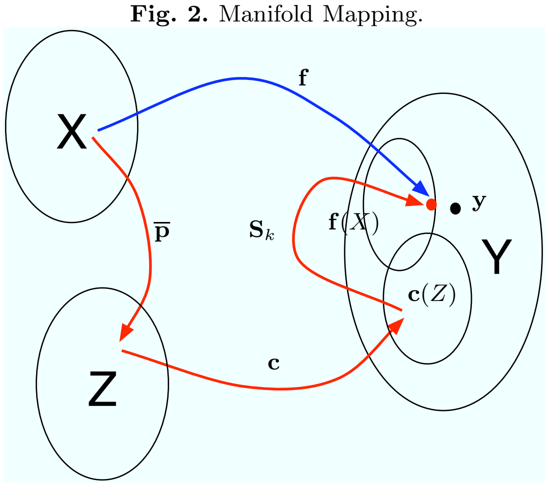
This, and the fact that $\mathbf{S}_k(\mathbf{c}(\overline{\mathbf{p}}(\mathbf{x}_k))) = \mathbf{f}(\mathbf{x}_k)$, makes that, under convergence to $\overline{\mathbf{x}}$, the fixed point is a (local) optimum of the fine model minimization.

The improved space-mapping scheme

$$\mathbf{x}_{k+1} = \underset{\mathbf{x}}{\mathrm{argmin}} \; \| \, \mathbf{S}_k(\mathbf{c}(\overline{\mathbf{p}}_k(\mathbf{x})))) - \mathbf{y} \| \; ,$$

can also be recognized as defect correction iteration with either $\widetilde{\mathcal{F}}_k = \mathbf{S}_k \circ \mathbf{c} \circ \overline{\mathbf{p}}_k$ and $\mathcal{F} = \mathbf{f}$ or with $\widetilde{\mathcal{F}}_k = \mathbf{S}_k \circ \mathbf{c}$ and $\mathcal{F} = \mathbf{f} \circ \overline{\mathbf{p}}_k^{-1}$.

Fig. 2. Manifold Mapping.



The surrogate model: $\mathbf{S}_k \circ \mathbf{c} \circ \overline{\mathbf{p}} \approx \mathbf{f}$.

6 Conclusion

By right preconditioning, manifold mapping improves traditional space mapping because it delivers the accurate optimum with the same computational efficiency as the space mapping algorithm.

References

1. M.H. Bakr, J.W. Bandler, K. Madsen, and J. Søndergaard. Review of the space mapping approach to engineering optimization and modeling. *Optimization and Engineering*, 1(3):241–276, 2000.
2. J.W. Bandler, Q.S. Cheng, A.S. Dakroury, A.S. Mohamed, M.H. Bakr, K. Madsen, and J. Søndergaard. Space mapping: The state of the art. *IEEE Transactions on Microwave Theory and Techniques*, 52:337–360, 2004.
3. K. Böhmer, P. W. Hemker, and H. J. Stetter. The defect correction approach. In K. Böhmer and H. J. Stetter, editors, *Defect Correction Methods: Theory and Applications*, Computing Suppl. 5, pages 1–32. Springer-Verlag, Berlin, Heidelberg, New York, Tokyo, 1984.
4. D. Echeverría and P.W. Hemker. Space mapping and defect correction. *Comp. Methods in Appl. Math.*, 5(2):107–136, 2005.