Multi-level optimization. Space mapping and manifold mapping

Echeverría, D.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 4

Manifold-Mapping Analysis

4.1 Introduction

A detailed study of the manifold-mapping concept was started in the last chapter. There, three basic algorithms and two main extensions were introduced. In this chapter we will prove that for the basic schemes, the stationary point of the iteration is a local minimizer of the fine model cost function. We will refer to this properly as the consistency of manifold mapping. Later, in Section 4.3, we will also give conditions for convergence. Analogous results for the extensions would require variations of the reasoning that are slightly more complex and need a substantial overload in notation. However, the ideas would be essentially the same as used in the proofs for the basic algorithms.

In general, manifold mapping yields linear convergence. The most important conditions needed for convergence are: (i) $f(X)$ and $c(X)$ being manifolds of class $C^2$, (ii) the models $f(x)$ and $c(x)$ showing a similar behavior in the neighborhood of the solution and (iii) the matrices $\Delta F$ and $\Delta C$ being sufficiently well-conditioned. In some cases it might be difficult to guarantee condition (iii). Trust-region manifold mapping was devised as a means to be used in those more difficult situations.

This chapter is structured as follows. In Section 4.2 consistency is proven for the basic manifold-mapping schemes. Section 4.3 is devoted to a detailed convergence study of those algorithms. Two simple problems illustrate in Section 4.4 some of the theoretical aspects dealt with along the chapter. They also show the convergence behavior of the trust-region manifold-mapping iteration. Chapters 5 and 6 represent the application of the manifold-mapping approach to design problems of practical relevance.
4.2 Consistency

In this section we will specify the conditions under which any of the basic algorithms presented in Section 3.2 is consistent, i.e., if a stationary point of the iteration is obtained, then this point coincides with the fine model optimum \( \mathbf{x}_t^* \).

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 7.** If \( \| f(\mathbf{x}) - \mathbf{y} \| \) is locally convex at \( \mathbf{x}_t^* \) then \( \| \mathbf{S} \mathbf{c}(\mathbf{x}) - \mathbf{y} \| \) is also locally convex at \( \mathbf{x}_t^* \).

**Assumption 8.** If \( \mathbf{x}_t^* \) is a local optimum of \( \| \mathbf{S} \mathbf{c}(\mathbf{x}) - \mathbf{y} \| \) then \( \mathbf{x}_t^* \) is the global optimum of \( \| \mathbf{S} \mathbf{c}(\mathbf{x}) - \mathbf{y} \| \).

In these assumptions \( \mathbf{S} \) is the affine mapping introduced in Section 3.2.1, i.e.,

\[
\mathbf{S} \mathbf{c}(\mathbf{x}) = f(\mathbf{x}_t^*) + \mathbf{S}(\mathbf{c}(\mathbf{x}) - \mathbf{c}(\mathbf{x}_t^*)),
\]

where

\[
\mathbf{S} = J_f(\mathbf{x}_t^*) J_s^T(\mathbf{x}_t^*). \tag{4.1}
\]

These are mild assumptions for the models used in practice. Assumption 7 specifies only a similar local behavior in the region of interest, i.e., in a neighborhood of the specifications \( \mathbf{y} \). Assumption 8 means that the surrogate optimization does not allow a spurious global optimum near the true minimizer \( \mathbf{x}_t^* \).

The manifold-mapping solution \( \mathbf{x}_{mm}^* \) was defined as

\[
\mathbf{x}_{mm}^* = \arg\min_{\mathbf{x} \in \mathcal{X}} \| \mathbf{S} \mathbf{c}(\mathbf{x}) - \mathbf{y} \|. \tag{4.3}
\]

The following two lemmas indicate that this manifold-mapping solution is a local minimizer of the fine model cost function \( \| f(\mathbf{x}) - \mathbf{y} \| \). This fact motivates the basic algorithms for constrained and unconstrained optimization presented in the previous chapter.

**Lemma 18.** Any (local) minimizer of the fine model cost function \( \| f(\mathbf{x}) - \mathbf{y} \| \) is a (local) minimizer of \( \| \mathbf{S} \mathbf{c}(\mathbf{x}) - \mathbf{y} \| \).

**Proof.** We denote a minimizer of the fine cost function by \( \mathbf{x}_t^* \). First we see that \( \mathbf{x}_t^* \) satisfies the KKT conditions associated with the minimization in (4.3). From (4.1) and (4.2) we have \( \mathbf{S} \mathbf{c}(\mathbf{x}_t^*) = f(\mathbf{x}_t^*) \) and \( J_{\mathbf{S} \mathbf{c}}(\mathbf{x}_t^*) = J_f(\mathbf{x}_t^*) \).

Thus, the first derivatives of \( F(\mathbf{x}) \) and of the surrogate cost function \( \| \mathbf{S} \mathbf{c}(\mathbf{x}) - \mathbf{y} \| \) coincide at \( \mathbf{x}_t^* \). Since the constraints are the same in both optimization problems\(^1\) and \( \mathbf{x}_t^* \) is a local optimum of \( F(\mathbf{x}) \) (i.e., the fine KKT conditions hold), we conclude that \( \mathbf{x}_t^* \) satisfies the surrogate KKT conditions. The fine model cost function is locally convex at \( \mathbf{x}_t^* \). Because of Assumption 7, the fine model optimum is also a local minimum of \( \| \mathbf{S} \mathbf{c}(\mathbf{x}) - \mathbf{y} \| \).

