Centrum voor Wiskunde en Informatica

Modelling, Analysis and Simulation

On the manifold-mapping optimization technique

D. Echeverría, P.W. Hemker

On the manifold-mapping optimization technique

ABSTRACT

In this paper, we study in some detail the manifold-mapping optimization technique introduced in an earlier paper. Manifold mapping aims at accelerating optimal design procedures that otherwise require many evaluations of time-expensive cost functions. We give a proof of convergence for the manifold-mapping iteration. By means of two simple optimization problems we illustrate the convergence results derived.Finally, the performances of several variants of the method are compared for some practical design problems in electromagnetics.

On the Manifold-Mapping Optimization Technique

D. Echeverría and P.W. Hemker

CWI, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands

ABSTRACT

In this paper, we study in some detail the manifold-mapping optimization technique introduced in [1]. Manifold mapping aims at accelerating optimal design procedures that otherwise require many evaluations of time-expensive cost functions. We give a proof of convergence for the manifold-mapping iteration. By means of two simple optimization problems we illustrate the convergence results derived. Finally, the performances of several variants of the method are compared for some practical design problems in electromagnetics.

2000 Mathematics Subject Classification: 90C30, 65K05, 49M99, 49Q99

Keywords and Phrases: nonlinear optimization, surrogate optimization, simulation-based optimization, two-level optimization, space mapping, convergence analysis

1. INTRODUCTION

Optimization problems in practice often need function evaluations that are very expensive to compute and, moreover, gradient information is in most of these situations *unavailable*, e.g., optimal design problems based on complex finite element simulations. As a consequence, many optimization processes may require very long computing times.

Surrogates have been used since long in analysis and design in engineering [2, 3, 4] and recently they have been successfully applied to the solving of very time-consuming optimization problems [5, 6, 7, 8]. In surrogate optimization the quality of the initial approximation is iteratively improved. This first surrogate yields the first iterant and with the use of it, the surrogate is improved. This iterative procedure is repeated until some stopping criteria is met.

The nature of the surrogates is problem dependent. We can clearly distinguish two different types of surrogates. If no information is a priori available, approximations can be obtained from scratch, in most cases combining experiment design strategies (e.g., latin hypercube sampling [9] or orthogonal arrays [10]) with interpolation/approximation techniques (e.g., low degree polynomials [11], kriging [12] or radial basis functions [13]). The second type of approximations is found in situations where, because of e.g., experience or simple rules

of thumb, the derivation of the surrogate is simplified. An example of this is the the use of lumped parameter models (e.g., magnetic [1], electric [14] or thermal [15] equivalent circuits).

Space mapping [14, 16] is a well-known surrogate-based optimization technique. Though it can be used in combination with any type of surrogate, it is traditionally applied with approximations of the second type described above. The space-mapping technique has been reported as an efficient minimization procedure in a large number of cases [14].

There are many variants of the original space-mapping algorithm: aggressive space mapping (ASM) [17], trust-region aggressive space mapping (TRASM) [18], neural space mapping (NSM) [19] and implicit space mapping (ISM) [20] are the most significant examples. Although all these schemes do not always converge to the right solution [1], the one obtained is generally acceptable for practical purposes. Recently, in [21] the original space-mapping approach is modified according to the framework proposed in [7]. At the expense of incorporating *exact* gradient information, the new scheme yields convergence to the accurate optimum.

Defect-correction theory [22] helps in understanding the space-mapping concept. Defect correction is the underlying basis of a great number of mathematical techniques that essentially solve a complex problem by the iterative use of a simpler one (e.g., Newton's method, relaxation procedures [23, 24], iterative refinement [23] or multigrid methods [24]). Though defect-correction theory was originally developed for nonlinear systems, it can also be applied in optimization problems. Many space-mapping procedures can be seen as special cases of the defect-correction iteration [1] and this fact can be used in analysis for explaining when the space-mapping approach may fail to find the right solution.

Manifold mapping [1] is an alternative surrogate-based optimization technique. Manifold mapping can be used without computing *exact* gradient information and it has provable convergence to the right solution. This paper presents a thorough analysis of the manifold-mapping approach and it is structured as follows. In Section 2 the basic terminology and some general assumptions are introduced. The manifold-mapping approach, together with the algorithms derived, are described in Section 3. It will be seen also in this section, that under mild assumptions, the fixed point for all the algorithms/procedures is the accurate optimum. Convergence theorems for all the schemes will be stated in Section 4 and they will be illustrated by means of two simple design problems in Section 5. Also in this section, the manifold-mapping technique will be eventually compared with other efficient methods, in the case of optimal design problems from the field of electromagnetics.

2. The Problem

The optimization problem   Let the specification of the aim (the data) in an optimization problem be denoted by $\mathbf{y} \in Y \subset \mathbb{R}^m$. Since the true mechanism how this aim can be approximated can be extremely complex, or even impossible to describe in all its details, we study it by mathematical models. Often, such models appear in several degrees of sophistication. It is the purpose of manifold mapping to exploit the simpler models by combining their efficiency in computation with the accuracy of the more complex ones. Thus, we distinguish two types of model: fine and coarse.

The fine model   The fine model response function is denoted by $\mathbf{f} : X \subset \mathbb{R}^n \to \mathbb{R}^m$, and $\mathbf{x} \in X$ is the control variable. The set X of possible control variables is usually a closed and

bounded subset of $\mathbb{R}^n$. The fine model is assumed to be *accurate* but *expensive* to evaluate. We assume that $\mathbf{f}(\mathbf{x})$ is differentiable but its Jacobian matrix $J_{\mathbf{f}}(\mathbf{x}) = \mathbf{df}/\mathbf{dx}$ is generally supposed to be unavailable.

For the optimization problem a *fine model cost function*, $F(\mathbf{x}) = \|\mathbf{f}(\mathbf{x}) - \mathbf{y}\|$ is defined, which is a measure for the discrepancy between the aim and a particular response of the mathematical model. This cost function should be minimized. So we look for

$$\mathbf{x}_{\mathbf{f}}^* = \operatorname{argmin}_{\mathbf{x} \in X} \|\mathbf{f}(\mathbf{x}) - \mathbf{y}\| . \tag{2.1}$$

For simplicity, in this paper we take for $\|\cdot\|$ the Euclidean norm on $\mathbb{R}^m$. Not every optimization problem is of this type, but most practical design situations can finally be reduced to this model-specification structure.

A design is called *reachable* if there exists an $\mathbf{x}_{\mathbf{f}}^* \in X$ such that $\mathbf{f}(\mathbf{x}_{\mathbf{f}}^*) = \mathbf{y}$. This situation can often be expected when $n \geq m$ since in that situation the number of degrees of freedom in the design is larger than or equal to the number of specifications. Reachable designs can be formulated as equations and, hence, they can generally be solved as nonlinear systems. The original defect-correction iteration [22] can be directly applied. For this reason, in this work we will particularly analyze the overdetermined case $n < m$. We formulate this as our first assumption.

Assumption 1 *The dimension of the space of possible aims exceeds the dimension of the control space: $n < m$.*

The coarse model   The coarse model is denoted by $\mathbf{c} : Z \subset \mathbb{R}^n \to \mathbb{R}^m$ with $\mathbf{z} \in Z$ the coarse model control variable. In this work, for simplicity, we consider $X = Z$, but the general case can be dealt with by the introduction of an additional mapping $\bar{\mathbf{p}} : X \to Z$, as used in [1]. In contrast to the fine model, the coarse model is assumed to be *cheap* to evaluate but *less accurate*. For the coarse model we have the *coarse model cost function*, $C(\mathbf{z}) = \|\mathbf{c}(\mathbf{z}) - \mathbf{y}\|$. We denote its minimizer by $\mathbf{x}_{\mathbf{c}}^* \in Z$,

$$\mathbf{x}_{\mathbf{c}}^* = \operatorname{argmin}_{\mathbf{z} \in Z} \|\mathbf{c}(\mathbf{z}) - \mathbf{y}\| . \tag{2.2}$$

Again, for this cost function we consider the Euclidean norm. The Jacobian matrix $J_{\mathbf{c}}(\mathbf{z}) = \mathbf{dc}/\mathbf{dz}$ is assumed to be available with no significant computational cost.

Manifolds and constraints   If the functions $\mathbf{f}(\mathbf{x})$ and $\mathbf{c}(\mathbf{x})$ are sufficiently smooth, the sets $\mathbf{f}(X) \subset \mathbb{R}^m$ and $\mathbf{c}(X) \subset \mathbb{R}^m$ can be considered as differential manifolds. In the appendix we summarize a number of manifold-related definitions (from [25]) relevant for our discussion. For simplicity and without loss of generality, we just consider one particular chart of the manifold that covers a sufficiently large neighborhood of the solution region. We thus make the following assumption.

Assumption 2 *The sets $\mathbf{f}(X)$ and $\mathbf{c}(X)$ are differentiable manifolds of class C^2.*

We need the concept of the tangent plane for a manifold at a point, which is used throughout this work.

Definition 1 *Let M be a differentiable manifold and $\mathbf{v} \in M$ a point in that manifold. The tangent plane for M at $\mathbf{v} \in M$ is defined as the affine space spanned by the Jacobian of φ at $\mathbf{x}$, being φ any chart in the atlas of the manifold such that $\varphi(\mathbf{x}) = \mathbf{v}$.*

Since we only consider one chart we will denote the tangent plane for $\mathbf{f}(X)$ at $\mathbf{f}(\mathbf{x})$ simply by the Jacobian of $\mathbf{f}$ at $\mathbf{x}$, i.e., $J_{\mathbf{f}}(\mathbf{x})$. Thus the tangent plane is well defined. Similarly, $J_{\mathbf{c}}(\mathbf{x})$ will denote the tangent plane for $\mathbf{c}(X)$ at $\mathbf{c}(\mathbf{x})$.

We can state a general *constrained* optimization problem as follows

$$
\begin{aligned}
\mathbf{x}_{\mathbf{f}}^* &= \operatorname{argmin}_{\mathbf{x} \in X} \|\mathbf{f}(\mathbf{x}) - \mathbf{y}\|, & (2.3) \\
X &= \left\{ \mathbf{x} \in \mathbb{R}^n; \ \mathbf{k_f}(\mathbf{x}) = 0, \quad \bar{\mathbf{k}}_{\mathbf{f}}(\mathbf{x}) \geq 0 \right\},
\end{aligned}
$$

where $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$, $\mathbf{k_f} : \mathbb{R}^n \to \mathbb{R}^{n_{\mathbf{k}}}$ and $\bar{\mathbf{k}}_{\mathbf{f}} : \mathbb{R}^n \to \mathbb{R}^{\bar{n}_{\mathbf{k}}}$ are assumed to be differentiable, and $n_{\mathbf{k}}$ and $\bar{n}_{\mathbf{k}}$ are the number of equality and inequality constraints respectively. In general we cannot expect the model and constraints to be defined over the entire $\mathbb{R}^n$. For example, it makes no sense in many cases to consider negative lengths. But it is common in practice that $\mathbf{f}$ and $\mathbf{k_f}$ are correctly defined in $\widehat{X}$, the set where the inequality constraints are feasible

$$
\widehat{X} = \{\mathbf{x} \in \mathbb{R}^n; \ \bar{\mathbf{k}}_{\mathbf{f}}(\mathbf{x}) \geq 0\}. \tag{2.4}
$$

It should be noted that box-constraints can be obtained with a proper choice for $\bar{\mathbf{k}}_{\mathbf{f}}(\mathbf{x})$. The inequality constraints do not generally reduce the dimensionality of the design space and therefore we prefer to rewrite (2.3) as

$$
\mathbf{x}_{\mathbf{f}}^* = \operatorname{argmin}_{\{\mathbf{x} \in \widehat{X}; \ \mathbf{k_f}(\mathbf{x}) = 0\}} \|\mathbf{f}(\mathbf{x}) - \mathbf{y}\|. \tag{2.5}
$$

In order to have degrees of freedom left for optimization, the number of equality constraints $n_{\mathbf{k}}$ should be smaller than the number of design variables n, i.e., $n_{\mathbf{k}} < n$. We also assume that the set X is a differentiable manifold in $\mathbb{R}^n$. We formalize this in the following assumption.

Assumption 3 *The set X is either a subset of $\mathbb{R}^n$ or a differentiable manifold in $\mathbb{R}^n$ of dimension $n - n_{\mathbf{k}} > 0$.*

Now we can write the constrained optimization problem (2.5) as (2.1), with X a differentiable manifold.

In this work we only analyze those cases in which the functions $\bar{\mathbf{k}}_{\mathbf{f}}(\mathbf{x})$ and $\mathbf{k_f}(\mathbf{x})$ are easy to compute (and thus it makes sense to take $Z = X$). Below in Remark 3, we indicate how more complex constraints can be handled.

