On Sequential Multiscale Inversion
and Data Assimilation

AAMIR NADEEM[2], ROLAND POTTHAST[1,3], AND ANDREAS RHODIN[3]

[1] *Department of Mathematics and Statistics, University of Reading, Whiteknights, PO Box 220, Berkshire RG6 6AX, United Kingdom*
[2] *University of Göttingen, Institute for Numerical and Applied Mathematics, Lotzestr. 16-18, 37083 Göttingen, Germany*
[3] *Deutscher Wetterdienst, Frankfurter Str. 135, 63067 Offenbach, Germany*

February 10, 2017

Abstract

Multiscale approaches are very popular for example for solving partial differential equations and in many applied fields dealing with phenomena which take place on different levels of detail. The broad idea of a multiscale approach is to decompose your problem into different scales or levels and to use these decompositions either for constructing appropriate approximations or to solve smaller problems on each of these levels, leading to increased stability or increased efficiency. The idea of *sequential multiscale* is to first solve the problem in a *large-scale subspace* and then successively move to *finer scale* spaces.

Our goal is to analyse the *sequential multiscale approach* applied to an inversion or state estimation problem. We work in a generic setup given by a Hilbert space environment. We work out the analysis both for an *unregularized* and a *regularized* sequential multiscale inversion. In general the sequential multiscale approach is not equivalent to a full solution, but we show that under appropriate assumptions we obtain convergence of an *iterative* sequential multiscale version of the method. For the *regularized case* we develop a strategy to appropriately adapt the regularization when an iterative approach is taken.

We demonstrate the validity of the iterative sequential multiscale approach by testing the method on an integral equation as it appears for atmospheric temperature retrieval from infrared satellite radiances.

1   Introduction

Multiscale methods are very popular for modeling natural phenomena, they have attracted a lot of attention over recent years in the scientific literature, both for *direct* and *inverse*

1

problems, compare for example [1, 2, 3, 4, 28, 34] for multiscale methods in simulation and [5, 12, 14, 16, 31, 38] for their use in inverse problems. For solving an inverse problem, the *singular value decomposition* (see [13, 18, 22, 30]) provides a traditional approach to study the different scales of the problem. Classical regularization theory can be viewed as a damping of higher modes of the inverse of the operator which maps the unknown states onto the observations. In the area of *data assimilation*, different versions of the multi-scale approach has been tested over the past years, for example in [23, 24, 26, 27, 32, 33, 36, 37, 39, 40, 41]. For many systems in the geosciences, we have something like a natural low-scale approach by models with some course resolution, which are complemented by high-resolution models which can be run locally on nested regions. Over the past 15 years, *ensemble methods* have been highly successful in meteorological or geophysical data assimilation. Here, different scales are implicitly chosen by *localization techniques*, see for example [6, 8, 10, 20, 25, 29, 35]. The relationship of multiscale methods to other iterative methods such as Gauss-Newton or Landweber methods, compare [21, 30] or homotopy perturbation methods, compare [11, 19, 7], which also work with a changing regularization parameter, but for generating a homotopy and not for purposes of equivalence, will be part of future work.

In this work, our goal is *first* a generic *theoretical understanding* of the role and effects which appear in multiscale approaches to *inverse* and *data assimilation* methods. *Second*, our goal is to analyse and test *iterative sequential multiscale*, with a numerical example from the field of integral equations as they appear in *atmospheric temperature retrieval* from *infrared radiances*. We will restrict our attention to *one inversion or data assimilation step*. In this case, the generic situation is described by an equation of the form

$$H\varphi = f \tag{1.1}$$

where $\varphi \in X$ is in some state space X and $f \in Y$ is the observation vector in some observation space Y. Here, we assume that X and Y are Hilbert spaces, usually for simplicity we restrict our attention to finite dimensional spaces, i.e. to the case where $X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$ with $n, m \in \mathbb{N}$, where X and Y each are equipped with some norm. Most of our arguments, though, will carry over to the general infinite-dimensional case. We assume that some covariance matrix $B \in \mathbb{R}^{n \times n}$ describes the knowledge about the relationships between the different components φ_j, $j = 1, ..., n$ of $\varphi \in X$, i.e.

$$\overline{\varphi} := \mathbb{E}\left\{\varphi\right\}, \qquad B = \mathbb{E}\left\{(\varphi - \overline{\varphi})(\varphi - \overline{\varphi})^T\right\}. \tag{1.2}$$

Let $R \in \mathbb{R}^{m \times m}$ be the covariance matrix of the observation error on $\mathbb{R}^m$.

We will study the situation where the multiscale approach obtains its input from some given multiscale basis of X, which might be *orthogonal* or *non-orthogonal*. This includes *classical multiscale basis sets* which are obtained from generating radial basis functions, but it also includes a multiscale basis coming from two ensembles, one covering the low scale, the other the finer scale. We will study the generic approximation properties of such an approach, showing that a *sequential multiscale decomposition* leads to typical

approximation errors when the basis is non-orthogonal. We then provide the analysis for the *iteration* of sequential multiscale, both for the case where the reconstruction is *regularized* as well as for the *unregularized* case.

For the *unregularized case* and *non-orthogonal case* we will show that in general the sequential multiscale approach cannot provide the same solution as if we would solve on all scales simultaneously. The error is given by the successive projections of the solution on the spaces perpendicular to the multiscale spaces under consideration. A generic two-dimensional example is given and visualized in Figure 1.

We then study *iterated sequential multiscale*, i.e. we repeat the squential multiscale method starting with the result of each previous iteration. For the *unregularized* case we obtain *linear* convergence of these iterations in the spirit of *alternating projections* [15] where the convergence speed is determined by $\cos(\alpha)$ where α is the smallest angle between the subspaces under consideration.

Inverse problems or data assimilation usually involves *regularization*, which causes further complications to the above convergence analysis. Carrying out the analysis step several times – either for the sequential steps of the multiscale method or for the iteration of the sequence of multiscale steps – the role of the *regularizing* or *background term*, respectively, needs to be taken carefully into account.

We first provide a generic example of iterated regularization which shows that ℓ steps of a sequential regularization with regularization parameter β are identical to one-step regularization with regularization parameter α, when the regularization parameters satisfy the relationship

$$\sqrt[\ell]{(1-\alpha)} = 1 - \beta. \tag{1.3}$$

The general theory is worked out for the *Kalman filter*, proving that the Kalman gain matrix $\tilde{K}$ for the iterations and the Kalman gain matrix $K = BH'(R + HBH')^{-1}$ for the *one-step* analysis need to satisfy the equation

$$(I - \tilde{K}H)^{\ell} = I - KH, \tag{1.4}$$

where H is our *observation operator* H under consideration. The solvability of (1.4) is shown by studying the operator in a particular weighted Hilbert space, where KH becomes self-adjoint.

After describing our general setup in Section 2 we will start with the solution of the inverse problem in a subspace and its reformulation into a basis transformation of the covariance matrix and the states under consideration.

We then study the question of sequential scale splitting in Section 3. We first clarify that in general scale splitting does not lead to equivalence of the sequential approach to the full problem when non-orthogonal spaces are employed. For orthogonal decompositions the problem can be equivalently split into multiscale spaces. The problem of aliasing when data on different scales is employed is discussed in Section 3.2. Here, aliasing means that

the high-frequency data may lead to large errors when the inversion in a low-frequency space is carried out.