\(^1\)The constraint function \( k_f \) in this analysis is assumed to be easy to compute (see Section 3.2).
Lemma 19. $x_{mn}^*$ is a local minimizer of $\|f(x) - y\|$.

Proof. Use Lemma 18 and Assumption 8.

Remark 4. We cannot directly conclude from Lemma 18 that $x_{mn}^*$ is a minimizer of the fine cost function because the point $x_f^*$ in the lemma could be just a local minimizer of $\|S c(x) - y\|$ (the manifold mapping is introduced as a local model correction).

4.2.1 Original manifold-mapping algorithm

We consider the original manifold-mapping (OMM) algorithm introduced in Section 3.2.1. The following lemma will be very useful for proving that, if the OMM scheme converges, it yields the fine model optimum.

Lemma 20. Let $\tilde{x} \in X$ be the minimizer of a surrogate model problem

$$\tilde{x} = \arg\min_{x \in X} \|\tilde{S} c(x) - y\|,$$

with

$$\tilde{S} c(x) = f(\tilde{x}) + J_f(\tilde{x}) J^t_{s e} (c(x) - c(\tilde{x})), \quad (4.5)$$

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

Proof. Clearly, $\tilde{x}$ satisfies the KKT conditions associated with $\|\tilde{S} c(x) - y\|$, and because of (4.5) we have $\tilde{S} c(\tilde{x}) = f(\tilde{x})$ and $J_{s e}(\tilde{x}) = J_f(\tilde{x})$. Proceeding as in Lemma 18, we see that the point $\tilde{x}$ satisfies also the fine KKT conditions and, because $\|f(x) - y\|$ is locally convex at $\tilde{x}$, this point is a (local) minimizer of the fine cost function $F(x)$.

Remark 5. Note that from Lemma 20 it also follows that $\tilde{S} = S$.

We can replace the requirement of $\|f(x) - y\|$ being locally convex at $\tilde{x}$ from Lemma 20 by an assumption, very similar in nature to Assumption 7, and also likely to hold in practice:

Assumption 9. If $\|\tilde{S} c(x) - y\|$ is locally convex at $\tilde{x}$, then $\|f(x) - y\|$ is locally convex at $\tilde{x}$.

Remark 6. The manifold-mapping theory is generally stated in terms of a 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 converges to a fixed point $\tilde{x}$, this fixed point is a (local) minimizer of the fine cost function. Since we are
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 \( \mathcal{X} \) (locally) minimizes \( \| f(\mathbf{x}) - \mathbf{y} \| \). Since the Jacobians \( J_f(\mathbf{x}) \) and \( J_c(\mathbf{x}) \) have both rank \( n \), we expect that the matrices \( \Delta F \) and \( \Delta C \) 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 can be made in the algorithm to cope with this problem. One possibility, e.g., is the introduction of a stabilization parameter as in trust-region MM (see the scheme in Figure 3.5). Such changes will have no real influence in the computation of the final results. So, in order to prevent minor details in the analysis making the discussion much more complex, we introduce the following assumption.

**Assumption 10.** For \( k \) large enough, the \( m \times n \) matrices \( \Delta F \) and \( \Delta C \) have rank \( n \) and there are constants \( K_1, K_2 > 0 \) independent of \( k \) such that

\[
\left( \max_{i=0, \ldots, n-1} \| \mathbf{x}_{k+1-i} - \mathcal{X} \|^2 \right) \| \Delta F^\dagger \|_2^2 \leq K_1, \tag{4.6}
\]

\[
\left( \max_{i=0, \ldots, n-1} \| \mathbf{x}_{k+1-i} - \mathcal{X} \|^2 \right) \| \Delta C^\dagger \|_2^2 \leq K_2. \tag{4.7}
\]

We will see in Lemma 21 that Assumption 10 together with Assumption 11 below guarantee, that \( \Delta F \Delta C^\dagger \) converges to \( J_f(\mathcal{X}) J_c^\dagger(\mathcal{X}) \) and thus, that Lemma 20 can be applied.

**Assumption 11.** For \( k \) large enough the matrix \( \Delta X_{k+1} \) defined by

\[
\Delta X_{k+1} = [\mathbf{x}_{k+1} - \mathbf{x}_k, \mathbf{x}_{k+1} - \mathbf{x}_{k-1}, \ldots, \mathbf{x}_{k+1} - \mathbf{x}_{k-n+1}] \tag{4.8}
\]

is regular and there is a constant \( K_3 > 0 \) independent of \( k \) such that

\[
\left( \max_{i=0, \ldots, n-1} \| \mathbf{x}_{k+1-i} - \mathcal{X} \|^2 \right) \| \Delta X_{k+1}^{-1} \|^2 \leq K_3. \tag{4.9}
\]