If the equality constraints $\mathbf{k_f}(\mathbf{x})$ can be evaluated easily and a chart φ_X for $\bar{X} = \{\mathbf{x} \in \mathbb{R}^n; \ \mathbf{k_f}(\mathbf{x}) = 0\}$ in the region of interest can be obtained with not significant computational cost, then the constrained optimization problem can be restated with $\mathbf{f} \circ \varphi_X$ as fine model and a subset of $\mathbb{R}^{n - n_{\mathbf{k}}}$ as the control space.

By the above argument we assume that the equality constraints can be eliminated and, hence, X can be considered as a subset of $\mathbb{R}^n$, understanding n as the number of design variables left after removing the equality constraints. As a consequence, both Jacobians $J_{\mathbf{f}}(\mathbf{x})$ and $J_{\mathbf{c}}(\mathbf{x})$ have both rank n.

If $\mathbf{x_f^*}$ is a local optimum of the constrained problem (2.3), then it satisfies the *Karush-Kuhn-Tucker* (KKT) conditions [26], i.e., there are two Lagrange multiplier vectors $\lambda_\mathbf{f}^* \in \mathbb{R}^{n_\mathbf{k}}$ and $\bar{\lambda}_\mathbf{f}^* \in \mathbb{R}^{\bar{n}_\mathbf{k}}$ such that

$$\mathbf{d/dx}\left(F - \lambda_\mathbf{f}^{*T}\,\mathbf{k_f} - \bar{\lambda}_\mathbf{f}^{*T}\,\bar{\mathbf{k}}_\mathbf{f}\right)(\mathbf{x_f^*}) = \mathbf{0} \tag{2.6}$$

$$\mathbf{k_f}(\mathbf{x_f^*}) = \mathbf{0} \tag{2.7}$$

$$\bar{\mathbf{k}}_\mathbf{f}(\mathbf{x_f^*}) \geq \mathbf{0} \tag{2.8}$$

$$\lambda_\mathbf{f}^*, \bar{\lambda}_\mathbf{f}^* \geq \mathbf{0} \tag{2.9}$$

$$\mathrm{diag}\left(\bar{\lambda}_\mathbf{f}^* \cdot \bar{\mathbf{k}}_\mathbf{f}(\mathbf{x_f^*})\right) = \mathbf{0}\,, \tag{2.10}$$

where $\cdot$ denotes here the vector direct product.

Remark 1 *In the case of an unconstrained optimization based on the Euclidean norm, the KKT conditions represent the orthogonality between the tangent plane for $\mathbf{f}(X)$ at $\mathbf{x_f^*}$ and the optimal model-specifications discrepancy $\mathbf{f}(\mathbf{x_f^*}) - \mathbf{y}$, i.e.,*

$$J_\mathbf{f}^T(\mathbf{x_f^*})\left(\mathbf{f}(\mathbf{x_f^*}) - \mathbf{y}\right) = 0\,. \tag{2.11}$$

The following concept will be very useful when formalizing the similarity between models.

Definition 2 *We say that an optimization problem is locally convex at $\mathbf{x} \in \mathbb{R}^n$ if and only if there is a neighborhood U of $\mathbf{x}$ such that every point in U satisfying the KKT conditions is a local minimum.*

The assumption for a well defined optimization problem   In order to be sure that the problems we want to solve make sense, we have to assume that the there exist solutions for these designs.

Assumption 4 *The fine and coarse optimization problems, characterized by $\mathbf{y}$, $\mathbf{f}$ and $\mathbf{c}$ and X, are uniquely solvable, i.e.,*

$$\forall \mathbf{y} \in Y \;\; \exists!\; \mathbf{x_f^*} \in X \qquad \mathbf{x_f^*}(\mathbf{y}) = \mathrm{argmin}_{\mathbf{x} \in X} \|\mathbf{f}(\mathbf{x}) - \mathbf{y}\|\,, \tag{2.12}$$

and

$$\forall \mathbf{y} \in Y \;\; \exists!\; \mathbf{x_c^*} \in X \qquad \mathbf{x_c^*}(\mathbf{y}) = \mathrm{argmin}_{\mathbf{x} \in X} \|\mathbf{c}(\mathbf{x}) - \mathbf{y}\|\,. \tag{2.13}$$

In most practical cases this assumption is reasonable. If X is a closed and bounded non-empty set and $\mathbf{f}$ and $\mathbf{c}$ are continuous functions, the existence of the solutions is guaranteed. Generally, uniqueness can be achieved by properly reducing the set X.

The assumption of well-posedness is particularly important for the coarse model and will be discussed later (Assumption 12).

3. MANIFOLD MAPPING
3.1 Original manifold mapping
In [1] the *manifold mapping* $\mathbf{S} : \mathbf{c}(X) \to \mathbf{f}(X)$ is introduced with the aim of correcting the misalignment between the manifolds $\mathbf{f}(X)$ and $\mathbf{c}(X)$. With $\mathbf{S}$, the point $\mathbf{c}(\mathbf{x}_f^*)$ is mapped to
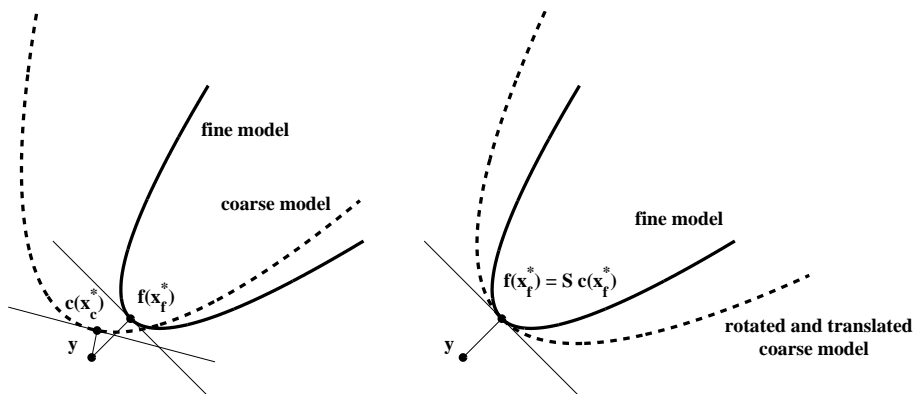
Figure 1: Manifold-mapping model alignment.

$\mathbf{f}(\mathbf{x}_f^*)$ and the tangent plane for $\mathbf{c}(X)$ at $\mathbf{c}(\mathbf{x}_f^*)$ to the tangent plane for $\mathbf{f}(X)$ at $\mathbf{f}(\mathbf{x}_f^*)$ (see Figure 1). Other approaches are possible but in this work we define $\mathbf{S}$ as the affine mapping

$$\mathbf{S}\,\mathbf{c}(\mathbf{x}) = \mathbf{f}(\mathbf{x}_f^*) + \bar{S}\left(\mathbf{c}(\mathbf{x}) - \mathbf{c}(\mathbf{x}_f^*)\right), \tag{3.1}$$

where

$$\bar{S} = J_{\mathbf{f}}(\mathbf{x}_{\mathbf{f}}^*)\,J_{\mathbf{c}}^{\dagger}(\mathbf{x}_{\mathbf{f}}^*). \tag{3.2}$$

Here the pseudoinverse † is defined as $J_{\mathbf{c}}^{\dagger}(\mathbf{x}_{\mathbf{f}}^*) = V_{\mathbf{c}}\,\Sigma_{\mathbf{c}}^{\dagger}\,U_{\mathbf{c}}^T$, being $U_{\mathbf{c}}$, $\Sigma_{\mathbf{c}}$ and $V_{\mathbf{c}}$ the factors in the singular value decomposition of $J_{\mathbf{c}}(\mathbf{x}_f^*) = U_{\mathbf{c}}\,\Sigma_{\mathbf{c}}\,V_{\mathbf{c}}^T$. The matrix $\Sigma_{\mathbf{c}}^{\dagger}$ is the result of inverting the nonzero entries in $\Sigma_{\mathbf{c}}$, leaving the zeroes invariant. It should be noted that, because $J_{\mathbf{f}}(\mathbf{x}_{\mathbf{f}}^*)$ and $J_{\mathbf{c}}(\mathbf{x}_{\mathbf{f}}^*)$ are both full rank, also the $m \times m$ matrix $\bar{S}$ has rank n.

The manifold-mapping solution $\mathbf{x}_{\mathbf{mm}}^*$ is defined as

$$\mathbf{x}_{\mathbf{mm}}^* = \mathrm{argmin}_{\mathbf{x} \in X}\|\mathbf{S}\,\mathbf{c}(\mathbf{x}) - \mathbf{y}\|. \tag{3.3}$$

Now the combination $\mathbf{S} \circ \mathbf{c}$ acts as the *surrogate model* for $\mathbf{f}$ and under some assumptions about the similarity between the fine and the coarse model that usually hold in practice, it will be shown below that the manifold-mapping solution $\mathbf{x}_{\mathbf{mm}}^*$ is a local minimizer of the fine cost function. As one may expect, not every coarse model can be successfully used within the manifold-mapping framework. We partly formalize this by two assumptions:

Assumption 5 *If $\|\mathbf{f}(\mathbf{x}) - \mathbf{y}\|$ is locally convex at $\mathbf{x}_{\mathbf{f}}^*$ then $\|\mathbf{S}\,\mathbf{c}(\mathbf{x}) - \mathbf{y}\|$ is also locally convex at $\mathbf{x}_{\mathbf{f}}^*$.*

Assumption 6 *If $\mathbf{x}_{\mathbf{f}}^*$ is a local optimum of $\|\mathbf{S}\,\mathbf{c}(\mathbf{x}) - \mathbf{y}\|$ then $\mathbf{x}_{\mathbf{f}}^*$ is the global optimum of $\|\mathbf{S}\,\mathbf{c}(\mathbf{x}) - \mathbf{y}\|$.*

These are mild assumptions for the models used in practice. Assumption 5 specifies only a similar *local* behavior in the region of interest, i.e., in a neighborhood of the specifications $\mathbf{y}$. Assumption 6 means that the surrogate optimization does not allow a spurious global optimum near the *true* minimum $\mathbf{x}_{\mathbf{f}}^*$.

Lemma 1 *Any (local) minimizer of the fine model cost function $\|\mathbf{f}(\mathbf{x}) - \mathbf{y}\|$ is a (local) minimizer of $\|\mathbf{S}\,\mathbf{c}(\mathbf{x}) - \mathbf{y}\|$.*

Proof We denote a minimizer of the fine cost function by $\mathbf{x}_\mathbf{f}^*$. First we see that $\mathbf{x}_\mathbf{f}^*$ satisfies the KKT conditions associated with (3.3). From (3.1) and (3.2) we have $\mathbf{S}\,\mathbf{c}(\mathbf{x}_\mathbf{f}^*) = \mathbf{f}(\mathbf{x}_\mathbf{f}^*)$ and $J_{\mathbf{S}\,\mathbf{c}}(\mathbf{x}_\mathbf{f}^*) = \bar{S}\,J_\mathbf{c}(\mathbf{x}_\mathbf{f}^*) = J_\mathbf{f}(\mathbf{x}_\mathbf{f}^*)$. Thus, the first derivatives of $F(\mathbf{x})$ and of the surrogate cost function $\|\mathbf{S}\,\mathbf{c}(\mathbf{x}) - \mathbf{y}\|$ coincide at $\mathbf{x}_\mathbf{f}^*$. Since the constraints are the same in both optimization problems and $\mathbf{x}_\mathbf{f}^*$ is a local optimum of $F(\mathbf{x})$ (i.e., the fine KKT conditions hold), we conclude that $\mathbf{x}_\mathbf{f}^*$ satisfies the surrogate KKT conditions. The fine model cost function is locally convex at $\mathbf{x}_\mathbf{f}^*$. Because of Assumption 5, the fine model optimum is also a local minimum of $\|\mathbf{S}\,\mathbf{c}(\mathbf{x}) - \mathbf{y}\|$. □

Lemma 2 $\mathbf{x}_\mathbf{mm}^*$ *is a local minimizer of* $\|\mathbf{f}(\mathbf{x}) - \mathbf{y}\|$.