We provide a numerical example in our Section 4, which demonstrates the feasibility of the multiscale inversion. We apply iterative sequential multiscale to an integral equation which is often seen as a generic example for atmospheric temperature or humidity retrieval from satellite radiances [30].

2 A Generic Multiscale Approach

The classical Tikhonov regularization for the solution of (1.1) is given by

$$\varphi_\alpha := (\alpha I + H^*H)^{-1}H^*f, \tag{2.1}$$

where $\alpha > 0$ is known as *regularization parameter*. Here, in a functional analytic framework, we will employ the letter φ to denote the states, the observations will be denoted by f.

Either starting with Tikhonov regularization in X equipped with a norm weighted by B^{-1} or using a Bayesian approach, it is well known (see e.g. [9], [17] or [30]), that the regularized solution of (1.1) can be obtained by a minimization of

$$J(\varphi) := ||\varphi - \varphi^{(b)}||^2_{B^{-1}} + ||f - H\varphi||^2_{R^{-1}}, \tag{2.2}$$

where the minimizer is given by

$$\varphi^{(a)} := \varphi^{(b)} + BH^T(R + HBH^T)^{-1}(f - H\varphi^{(b)}). \tag{2.3}$$

For ensemble methods, often the covariance matrix B is given by the standard covariance estimator

$$B = QQ^T \tag{2.4}$$

with Q defined by

$$Q = \frac{1}{\sqrt{L-1}}(\varphi^{(1)} - \mu, ..., \varphi^{(L)} - \mu), \tag{2.5}$$

where $\varphi^{(\xi)}$ denotes an ensemble of states under consideration and μ is its mean

$$\mu := \frac{1}{L}\sum_{\xi=1}^{L}\varphi^{(\xi)}. \tag{2.6}$$

A key idea of a *multi-scale inversion method* or *data assimilation method* is to carry out an inversion on a lower scale first and then move to the next finer scale. Let

$$U = \tilde{U}_1 \oplus \tilde{U}_2 \oplus \tilde{U}_3 \oplus \cdots \oplus \tilde{U}_M \tag{2.7}$$

be a multi-scale decomposition of U into a hierarchy of spaces defined by

$$U_j := \bigoplus_{\xi=1}^{j} \tilde{U}_\xi, \ \ j = 1, 2, ..., M. \tag{2.8}$$

We assume that the spaces are mutual linearly independent. Instead of solving the full minimization problem minimizing

$$J(\varphi) := ||f - H\varphi^{(b)} - H\varphi||_{R^{-1}}^2, \ \ \varphi \in U, \tag{2.9}$$

we solve problems with $\tilde{\varphi} \in \tilde{U}_\xi$ successively. In each step $\xi = 1, 2, 3, ..., M$ we solve the equation

$$H\varphi = f^{(\xi)}, \ \ \varphi \in \tilde{U}_\xi, \tag{2.10}$$

where we start with $f^{(1)} := f - H\varphi^{(b)}$. Starting with $\varphi^{(0)} = \varphi^{(b)}$ for $\xi = 1, 2, 3, ..., M$ we denote the minimizer of

$$J_\xi(\varphi) := \alpha||\varphi||_{B^{-1}}^2 + ||H\varphi - f^{(\xi)}||_{R^{-1}}^2. \ \ \varphi \in \tilde{U}_\xi, \tag{2.11}$$

by $\tilde{\varphi}^{(\xi)}$, where

$$f^{(\xi)} := f^{(\xi-1)} - H\tilde{\varphi}^{(\xi-1)} \tag{2.12}$$

and $\alpha > 0$ is our regularization parameter. We also allow $\alpha = 0$, in this case the problem is called *unregularized*. We carry out the loop of (2.11) - (2.12) for $\xi = 1, ..., M$. This leads to successive elements $\tilde{\varphi}^{(1)}, \tilde{\varphi}^{(2)}, ..., \tilde{\varphi}^{(M)}$ which are in the spaces $\tilde{U}_1, ..., \tilde{U}_M$. The corresponding solutions up to level j are given by

$$\varphi^{(j)} := \varphi^{(b)} + \sum_{\xi=1}^{j} \tilde{\varphi}^{(\xi)}, \ \ j = 1, 2, 3, ..., M, \tag{2.13}$$

which is an element of U_j defined in (2.8).

A Multiscale Step in State Space. Here, we study one reconstruction step of (2.11) - (2.12). Let the space under consideration be denoted by $U \subset X$. We assume that a set of linearly independent *ensemble members* or a *multiscale basis* $\varphi^{(1)}, ..., \varphi^{(L)}$ of U is given. We drop unnecessary indices for simplicity.

THEOREM 2.1 *Let $\Phi = (\varphi^{(1)}, ..., \varphi^{(L)})$ be a basis of $U \subset \mathbb{R}^n$ for $L \leq n$, i.e. it is a matrix with columns being linearly independent vectors spanning U. Then, the minimization of $J(\varphi)$ given in (2.2) on the affine subspace $\varphi^{(b)} + U$ defined by*

$$\varphi = \varphi^{(b)} + \sum_{\ell=1}^{L} \alpha_\ell \varphi^{(\ell)} \tag{2.14}$$

for $\alpha = (\alpha_1, ..., \alpha_L)$, is carried out by

$$\alpha^{(a)} = \breve{B}\breve{H}^T(R + \breve{H}\breve{B}\breve{H}^T)^{-1}\breve{f}, \tag{2.15}$$

or alternatively by

$$\alpha^{(a)} = (\alpha I + \breve{B}\breve{H}^T R^{-1}\breve{H})^{-1}\breve{B}\breve{H}^T R^{-1}\breve{f}, \tag{2.16}$$

and

$$\varphi^{(a)} := \varphi^{(b)} + \Phi\alpha^{(a)}, \tag{2.17}$$

using the abbreviations

$$\breve{f} = f - H\varphi^{(b)}, \quad \breve{B}^{-1} = \Phi^T B^{-1}\Phi \quad and \quad \breve{H} = H\Phi. \tag{2.18}$$

Remark. We note that when the dimension of U is L and the dimension of Y is m, the matrix $\breve{B}^{-1}$ is an $L \times L$ matrix and $\breve{H}$ is an $m \times L$ matrix. Thus, the inversion of (2.16) is taking place in the L-dimensional space only, where we might have $L \ll m$. In the case where $L > m$ we will prefer to use (2.15).