**Remark 7.** Assumption 11 refers to the condition of the matrix \( \Delta X_{k+1} \) 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 by introducing a stabilization parameter in order to alleviate that, cf. the scheme in Figure 3.5. Assumption 10 is related to Assumption 11 and to the well-posedness of the inverse model operators (see Assumption 12 concerning the coarse model). Because \( \mathcal{C}(\mathcal{X}) \) is a differentiable manifold we have \( \Delta C \approx J_c(\mathcal{X}) \Delta X_{k+1} \) in a neighborhood of \( \mathcal{X} \), and thus, as it will become clear in the proof for Lemma 21, that \( \Delta C^\dagger \approx \Delta X_{k+1}^{-1} J_c^\dagger(\mathcal{X}) \). Thus, we can expect Assumption 10 to be satisfied if Assumption 11 holds and \( \| J_c^\dagger(\mathcal{X}) \|^2 \) is bounded. This last fact can be expressed as the inverse coarse model operator being Lipschitz in the region of interest. The inequality (4.6) is the analogous relation with respect to the fine model.
Lemma 21. Let the sequence of iterants \( x_h \) and operators \( S_{k+1} \) be defined by the original manifold-mapping algorithm (OMM). Then, under Assumptions 10 and 11 the operators \( S_{k+1} \) converge to \( J_r(\bar{x}) J^*_r(\bar{x}) \), where \( \bar{x} \) is the fixed point of the iteration.

Proof. By Assumption 10 and because \( f \) and \( c \) are differentiable, we have

\[
\Delta F = J_r(\bar{x}) \Delta X_{k+1} + M_f O(\max_{i=1, \ldots, n} \|x_{k+1-i} - \bar{x}\|^2),
\]

(4.10)

\[
\Delta C = J_c(\bar{x}) \Delta X_{k+1} + M_c O(\max_{i=1, \ldots, n} \|x_{k+1-i} - \bar{x}\|^2),
\]

(4.11)

where \( M_f \) and \( M_c \) are some \( m \times n \) matrices that depend on the smoothness of the manifolds \( f(X) \) and \( c(X) \) but not on \( k \). We can use a generalization of the Banach Lemma for the inverse of a perturbed matrix [45, Theorem 6.1-2] applied to \( \Delta C \) and conclude that

\[
\|\Delta C^\dagger - \Delta X_{k+1} J^*_r(\bar{x})\|_2 \leq 2 \|M_c\|_2 \max \{K_2, K_3\} \|J^*_r(\bar{x})\|_2^2.
\]

(4.12)

Because of (4.10) and the fact that the norm of \( \Delta C^\dagger - \Delta X_{k+1} J^*_r(\bar{x}) \) is bounded by a constant independent of \( k \), we obtain that \( S_{k+1} = \Delta F \Delta C^\dagger \) converges to \( J_r(\bar{x}) J^*_r(\bar{x}) \).

Using this result, we can apply Lemma 20 and conclude that if the OMM algorithm converges, then the fixed point of the iteration is a local minimizer of the finite model cost function \( \|f(x) - y\| \). This is summarized in the following theorem.

Theorem 22. Let \( \bar{x} \) be the fixed point of the original manifold-mapping (OMM, Fig. 3.2) iteration and let the finite model cost function \( F(x) = \|f(x) - y\| \) be locally convex at \( \bar{x} \), then under Assumptions 1, 2, 3, 5, 10 and 11, the point \( \bar{x} \) is a local minimizer of \( F(x) \).

Remark 8. The assumption of local convexity of \( F(x) \) can be replaced by Assumption 9 (model similarity) with \( S \) as in (4.1)-(4.2).

Remark 9. By similar arguments as those used here we can obtain similar consistency results for those OMM general algorithms based on mappings \( S \) satisfying \( S J_c(\bar{x}^2) = J_r(\bar{x}^2) \) (see Remark 2, Chapter 3).

Remark 10. We can make a similar consistency reasoning for the constraint manifold-mapping algorithm described in Section 3.3.2. We will not analyze it in detail because in essence the proofs are similar to those formulated above. The key point is again that for the stationary point, the KKT conditions related to the
surrogate optimization, associated with the mappings $S$ and $K$, reproduce those for the fine model optimization. I.e., we have that

$$\begin{align*}
S c(x^*_f) &= f(x^*_f), \\
J_{S c}(x^*_f) &= J_f(x^*_f), \\
K k_c(x^*_f) &= k_f(x^*_f), \\
J_{K k_c}(x^*_f) &= J_{k_f}(x^*_f).
\end{align*}$$

(4.13)  \hfill (4.14)  \hfill (4.15)  \hfill (4.16)

4.2.2 Manifold-Mapping algorithm

We will consider now the manifold-mapping (MM) algorithm introduced in Section 3.2.2. We can prove a theorem similar to Theorem 22 if Assumption 5 is replaced by Assumption 6.

**Theorem 23.** Let $\bar{x}$ be the fixed point of the manifold-mapping (MM, Fig. 3.3) iteration, and let the fine model cost function $F(x) = \|f(x) - y\|$ be locally convex at $\bar{x}$, then under Assumptions 1, 2, 3, 6, 10 and 11, the point $\bar{x}$ is a local minimizer of $F(x)$.