Proof Use Lemma 1 and Assumption 6. □

Remark 2 *We cannot directly conclude from Lemma 1 that $\mathbf{x}_\mathbf{mm}^*$ is a minimizer of the fine cost function because the point $\mathbf{x}_\mathbf{f}^*$ in the lemma, could be just a local minimizer of $\|\mathbf{S}\,\mathbf{c}(\mathbf{x}) - \mathbf{y}\|$. This agrees with the way the two-level approach is taken into practice, as a local manifold correction.*

Remark 3 *Based on the proof for Lemma 1 we can think of an strategy for dealing with the expensive constraints $\mathbf{k_f}(\mathbf{x})$ and $\bar{\mathbf{k}}_\mathbf{f}(\mathbf{x})$, provided some fast-to-compute approximations $\mathbf{k_c}(\mathbf{x})$ and $\bar{\mathbf{k}}_\mathbf{c}(\mathbf{x})$ are available. Analog manifold mappings for the constraints $\mathbf{S_k} : \mathbf{k_c}(X) \to \mathbf{k_f}(X)$ and $\mathbf{S_{\bar{k}}} : \bar{\mathbf{k}}_\mathbf{c}(X) \to \bar{\mathbf{k}}_\mathbf{f}(X)$ can be defined so that*

$$\mathbf{S_k}\,\mathbf{k_c}(\mathbf{x}_\mathbf{f}^*) = \mathbf{k_f}(\mathbf{x}_\mathbf{f}^*) \tag{3.4}$$

$$\mathbf{S_{\bar{k}}}\,\bar{\mathbf{k}}_\mathbf{c}(\mathbf{x}_\mathbf{f}^*) = \bar{\mathbf{k}}_\mathbf{f}(\mathbf{x}_\mathbf{f}^*) \tag{3.5}$$

$$J_{\mathbf{S_k}\,\mathbf{k_c}}(\mathbf{x}_\mathbf{f}^*) = J_{\mathbf{k_f}}(\mathbf{x}_\mathbf{f}^*) \tag{3.6}$$

$$J_{\mathbf{S_{\bar{k}}}\,\bar{\mathbf{k}}_\mathbf{c}}(\mathbf{x}_\mathbf{f}^*) = J_{\bar{\mathbf{k}}_\mathbf{f}}(\mathbf{x}_\mathbf{f}^*) \tag{3.7}$$

and thus again, the surrogate KKT conditions at $\mathbf{x}_\mathbf{f}^*$ reproduce the fine ones at the same point. This constraint manifold mapping is the approach taken in [27].

The mapping $\mathbf{S}$ is not known a priori, because it depends on the solution of the optimization problem. We propose an algorithm (see Figure 2) that, when it converges, yields —as we shall see— both the mapping $\mathbf{S}$ and the desired fine model optimum $\mathbf{x}_f^*$. We will refer to this scheme as the *original manifold-mapping* (OMM) algorithm. The model alignment can be improved by an additional (right-preconditioning) mapping $\bar{\mathbf{p}} : X \to Z$. This mapping $\bar{\mathbf{p}}$ is optional in the present case where the coarse and fine control spaces Z and X coincide, but it is obligatory when those spaces differ. For simplicity, in the algorithm in Figure 2 we take $\bar{\mathbf{p}} = I$, the identity. The optimization procedure needed to compute $\mathbf{x}_{k+1}$ is not essentially different from the one to obtain the coarse model optimum $\mathbf{x_c}^*$. Therefore, we may expect that optimization problem to be well defined in each iteration step. However, formally this has to be introduced as an assumption.

$$
\boxed{
\begin{aligned}
&\mathbf{x}_0 = \mathbf{x}_{\mathbf{c}}^* = \operatorname{argmin}_{\mathbf{x} \in X} \| \mathbf{c}(\mathbf{x}) - \mathbf{y} \| \, ; \\
&\mathbf{S}_0 \bullet = \mathbf{f}(\mathbf{x}_0) + (\bullet - \mathbf{c}(\mathbf{x}_0)) \, ; \\
&\mathbf{for} \quad k = 0, 1, \ldots, \mathbf{while} \ldots \\
&\mathbf{do} \\
&\qquad \mathbf{x}_{k+1} = \operatorname{argmin}_{\mathbf{x} \in X} \| \mathbf{S}_k(\mathbf{c}(\mathbf{x})) - \mathbf{y} \| \, ; \\
&\qquad \mathbf{break\ if} \ldots \\
&\qquad \Delta F = [\mathbf{f}(\mathbf{x}_{k+1}) - \mathbf{f}(\mathbf{x}_k), \cdots, \mathbf{f}(\mathbf{x}_{k+1}) - \mathbf{f}(\mathbf{x}_{\max(k+1-n,0)})] \, ; \\
&\qquad \Delta C = [\mathbf{c}(\mathbf{x}_{k+1}) - \mathbf{c}(\mathbf{x}_k), \cdots, \mathbf{c}(\mathbf{x}_{k+1}) - \mathbf{c}(\mathbf{x}_{\max(k+1-n,0)})] \, ; \\
&\qquad \Delta C = U_{\mathbf{c}} \, \Sigma_{\mathbf{c}} \, V_{\mathbf{c}}^T \ \text{singular value decomposition} \, ; \\
&\qquad \Delta C^{\dagger} = V_{\mathbf{c}} \, \Sigma_{\mathbf{c}}^{\dagger} \, U_{\mathbf{c}}^T \, ; \\
&\qquad S_{k+1} = \Delta F \, \Delta C^{\dagger} \, ; \\
&\qquad \mathbf{S}_{k+1} \bullet = \mathbf{f}(\mathbf{x}_{k+1}) + S_{k+1} \, (\bullet - \mathbf{c}(\mathbf{x}_{k+1})) \, ; \\
&\mathbf{enddo}
\end{aligned}
}
$$

Figure 2: The original manifold-mapping (OMM) algorithm.

Assumption 7 *The minimization*

$$
\mathbf{x}_{k+1} = \operatorname{argmin}_{\mathbf{x} \in X} \| \mathbf{S}_k(\mathbf{c}(\mathbf{x})) - \mathbf{y} \| \tag{3.8}
$$

is well defined for every k.

This assumption is the surrogate equivalent of Assumption 4.

For the proof that, if it converges, the OMM algorithm yields the fine model optimum, the following lemma will be very useful.

Lemma 3 *Let $\widetilde{\mathbf{x}} \in X$ be the minimizer of a surrogate model problem*

$$
\widetilde{\mathbf{x}} = \operatorname{argmin}_{\mathbf{x} \in X} \| \widetilde{\mathbf{S}} \, \mathbf{c}(\mathbf{x}) - \mathbf{y} \| \, , \tag{3.9}
$$

with

$$
\widetilde{\mathbf{S}} \, \mathbf{c}(\mathbf{x}) = \mathbf{f}(\widetilde{\mathbf{x}}) + J_{\mathbf{f}}(\widetilde{\mathbf{x}}) \, J_{\mathbf{c}}^{\dagger}(\widetilde{\mathbf{x}}) \, (\mathbf{c}(\mathbf{x}) - \mathbf{c}(\widetilde{\mathbf{x}})) \, , \tag{3.10}
$$

where $\| \mathbf{f}(\mathbf{x}) - \mathbf{y} \|$ is locally convex at $\widetilde{\mathbf{x}}$, then $\widetilde{\mathbf{x}}$ is a (local) minimizer of $\| \mathbf{f}(\mathbf{x}) - \mathbf{y} \|$.

Proof Clearly, $\widetilde{\mathbf{x}}$ satisfies the KKT conditions associated with $\| \widetilde{\mathbf{S}} \, \mathbf{c}(\mathbf{x}) - \mathbf{y} \|$, and because of (3.10) we have $\widetilde{\mathbf{S}} \, \mathbf{c}(\widetilde{\mathbf{x}}) = \mathbf{f}(\widetilde{\mathbf{x}})$ and $J_{\widetilde{\mathbf{S}} \, \mathbf{c}}(\widetilde{\mathbf{x}}) = J_{\mathbf{f}}(\widetilde{\mathbf{x}})$. Proceeding as in Lemma 1 we see that the point $\widetilde{\mathbf{x}}$ satisfies also the fine KKT conditions and, because $\| \mathbf{f}(\mathbf{x}) - \mathbf{y} \|$ is locally convex at $\widetilde{\mathbf{x}}$, this point is a (local) minimizer of the fine cost function $F(\mathbf{x})$. $\square$

Remark 4 *Note that from Lemma 3 it also follows that $\widetilde{\mathbf{S}} = \mathbf{S}$.*

We can replace the requirement of $\| \mathbf{f}(\mathbf{x}) - \mathbf{y} \|$ being locally convex at $\widetilde{\mathbf{x}}$ from Lemma 3 by an assumption, very similar in nature to Assumption 5, and also likely to hold in practice:

Assumption 8 *If $\| \widetilde{\mathbf{S}} \, \mathbf{c}(\mathbf{x}) - \mathbf{y} \|$ is locally convex at $\widetilde{\mathbf{x}}$, then $\| \mathbf{f}(\mathbf{x}) - \mathbf{y} \|$ is locally convex at $\widetilde{\mathbf{x}}$.*

Remark 5 *The manifold-mapping theory is generally stated in terms of local alignment between the surrogate model and the fine model. As a consequence, we can only state results concerning local optima of the fine cost function.*

Now we will show that, if the OMM algorithm in Figure 2 converges to a fixed point $\bar{\mathbf{x}}$, this fixed point is a (local) minimizer of the fine cost function. Studying the fixed point situation, we may assume $k > n$. The iterants of the OMM algorithm are denoted by $\mathbf{x}_k$.

Further, some more mild additional assumptions are needed for proving that $\bar{\mathbf{x}}$ (locally) minimizes $\|\mathbf{f}(\mathbf{x}) - \mathbf{y}\|$. Since the Jacobians $J_{\mathbf{f}}(\mathbf{x})$ and $J_{\mathbf{c}}(\mathbf{x})$ have both rank n, we expect that the matrices ΔF and ΔC in the OMM algorithm are also full-rank. In practice, this will generally be the case and for the exceptional situation where it is not, minor changes in the algorithm with not real influence in the results can be made. So, to prevent minor details in the discussion making the analysis much more complex, we introduce the following assumption.

Assumption 9 *For k large enough, the $m \times n$ matrices ΔF and ΔC have rank n and there are constants K_1, $K_2 > 0$ independent of k such that*

$$\left(\max_{i=0,\,\ldots,\,n-1} \|\mathbf{x}_{k+1-i} - \bar{\mathbf{x}}\|^2 \right) \|\Delta F^{\dagger}\|_2^2 \ \leq \ K_1 \tag{3.11}$$

$$\left(\max_{i=0,\,\ldots,\,n-1} \|\mathbf{x}_{k+1-i} - \bar{\mathbf{x}}\|^2 \right) \|\Delta C^{\dagger}\|_2^2 \ \leq \ K_2 \,. \tag{3.12}$$

We will see in Lemma 4 that Assumption 9 together with the next one guarantee, that $\Delta F \, \Delta C^{\dagger}$ converges to $J_{\mathbf{f}}(\bar{\mathbf{x}}) \, J_{\mathbf{c}}^{\dagger}(\bar{\mathbf{x}})$ and thus that Lemma 3 can be applied.