Proof. With $\varphi = \varphi^{(b)} + \Phi\alpha$ for $\alpha \in \mathbb{R}^L$ we rewrite (2.2) into

$$J(\alpha) := \|\Phi\alpha\|_{B^{-1}}^2 + \|(f - H\varphi^{(b)}) - H\Phi\alpha\|_{R^{-1}}^2, \quad \alpha \in \mathbb{R}^L. \tag{2.19}$$

We remark that if B is invertible and Φ is of maximal rank, then $\Phi^T B^{-1}\Phi$ is invertible in $\mathbb{R}^{L \times L}$, since

$$\Phi^T B^{-1}\Phi\alpha = 0 \implies \langle \Phi\alpha, B^{-1}\Phi\alpha \rangle = 0 \implies \Phi\alpha = 0 \implies \alpha = 0, \tag{2.20}$$

such that $\Phi^T B^{-1}\Phi$ is injective and thus surjective on $\mathbb{R}^L$. Then, with the definitions of $\breve{H}$, $\breve{B}$ and $\breve{f}$ from above we transform (2.19) into

$$J(\alpha) := \|\alpha\|_{\breve{B}^{-1}}^2 + \|\breve{f} - \breve{H}\alpha\|_{R^{-1}}^2, \quad \alpha \in \mathbb{R}^L. \tag{2.21}$$

Minimization of (2.21) with respect to α leads to (2.15), such that $\alpha^{(a)}$ is the minimizer of (2.19) and (2.17) is a minimizer of (2.2) over $\varphi^{(b)} + U$. For the different forms of the update operator we refer to the recent book [30], Chapter 5. $\qquad \square$

In the case where U is the whole space and Φ is linearly independent, we obtain

$$\breve{B} = \Phi^{-1}B(\Phi^T)^{-1} \tag{2.22}$$

with the inverse Φ^{-1} of Φ and the inverse $(\Phi^T)^{-1}$ of the transpose $n \times n$-matrix Φ. In the case where Φ is an orthonormal basis of X, the transform of the B matrix is obtained by taking the inverse of $\breve{B}^{-1}$, which according to (2.18) leads to

$$\begin{aligned} \breve{B} &= \Phi^{-1}B(\Phi^T)^{-1} \\ &= \Phi^T B\Phi. \end{aligned} \tag{2.23}$$

In the orthonormal case and when B is given by the covariance estimator (2.4), we obtain

$$\breve{B} = \Phi^T B \Phi = \Phi^T Q Q^T \Phi = (\Phi^T Q)(\Phi^T Q)^T, \tag{2.24}$$

which is the estimated covariance for the *multi-scale coefficients*. If we are in the non-orthonormal case, we obtain

$$\breve{B} = \Phi^{-1} B (\Phi^T)^{-1} = (\Phi^{-1} Q)(\Phi^{-1} Q)^T \tag{2.25}$$

based on $(\Phi^T)^{-1} = (\Phi^{-1})^T$, which holds for invertible matrices on $\mathbb{R}^n$.

We need to extend our tansformation to the case where Φ is a matrix of vectors of dimension $n \times L$ and where $B \in \mathbb{R}^{n \times n}$ is rank deficient. Here, we can replace B^{-1} by its *pseudo-inverse*

$$B^\dagger := \lim_{\alpha \to 0} (\alpha I + B^T B)^{-1} B^T, \tag{2.26}$$

compare [30], Section 3.2. If $\text{span}\{\Phi\} \cap N(B) = \{0\}$ the matrix $\Phi^T B^\dagger \Phi$ is invertible and (2.18) with B^{-1} replaced by $B^\dagger$ is well-defined and we obtain

$$\breve{B} = \left(\Phi^T B^\dagger \Phi \right)^{-1} = \Phi^\dagger (B^\dagger)^\dagger (\Phi^T)^\dagger = \Phi^\dagger B (\Phi^T)^\dagger. \tag{2.27}$$

If we search for the solution in ensemble space, i.e. if U is the space defined by the columns of Q defined in (2.5), then (2.25) leads to

$$\breve{B} = Q^\dagger Q Q^T (Q^T)^\dagger = I. \tag{2.28}$$

3 Non-Equivalence and Iterations of the Sequential Multiscale Approach

In this part we first study the non-equivalence of sequential multiscale to the full inversion step, for both the unregularized and the regularized case. We show that an iteration of the sequential steps will converge under appropriate assumptions. We will interpret our analysis as a type of *aliasing* in the framework of well-posed or ill-posed inversion.

3.1 On the Equivalence and Non-Equivalence of Sequential Multiscale

In general, the sequential multiscale approach does not lead to the correct solution of the full inverse problem under consideration.

THEOREM 3.1 *Consider the sequential multiscale method defined in (2.11)-(2.12). If the spaces $\tilde{U}_1, ..., \tilde{U}_M$ are not orthogonal to each other with respect to $\langle \cdot, \cdot \rangle_{H^* R^{-1} H}$, then in general the sequential solution $\varphi^{(M)}$ defined in (2.13) will not coincide with the solution to the full inverse problem given by (2.1).*

Proof. Let us study the generic case $M = 2$ and first consider the unregularized solution and an injective operator H first, such that

$$\langle \varphi, \psi \rangle_{H^*R^{-1}H} := \langle H\varphi, R^{-1}H\psi \rangle, \quad \varphi, \psi \in X \tag{3.1}$$

defines a scalar product on the state space X. For the case $M = 2$ we assume that the subspaces $\tilde{U}_1$ and $\tilde{U}_2$ satisfy $X = \tilde{U}_1 \oplus \tilde{U}_2$. In this case, if $\tilde{U}_1 \perp \tilde{U}_2$, then $\tilde{U}_1 = \tilde{U}_2^\perp$ and $\tilde{U}_2 = \tilde{U}_1^\perp$. Let P_1 be the orthogonal projector of X onto $\tilde{U}_1$ and P_2 the corresponding projector onto $\tilde{U}_2$. Then, $I - P_2$ is the orthogonal projector onto $\tilde{U}_2^\perp = \tilde{U}_1$ and $I - P_1$ is the orthogonal projector onto $\tilde{U}_1^\perp = \tilde{U}_2$.

LEMMA 3.2 *If $\tilde{U}_1$ and $\tilde{U}_2$ are* not *perpendicular with respect to $\langle \cdot, \cdot \rangle_{H^*R^{-1}H}$, then the projector $(I - P_2)(I - P_1)$ is a non-zero mapping, i.e. there is $z \in X$ such that*

$$(I - P_2)(I - P_1)z \neq 0. \tag{3.2}$$

Proof. We remark that if $\tilde{U}_1$ is not perpendicular to $\tilde{U}_2$, then we have $\tilde{U}_1^\perp$ is not perpendicular to $\tilde{U}_2^\perp$ just since $\tilde{U}_1 = \tilde{U}_2^\perp$ and $\tilde{U}_2 = \tilde{U}_1^\perp$. This means that there are elements in $\tilde{U}_1^\perp$ whose projection onto $\tilde{U}_2^\perp$ is non-zero, which proves (3.2). $\square$

We now study the difference between the full unregularized solution of the problem (2.2) and the sequential process given by (2.11)-(2.12). We note that if the unregularized sequential solutions are not identical to the full unregularized solution, this cannot be the case for the regularized solutions, since for large covariance matrices B the regularization term is very small and the regularized solution will be close to the unregularized solutions.