**Proof.** Proceeding as in Lemma 21 we can see that the sequence of operators $T_{k+1} = S_{k+1}^{\dagger}$ converges to $(J_f(\bar{x}) J_c^\dagger(\bar{x}))^{\dagger} = J_c(\bar{x}) J_f^\dagger(\bar{x})$. Also in the limit

$$\bar{x} = \arg\min_{x \in X} \| c(x) - c(\bar{x}) + (J_f(\bar{x}) J_c^\dagger(\bar{x}))^{\dagger} (f(\bar{x}) - y) \|. \quad (4.17)$$

Since $c(X)$ is a manifold of class $C^2$, (4.17) is equivalent to

$$\bar{x} = \arg\min_{x \in X} \| J_c(\bar{x}) (x - \bar{x}) + (J_f(\bar{x}) J_c^\dagger(\bar{x}))^{\dagger} (f(\bar{x}) - y) \|. \quad (4.18)$$

Because $J_c(\bar{x})$ and $J_f(\bar{x})$ are full rank, we can write the former equality as

$$\bar{x} = \arg\min_{x \in X} \| J_f(\bar{x}) J_c^\dagger(\bar{x}) J_c(\bar{x}) (x - \bar{x}) + f(\bar{x}) - y \|, \quad (4.19)$$

and with the same reasoning as in (4.17)-(4.18), equation (4.19) is equivalent to

$$\bar{x} = \arg\min_{x \in X} \| J_f(\bar{x}) J_c^\dagger(\bar{x}) (c(x) - c(\bar{x})) + f(\bar{x}) - y \|. \quad (4.20)$$

Then Lemma 20 is applied to show that $\bar{x}$ is a local minimizer of $\|f(x) - y\|$.

As a consequence of this theorem and by just rewriting (4.17) with $\bar{x}$ a local minimizer of the fine cost function $F(x)$, the following interesting property for the fine model optimum can be formulated.

**Corollary 1.** The fixed point of the manifold-mapping (MM) iteration $x^*_f$ satisfies

$$x^*_f = \arg\min_{x \in X} \| c(x) - c(x^*_f) + \nabla f(x^*_f) - y \|. \quad (4.21)$$
Remark 11. Proceeding analogously we can extend these results to those MM algorithms based on mappings $\mathcal{S}$ satisfying $J_1^b(x^*_\ell) \mathcal{S} = J_1^b(x^*_\ell)$ (see Remark 3, Chapter 3).

Remark 12. If in the MM trust-region strategy presented in Section 3.3.1 we assume that $\lambda_k \to 0$, then consistency can be obtained with similar arguments as those in this section. We will not go further and state the strict conditions that guarantee $\lambda_k \to 0$. Roughly, $\lambda_k$ will vanish for large $k$ if the cost function has a regular behavior and it can locally be approximated by linearization. In the examples described in Section 4.4.3, it will be checked that the value of $\lambda_k$ tends to zero in a neighborhood of the stationary point.

### 4.2.3 Generalized manifold-mapping algorithm

Under convergence of the GMM scheme (Section 3.2.3), we can also see that the fixed point for the iteration is again the accurate solution of the optimization problem, $x^*_\ell$. The proof for the following theorem is completely analogous to those given for Theorems 22 and 23.

**Theorem 24.** Let $x$ be the fixed point of the generalized manifold-mapping algorithms (GMM, Fig. 3.4) and let the fine model cost function $F(x) = \|f(x) - y\|$ be locally convex at $x$, then under Assumption 1, 2, 3, 6, 10 and 11, the point $x$ is a local minimizer of $F(x)$.

### 4.3 Convergence

In this section we will give conditions for convergence, i.e., for assuring the existence of a stationary point. The consistency theorems in Section 4.2 guarantee that this point will be the right optimum. We will first present conditions for convergence of the manifold-mapping (MM) algorithm. Then we will show that the original manifold-mapping (OMM) iteration is asymptotically equivalent to the MM iteration and thus, a convergence theorem can be stated with a proof analogous to that for MM.

Convergence depends strongly on the well-posedness of the generalized inverse of the coarse model operator. The generalized inverse coarse model operator $c^l : Y \subset \mathbb{R}^m \to X \subset \mathbb{R}^n$ is defined as

$$c^l \tilde{y} = \arg\min_{x \in X} \|c(x) - \tilde{y}\|.$$  \hspace{1cm} (4.22)

This operator is well defined because of Assumption 3. 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 generalized inverse coarse model operator, $c^l$, is Lipschitz with a Lipschitz constant bounded by $L_{c^l}$. 
4.3.1 Manifold-Mapping algorithm

We start the convergence analysis by rewriting (4.21) and the expression for iterant \( x_{k+1} \) in the MM algorithm in terms of the inverse coarse model operator:

\[
x_k^* = c^\dagger (c(x_k^*) - S_k^\dagger (f(x_k^*) - y)),
\]

\[
x_{k+1} = c^\dagger (c(x_k) - S_k^\dagger (f(x_k) - y)).
\]

Subtracting (4.23) from (4.24) and using Assumption 12, we obtain

\[
\|x_{k+1} - x_k^*\| \leq L c_1 \|c(x_k) - S_k^\dagger (f(x_k) - y) - c(x_k^*) + S_k^\dagger (f(x_k^*) - y)\|. \tag{4.25}
\]

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

\[
c(x_k) - c(x_k^*) - S_k^\dagger (f(x_k) - f(x_k^*)) = (J_c(x_k) - S_k^\dagger J_f(x_k^*)) (x_k - x_k^*) + (S_k^\dagger - S_k^\dagger_0) (f(x_k^*) - y) + O(\|x_k - x_k^*\|^2). \tag{4.26}
\]

We now analyze the term \((S_k^\dagger - S_k^\dagger_0)\) in (4.26). We proceed as in the proof of Lemma 21, 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 21 (Assumptions 10 and 11), we will be able to establish a relation between \(\|S_k^\dagger - S_k^\dagger_0\|\) and \(\|x_k^* - x_k\|\).