Assumption 10 *For k large enough the matrix ΔX_{k+1} defined by*

$$\Delta X_{k+1} = [\mathbf{x}_{k+1} - \mathbf{x}_k, \, \mathbf{x}_{k+1} - \mathbf{x}_{k-1}, \, \ldots, \, \mathbf{x}_{k+1} - \mathbf{x}_{k-n+1}] \,. \tag{3.13}$$

is regular and there is a constant $K_3 > 0$ independent of k such that

$$\left(\max_{i=0,\,\ldots,\,n-1} \|\mathbf{x}_{k+1-i} - \bar{\mathbf{x}}\|^2 \right) \|\Delta X_{k+1}^{-1}\|_2^2 \leq K_3 \,. \tag{3.14}$$

Remark 6 *Assumption 10 refers to the condition of the matrix ΔX_{k+1} and equivalently to the scaled step directions $(\mathbf{x}_{k+1-i} - \mathbf{x}_{k-i}$, with $i = 0, \, \ldots, \, n-1)$. In the exceptional situations where the condition becomes too bad, the algorithm can be easily modified in order to alleviate that. Assumption 9 is related to Assumption 10 and to the well-posedness of the inverse model operators (see Assumption 12 concerning the coarse model). Because $\mathbf{c}(X)$ is a differentiable manifold we have $\Delta C \approx J_{\mathbf{c}}(\bar{\mathbf{x}}) \, \Delta X_{k+1}$ in a neighborhood of $\bar{\mathbf{x}}$, and thus, as it will become clear in the proof for Lemma 4, that $\Delta C^{\dagger} \approx \Delta X_{k+1}^{-1} \, J_{\mathbf{c}}^{\dagger}(\bar{\mathbf{x}})$. Thus we can expect Assumption 9 to be satisfied if Assumption 10 holds and $\|J_{\mathbf{c}}^{\dagger}(\bar{\mathbf{x}})\|_2^2$ is bounded. This last fact can be expressed as the inverse coarse model operator being Lipschitz in the region of interest. The inequality (3.11) is the analogous relation with respect to the fine model.*

Lemma 4 *Let the sequence of iterants $\mathbf{x}_k$ and operators S_{k+1} be defined by the original manifold-mapping algorithm (OMM). Then, under Assumptions 9 and 10 the operators S_{k+1} converge to $J_{\mathbf{f}}(\bar{\mathbf{x}}) \, J_{\mathbf{c}}^{\dagger}(\bar{\mathbf{x}})$, where $\bar{\mathbf{x}}$ is the fixed point of the iteration.*

Proof By Assumption 9 and because $\mathbf{f}$ and $\mathbf{c}$ are differentiable, we have

$$\Delta F = J_{\mathbf{f}}(\bar{\mathbf{x}}) \Delta X_{k+1} + M_{\mathbf{f}} O(\max_{i=1,\ldots,n} \|\mathbf{x}_{k+1-i} - \bar{\mathbf{x}}\|^2) \qquad (3.15)$$

$$\Delta C = J_{\mathbf{c}}(\bar{\mathbf{x}}) \Delta X_{k+1} + M_{\mathbf{c}} O(\max_{i=1,\ldots,n} \|\mathbf{x}_{k+1-i} - \bar{\mathbf{x}}\|^2), \qquad (3.16)$$

where $M_{\mathbf{f}}$ and $M_{\mathbf{c}}$ are some $m \times n$ matrices that depend on the smoothness of the manifolds $\mathbf{f}(X)$ and $\mathbf{c}(X)$ but not on k. We can use a generalization of the Banach Lemma for the inverse of a perturbed matrix [28, Theorem 6.1-2] applied to ΔC and conclude that

$$\|\Delta C^\dagger - \Delta X_{k+1}^{-1} J_{\mathbf{c}}^\dagger(\bar{\mathbf{x}})\|_2 \leq 2 \|M_{\mathbf{c}}\|_2 \max \left\{ K_2, K_3 \|J_{\mathbf{c}}^\dagger(\bar{\mathbf{x}})\|_2^2 \right\}. \qquad (3.17)$$

Because of (3.15) and the fact that the norm of $\Delta C^\dagger - \Delta X_{k+1}^{-1} J_{\mathbf{c}}^\dagger(\bar{\mathbf{x}})$ is bounded by a constant independent of k, we obtain that $S_{k+1} = \Delta F \Delta C^\dagger$ converges to $J_{\mathbf{f}}(\bar{\mathbf{x}}) J_{\mathbf{c}}^\dagger(\bar{\mathbf{x}})$. $\square$

Using this result, we can apply Lemma 3 and conclude that if the OMM algorithm converges, then the fixed point of the iteration is a local minimizer of the fine model cost function $\|\mathbf{f}(\mathbf{x}) - \mathbf{y}\|$. This is summarized in the following theorem.

Theorem 1 *Let $\bar{\mathbf{x}}$ be the fixed point of the original manifold-mapping (OMM, Fig. 2) iteration and let the fine model cost function $F(\mathbf{x}) = \|\mathbf{f}(\mathbf{x}) - \mathbf{y}\|$ be locally convex at $\bar{\mathbf{x}}$, then under Assumptions 2, 3, 4, 7, 9 and 10 the point $\bar{\mathbf{x}}$ is a local minimizer of $F(\mathbf{x})$.*

Remark 7 *The assumption of local convexity of $F(\mathbf{x})$ can be replaced by Assumption 8 (model similarity) with $\mathbf{S}$ as in (3.1)-(3.2).*

Remark 8 *The results thus far presented for the mapping proposed in the original manifold-mapping algorithm can be extended, by the same arguments, to any such algorithm based on a matrix $\bar{S}$ satisfying $\bar{S} J_{\mathbf{c}}(\mathbf{x}_{\mathbf{f}}^*) = J_{\mathbf{f}}(\mathbf{x}_{\mathbf{f}}^*)$. In fact there is complete freedom how $\bar{S}$ handles components in the complement of the range of $U_{\mathbf{c}}$. The general case is $\bar{S} = J_{\mathbf{f}}(\mathbf{x}_{\mathbf{f}}^*) J_{\mathbf{c}}^\dagger(\mathbf{x}_{\mathbf{f}}^*) + A(I - U_{\mathbf{c}} U_{\mathbf{c}}^T)$ with A any $m \times m$ matrix. This freedom can be used to stabilize the algorithm.*

3.2 Manifold mapping

In the OMM algorithm, $\mathbf{S}_k \circ \mathbf{c}$ is used as the surrogate model, i.e., it is updated during iteration, and the aim $\mathbf{y}$ is kept constant. From an implementational point of view, it is interesting to proceed the other way: the model is kept fixed and the aim is updated in each step. Then it is particularly attractive to take for this model the available coarse model, which is easily solved by assumption. This leads to a modification of the OMM algorithm. The procedure is shown in Figure 3 and we denote it simply as the *manifold-mapping* (MM) *algorithm*. As in the OMM algorithm we need an assumption to assure that $\mathbf{x}_{k+1}$ is well-defined:

Assumption 11 *In the MM algorithm, the updated aims satisfy $\mathbf{y}_k \in Y \; \forall k$.*

With this assumption replacing Assumption 7, we can prove for MM a theorem similar to Theorem 1.

$$
\begin{aligned}
&\mathbf{x}_0 = \mathbf{x}_{\mathbf{c}}^* = \operatorname{argmin}_{\mathbf{x} \in X} \|\mathbf{c}(\mathbf{x}) - \mathbf{y}\| \, ; \\
&T_0 = I_{m \times m} \, ; \\
&\mathbf{for} \quad k = 0, 1, \ldots, \, \mathbf{while} \ldots \\
&\mathbf{do} \\
&\qquad \mathbf{y}_k = \mathbf{c}(\mathbf{x}_k) - T_k \left(\mathbf{f}(\mathbf{x}_k) - \mathbf{y} \right) ; \\
&\qquad \mathbf{x}_{k+1} = \operatorname{argmin}_{\mathbf{x} \in X} \|\mathbf{c}(\mathbf{x}) - \mathbf{y}_k\| \, ; \\
&\qquad \mathbf{break \ if} \ldots \\
&\qquad \Delta F = [\mathbf{f}(\mathbf{x}_{k+1}) - \mathbf{f}(\mathbf{x}_k), \cdots, \mathbf{f}(\mathbf{x}_{k+1}) - \mathbf{f}(\mathbf{x}_{\max(k+1-n,0)})] \, ; \\
&\qquad \Delta C = [\mathbf{c}(\mathbf{x}_{k+1}) - \mathbf{c}(\mathbf{x}_k), \cdots, \mathbf{c}(\mathbf{x}_{k+1}) - \mathbf{c}(\mathbf{x}_{\max(k+1-n,0)})] \, ; \\
&\qquad \Delta F = U_{\mathbf{f}} \, \Sigma_{\mathbf{f}} \, V_{\mathbf{f}}^T \ \text{singular value decomposition} \, ; \\
&\qquad \Delta F^{\dagger} = V_{\mathbf{f}} \, \Sigma_{\mathbf{f}}^{\dagger} \, U_{\mathbf{f}}^T \, ; \\
&\qquad T_{k+1} = \Delta C \, \Delta F^{\dagger} \, ; \\
&\mathbf{enddo}
\end{aligned}
$$

Figure 3: The manifold-mapping (MM) algorithm. Note that $T_k = S_k^{\dagger}$ for all $k > 0$.

Theorem 2 *Let $\bar{\mathbf{x}}$ be the fixed point of the manifold-mapping (MM, Fig. 3) iteration, and let the fine model cost function $F(\mathbf{x}) = \|\mathbf{f}(\mathbf{x}) - \mathbf{y}\|$ be locally convex at $\bar{\mathbf{x}}$, then under Assumptions 2, 3, 4, 9, 10 and 11 the point $\bar{\mathbf{x}}$ is a local minimizer of $F(\mathbf{x})$.*

Proof Proceeding as in Lemma 4 we can see that the sequence of operators $T_{k+1} = S_{k+1}^{\dagger}$ converges to $(J_{\mathbf{f}}(\bar{\mathbf{x}}) \, J_{\mathbf{c}}^{\dagger}(\bar{\mathbf{x}}))^{\dagger} = J_{\mathbf{c}}(\bar{\mathbf{x}}) \, J_{\mathbf{f}}^{\dagger}(\bar{\mathbf{x}})$. Also in the limit

$$
\bar{\mathbf{x}} = \operatorname{argmin}_{\mathbf{x} \in X} \|\mathbf{c}(\mathbf{x}) - \mathbf{c}(\bar{\mathbf{x}}) + (J_{\mathbf{f}}(\bar{\mathbf{x}}) \, J_{\mathbf{c}}^{\dagger}(\bar{\mathbf{x}}))^{\dagger} \, (\mathbf{f}(\bar{\mathbf{x}}) - \mathbf{y})\| \, . \tag{3.18}
$$

Since $\mathbf{c}(X)$ is a manifold of class C^2, (3.18) is equivalent to

$$
\bar{\mathbf{x}} = \operatorname{argmin}_{\mathbf{x} \in X} \|J_{\mathbf{c}}(\bar{\mathbf{x}}) \, (\mathbf{x} - \bar{\mathbf{x}}) + (J_{\mathbf{f}}(\bar{\mathbf{x}}) \, J_{\mathbf{c}}^{\dagger}(\bar{\mathbf{x}}))^{\dagger} \, (\mathbf{f}(\bar{\mathbf{x}}) - \mathbf{y})\| \, . \tag{3.19}
$$

Because $J_{\mathbf{c}}(\bar{\mathbf{x}})$ and $J_{\mathbf{f}}(\bar{\mathbf{x}})$ are full rank, we can write the former equality as

$$
\bar{\mathbf{x}} = \operatorname{argmin}_{\mathbf{x} \in X} \|J_{\mathbf{f}}(\bar{\mathbf{x}}) \, J_{\mathbf{c}}^{\dagger}(\bar{\mathbf{x}}) \, J_{\mathbf{c}}(\bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}}) + \mathbf{f}(\bar{\mathbf{x}}) - \mathbf{y}\| \, . \tag{3.20}
$$

And with the same reasoning as in (3.18)-(3.19), equation (3.20) is equivalent to

$$
\bar{\mathbf{x}} = \operatorname{argmin}_{\mathbf{x} \in X} \|J_{\mathbf{f}}(\bar{\mathbf{x}}) \, J_{\mathbf{c}}^{\dagger}(\bar{\mathbf{x}}) \, (\mathbf{c}(\mathbf{x}) - \mathbf{c}(\bar{\mathbf{x}})) + \mathbf{f}(\bar{\mathbf{x}}) - \mathbf{y}\| \, . \tag{3.21}
$$

Then Lemma 3 is applied in order to show that $\bar{\mathbf{x}}$ is a local minimizer of $\|\mathbf{f}(\mathbf{x}) - \mathbf{y}\|$. $\square$

As a consequence of this theorem and by just rewriting (3.18) with $\bar{\mathbf{x}}$ a local minimizer of the fine cost function $F(\mathbf{x})$, the following interesting property for the fine model optimum can be formulated.

Corollary 1 *The fixed point of the manifold-mapping (MM) iteration $\mathbf{x}_{\mathbf{f}}^*$ satisfies*

$$
\mathbf{x}_{\mathbf{f}}^* = \operatorname{argmin}_{\mathbf{x} \in X} \|\mathbf{c}(\mathbf{x}) - \mathbf{c}(\mathbf{x}_{\mathbf{f}}^*) + \bar{S}^{\dagger} \, (\mathbf{f}(\mathbf{x}_{\mathbf{f}}^*) - \mathbf{y})\| \, . \tag{3.22}
$$

In Section 4 we show that both schemes, OMM and MM, behave asymptotically the same and that the convergence study for these two algorithms coincides.

$$\begin{aligned}
&\mathbf{x}_0 = \mathbf{x_c^*} = \mathrm{argmin}_{\mathbf{x} \in X} \|\mathbf{c}(\mathbf{x}) - \mathbf{y}\| \, ; \\
&T_0 = I_{m \times m} \, ; \\
&\textbf{for} \quad k = 0, 1, \ldots, \textbf{while} \ldots \\
&\textbf{do} \\
&\qquad \mathbf{y}_k = \mathbf{c}(\mathbf{x}_k) - T_k \left(\mathbf{f}(\mathbf{x}_k) - \mathbf{y}\right) ; \\
&\qquad \mathbf{x}_{k+1} = \mathrm{argmin}_{\mathbf{x} \in X} \|\mathbf{c}(\mathbf{x}) - \mathbf{y}_k\| \, ; \\
&\qquad \textbf{break if} \ldots \\
&\qquad T_{k+1} = J_\mathbf{c}(\mathbf{x}_{k+1}) \, \widetilde{J}_\mathbf{f}^{\dagger}(\mathbf{x}_{k+1}) \, ; \\
&\textbf{enddo}
\end{aligned}$$

Figure 4: The Generalized Manifold-Mapping (GMM) algorithm. Note that $T_k = S_k^{\dagger}$ for all $k > 0$. $\widetilde{J}_\mathbf{f}(\mathbf{x}_{k+1})$ is either $J_\mathbf{f}(\mathbf{x}_{k+1})$ in case it is available or, in other cases, it is an estimate of it.