For the case of true data we have $f = H\varphi^*$ we write the first step as

$$\tilde{\varphi}^{(1)} = \arg \min_{\varphi \in \tilde{U}_1} ||\varphi - \varphi^*||^2_{H^*R^{-1}H}, \tag{3.3}$$

the second step as

$$\tilde{\varphi}^{(2)} = \arg \min_{\varphi \in \tilde{U}_2} ||\varphi - (\varphi^* - \tilde{\varphi}^{(1)})||^2_{H^*R^{-1}H} \tag{3.4}$$

and the full solution as

$$\varphi^{(tot)} = \arg \min_{\varphi \in X} ||\varphi - \varphi^*||^2_{H^*R^{-1}H} \tag{3.5}$$

The solutions can be understood as orthogonal projections, in particular we have

$$\tilde{\varphi}^{(1)} = P_1\varphi^*, \tag{3.6}$$
$$\tilde{\varphi}^{(2)} = P_2(\varphi^* - \tilde{\varphi}^{(1)})$$
$$= P_2\varphi^* - P_2P_1\varphi^*. \tag{3.7}$$

This leads to

$$\varphi^{(2)} = (P_1 + P_2 - P_2P_1)\varphi^*. \tag{3.8}$$

The difference between the full solution and the sequential solution in the case of injective H is given by

$$
\begin{aligned}
\varphi^{(tot)} - \varphi^{(2)} &= \varphi^* - (P_1 + P_2 - P_2 P_1)\varphi^* \\
&= \left\{ I - (P_1 + P_2 - P_2 P_1) \right\} \varphi^*.
\end{aligned} \tag{3.9}
$$

Finally, we note that

$$
\begin{aligned}
I &= P_1 + (I - P_1) \tag{3.10} \\
&= P_1 + [P_2 + (I - P_2)](1 - P_1) \tag{3.11} \\
&= P_1 + P_2 - P_2 P_1 + (I - P_2)(I - P_1), \tag{3.12}
\end{aligned}
$$

such that

$$
\varphi^{(tot)} - \varphi^{(2)} = (I - P_2)(I - P_1)\varphi^*. \tag{3.13}
$$

By Lemma 3.2 the right-hand side of the above term is non-zero in general, and thus we have shown that the sequential method does not coincide with the full minimization procedure under the above conditions. We note that equation (3.13) can also be used to estimate the error between the two methods.

In the case where H is not injective, the minimizer with minimal norm is the pseudo-inverse with respect to the scalar product (3.1). In this case the above arguments will still work when we replace $\tilde{U}_\xi$ by $\tilde{U}_\xi \cap N(H^* R^{-1} H)^\perp$, on which the operator is injective and all arguments apply.

Finally, the case *with regularization* has the same property that the sequential solution and the full solution will not conincide in general. This is clear since for small regularization parameters the regularized solution is close to the unregularized solution (if it exists), and thus they must be different in the two cases under consideration. □

We demostrate the situation by a two-dimensional unregularized example to visualize the situation. For $M = 2$ and $\varphi^{(b)} = 0$ we show that in general the successive solution is not the same as a solution which is calculated in one big step. Assume that $R = I$ and $H = I$. In this case the solution of (2.10) in U_1 corresponds to an orthogonal projection $P^{(1)} f$ of f onto U_1. Then, $f^{(2)} = f^{(1)} - P^{(1)} f$ is perpendicular to $U^{(1)}$. In the second step, the solution of (2.10) corresponds to an orthogonal projection of $f^{(2)}$ onto $U^{(2)}$. We visualize this situation in Figure 1. In particular, the result for the full minimization would be the measured true solution $\varphi = f$. But in the successive case, we end up with the gray solution $\varphi^{(2)}$ in the second step.

The above result shows that in general we need to be careful with the successive solution of the minimization problem (2.10). We will see next that when we guarantee orthogonality of the subspaces, we are fine.

We now come to a generic result which shows the validity and convergence of the multi-scale minimization problem.
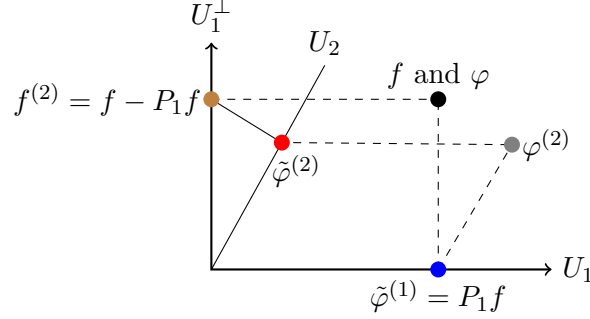
Figure 1: The successive solution of the equations leads first to the point $P_1 f$ shown in blue. The reduced data for step two $f^{(2)}$ is shown as a brown dot. This is projected onto U_2 in step two, see the red dot visualizing $\tilde{\varphi}^{(2)}$, leading to the solution $\varphi^{(2)}$ shown by the gray point. However, the true solution φ would be the point f itself in this example. Thus, we see that in the general case the successive solution of the minimization problem cannot lead to correct results.

THEOREM 3.3 *Assume that H is linear and injective on U and assume that the spaces $\tilde{U}_1, ..., \tilde{U}_M$ are mutual orthogonal with respect to the scalar product defined by*

$$\langle \varphi, \psi \rangle_{H^T R^{-1} H} := \langle \varphi, H^T R^{-1} H \psi \rangle, \quad \varphi, \psi \in X. \tag{3.14}$$

Then, for the unregularized approach the full minimizer φ of (2.9) and the iterative minimizer $\varphi^{(M)}$ coincide.

Proof. We have worked out the error estimate for the unregularized approach for the case $M = 2$ in the equations (3.1) to (3.13). When we have orthogonality of $\tilde{U}_1$ and $\tilde{U}_2$, we obtain $(I - P_2)(I - P_1) = 0$, such that the difference between the full solution and the successive solution vanishes. This applies inductively to all subspaces $\tilde{U}_1, \tilde{U}_2, ...$, such that we obtain full equivalence. $\square$

We also not at this point that the regularized successive solutions will not be identical even if we have orthogonality in the spaces. Differences come from the role of the regularizing term, since in each regularization step the regularized solution is between the background and the unregularized solution, but does not use the same norm but a weight which involves the covariance matrix B.

3.2 General Form of Aliasing Problems

One important phenomenon of signal processing is *aliasing*. It means that if you sample a signal with high frequency components and you try to reconstruct its low frequency parts from a low number of sampling points, the high frequency parts of your signal will be

transformed into spurious lower frequency signal. It is possible to avoid aliasing either by using more sampling points even when you reconstruct the low modes only, or you need to run a *filter* on your signal before sampling and then only sample the filtered signal and transmit it.

The aliasing problem also appears when you try to solve multiscale inverse problems by using a reduced number of measurements coming from a signal which contains higher modes. In general, the aliasing problem is a null-space phenomenon on the signal or state reconstruction operator.

Let $H : X \to Y$ be our observation operator in the finite dimensional case, i.e. $X = \mathbb{R}^n$, $Y = \mathbb{R}^m$, and let $\hat{H} : X \to \hat{Y}$ be a selection of the rows of H, i.e. $\hat{Y} = \mathbb{R}^{\hat{m}}$ with $\hat{m} < m$. Let us assume that $m = n$ and H is injective. Then, $\hat{H}$ will not be injective, and there is some nullspace $U := N(\hat{H})$ of $\hat{H}$ in X. The solution of the reduced data equation

$$\hat{H}\varphi = \hat{f}, \quad \varphi \in V, \tag{3.15}$$

in some reduced subspace $V \subset X$ does not need to coincide with the orthogonal or any other particular projection Px_0 of the true solution x_0 of

$$H\varphi = f, \quad \varphi \in X \tag{3.16}$$

onto the space V. If the nullspace U of $\hat{H}$ does not coincide with $V^\perp$ or the space along which P projects, there are many solutions $\varphi_0 + \Delta\varphi$ of (3.15) which are projected onto different solutions $\varphi \in V$ by P. If, however, we have $U = V^\perp$ or more general U is the space along which P projects, then the projection of any solution $\varphi_0 + \Delta\varphi$ with $\Delta\varphi \in U$ will lead to the projection $P\varphi = P\varphi_0$ in V, i.e. we obtain the correct projection in the low-dimensional subspace V.