**Assumption 13.** For \(k\) large enough, there are constants \(K_4, K_5 > 0\) independent of \(k\) such that

\[
\left( \max_{i=1, \ldots, n} \|x_{k+1-i} - x_{k+1}\|^2 \right) \|\Delta F\|^2_2 \leq K_4, \tag{4.27}
\]

\[
\left( \max_{i=1, \ldots, n} \|x_{k+1-i} - x_{k+1}\|^2 \right) \|\Delta C\|^2_2 \leq K_5. \tag{4.28}
\]

**Assumption 14.** For \(k\) large enough, there is a constant \(K_6 > 0\) independent of \(k\) such that

\[
\left( \max_{i=1, \ldots, n} \|x_{k+1-i} - x_{k+1}\|^2 \right) \|\Delta X^{-1}_k\|^2_2 \leq K_6, \tag{4.29}
\]

where \(\Delta X_{k+1}\) is the square matrix defined in Assumption 11.

**Remark 13.** We recognize in \(\left( \max_{i=1, \ldots, n} \|x_{k+1-i} - x_{k+1}\|^2 \right)^{1/2}\) a matrix-norm for \(\Delta X_{k+1}\). Thus, Assumption 14 can be stated in terms of \(\kappa(\Delta X_{k+1})\) the condition number of \(\Delta X_{k+1}\), 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...
obtained, the algorithm can be slightly modified to cope with that issue, with no significant influence in the final result. Assumption 13 can be related to the condition number of \( \Delta X_{k+1} \) and the well-posedness of the inverse model operators in the region of interest.

**Lemma 25.** Under Assumptions 1, 2, 3, 6, 9, 10, 11, 13 and 14,

\[
\| S_k^T - J_c(x_k) J_f^T(x_k) \| \leq M_1 \max_{i=0, \ldots, n-1} \| x_{k-i} - x_f^* \|, \tag{4.30}
\]

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

**Proof.** As in Lemma 21 we can write

\[
\Delta F = J_f(x_{k+1}) \Delta X_{k+1} + M_f O( \max_{i=1, \ldots, n} \| x_{k+1-i} - x_{k+1} \|^2 ), \tag{4.31}
\]

\[
\Delta C = J_c(x_{k+1}) \Delta X_{k+1} + M_c O( \max_{i=1, \ldots, n} \| x_{k+1-i} - x_{k+1} \|^2 ), \tag{4.32}
\]

where \( M_f \) and \( M_c \) are some \( m \times n \) matrices that depend on the smoothness of the manifolds \( f(X) \) and \( c(X) \) but not on \( k \). Again with [45, Theorem 6.1-2] and by Assumptions 13 and 14 we can conclude that \( S_k^T - J_c(x_k) J_f^T(x_k) \) is bounded by a constant multiplied by \( \max_{i=1, \ldots, n} \| x_{k+1-i} - x_k \| \). Clearly, this is equivalent to \( S_k^T - J_c(x_k) J_f^T(x_k) \) being bounded by \( M_1 \max_{i=0, \ldots, n-1} \| x_{k-i} - x_f^* \| \) where \( M_1 > 0 \) is a constant that depends on the smoothness of \( f(X) \) and \( c(X) \) but not on \( k \).

**Remark 14.** The constant \( M_1 \) depends on the smoothness of \( f(X) \) and \( c(X) \). (If both manifolds are linear in the neighborhood of the solution then \( M_1 = 0 \).)

**Lemma 26.** Under Assumptions 1, 2, 3, 6, 9, 10, 11, 13 and 14,

\[
\| \bar{S}_k^T - J_c(x_k) J_f^T(x_k) \| \leq M_2 \| x_k - x_f^* \|, \tag{4.33}
\]

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

**Proof.** We can write

\[
\bar{S}_k^T - J_c(x_k) J_f^T(x_k) = J_c(x_f^*) J_f^T(x_f^*) - J_c(x_k) J_f^T(x_k) \tag{4.34}
\]

\[
= J_c(x_f^*) J_f^T(x_f^*) - J_c(x_k) J_f^T(x_f^*) + J_c(x_k) J_f^T(x_f^*) - J_c(x_k) J_f^T(x_k) \]

\[
= (J_c(x_f^*) - J_c(x_k)) J_f^T(x_f^*) + J_c(x_k) (J_f^T(x_f^*) - J_f^T(x_k)).
\]
Since \( f(X) \) and \( c(X) \) are manifolds of class \( C^2 \), we can bound 
\[
\| S^\dag - J_c(x_k) J_f^\dag(x_k) \| \quad \text{by} \quad M_2 \| x_k - x^*_f \|, \quad \text{where} \quad M_2 > 0 \quad \text{is a constant that depends on the smoothness of the two manifolds but not on} \ k.
\]

**Corollary 2.** Under Assumptions 1, 2, 3, 6, 9, 10, 11, 13 and 14,
\[
\| S^\dag - S_k^\dag \| \leq M \max_{i=0, \ldots, n-1} \| x_{k-i} - x^*_f \|, \tag{4.35}
\]
where \( M > 0 \) is a constant that depends on the smoothness of the manifolds \( f(X) \) and \( c(X) \) but not on \( k \).