Remark 9 *By a reasoning similar to that in Remark 8 it can be shown that the manifold-mapping algorithm can be based on any matrix $\bar{S}$ satisfying $J_\mathbf{c}^{\dagger}(\mathbf{x_f^*}) \, \bar{S}^{\dagger} = J_\mathbf{f}^{\dagger}(\mathbf{x_f^*})$. In fact there is complete freedom with respect to how $\bar{S}$ handles components in the complement of the range of $U_\mathbf{c}$. The general case is $\bar{S}^{\dagger} = J_\mathbf{c}(\mathbf{x_f^*}) \, J_\mathbf{f}^{\dagger}(\mathbf{x_f^*}) + (I - U_\mathbf{c} \, U_\mathbf{c}^T) \, A$ with A any $m \times m$ matrix. In [1], the particular cases $A = (I - U_\mathbf{f} \, U_\mathbf{f}^T)$ and $A = I$ are considered.*

3.3 Generalized Manifold Mapping

The two algorithms, OMM and MM, can be generalized by selecting any S_k for which Lemma 4 holds (i.e., S_k converges to $J_\mathbf{f}(\bar{\mathbf{x}}) \, J_\mathbf{c}^{\dagger}(\bar{\mathbf{x}})$). We call this more general scheme Generalized Manifold Mapping (GMM). A natural choice for S_k would be $J_\mathbf{f}(\mathbf{x}_k) \, J_\mathbf{c}^{\dagger}(\mathbf{x}_k)$. This makes sense in particular in case the Jacobian of the fine model is available, e.g., via automatic differentiation [29]. But in most cases that information cannot be obtained efficiently and, therefore, we rather write $S_k = \widetilde{J}_\mathbf{f}(\mathbf{x}_k) \, J_\mathbf{c}^{\dagger}(\mathbf{x}_k)$, with $\widetilde{J}_\mathbf{f}(\mathbf{x}_k)$ an approximation of $J_\mathbf{f}(\mathbf{x}_k)$, e.g., computed by means of Broyden's method [30]. In Figure 4 the GMM scheme corresponding to the MM algorithm is shown. If an approximation for $\widetilde{J}_\mathbf{f}(\mathbf{x}_0)$ is available, $T_0 = J_\mathbf{c}(\mathbf{x}_0) \, \widetilde{J}_\mathbf{f}(\mathbf{x}_0)$ could be used as initial guess instead of the identity matrix. We consider Assumption 11 to be also applicable to GMM.

Under convergence of the GMM scheme, we can also see that the fixed point for the iteration is again the accurate solution of the optimization problem, $\mathbf{x_f^*}$. The proof for the following lemma is completely analogous to those given for Theorems 1 and 2 and is, therefore, omitted.

Theorem 3 *Let $\bar{\mathbf{x}}$ be the fixed point of the generalized manifold-mapping algorithms (GMM, Figure 4) and let the fine model cost function $F(\mathbf{x}) = \|\mathbf{f}(\mathbf{x}) - \mathbf{y}\|$ be locally convex at $\bar{\mathbf{x}}$, then under Assumption 2, 3, 4, 9, 10 and 11 the point $\bar{\mathbf{x}}$ is a local minimizer of $F(\mathbf{x})$.*

4. MANIFOLD-MAPPING CONVERGENCE PROOF

Here we first present conditions for convergence of the manifold-mapping (MM) algorithm and then we show that the original manifold-mapping (OMM) iteration can be written in such a way that an equivalent convergence theorem can be stated for it, with a proof analogous to that for MM.

We define the general inverse coarse operator $\mathbf{c}^\dagger : Y \subset \mathbb{R}^m \to X \subset \mathbb{R}^n$ by

$$\mathbf{c}^\dagger \, \bar{\mathbf{y}} = \mathrm{argmin}_{\mathbf{x} \in X} \|\mathbf{c}(\mathbf{x}) - \bar{\mathbf{y}}\| . \tag{4.1}$$

We notice that this operator is well-defined because of Assumption 4. The coarse model is an essential component of any manifold-mapping technique. We are interested in coarse models whose associated inverse operators are well-posed. This is formalized in the next condition that should be satisfied by any coarse model used in practice.

Assumption 12 *The general inverse coarse model operator, $\mathbf{c}^\dagger$, is Lipschitz with a Lipschitz constant bounded by $L_{\mathbf{c}^\dagger}$.*

We can write (3.22) and the expression for iterant $\mathbf{x}_{k+1}$ in the MM algorithm as

$$\mathbf{x}_{\mathbf{f}}^* = \mathbf{c}^\dagger \left(\mathbf{c}(\mathbf{x}_{\mathbf{f}}^*) - \bar{S}^\dagger \left(\mathbf{f}(\mathbf{x}_{\mathbf{f}}^*) - \mathbf{y} \right) \right), \tag{4.2}$$

$$\mathbf{x}_{k+1}^* = \mathbf{c}^\dagger \left(\mathbf{c}(\mathbf{x}_k) - S_k^\dagger \left(\mathbf{f}(\mathbf{x}_k) - \mathbf{y} \right) \right). \tag{4.3}$$

Subtracting (4.2) from (4.3) and using Assumption 12, we get

$$\|\mathbf{x}_{k+1} - \mathbf{x}_{\mathbf{f}}^*\| \le L_{\mathbf{c}^\dagger} \left\| \mathbf{c}(\mathbf{x}_k) - S_k^\dagger \left(\mathbf{f}(\mathbf{x}_k) - \mathbf{y} \right) - \mathbf{c}(\mathbf{x}_{\mathbf{f}}^*) + \bar{S}^\dagger \left(\mathbf{f}(\mathbf{x}_{\mathbf{f}}^*) - \mathbf{y} \right) \right\| . \tag{4.4}$$

We can write the expresion in the norm at the right-hand side as

$$\mathbf{c}(\mathbf{x}_k) - S_k^\dagger \left(\mathbf{f}(\mathbf{x}_k) - \mathbf{y} \right) - \mathbf{c}(\mathbf{x}_{\mathbf{f}}^*) + \bar{S}^\dagger \left(\mathbf{f}(\mathbf{x}_{\mathbf{f}}^*) - \mathbf{y} \right)) \ =$$

$$\mathbf{c}(\mathbf{x}_k) - \mathbf{c}(\mathbf{x}_{\mathbf{f}}^*) - S_k^\dagger \left(\mathbf{f}(\mathbf{x}_k) - \mathbf{f}(\mathbf{x}_{\mathbf{f}}^*) \right) + (\bar{S}^\dagger - S_k^\dagger) \left(\mathbf{f}(\mathbf{x}_{\mathbf{f}}^*) - \mathbf{y} \right) \ =$$

$$(J_{\mathbf{c}}(\mathbf{x}_{\mathbf{f}}^*) - S_k^\dagger J_{\mathbf{f}}(\mathbf{x}_{\mathbf{f}}^*)) (\mathbf{x}_k - \mathbf{x}_{\mathbf{f}}^*) + (\bar{S}^\dagger - S_k^\dagger) \left(\mathbf{f}(\mathbf{x}_{\mathbf{f}}^*) - \mathbf{y} \right) + O(\|\mathbf{x}_k - \mathbf{x}_{\mathbf{f}}^*\|^2). \tag{4.5}$$

We now analyze the term $(\bar{S}^\dagger - S_k^\dagger)$ in (4.5). We proceed as in the proof of Lemma 4, with the difference that we cannot assume convergence of the algorithm. Since we know the *possible* fixed point of the iteration, with assumptions analogous to those for Lemma 4 (Assumptions 9 and 10), we will be able to establish a relation between $\|\bar{S}^\dagger - S_k^\dagger\|$ and $\|\mathbf{x}_{\mathbf{f}}^* - \mathbf{x}_k\|$.

Assumption 13 *For k large enough, there are constants K_4, $K_5 > 0$ independent of k such that*

$$\left(\max_{i=1,\dots,n} \|\mathbf{x}_{k+1-i} - \mathbf{x}_{k+1}\|^2 \right) \|\Delta F^\dagger\|_2^2 \ \le \ K_4 \tag{4.6}$$

$$\left(\max_{i=1,\dots,n} \|\mathbf{x}_{k+1-i} - \mathbf{x}_{k+1}\|^2 \right) \|\Delta C^\dagger\|_2^2 \ \le \ K_5 . \tag{4.7}$$

Assumption 14 *For k large enough, there is a constant $K_6 > 0$ independent of k such that*

$$\left(\max_{i=1,\dots,n} \|\mathbf{x}_{k+1-i} - \mathbf{x}_{k+1}\|^2 \right) \|\Delta X_{k+1}^{-1}\|_2^2 \le K_6 , \tag{4.8}$$

where ΔX_{k+1} is the square matrix defined in Assumption 10.

Remark 10 *We recognize in $\left(\max_{i=1,\dots,n} \|\mathbf{x}_{k+1-i} - \mathbf{x}_{k+1}\|^2\right)^{1/2}$ a matrix-norm for ΔX_{k+1}. Thus, Assumption 14 can be stated in terms of $\kappa(\Delta X_{k+1})$ the condition number of ΔX_{k+1}, i.e., $\kappa(\Delta X_{k+1}) \leq K_6$, with $K_6 > 0$ a constant independent of k. In the rare situations where linear dependence in the step directions is obtained, the algorithm can be slightly modified, with no significant influence in the final result, to cope with that issue. Assumption 13 can be related to the condition number of ΔX_{k+1} and the well-posedness of the inverse model operators in the region of interest.*

Lemma 5 *Under Assumptions 2, 3, 4, 8, 9, 10, 11, 13 and 14,*

$$\|S_k^\dagger - J_{\mathbf{c}}(\mathbf{x}_k)\, J_{\mathbf{f}}^\dagger(\mathbf{x}_k)\| \leq M_1 \max_{i=0,\dots,n-1} \|\mathbf{x}_{k-i} - \mathbf{x}_{\mathbf{f}}^*\|, \tag{4.9}$$

where $M_1 > 0$ is a constant that depends on the smoothness of the manifolds $\mathbf{f}(X)$ and $\mathbf{c}(X)$ but not on k.

Proof As in Lemma 4 we can write

$$\Delta F = J_{\mathbf{f}}(\mathbf{x}_{k+1})\, \Delta X_{k+1} + M_{\mathbf{f}_k}\, O\left(\max_{i=1,\dots,n} \|\mathbf{x}_{k+1-i} - \mathbf{x}_{k+1}\|^2\right) \tag{4.10}$$

$$\Delta C = J_{\mathbf{c}}(\mathbf{x}_{k+1})\, \Delta X_{k+1} + M_{\mathbf{c}_k}\, O\left(\max_{i=1,\dots,n} \|\mathbf{x}_{k+1-i} - \mathbf{x}_{k+1}\|^2\right), \tag{4.11}$$

where $M_{\mathbf{f}_k}$ and $M_{\mathbf{c}_k}$ are some $m \times n$ matrices that depend on the smoothness of the manifolds $\mathbf{f}(X)$ and $\mathbf{c}(X)$ but not on k. Again with [28, Theorem 6.1-2] and by Assumptions 13 and 14 we can conclude that $S_k^\dagger - J_{\mathbf{c}}(\mathbf{x}_k)\, J_{\mathbf{f}}^\dagger(\mathbf{x}_k)$ is bounded by a constant multiplied by $\max_{i=1,\dots,n} \|\mathbf{x}_{k-i} - \mathbf{x}_k\|$. Clearly, this is equivalent to $S_k^\dagger - J_{\mathbf{c}}(\mathbf{x}_k)\, J_{\mathbf{f}}^\dagger(\mathbf{x}_k)$ being bounded by $M_1 \max_{i=0,\dots,n-1} \|\mathbf{x}_{k-i} - \mathbf{x}_{\mathbf{f}}^*\|$ where $M_1 > 0$ is a constant that depends on the smoothness of $\mathbf{f}(X)$ and $\mathbf{c}(X)$ but not on k. $\square$

Remark 11 *The constant M_1 depends on the smoothness of $\mathbf{f}(X)$ and $\mathbf{c}(X)$ ($M_1 = 0$ if both manifolds are linear in the neighborhood of the solution).*

Lemma 6 *Under Assumptions 2, 3, 4, 8, 9, 10, 11, 13 and 14,*

$$\|\bar{S}^\dagger - J_{\mathbf{c}}(\mathbf{x}_k)\, J_{\mathbf{f}}^\dagger(\mathbf{x}_k)\| \leq M_2 \|\mathbf{x}_k - \mathbf{x}_{\mathbf{f}}^*\|, \tag{4.12}$$

where $M_2 > 0$ is a constant that depends on the smoothness of the manifolds $\mathbf{f}(X)$ and $\mathbf{c}(X)$ but not on k.