In the framework of (3.15) filtering of the data corresponds to the transformation of $\hat{f}$ into data which are in the image space of V under $\hat{H}$, i.e. the filter needs to construct $\hat{f}^{filter} := \hat{H}(P\varphi_0)$ as data for which the correct state estimate $P\varphi_0 \in V$ is obtained when a solution to (3.15) in V is calculated. As argued in Lemma 3.2, the use of the data $\hat{f}$ can lead to arbitrarily large errors when solving the unregularized reduced equation.

3.3 Convergence for Iterated Sequential Multiscale

We now define an *iterated* approach to *sequential multiscale*. The basic idea is to carry out the multiscale approach several times to improve on the solution in the case where orthogonality of the spaces is not given.

DEFINITION 3.4 (ITERATIVE SEQUENTIAL MULTISCALE) *We carry out the sequential multiscale approach iteratively by the following steps.*

1. *We start with a first application of the sequential multiscale as given by equations (2.11) - (2.12). For $\ell = 1$ we define $\varphi^{(\xi,\ell)} := \varphi^{(\xi)}$ and $f^{(\xi,\ell)} := f^{(\xi)}$ for $\xi = 1, ..., M$.*

2. *In each iteration step for $\ell = 1, 2, 3, ...$ we define new data $f^{(1,\ell+1)}$ by*

$$f^{(1,\ell+1)} := f - H\varphi^{(M,\ell)} \tag{3.17}$$

and

$$\varphi^{(b,\ell+1)} = \varphi^{(M,\ell)}. \tag{3.18}$$

3. *Then, the sequential multiscale method is iterated until some convergence criterion is satisfied or stopping rule fires.*

The goal of the following paragraphs is to carry out the convergence analysis of the iterative sequential multiscale algorithm for the *unregularized* case and to show convergence when true data $f = H\varphi^*$ with an injective operator H are given. The regularized case will be discussed in the following Section 3.4

LEMMA 3.5 *The error vector for the M-th iteration of the sequential multiscale is given by*

$$\varphi^* - \varphi^{(M)} = (I - P_M)(I - P_{M-1}) \cdots (I - P_1)\varphi^*. \tag{3.19}$$

Proof. We carry out a proof by induction. Assume that we have already shown

$$\varphi^* - \varphi^{(M-1)} = (I - P_{M-1}) \cdots (I - P_1)\varphi^*. \tag{3.20}$$

Then, we obtain

$$
\begin{aligned}
\varphi^* - \varphi^{(M)} &= \varphi^* - (\varphi^{(M-1)} + \tilde{\varphi}^{(M)}) \\
&= \varphi^* - \varphi^{(M-1)} - \tilde{\varphi}^{(M)} \\
&= \varphi^* - \varphi^{(M-1)} - P_M(\varphi^* - \varphi^{(M-1)}) \\
&= (I - P_M)(\varphi^* - \varphi^{(M-1)}) \\
&= (I - P_M)(I - P_{M-1}) \cdots (I - P_1)\varphi^*
\end{aligned} \tag{3.21}
$$

and the proof is complete. □

We can now estimate the error for the iterated sequential multiscale as follows.

THEOREM 3.6 *The error vector for the iterated sequential multiscale in the unregularized case can be calculated by*

$$\varphi^* - \varphi^{(M,\ell)} = \left[(I - P_M)(I - P_{M-1}) \cdots (I - P_1) \right]^\ell \varphi^* \tag{3.22}$$

for $\ell = 1, 2, 3, ...$

Proof. For $\ell = 1$ the result has been shown in Lemma 3.5. If we now carry out the iterative procedure, by definition (3.17) we start with the difference between the true data and the last outcome of the sequential multiscale, which is

$$
\begin{aligned}
f^{(1,\ell+1)} &= f - H\varphi^{(M,\ell)} \\
&= H\left(\varphi^* - \varphi^{(M,\ell)}\right).
\end{aligned}
\tag{3.23}
$$

This means that by applying Lemma 3.5 inductively, in each application of the multiscale method we increase ℓ by one in the error estimate (3.22). $\square$

In each step of the method, the error is reduced by a factor $\cos(\alpha)$, where α is the angle between the subspaces under consideration. This is linear convergence for the iterative sequential multiscale method, as a consequence of its understanding as a sequential projection method. We summarize this immediate consequence in the following corollary.

COROLLARY 3.7 (CONVERGENCE ITERATED SEQUENTIAL MULTISCALE) *The sequential multiscale given by (3.17)-(3.18) in the* unregularized case *shows linear convergence with factor $c = \cos(\alpha)$ where α is a lower estimate for the angles between the subspaces $\tilde{U}_\xi$, i.e. the error $e_{\xi,\ell} = ||\varphi^{\xi,\ell} - \varphi^*||_{H^*R^{-1}H}$ in the ℓ-th iteration and after step ξ of this iteration is estimated by*

$$
e_{\xi,\ell} \le \left(\cos(\alpha)\right)^{\xi + M(\ell-1)} ||\varphi^{(b)} - \varphi^*||_{H^*R^{-1}H}
\tag{3.24}
$$

3.4 Regularized Sequential Multiscale

We have seen that we obtain convergence in the unregularized case when we iterate the sequential multiscale, with a linear error estimate which is determined by the angle between the multi-scale subspaces.

The *regularized* case is much more involved, since the background term of the regularizer needs to be taken care of in a special way. To see the challenge let us first study the generic but simplified situation where we start some iteration with

$$
\gamma_0 := \varphi^{(b)}
\tag{3.25}
$$

and then iterate by

$$
\gamma_\ell = \gamma_{\ell-1} + \beta(\varphi^* - \gamma_{\ell-1})
\tag{3.26}
$$

for $\ell = 1, 2, 3, \ldots$ with some parameter $\beta \in (0,1)$. This corresponds to the increment (2.3) where perfect data φ^* are given and where we have an observation operator $H = I$ and covariance matrices $B = \frac{\beta}{1-\beta}I$ and $R = I$. It is the case where we do not have different multi-scale spaces, but carry out an iteration in one single space only. In this case, the iterations can be calculated exactly as follows.

LEMMA 3.8 *The iteration (3.25)-(3.26) leads to*

$$\gamma_\ell = \gamma_0 + \left\{ \sum_{\xi=1}^{\ell} \binom{\ell}{\xi} (-1)^{\xi+1} \beta^\xi \right\} (\varphi^* - \gamma_0) \tag{3.27}$$

$$= \gamma_0 + \left\{ 1 - (1-\beta)^\ell \right\} (\varphi^* - \gamma_0) \tag{3.28}$$

$$= \varphi^* - (1-\beta)^\ell (\varphi^* - \gamma_0) \tag{3.29}$$

for $\ell = 1, 2, 3, \ldots$.