**Proof.** We apply Lemmas 25 and 26 and set \( M = \max (M_1, M_2) \).

Now, combining (4.25) and (4.26) we get
\[
\| x_{k+1} - x^*_f \| \leq L_{c1} \| J_c(x^*_f) - S_k^\dag J_f(x^*_f) \| \| x_k - x^*_f \| + \\
+ \| S^\dag - S_k^\dag \| \| f(x^*_f) - y \| + O(\| x_k - x^*_f \|^2). \tag{4.36}
\]
Due to (4.35) we can finally write
\[
\| x_{k+1} - x^*_f \| \leq L_{c1} (\| J_c(x^*_f) - S_k^\dag J_f(x^*_f) \| + \\
+ M \max_{i=0, \ldots, n-1} \| x_{k-i} - x^*_f \| + \\
+ O(\| x_k - x^*_f \|^2). \tag{4.37}
\]
We formulate this result in the following theorem.

**Theorem 27.** Under Assumptions 1, 2, 3, 6, 9, 10, 11, 12, 13 and 14, and the condition
\[
L_{c1} (\| J_c(x^*_f) - S_k^\dag J_f(x^*_f) \| + M \| f(x^*_f) - y \|) < 1 \quad \text{for} \quad k \geq k_0, \tag{4.38}
\]
where \( M > 0 \) is a constant that depends on the smoothness of the manifolds \( f(X) \) and \( c(X) \) but not on \( k \), the manifold-mapping algorithm (MM, Fig. 3.3) yields (linear) convergence to \( x^*_f \).

**Corollary 3.** If, in addition to the assumptions for Theorem 27, we have \( f(x^*_f) = y \) (i.e., a reachable design), then the convergence of the MM algorithm is superlinear.

**Corollary 4.** If in addition to the assumptions for Theorem 27 we have \( S_k = S = J_f(x^*_f) J_f^\dag(x^*_f) \) for every \( k \geq k_0 \), then the convergence of the MM algorithm is quadratic.
4.3.2 Original manifold-mapping algorithm

OMM convergence results will be obtained from those derived for MM in the previous subsection. For that purpose we will just see that —under convergence—the original (OMM) and the manifold-mapping (MM) iterations are asymptotically equivalent.

**Lemma 28.** If the iteration in the original manifold-mapping (OMM) algorithm converges, for a large enough \( k \) we find for its iterant \( x_{k+1} \)

\[
x_{k+1} = \arg\min_{x \in X} \|c(x) - c(x_k) + S_k^k (f(x_k) - y) + O(\|x - x_k\|^2)\|.
\]

(4.39)

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

\[
x_{k+1} = \arg\min_{x \in X} \|S_k c(x) - S_k c(x_k) + f(x_k) - y\| \quad \text{(4.40)}
\]

\[
= \arg\min_{x \in X} \|S_k J_c(x_k) (x - x_k) + f(x_k) - y + S_k O(\|x - x_k\|^2)\|
\]

and for the last equality we remember that \( S_k \) converges to \( S \) (Lemma 21).

The iterant \( x_{k+1} \) can be expressed as

\[
x_{k+1} = \arg\min_{x \in X} \|S_k \Delta C \Delta X_k^{-1} (x - x_k) + f(x_k) - y + O(\|x - x_k\|^2)\| \quad \text{(4.41)}
\]

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

\[
x_{k+1} = \arg\min_{x \in X} \|\Delta F \Delta X_k^{-1} (x - x_k) + f(x_k) - y + O(\|x - x_k\|^2)\| \quad \text{(4.42)}
\]

Further, since \( \Delta F \) and \( \Delta C \) are full-rank and \( \Delta X_k \) is regular

\[
x_{k+1} = \arg\min_{x \in X} \|\Delta C \Delta X_k^{-1} (x - x_k) + f(x_k) - y + O(\|x - x_k\|^2)\|.
\]

(4.43)

The lemma follows immediately from this last equation.

**Lemma 29.** If the iteration in the manifold-mapping (MM) algorithm converges, for a large enough \( k \) we find for its iterant \( x_{k+1} \)

\[
x_{k+1} = \arg\min_{x \in X} \|S_k (c(x) - c(x_k)) + f(x_k) - y + O(\|x - x_k\|^2)\|.
\]

(4.44)

**Proof.** Analogous to that for Lemma 28.
Corollary 5. Under convergence, the OMM algorithm and the MM algorithm are asymptotically equivalent.

Remark 15. Due to Corollary 5, Theorem 27 and Corollaries 3 and 4 are also valid for the original manifold-mapping algorithm (OMM, Fig. 3.2) when Assumption 6 is replaced by Assumption 5.

Remark 16. By similar arguments as those given for mappings based on \( S = J_f(x_f)J_i^t(x_i) \), analogous convergence results can be obtained for the other versions of the algorithms introduced in Remarks 2 and 3 in Chapter 3.

Remark 17. In the case of the constrained version of the MM scheme we can write the expression for iterant \( x_{k+1} \) as

\[
x_{k+1} = c_k^t (c(x_k) - S_k^t (f(x_k) - y)),
\]

where \( c_k^t \) denotes the general inverse operator of the coarse model with a constraint function given by \( K_k \circ k_c \). We can assume that for \( k \) large enough and in a neighborhood of the problem solution

\[
|c^t(y_1) - c_k^t(y_2)| \leq L_{c1} \|y_1 - y_2\|,
\]

with \( L_{c1} \) a constant independent of \( k \). A similar expression to (4.25) can be obtained and then, the rest of the proof would be essentially the same as that for the MM algorithm. The additional condition (4.46) measures again the similarity between the fine and the coarse constraints and in many practical cases is a reasonable assumption.