Proof We can write

$$\begin{aligned}
\bar{S}^\dagger - J_{\mathbf{c}}(\mathbf{x}_k)\, J_{\mathbf{f}}^\dagger(\mathbf{x}_k) &= J_{\mathbf{c}}(\mathbf{x}_{\mathbf{f}}^*)\, J_{\mathbf{f}}^\dagger(\mathbf{x}_{\mathbf{f}}^*) - J_{\mathbf{c}}(\mathbf{x}_k)\, J_{\mathbf{f}}^\dagger(\mathbf{x}_k) \\
&= J_{\mathbf{c}}(\mathbf{x}_{\mathbf{f}}^*)\, J_{\mathbf{f}}^\dagger(\mathbf{x}_{\mathbf{f}}^*) - J_{\mathbf{c}}(\mathbf{x}_k)\, J_{\mathbf{f}}^\dagger(\mathbf{x}_{\mathbf{f}}^*) + \\
&\quad + J_{\mathbf{c}}(\mathbf{x}_k)\, J_{\mathbf{f}}^\dagger(\mathbf{x}_{\mathbf{f}}^*) - J_{\mathbf{c}}(\mathbf{x}_k)\, J_{\mathbf{f}}^\dagger(\mathbf{x}_k) \\
&= (J_{\mathbf{c}}(\mathbf{x}_{\mathbf{f}}^*) - J_{\mathbf{c}}(\mathbf{x}_k))\, J_{\mathbf{f}}^\dagger(\mathbf{x}_{\mathbf{f}}^*) + \\
&\quad + J_{\mathbf{c}}(\mathbf{x}_k)\, (J_{\mathbf{f}}^\dagger(\mathbf{x}_{\mathbf{f}}^*) - J_{\mathbf{f}}^\dagger(\mathbf{x}_k)).
\end{aligned} \tag{4.13}$$

Since $\mathbf{f}(X)$ and $\mathbf{c}(X)$ are manifolds of class C^2, we can bound $\|\bar{S}^\dagger - J_{\mathbf{c}}(\mathbf{x}_k)\, J_{\mathbf{f}}^\dagger(\mathbf{x}_k)\|$ by $M_2 \|\mathbf{x}_k - \mathbf{x}_{\mathbf{f}}^*\|$, where $M_2 > 0$ is a constant that depends on the smoothness of the two manifolds but not on k. $\square$

Corollary 2 *Under Assumptions 2, 3, 4, 8, 9, 10, 11, 13 and 14,*

$$\|\bar{S}^{\dagger} - S_k^{\dagger}\| \leq M \max_{i=0,\dots,n-1} \|\mathbf{x}_{k-i} - \mathbf{x}_{\mathbf{f}}^*\|, \tag{4.14}$$

where $M > 0$ is a constant that depends on the smoothness of the manifolds $\mathbf{f}(X)$ and $\mathbf{c}(X)$ but not on k.

Proof We apply Lemmas 5 and 6 and set $M = \max(M_1, M_2)$. $\square$

Now, combining (4.4) and (4.5) we get

$$\begin{aligned}
\|\mathbf{x}_{k+1} - \mathbf{x}_{\mathbf{f}}^*\| &\leq L_{\mathbf{c}\dagger} \|J_{\mathbf{c}}(\mathbf{x}_{\mathbf{f}}^*) - S_k^{\dagger} J_{\mathbf{f}}(\mathbf{x}_{\mathbf{f}}^*)\| \|\mathbf{x}_k - \mathbf{x}_{\mathbf{f}}^*\| + \\
&\quad + \|\bar{S}^{\dagger} - S_k^{\dagger}\| \|\mathbf{f}(\mathbf{x}_{\mathbf{f}}^*) - \mathbf{y}\| + O(\|\mathbf{x}_k - \mathbf{x}_{\mathbf{f}}^*\|^2).
\end{aligned} \tag{4.15}$$

Due to (4.14) we can finally write

$$\begin{aligned}
\|\mathbf{x}_{k+1} - \mathbf{x}_{\mathbf{f}}^*\| &\leq L_{\mathbf{c}\dagger} (\|J_{\mathbf{c}}(\mathbf{x}_{\mathbf{f}}^*) - S_k^{\dagger} J_{\mathbf{f}}(\mathbf{x}_{\mathbf{f}}^*)\| + \\
&\quad + M \|\mathbf{f}(\mathbf{x}_{\mathbf{f}}^*) - \mathbf{y}\|) \max_{i=0,\dots,n-1} \|\mathbf{x}_{k-i} - \mathbf{x}_{\mathbf{f}}^*\| + \\
&\quad + O(\|\mathbf{x}_k - \mathbf{x}_{\mathbf{f}}^*\|^2).
\end{aligned} \tag{4.16}$$

We formulate this result in the following theorem.

Theorem 4 *Under Assumptions 2, 3, 4, 8, 9, 10, 11, 12, 13 and 14, and the condition*

$$L_{\mathbf{c}\dagger} (\|J_{\mathbf{c}}(\mathbf{x}_{\mathbf{f}}^*) - S_k^{\dagger} J_{\mathbf{f}}(\mathbf{x}_{\mathbf{f}}^*)\| + M \|\mathbf{f}(\mathbf{x}_{\mathbf{f}}^*) - \mathbf{y}\|) < 1 \qquad \text{for} \quad k \geq k_0, \tag{4.17}$$

where $M > 0$ is a constant that depends on the smoothness of the manifolds $\mathbf{f}(X)$ and $\mathbf{c}(Z)$ but not on k, the manifold-mapping algorithm (MM, Fig. 3) yields (linear) convergence to $\mathbf{x}_{\mathbf{f}}^*$.

Corollary 3 *If in addition to the assumptions for Theorem 4 we have $\mathbf{f}(\mathbf{x}_{\mathbf{f}}^*) = \mathbf{y}$ (i.e., a reachable design), then the convergence of the MM algorithm is superlinear.*

Corollary 4 *If in addition to the assumptions for Theorem 4 we have $S_k = \bar{S} = J_{\mathbf{f}}(\mathbf{x}_{\mathbf{f}}^*) J_{\mathbf{c}}^{\dagger}(\mathbf{x}_{\mathbf{f}}^*)$ for every $k \geq k_0$, then the convergence of the MM algorithm is quadratic.*

As a corollary of the following two lemmas, we will see that —under convergence— the original (OMM) and the manifold-mapping (MM) iterations are asymptotically equivalent.

Lemma 7 *If the iteration in the original manifold-mapping (OMM) algorithm converges, for a large enough k we find for its iterant $\mathbf{x}_{k+1}$*

$$\mathbf{x}_{k+1} = \operatorname{argmin}_{\mathbf{x} \in X} \|\mathbf{c}(\mathbf{x}) - \mathbf{c}(\mathbf{x}_k) + S_k^{\dagger} (\mathbf{f}(\mathbf{x}_k) - \mathbf{y}) + O(\|\mathbf{x} - \mathbf{x}_k\|^2)\|. \tag{4.18}$$

Proof Because $\mathbf{c}(X)$ is differentiable in a neighborhood of the fixed point, for a large enough k we can write

$$
\begin{aligned}
\mathbf{x}_{k+1} \;=\;& \operatorname{argmin}_{\mathbf{x}\in X}\|S_k \mathbf{c}(\mathbf{x}) - S_k \mathbf{c}(\mathbf{x}_k) + \mathbf{f}(\mathbf{x}_k) - \mathbf{y}\| && (4.19) \\
=\;& \operatorname{argmin}_{\mathbf{x}\in X}\|S_k J_{\mathbf{c}}(\mathbf{x}_k)\,(\mathbf{x}-\mathbf{x}_k) + \mathbf{f}(\mathbf{x}_k) - \mathbf{y} + S_k\,O(\|\mathbf{x}-\mathbf{x}_k\|^2)\| \\
=\;& \operatorname{argmin}_{\mathbf{x}\in X}\|S_k J_{\mathbf{c}}(\mathbf{x}_k)\,(\mathbf{x}-\mathbf{x}_k) + \mathbf{f}(\mathbf{x}_k) - \mathbf{y} + O(\|\mathbf{x}-\mathbf{x}_k\|^2)\|.
\end{aligned}
$$

and for the last equality we remember that S_k converges to $\bar{S}$ (Lemma 4).
The iterant $\mathbf{x}_{k+1}$ can be expressed as

$$
\begin{aligned}
\mathbf{x}_{k+1} \;=\;& \operatorname{argmin}_{\mathbf{x}\in X}\|S_k\,\Delta C\,\Delta X_k^{-1}\,(\mathbf{x}-\mathbf{x}_k) + && (4.20) \\
& +\mathbf{f}(\mathbf{x}_k) - \mathbf{y} + O(\|\mathbf{x}-\mathbf{x}_k\|^2)\|,
\end{aligned}
$$

where, for a large enough k we have $O(\|\mathbf{x}-\mathbf{x}_k\|) = O(\max_{i=1,\dots,n}\|\mathbf{x}_{k-i}-\mathbf{x}_k\|)$ in the Taylor expansions, since there is convergence. Because $S_k = \Delta F\,\Delta C^{\dagger}$ and $\Delta C^{\dagger}\,\Delta C = I$, we have

$$
\begin{aligned}
\mathbf{x}_{k+1} \;=\;& \operatorname{argmin}_{\mathbf{x}\in X}\|\Delta F\,\Delta X_k^{-1}\,(\mathbf{x}-\mathbf{x}_k) + && (4.21) \\
& +\mathbf{f}(\mathbf{x}_k) - \mathbf{y} + O(\|\mathbf{x}-\mathbf{x}_k\|^2)\|.
\end{aligned}
$$

Further, since ΔF and ΔC are full-rank and ΔX_k is regular

$$
\begin{aligned}
\mathbf{x}_{k+1} \;=\;& \operatorname{argmin}_{\mathbf{x}\in X}\|\Delta C\,\Delta X_k^{-1}\,(\mathbf{x}-\mathbf{x}_k) + && (4.22) \\
& +S_k^{\dagger}\,(\mathbf{f}(\mathbf{x}_k) - \mathbf{y}) + O(\|\mathbf{x}-\mathbf{x}_k\|^2)\|.
\end{aligned}
$$

The lemma follows immediately from this last equation. □

Lemma 8 *If the iteration in the manifold-mapping (MM) algorithm converges, for a large enough k we find for its iterant $\mathbf{x}_{k+1}$*

$$
\mathbf{x}_{k+1} = \operatorname{argmin}_{\mathbf{x}\in X}\|S_k\,(\mathbf{c}(\mathbf{x}) - \mathbf{c}(\mathbf{x}_k)) + \mathbf{f}(\mathbf{x}_k) - \mathbf{y} + O(\|\mathbf{x}-\mathbf{x}_k\|^2)\|. \tag{4.23}
$$

Proof Analogous to that for Lemma 7. □

Corollary 5 *Under convergence, the OMM algorithm and the MM algorithm are asymptotically equivalent.*

Remark 12 *Due to Corollary 5, Theorem 4 and Corollaries 3 and 4 are also valid for the original manifold-mapping algorithm (OMM, Fig. 2) when Assumption 11 is replaced by Assumption 7.*

Remark 13 *By similar arguments as given for mappings based on $\bar{S} = J_{\mathbf{f}}(\mathbf{x}_{\mathbf{f}}^{*})\,J_{\mathbf{c}}^{\dagger}(\mathbf{x}_{\mathbf{f}}^{*})$, analogous convergence results can be obtained for the other versions of the algorithms as introduced in Remarks 8 and 9.*