Proof. We use induction for the proof. For $\ell = 1$ it is

$$\gamma_1 = \gamma_0 + \left\{ \sum_{\xi=1}^{1} \binom{1}{\xi} (-1)^{\xi+1} \beta^\xi \right\} (\varphi^* - \gamma_0)$$

$$= \gamma_0 + \beta(\varphi^* - \gamma_0), \tag{3.30}$$

which coincides with our claim. Suppose the formula is true for $\ell = n$

$$\gamma_n = \gamma_0 + \left\{ \sum_{\xi=1}^{n} \binom{n}{\xi} (-1)^{\xi+1} \beta^\xi \right\} (\varphi^* - \gamma_0) \tag{3.31}$$

Then we obtain for $\ell = n + 1$

$$\begin{aligned}
\gamma_{n+1} &= \gamma_n + \beta(\varphi^* - \gamma_n) \\
&= \gamma_0 + \left\{ \sum_{\xi=1}^{n} \binom{n}{\xi} (-1)^{\xi+1} \beta^\xi \right\} (\varphi^* - \gamma_0) \\
&\quad + \beta \left(\varphi^* - \left(\gamma_0 + \left\{ \sum_{\xi=1}^{n} \binom{n}{\xi} (-1)^{\xi+1} \beta^\xi \right\} (\varphi^* - \gamma_0) \right) \right) \\
&= \gamma_0 + \beta(\varphi^* - \gamma_0) + \left\{ \sum_{\xi=1}^{n} \binom{n}{\xi} (-1)^{\xi+1} \beta^\xi \right\} (\varphi^* - \gamma_0) \\
&\quad - \beta \left\{ \sum_{\xi=1}^{n} \binom{n}{\xi} (-1)^{\xi+1} \beta^\xi \right\} (\varphi^* - \gamma_0) \\
&= \gamma_0 + \beta(\varphi^* - \gamma_0) + \left\{ \sum_{\xi=1}^{n} \binom{n}{\xi} (-1)^{\xi+1} \beta^\xi \right\} (\varphi^* - \gamma_0) \\
&\quad + \left\{ \sum_{\xi=1}^{n} \binom{n}{\xi} (-1)^{\xi+2} \beta^{\xi+1} \right\} (\varphi^* - \gamma_0)
\end{aligned}$$

$$= \gamma_0 + \beta(\varphi^* - \gamma_0) + \left\{ \sum_{\xi=1}^{n} \binom{n}{\xi} (-1)^{\xi+1} \beta^\xi \right\} (\varphi^* - \gamma_0)$$

$$+ \left\{ \sum_{\xi=2}^{n+1} \binom{n}{\xi-1} (-1)^{\xi+1} \beta^\xi \right\} (\varphi^* - \gamma_0)$$

$$= \gamma_0 + \left\{ \sum_{\xi=1}^{n} \binom{n}{\xi} (-1)^{\xi+1} \beta^\xi \right\} (\varphi^* - \gamma_0) + \beta(\varphi^* - \gamma_0)$$

$$+ \left\{ \sum_{\xi=2}^{n} \binom{n}{\xi-1} (-1)^{\xi+1} \beta^\xi \right\} (\varphi^* - \gamma_0) + \binom{n}{n} (-1)^{n+2} \beta^{n+1} (\varphi^* - \gamma_0)$$

$$= \gamma_0 + \left\{ \sum_{\xi=1}^{n} \binom{n}{\xi} (-1)^{\xi+1} \beta^\xi \right\} (\varphi^* - \gamma_0)$$

$$+ \left\{ \sum_{\xi=1}^{n} \binom{n}{\xi-1} (-1)^{\xi+1} \beta^\xi \right\} (\varphi^* - \gamma_0) + \binom{n}{n} (-1)^{n+2} \beta^{n+1} (\varphi^* - \gamma_0)$$

$$= \gamma_0 + \sum_{\xi=1}^{n} \left\{ \binom{n}{\xi} + \binom{n}{\xi-1} \right\} (-1)^{\xi+1} \beta^\xi (\varphi^* - \gamma_0) + \binom{n}{n} (-1)^{n+2} \beta^{n+1} (\varphi^* - \gamma_0)$$

$$= \gamma_0 + \sum_{\xi=1}^{n} \left\{ \binom{n+1}{\xi} (-1)^{\xi+1} \beta^\xi \right\} (\varphi^* - \gamma_0) + \binom{n+1}{n+1} (-1)^{n+2} \beta^{n+1} (\varphi^* - \gamma_0)$$

$$= \gamma_0 + \sum_{\xi=1}^{n+1} \left\{ \binom{n+1}{\xi} (-1)^{\xi+1} \beta^\xi \right\} (\varphi^* - \gamma_0), \tag{3.32}$$

such that the formula is true for $\ell = n + 1$. Hence, it is true for all $n \in \mathbb{N}$. By using the binomial expansion we have

$$(1 - X)^\ell = \left\{ \sum_{\xi=0}^{\ell} \binom{\ell}{\xi} (-1)^\xi (X)^\xi \right\}$$

$$(1 - X)^\ell = 1 + \left\{ \sum_{\xi=1}^{\ell} \binom{\ell}{\xi} (-1)^\xi (X)^\xi \right\} \tag{3.33}$$

So the equation (3.31) becomes

$$\gamma_\ell = \gamma_0 + \left\{ 1 - (1 - \beta)^\ell \right\} (\varphi^* - \gamma_0)$$

$$= \varphi^* - (1 - \beta)^\ell (\varphi^* - \gamma_0)$$

$$\tag{3.34}$$

for $\ell = 1, 2, 3, \ldots$ and the proof is complete. $\qquad \square$

Let us compare the ℓ-th iteration in the form (3.28) with the one-step regularized increment given by

$$\varphi^{(a)} = \gamma_0 + \alpha(\varphi^* - \gamma_0) \tag{3.35}$$

with replacing β by the regularization parameter α. To achieve equivalence of a ℓ-step method with the one-step method, we need to choose

$$\alpha = 1 - (1-\beta)^\ell \quad \Leftrightarrow \quad \sqrt[\ell]{1-\alpha} = 1 - \beta. \tag{3.36}$$

In this simplified case, when we use iterations for the regularized case, to achieve an appropriate balance between the background and the data, we need to adapt the relative weight using the ℓ-th square root of the regularization parameter $1 - \alpha$.
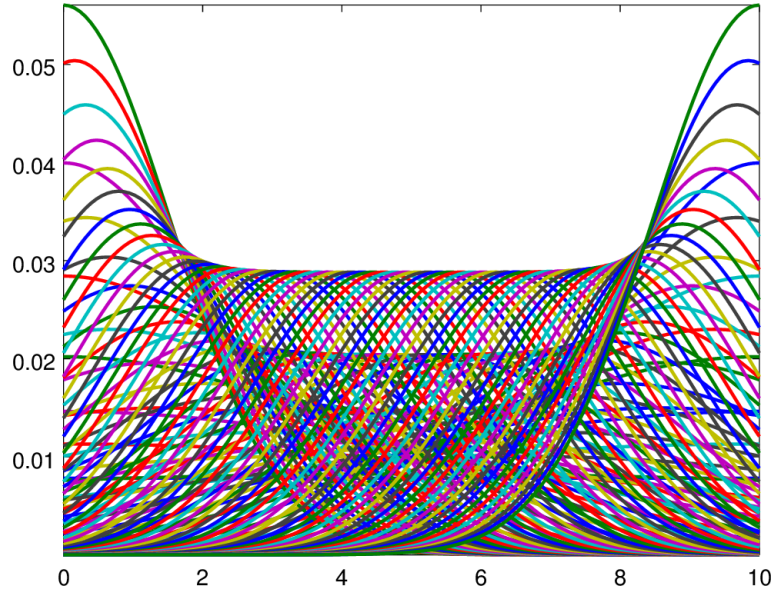

Figure 2: We show the basis functions used for the demonstration of the iterative multiscale method on the interval $[a, b] = [0, 10]$. The functions are generated by scaling and translating a Gaussian basis wavelet. All functions are scaled to have the L^2-norm $||\varphi||_{L^2([a,b])} = 1$, leading to some amplification at the edges. In this image, we display $n = 2^7 = 128$ basis functions.