4.3.3 Generalized manifold-mapping algorithm

The proof for generalized manifold mapping (GMM) is analogous to that for MM. The two possible GMM algorithms, corresponding to OMM and MM, are also asymptotically equivalent. We formulate this in the following theorem.

Theorem 30. Under Assumptions 1, 2, 3, 6, 9, 10, 11, 12, 13 and 14, and the condition

\[
L_{c1} (\|J_c(x_f^i) - S_k^t J_f(x_i^*^t)\| + \bar{M} \|f(x_f^i) - y\|) < 1 \quad \text{for} \quad k \geq k_0, \tag{4.47}
\]

where \( \bar{M} > 0 \) is a constant that depends on the smoothness of the manifolds \( f(X) \) and \( c(X) \) but not on \( k \), the generalized manifold-mapping algorithms (GMM, Fig. 3.4) yield (linear) convergence to \( x_i^* \).

Remark 18. In the case \( S_{k+1} = J_f(x_{k+1})J_i^t(x_{k+1}) \) an analog to Lemma 25 can be trivially stated. The constant \( \bar{M} \) introduced in the theorem for GMM could be in some situations smaller than \( M \), 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 Sections 4.4.1 and 4.4.2).
Figure 4.1: Left: The sets \( f(X) \) and \( c(X) \), specifications \( y \) and \( f(x_f^*) \) and \( c(x_c^*) \) for the example in Section 4.4.1. Right: the fine and coarse cost functions, \( \|f(x) - y\|_2 \) and \( \|c(x) - y\|_2 \), respectively, for the same example.

4.4 Examples

In Section 4.4.1 and 4.4.2 we illustrate the most significant results obtained in this chapter by means of two simple optimization problems. We also include a third section showing the convergence behavior of the TRMM scheme for different situations of the second of these two examples.

4.4.1 Linear and superlinear convergence

We illustrate the convergence Theorems 27 and 30 with this example. The fine model is defined over \( X = [-1, 1] \) by

\[
    f(x) = [x, x^2].
\]

The set \( f(X) \subset \mathbb{R}^2 \) is part of a parabola and we want to find the point in that set closest in Euclidean norm to the specifications \( y = [3/4, 0] \). The coarse model is defined over \( Z = X \) and is the linear \( c(x) = [x, (1 + x)/2] \). Figure 4.1 shows a representation of the problem and the fine and coarse cost functions, \( \|f(x) - y\|_2 \) and \( \|c(x) - y\|_2 \), respectively. The corresponding optima are \( x_f^* = 0.5 \) and \( x_c^* = 0.4 \).

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 \( f(X) \) and \( c(X) \) are smooth, we expect a reasonable small constant \( M \) in (4.38) for Theorem 27. The design is not reachable and convergence is linear for the manifold-mapping (MM) and the generalized manifold-mapping (GMM) iterations (see Figure 4.2). 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 twelve iterations for getting $|x_h - x_0^*|$ smaller than $10^{-6}$. The GMM scheme, using the exact Jacobian, yields the same accuracy in ten iterations. In Figure 4.2 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 $\Delta F$ and $\Delta C$ 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 4.2 we also observe two cases of superlinear convergence for MM (cf., Corollaries 3 and 4). If we apply MM with $S_k = S$ (denoted by MM* in Figure 4.2) we obtain a solution with the same accuracy of $10^{-6}$ 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 $y = [1/2, 1/4]$ (yielding again $x_0^* = 0.5$), the solution is obtained with an accuracy of $10^{-6}$ in five iterations. The same superlinear convergence is observed for the GMM algorithm.
4.4.2 Different types of linear convergence

With this example we show that different choices for the fine model Jacobian estimation at the $k$-th iteration, $\hat{J}_f(x_k)$, yield distinct convergence histories: the better $\hat{J}_f(x_k)$ approximates $J_f(x_k)$, the smaller the constant associated with the linear convergence. The example has been introduced in Section 2.6 (see (3) in Table 2.1) as a least squares best approximation of the data vector $y = [0, -0.4, 0.1]$ by the fine model

$$f(x) = f(x_1, x_2) = [x_1 (x_2 - 1)^2, x_1, x_1 (x_2 + 1)^2]$$

(4.49)

defined over $X = \mathbb{R}^2$. The design is not reachable. The coarse model

$$c(x) = c(x_1, x_2) = [-x_1 + x_2, x_2, x_1 + x_2]$$

(4.50)

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. In all these schemes the matrix $A$ in Remark 3 (Chapter 3) is the null matrix. 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 $x^*_f = [-0.101, -0.141]$. The convergence history for the three methods is shown in Figure 4.3. We clearly observe that the convergence is linear in all cases and that the constant $M$ in the convergence theorems (indicating the 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 an elevated number of iterations needed, when compared with the previous example. In practice, fine and coarse models are much more similar and, hence, convergence is generally achieved with much less function evaluations.