4.1 Generalized Manifold-Mapping Convergence Proof

The proof for generalized manifold mapping (GMM) is analogous to that for MM (by construction of S_{k+1} we fulfill the conditions for Lemma 5). The two possible GMM algorithms, corresponding to OMM and MM, are also asymptotically equivalent. We formulate this in the following theorem.
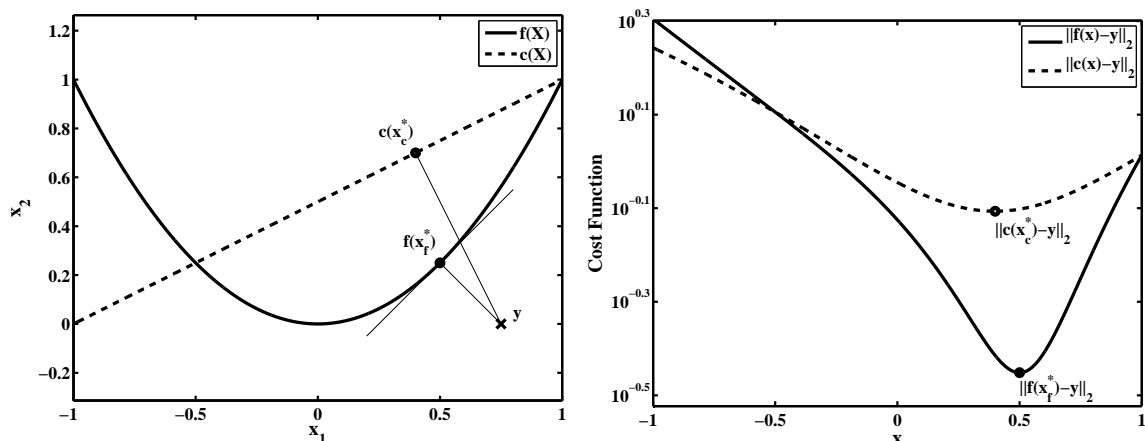
Figure 5: Top: The sets $\mathbf{f}(X)$ and $\mathbf{c}(X)$, specifications $\mathbf{y}$ and $\mathbf{f}(x_{\mathbf{f}}^*)$ and $\mathbf{c}(x_{\mathbf{c}}^*)$ for the first example. Bottom: the fine and coarse cost functions, $\|\mathbf{f}(x) - \mathbf{y}\|_2$ and $\|\mathbf{c}(x) - \mathbf{y}\|_2$ respectively, for the same example.

Theorem 5 *Under Assumptions 2, 3, 4, 8, 9, 10, 11, 12, 13 and 14, and the condition*

$$L_{\mathbf{c}\dagger} \left(\|J_{\mathbf{c}}(\mathbf{x}_{\mathbf{f}}^*) - S_k^\dagger J_{\mathbf{f}}(\mathbf{x}_{\mathbf{f}}^*)\| + \bar{M} \|\mathbf{f}(\mathbf{x}_{\mathbf{f}}^*) - \mathbf{y}\| \right) < 1 \qquad for \quad k \geq k_0 \, , \qquad (4.24)$$

where $\bar{M} > 0$ is a constant that depends on the smoothness of the manifolds $\mathbf{f}(X)$ and $\mathbf{c}(X)$ but not on k, the generalized manifold-mapping algorithms (GMM) yield (linear) convergence to $\mathbf{x}_{\mathbf{f}}^*$.

Remark 14 *In the case $S_{k+1} = J_{\mathbf{f}}(\mathbf{x}_{k+1}) J_{\mathbf{c}}^\dagger(\mathbf{x}_{k+1})$ an analog to Lemma 5 can be trivially stated. The constant introduced in that lemma for GMM could be in some situations smaller than the one for MM. As a consequence, the (linear) convergence for the GMM schemes may have smaller associated constants than that for the MM iterations (see Section 5.1).*

5. EXAMPLES

5.1 Two Simple Examples

First simple problem With this simple example we illustrate the convergence Theorems 4 and 5. The fine model is defined over $X = [-1, 1]$ by

$$\mathbf{f}(x) = [x, x^2] \, . \qquad (5.1)$$

The set $\mathbf{f}(X) \subset \mathbb{R}^2$ is part of a parabola and we want to find the point in that set closest in Euclidean norm to the specifications $\mathbf{y} = [3/4, 0]$. The coarse model is defined over $Z = X$ and is the linear $\mathbf{c}(x) = [x, (1+x)/2]$. Figure 5 shows a representation of the problem and the fine and coarse cost functions, $\|\mathbf{f}(x) - \mathbf{y}\|_2$ and $\|\mathbf{c}(x) - \mathbf{y}\|_2$ respectively. The corresponding optima are $x_{\mathbf{f}}^* = 0.5$ and $x_{\mathbf{c}}^* = 0.4$.

Though the two models are not specially similar around the solution region the similarity between them is enough for obtaining convergence with the manifold-mapping approach. Since both manifolds $\mathbf{f}(X)$ and $\mathbf{c}(X)$ are smooth, we expect a reasonable small constant M
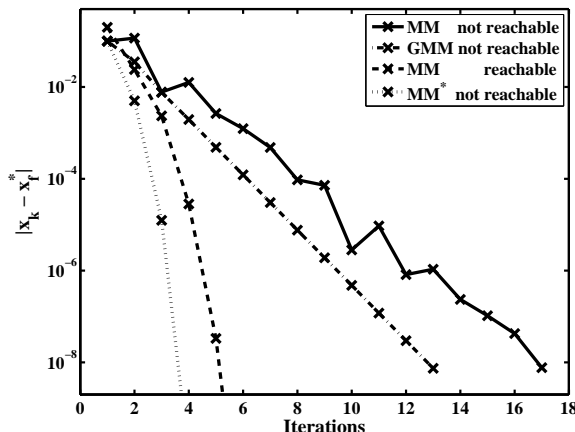
Figure 6: Convergence history for $|x_k - x_{\mathbf{f}}^*|$ for the manifold-mapping (MM) and generalized manifold-mapping (GMM) iterations for a reachable and not reachable designs in the first simple mathematical example. MM* denotes the MM algorithm with $S_k = \bar{S}$.

in (4.17) for Theorem 4. The design is not reachable and therefore convergence is linear for the manifold-mapping (MM) and the generalized manifold-mapping (GMM) iterations (see Figure 6). In this problem it is easy to check that both original manifold mapping (OMM) and MM coincide, iterant by iterant. We see that MM needs 17 iterations for getting $|x_k - x_{\mathbf{f}}^*|$ smaller than 10^{-8}. The GMM scheme, using the exact Jacobian, yields the same accuracy in 13 iterations. In Figure 6 we also check that the constant in the linear convergence rate is smaller for the latter algorithm. If the Jacobian is estimated by Broyden's method, the complete iteration history coincides with that for MM. The reason is that for a function of one variable, Broyden's method coincides with the secant algorithm for approximating a derivative, and that procedure is essentially the one followed in the computation of ΔF and ΔC in the MM algorithm (not only MM and GMM use the same Jacobian estimation, also the iterants x_0 and x_1 coincide). In the next example, when we consider a function of two variables, we see that the Broyden-based GMM algorithm differs significantly from the MM one.

In Figure 6 we also observe two cases of superlinear convergence for MM (cf, Corollaries 3 and 4). If we apply MM with $S_k = \bar{S}$ (denoted by MM* in Figure 6) we obtain a solution with the same accuracy of 10^{-8} in only four iterations. Nevertheless that situation is unrealistic because the necessary information is not available before the optimization problem has been solved. In the case of the reachable design, given by $\mathbf{y} = [1/2, \, 1/4]$ (yielding again $x_{\mathbf{f}}^* = 0.5$), the solution is obtained with an accuracy of 10^{-8} in six iterations. The same superlinear convergence is observed for the GMM algorithm.

Second simple problem By means of this example with two design variables we show that different choices for the fine model Jacobian estimation at the k-th iteration, $\widetilde{J}_{\mathbf{f}}(\mathbf{x}_k)$, yield distinct convergence histories: the better $\widetilde{J}_{\mathbf{f}}(\mathbf{x}_k)$ approximates $J_{\mathbf{f}}(\mathbf{x}_k)$, the smaller the constant associated with the linear convergence. The example was introduced in [1] as a least
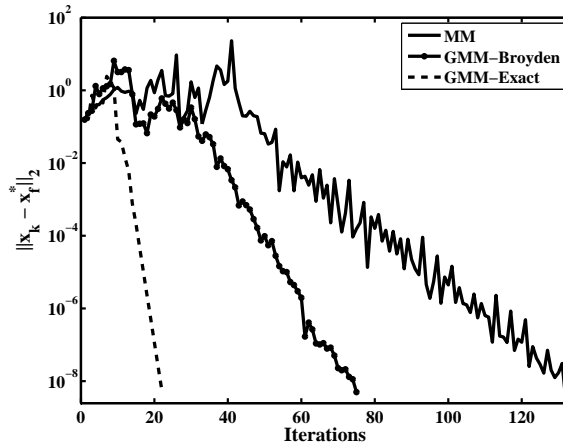
Figure 7: Convergence history for $\|\mathbf{x}_k - \mathbf{x}_f^*\|$ for the manifold-mapping (MM) and two different generalized manifold-mapping (GMM) iterations in the second simple example (non-reachable design and a bad correspondence between the fine and the coarse model). GMM-Exact denotes the GMM scheme with the exact Jacobian $J_\mathbf{f}(\mathbf{x}_k)$. GMM-Broyden approximates that Jacobian via Broyden's method.

squares best approximation of the data vector $\mathbf{y} = [0, -0.4, 0.1]$ by the fine model

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(x_1, x_2) = [x_1 \, (x_2 - 1)^2, \ x_1, \ x_1 \, (x_2 + 1)^2] \tag{5.2}$$

defined over $X = \mathbb{R}^2$. It can be seen that the design is not reachable. The coarse model

$$\mathbf{c}(\mathbf{x}) = \mathbf{c}(x_1, x_2) = [-x_1 + x_2, \ x_2, \ x_1 + x_2] \tag{5.3}$$

is again linear and it is also defined over $Z = X$.

We solve the problem with the manifold-mapping (MM) and the generalized manifold-mapping (GMM) algorithms. For GMM, two variants are compared, one with the exact Jacobian for the fine model and the other with an approximation based on Broyden's method. It should be noticed that for most time-expensive fine models, the availability of the exact Jacobian is an unrealistic assumption. All the schemes yield the fine optimum $\mathbf{x}_f^* = [-0.101, -0.141]$. The convergence history for the three methods is shown in Figure 7. We clearly observe that the convergence is linear in all cases and that the constant M in the convergence theorems (mean slope of the convergence history) is different for each algorithm. In this problem the discrepancy between the fine and coarse models in the solution region is large and this fact is recognized in a large number of iterations compared with the previous example. In practice, fine and coarse models are much more similar and, hence, convergence is generally achieved in much less iterations.

5.2 Practical Examples

The next two examples are design problems from practice [31, 27]. We show that that they can be solved efficiently by the manifold-mapping approach. Since, in both cases, Jacobians for the fine model are not available, Broyden's method is used to estimate $J_\mathbf{f}(\mathbf{x}_k)$ when the generalized manifold-mapping iteration is used. In the two optimal design problems, the inequality constraints $\bar{\mathbf{k}}(\mathbf{x})$ are easy to compute (box constraints).
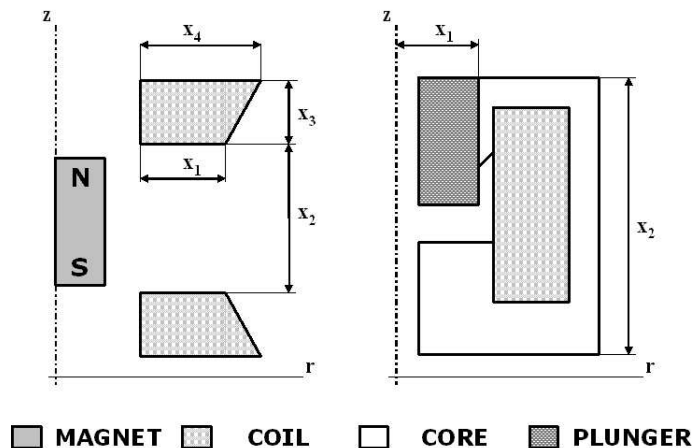
Figure 8: Geometries of the coreless (left) and automotive (right) actuators. The axis of rotational symmetry is denoted by **z**.

Coreless Actuator [31]   A magnetic actuator is a device that converts magnetic power into mechanical force and motion. The one in Figure 8 (left) consists of a moving cylindrical magnet and two fixed toroidal coils. It is called *coreless* because actuators usually have ferromagnetic cores. The specification **y** is a force response exerted on the magnet. This is specified as the force at a finite number of displacements of the magnet along the symmetry axis. The actuator geometry $\mathbf{x} = [x_1, x_2, x_3, x_4]$ (see Figure 8) has to be determined so that the discrepancy in the Euclidean norm between the computed and specified force response is minimized. The fine model is based on a calculation of the forces after a finite-element computation of the vector magnetic potential formulation [32] of the axisymmetric magnetostatic equations. In the coarse model, the force is computed for a much simplified model, where the actuator is represented by only a (small) number of coils. Depending on the number of coils in the simplification, several coarse models are obtained. The coarse model computation used here is around 300 times faster than the fine model one. More details on this particular problem can be found in [31].