To complete our analysis, we now study the general but still complete-scale case, where the iteration is given by

$$\gamma_0 \;=\; \varphi^{(b)} \tag{3.37}$$

$$\gamma_\ell \;=\; \gamma_{\ell-1} + \tilde{K}(H\varphi^* - H\gamma_{\ell-1}) \tag{3.38}$$

with some Kalman gain operator $\tilde{K}$ which we need to determine appropriately. With the analysis given in the proof of Lemma 3.8 we now obtain the following result.

THEOREM 3.9 *The iterations (3.37)-(3.38) lead to*

$$\gamma_\ell \;\; = \;\; \gamma_0 + \left(1 - (1 - \tilde{K}H)^\ell \right)(\varphi^* - \gamma_0) \tag{3.39}$$

for $\ell = 1, 2, 3, ...$

Proof. We basically use a modification of the proof of Lemma 3.8, which was the simplified case in which an observation operator $H = I$ and covariance matrices $B = \frac{\beta}{1-\beta} I$ and $R = I$ were studied. Here, we need to be careful with the operator nature of $\tilde{K}H$. We have

$$(I - \tilde{K}H)^\ell = \left\{ \sum_{r=1}^{\ell} \binom{\ell}{r} (-1)^r (\tilde{K}H)^r \right\} \tag{3.40}$$

since there is no difficulty with the comutative law of powers of an operator $A = \tilde{K}H$. So, induction is carried out following the steps of Lemma 3.8 to obtain

$$\gamma_\ell = \gamma_0 + \sum_{\xi=1}^{n+1} \left\{ \binom{n+1}{\xi} (-1)^{\xi+1} (\tilde{K}H)^\xi \right\}(\varphi^* - \gamma_0) \tag{3.41}$$

Now again, by using the binomial theorem representaion, we derive

$$\gamma_\ell \;\; = \;\; \gamma_0 + \left\{ I - (I - \tilde{K}H)^\ell \right\}(\varphi^* - \gamma_0) \tag{3.42}$$

$$= \;\; \varphi^* - (1 - \tilde{K}H)^\ell (\varphi^* - \gamma_0) \tag{3.43}$$

and the proof is complete. $\square$

To achieve equivalence of an ℓ-step reconstruction with the one-step reconstruction

$$\varphi^{(a)} = \varphi^{(b)} + KH(\varphi^* - \varphi^{(b)})), \tag{3.44}$$

we need to determine $\tilde{K}$ such that

$$(I - \tilde{K}H)^\ell = I - KH \tag{3.45}$$

as the counterpart of the postulate (3.36). It corresponds to taking the ℓ-th root of the operator $I - KH = I - BH^*(R + HBH^*)^{-1}H$, with K defined by the second term of (2.3). For symmetric positive operators, the square root is well-defined. Here, we obtain this property by using the weighted scalar products

$$\langle \varphi, \tilde{\varphi} \rangle_{B^{-1}} := \langle \varphi, B^{-1}\tilde{\varphi} \rangle_{L^2} \tag{3.46}$$
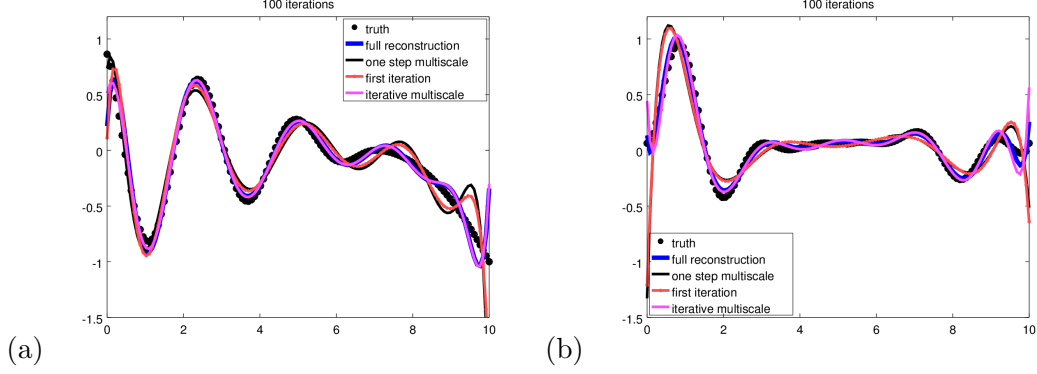
Figure 3: Two reconstruction results of the *multiscale method* and the *iterative multiscale method* after $\ell = 100$ iterations is displayed here. The blue curve shows the regularized reconstruction when the full basis is employed. The classical multiscale method leads to the black curve. The red curve shows the first step of the iteration, which is similar to the black one, but with a modified regularization parameter for the two steps. Then, iteration leads to the convergence as shown by the magenta line, where we stopped after ℓ iterations, we chose $\alpha = 10^{-9}$. For a convergence study see Figure 4.

and

$$\left\langle \psi, \tilde{\psi} \right\rangle_{R^{-1}} := \left\langle \psi, R^{-1}\tilde{\psi} \right\rangle_{L^2} \tag{3.47}$$

for functions $\varphi, \tilde{\varphi} \in X$ and $\psi, \tilde{\psi} \in Y$. Using equation (5.2.11) as in (5.2.14) of [30] with $H^{*,w}$ being the adjoint with respect to the scalar products (3.46) and (3.47), we see that we have $BH^*R^{-1} = H^{*,w}$ and thus

$$
\begin{aligned}
I - KH &= I - BH^*(R + HBH^*)^{-1}H \\
&= I - H^{*,w}(I + HH^{*,w})^{-1}H.
\end{aligned}
\tag{3.48}
$$

Clearly, we have

$$\left(H^{*,w}(I + HH^{*,w})^{-1}H \right)^{*,w} = H^{*,w}(I + HH^{*,w})^{-1}H, \tag{3.49}$$

i.e. we have a self-adjoint positive operator. Thus, the l-th square root of $I - KH$ exists and can be calculated using the singular value decomposition of the operator in the space with weighted scalar product. We now collect the above results into the following theorem.

COROLLARY 3.10 *Assume that for $\ell \in \mathbb{N}$ we determine $\tilde{K}$ such that (3.45) is satisfied. Then, after ℓ steps the iteration (3.37) - (3.38) is identical to the regularized solution (3.44).*

For practical applications of the *sequential multiscale method* taking the ℓ-th root of the operator K is not feasible. Usually, the reason why we consider multiscale is the complexity of the task under consideration. We suggest to employ the solution algorithms as they are available on each of the scales under consideration, adapt the regularization parameter according to (3.36) and then iterate the multiscale algorithms. Numerical tests confirm the feasibility and convergence of this approach in our final section.