### 4.4.3 TRMM convergence

As an example of the convergence behavior of the TRMM algorithm we consider all the different cases for the problem discussed in the previous subsection and introduced in Chapter 2. Different specifications $y$ created essentially distinct situations: (1) a reachable design, (2) a perfect mapping but a non-reachable design, (3) a non-perfect mapping, and (4) a design with multiple minima. In the four cases the fine and the coarse model are as in (4.49) and (4.50), respectively. The specifications $y$ are indicated in Table 2.1.

The values for the parameters in the TRMM scheme are as they appear in Figure 3.5: $\alpha = 1 + \tau > 1$, $\tau = 10^{-10}$, $\beta = 1/10$, $\lambda_{TR} = 1$ and $R_A = r_A = 2$. We set the maximum number of iterations to $N_{MAX} = 100$, and the damping parameter is not used ($\delta = 0$), except for the last example. Convergence results for the TRMM scheme are compared to those for the MM algorithm (no trust-region strategy). In all cases, the matrix $A$ introduced in Remark 3 (Chapter 3) is now the identity. As commented in Section 4.4.2, the coarse model is not specially adapted to the fine one, and particularly the last example shows what adverse effects can be expected in a case like this one.

1. **Reachable design.** The TRMM scheme takes 12 iterations (13 $f$-evaluations) until $\|x_k - x^*_f\|_2$ is smaller than $10^{-6}$. Figure 4.4 represents the history of $\|x_k - x^*_f\|_2$ (left) and of $\lambda_k$ (right). In contrast with all other examples shown, for this simple problem the trust-region strategy has no positive effect. Without the trust-region strategy a solution with similar quality is also obtained after 12 iterations.

2. **Perfect mapping.** The TRMM scheme takes 13 iterations (16 $f$-evaluations) until $\|x_k - x^*_f\|_2$ is smaller than $10^{-6}$. The convergence history is shown in Figure 4.5. The behavior is similar to that for the reachable design: first the parameter $\lambda$ does not decrease because of the initial dissimilarity between the fine and the coarse model and then, the reduction is monotonous. The matrices $\Delta F$ and $\Delta C$ remain well-conditioned. Without the trust-region strategy a solution with similar quality is obtained after 48 iterations.
3. Non-perfect mapping. The TRMM scheme takes 21 iterations (29 f-evaluations) until $\|x_k - x^*_f\|_2$ is smaller than $10^{-6}$. The convergence history is shown in Figure 4.6. In this case $\Delta F$ becomes very ill-conditioned and during the iteration process $\lambda$ has to be significantly increased several times. Without the trust-region strategy a solution with similar quality is obtained after 63 iterations\(^2\). If we compare Figure 4.6 with Figure 4.3 we can appreciate that the TRMM algorithm yields a much more stable convergence history than the MM scheme.

\(^2\)If the matrix $A$ in the MM scheme is the null-matrix, the number of iterations is larger than 300 (see Figure 4.3).
4. Multiple minima. In this example the fine model optimization has two local minima. The behavior of the fine model near the global minimum is much different from that of the coarse model, which is more similar to the (other) local minimum. In this example TRMM without damping ($\delta = 0$) takes 53 iterations (66 $f$-evaluations) until $||x_k - x^*_f||_2$ is smaller than $10^{-6}$. The stationary point obtained $x^*_f$ is the local minimum (not the global one). The convergence history is shown in Figure 4.7.

Using the damping parameter $\delta$ we can force the algorithm to find the global minimum. By taking $\delta > 0$ the method selects a path with smaller steps. The effect is stronger for larger $\delta$. The iteration process arrives in the attraction area of the global minimum, but because of the large discrepancy between the behavior of $||c(x) - y||$ and $||f(x) - y||$ in that region, the convergence is relatively slow. TRMM with damping ($\delta = 100$) takes 89 iterations (99 $f$-evaluations) until $||x_k - x^*_f||_2$ is smaller than $10^{-6}$. Now, $x^*_f$ represents the global optimum and the convergence history is shown in Figure 4.8. Without the trust-region strategy the global optimum is approximated with a similar quality within 95 iterations (there is no solid theoretical reasoning that explains this convergence to the global minimum for this particular example).

4.5 Conclusions

With this chapter we finish the detailed study of manifold mapping. In Chapter 3 we introduced a number of possible algorithmic variants: original manifold mapping (OMM), manifold mapping (MM) and generalized manifold map-
Figure 4.7: Multiple minima, damping parameter \( \delta = 0 \). Left: Convergence history for \( \| \mathbf{x}_k - \mathbf{x}_\star \|_2 \) for the trust-region manifold-mapping (TRMM) algorithm. Right: history of \( \lambda_k \) for the same algorithm.

Figure 4.8: Multiple minima, damping parameter \( \delta = 100 \). Left: Convergence history for \( \| \mathbf{x}_k - \mathbf{x}_\star \|_2 \) for the trust-region manifold-mapping (TRMM) algorithm. Right: history of \( \lambda_k \) for the same algorithm.

ping (GMM). Moreover, two scheme extensions were considered: trust region manifold-mapping (TRMM) and constrained optimization with manifold mapping. In this chapter we have seen that the stationary points of all these schemes are local minimizers of the fine model cost function. Conditions for convergence have also been formulated for OMM, MM and GMM. By simple examples some of the theoretical aspects dealt with in the chapter have eventually been illustrated. In the following two chapters, the basic algorithms and extensions will be applied to a number of design problems of practical relevance.