The coreless actuator optimization results are shown in Table 1. Both the manifold and the Broyden-based generalized manifold-mapping algorithms yield an acceptable design solution after only two fine model evaluations. The number of coarse model evaluations is larger for the second scheme because the coarse model Jacobian is approximated by finite differences in each iteration step. The space-mapping technique was reported to be the best efficient solver for this problem before [31]. In order for space mapping to obtain a comparable value for the cost function as the one for the two manifold-mapping algorithms in Table 1, all coarse models in [31] are considered. The best option found requires three fine model evaluations.

Automotive Actuator [27]   Automotive actuators are used in devices such as electromagnetic switches, relays, valves, etc. and they typically generate high levels of force. The one in Figure 8 (right) consists of a core and a plunger, both made of iron, and a copper coil. The design variables are the sizes $\mathbf{x} = [x_1, x_2]$ indicated in the figure. Again, the specification **y** is a force response exerted over the plunger, when it moves vertically. The Euclidean

Table 1: Optimizing the coreless actuator.

	# (f, c) evals.	Cost function
MM	(2, 132)	2.2
GMM	(2, 148)	2.2
SM	(3, 150)	2.3

MM: manifold mapping; GMM: generalized manifold mapping; SM: space mapping. The second column shows the total amount of fine and coarse model evaluations needed in the optimization. The coarse model used is approximately 300 times faster than the fine one.

Table 2: Optimizing the automotive actuator.

	# (f, c) evals.	Cost function
MM	(3, 66)	45.0
GMM	(3, 78)	45.0
SM	(6, 296)	45.3
SQP	(12, 47)	45.0

MM: manifold mapping; GMM: generalized manifold mapping; SM: space mapping; SQP: Sequential quadratic programming. The second column shows the total amount of fine and coarse model evaluations needed in the optimization. The coarse model used is approximately 400 times faster than the fine one. SQP takes the coarse model optimum $\mathbf{x_c^*}$ as its initial guess.

norm of the difference between the computed force for a given $\mathbf{x}$ and the specification $\mathbf{y}$ is minimized, keeping the device volume constant. (Elimination of this easy-to-compute equality constraint leads to the two mentioned design variables.) Finite elements for the axisymmetric magnetostatic equations plus a post-processing stage make the fine model. The coarse model is based on an equivalent magnetic circuit [33] of the actuator (analogous to an electrical circuit) and it is around 400 times faster than the fine one. It should be noticed that the fine model optimum $\mathbf{x_f^*}$ lies on the boundary of the set X. Additional problem details can be found in [27].

Table 2 shows the results for the automotive actuator optimization. The cost function corresponding to the solutions has a considerably larger value than in the previous example because the specification $\mathbf{y}$ is significantly higher and this cost function is not a relative measure of the force discrepancy. When these facts are taken into account, the optimal cost function values are similar for both design problems. The manifold-mapping algorithm yields the solution after three fine model function evaluations. Generalized manifold mapping, with the fine model Jacobian estimated via Broyden's method, performs almost identically. (The coarse model Jacobian computations are recognized in a small increase in the coarse model function evaluations for the generalized manifold-mapping algorithm). The usual (standard) optimization method for these problems, sequential quadratic programming (SQP) [26], is applied to the fine model cost function with the coarse model optimum $\mathbf{x_c^*}$ as initial guess.

A design with a cost function value comparable with the one obtained with the manifold-mapping approach is obtained with almost four times the computational cost. Space mapping, on the other side, performs faster than sequential quadratic programming but it is less efficient than manifold mapping. In this case we observe the well-know fact that SM does not converge to the accurate solution [1].

6. Conclusion

Manifold mapping was introduced in [1] as an efficient two-level approach for optimization, to be considered for very time-demanding design problems. In this paper a number of possible variants are identified: original manifold mapping (OMM), manifold mapping (MM) and generalized manifold mapping (GMM). OMM and MM are asymptotically equivalent and they perform identically in most practical situations. MM is preferred to OMM because it is easier to implement. GMM represents a whole family of schemes based on the manifold-mapping approach. If a better approximation for $J_{\mathbf{f}}(\mathbf{x}_k)$ than $\Delta F \, \Delta X_k^{-1}$ is available, then GMM is recommended. In all cases, if the iteration converges, the solution of the accurate model is the fixed point of the iteration. Conditions for convergence have been given for all the algorithms described. By two simple examples we illustrated some of the theoretical aspects dealt with in the paper. The practical performance of the presented algorithms is shown with two cases of optimal design in electromagnetics.

References

1. D. Echeverría and P.W. Hemker, *Space mapping and defect correction*, Comp. Methods in Appl. Math., vol. 5(2), pp. 107–136, 2005.

2. L.A. Schmit, Jr. and B. Farshi, *Some approximation concepts for structural synthesis*, AIAA Journal, 12, pp. 692–699, 1974.

3. V. Braibant and C. Fleury, *An approximation-concepts approach to shape-optimal design*, Computer Methods in Applied Mechanics and Engineering, 53, pp. 119–148, 1985.

4. J.-F.M. Barthelemy and R.T. Haftka, *Approximation concepts for optimum structural design - a review*, Structural Optimization, 5, pp. 129–144, 1993.

5. J.E. Dennis, Jr. and V. Torczon, *Managing approximation models in optimization*, in Multidisciplinary Design Optimization: State of the Art, edited by N. M. Alexandrov and M. Y. Hussaini, SIAM, pp. 330–347, 1997.

6. A.J. Booker, J.E. Dennis, Jr., P.D. Frank, D.B. Serafini, V. Torczon and M.W. Trosset, *A rigorous framework for optimization of expensive functions by surrogates*, Structural and Multidisciplinary Optimization, 17(1), pp. 1–13, 1999.

7. N. Alexandrov, J. E. Dennis, Jr., R. M. Lewis and V. Torczon, *A trust region framework for managing the use of approximation models in optimization*, pp. 16–23, Structural and Multidisciplinary Optimization, 15(1), 1998.

8. A.A. Giunta and M.S. Eldred, *Implementation of a Trust Region Model Management Strategy in the DAKOTA Optimization Toolkit*, paper AIAA-2000-4935 in Proceedings of the 8th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, Long Beach, CA, September 6–8, 2000.

9. M. Stein, *Large sample properties of simulating using latin hypercube sampling*, Technometrics, 29(2), pp. 143–151, 1987.

10. A.B. Owen, *Orthogonal arrays for computer experiments, integration and visualization*, Statistica Sinica, 2, pp. 439–452, 1992.

11. C. DE BOOR AND A. RON, *Computational aspects of polynomial interpolation in several variables*, Mathematics of Computation, 58, pp. 705–727, 1992.

12. A.G. JOURNEL AND CH.J. HUIJBREGTS, *Mining Geostatistics*, Academic Press, 1981.

13. M.D. BUHMANN AND M.J. ABLOWITZ, *Radial Basis Functions : Theory and Implementations*, Cambridge University, ISBN 0-521-63338-9, 2003.

14. 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. on Microwave Theory and Techniques, 52(1), pp. 337–361, 2004.

15. F. PEDERSEN, P. WEITZMANN AND S. SVENDSEN, *Modeling thermally active building components using space mapping* in the Proceedings of the 7th Symposium on Building Physics in the Nordic Countries (ISBN: 9979-9174-6-6), pp. 896–903, 2005.

16. J.W. BANDLER, R.M. BIERNACKI, C.H. CHEN, P.A. GROBELNY AND R.H. HEMMERS, *Space Mapping Technique for Electromagnetic Optimization*, IEEE Trans. on Microwave Theory and Techniques, 42(12), pp. 2536–2544, 1994.

17. J.W. BANDLER, R.M. BIERNACKI, S.H. CHEN, R.H. HEMMERS AND K. MADSEN, *Electromagnetic optimization exploiting aggressive space mapping*, IEEE Trans. on Microwave Theory and Techniques, 43, pp. 2874–2882, 1995.

18. 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. on Microwave Theory and Techniques, 46, pp. 2412–2425, 1998.

19. M.H. BAKR, J.W. BANDLER, M.A. ISMAIL , J.E. RAYAS-SÁNCHEZ AND Q.J. ZHANG, *Neural space-mapping optimization for EM-based design*, IEEE Trans. on Microwave Theory and Techniques, 48, pp. 2307–2315, 2000.

20. J.W. BANDLER, Q.S. CHENG, N.K. NIKOLOVA AND M.A. ISMAIL, *Implicit space mapping optimization exploiting preassigned parameters*, IEEE Trans. on Microwave Theory and Techniques, 52, pp. 378–385, 2004.

21. 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. on Microwave Theory and Techniques, 52, pp. 2593–2600, 2004.

22. K. BÖHMER, P.W. HEMKER AND H.J. STETTER, *Defect Correction Methods: Theory and Applications*, The defect correction approach, Computing Suppl. 5, pp. 1–32, K. Böhmer and H.J. Stetter ed., Springer-Verlag, Berlin, Heidelberg, New York, Tokyo, 1984.

23. K. BÖHMER AND H.J. STETTER, *Defect Correction Methods: Theory and Applications*, Springer, Berlin, 1984.

24. W. HACKBUSH, *Multigrid Methods and Applications*, Series in Computational Mathematics, Springer-Verlag, Berlin, 1985.

25. F.W. WARNER, *Foundations of Differentiable Manifolds and Lie Groups*, Graduate Texts in Mathematics, Springer-Verlag, New York-Berlin-Heidelberg-Tokyo, ISBN 0-387-90894-3, 1983.

26. J. NOCEDAL AND S.J. WRIGHT, *Numerical Optimization*, Springer Series in Operations Research, Springer, 1999.

27. D. ECHEVERRÍA, D. LAHAYE, L. ENCICA, E.A. LOMONOVA, P.W. HEMKER AND A.J.A. VANDENPUT, *Manifold-Mapping Optimization Applied to Linear Actuator Design*, IEEE Transactions on Magnetics, 42(4), pp. 1183–1186, 2006.

28. G.H. GOLUB AND C.F. VAN LOAN, *Matrix Computations*, North Oxford Academic, Oxford, England, 1983.

29. A. GRIEWANK, *Evaluating Derivatives. Principles and Techniques of Algorithmic Differentiation*, SIAM Frontiers in Applied Mathematics 19, Philadelphia, 2000.

30. C.G. BROYDEN, *A class of methods for solving nonlinear simultaneous equations*, Math. Comp., 19, pp. 577–593, 1965.

31. D. ECHEVERRÍA, D. LAHAYE, L. ENCICA AND P.W. HEMKER, *Optimisation in Electromagnetics with the Space-Mapping Technique*, COMPEL The International Journal for Computation and Mathematics in Electrical and Electronic Engineering, 24(3), pp. 952–966, 2005.

32. P.P. SILVESTER AND R.L. FERRARI, *Finite Element for Electrical Engineers*, Cambridge University Press, New York, 1996.

33. D.K. CHENG, *Field and Wave Electromagnetics*, Addison-Wesley, 1989.

Appendix I

1. Definitions for differentiable manifolds

Because the concept of manifold is central in our description of the optimization methods studied in this paper, in this appendix we summarize a number of basic notions, from [25], related with manifolds.

Definition 1 *An n-dimensional differentiable manifold of class C^k with $1 \leq k \leq \infty$ is a pair $(M, \mathcal{F})$ consisting of an n-dimensional, second countable, locally Euclidean space M together with a differentiable structure of class C^k. Usually, the differentiable manifold $(M, \mathcal{F})$ is denoted as M.*

Definition 2 *A locally Euclidean space M of dimension n is a Hausdorff topological space M for which each point has a neighborhood homeomorphic to an open subset of the Euclidean space $\mathbb{R}^n$.*

Definition 3 *A differentiable structure $\mathcal{F}$ of class C^k $(1 \leq k \leq \infty)$ on a locally Euclidean space M is a collection of coordinate systems (also called an atlas of charts) $\{(U_\alpha, \varphi_\alpha) : \alpha \in A\}$ satisfying:*

(a) $\bigcup_{\alpha \in A} U_\alpha = M$.

(b) $\varphi_\alpha \circ \varphi_\beta^{-1}$ is C^k for every α, $\beta \in A$.

(c) The collection $\mathcal{F}$ is maximal with respect to (b); that is, if (U, φ) is a coordinate system such that $\varphi \circ \varphi_\alpha^{-1}$ and $\varphi_\alpha \circ \varphi^{-1}$ are differentiable for all $\alpha \in A$, then $(U, \varphi) \in \mathcal{F}$.