4 Numerical Examples

This last section serves to test the above method of *iterative sequential multiscale* for solving integral equations of the form

$$f(x) = \int_a^b k(x,y)\varphi(y) \; ds(y), \quad x \in [a,b] \tag{4.1}$$

where $b \geq a$ are real numbers, k is a continuous kernel

$$k : [a,b] \times [a,b] \to \mathbb{R} \tag{4.2}$$

and we work in $L^2(a,b)$. For numerical testing we choose the example of a Gaussian kernel

$$k(x,y) := e^{-\sigma|x-y|^2}, \quad x,y \in [a,b], \tag{4.3}$$

with $\sigma = 3$, which serves as a generic example of atmospheric temperature or humidity retrieval from satellite radiances, compare [30, 29]. Here, we choose some example densities

$$\varphi_1(y) \quad := \quad \sin(c_6 \cdot (y - y_1)) \cdot |y - y_1| \cdot (y < y_1) + c_7 \cdot (y - y_1)^2 \cdot (y > y_1) \tag{4.4}$$

$$\varphi_2(y) \quad := \quad c_5 \cdot ((y - y_0)^{c_3} \cdot \sin(c_4 * y) \cdot \sin(y) \cdot (b - y) \cdot (b/2)^{-1/2} + 1) \tag{4.5}$$

$$\varphi_3(y) \quad := \quad y^{c_1} \cdot \sin(y) \cdot \cos(y) \cdot (b - y)^{c_2} \tag{4.6}$$

for $y \in [a,b]$ with $a = 0$, $b = 10$, $c_1 = 1.5$, $c_2 = 0.8$, $c_3 = 1.9$, $c_4 = 2.1$, $c_5 = 2$, $c_6 = 2.4$, $c_7 = -0.8$, $y_0 = 5$ and $y_1 = 7$. The functions are not in the span of the basis functions. Examples with φ_1 and φ_2 are shown in Figure 3 (a) and (b), the reconstruction of φ_3 is displayed in Figure 4. We calculate the simulated measured data $f(x)$ by numerical quadrature with trapezoidal or rectangular rule. The observation operator H is given by the integral (4.1), i.e. we have

$$(H\varphi)(x) := \int_a^b k(x,y)\varphi(y) \; ds(y), \quad x \in [a,b] \tag{4.7}$$

with kernel k given by (4.2) and (4.3). For the numerical simulation of H we have used a collocation scheme with a regular grid on $[a,b]$ with n points, defined by

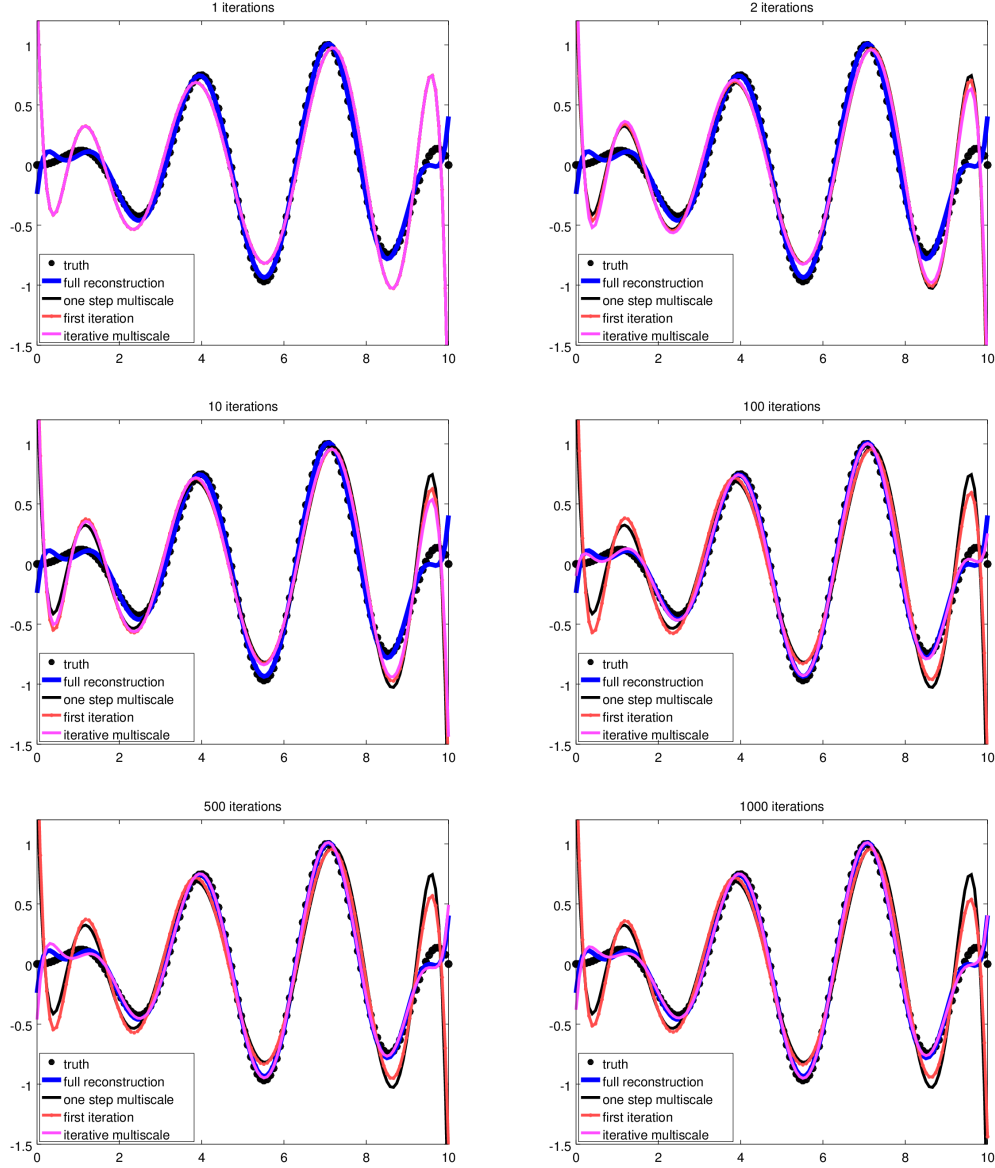$$x_k := a + k \cdot \frac{b-a}{n-1}, \quad k = 0, ..., n-1 \tag{4.8}$$

Figure 4:   The results of the *multiscale method* and the *iterative multiscale method* after $\ell = 1, 2, 10, 100, 500, 1000$ iterations is displayed here. The blue curve shows the regularized reconstruction of the truth in a space with $\ell = 128$ degrees of freedom, when the basis with $\ell = 32$ elements is employed. The classical multiscale method with dimension $\ell = 16$ of the course space and $\ell = 16$ for the fine space leads to the black curve. The red curve shows the first step of the iteration, which is similar to the black one, but with a modified regularization parameter for the two steps. The iterative multiscale solution converges to the blue solution, i.e. to the solution when all its basis functions are used in one step.

Since we want to investigate convergence, we have used *true data* which were generated by numerical collocation. We remark that the problem is *exponentially ill-posed* such that even when the dimension n of our space is just modestly high an algorithm without regularization does not lead to any reasonable solution.

There are many possibilities to define and test a multiscale basis on the interval $[a, b]$. Here, we carry out tests with a *Gaussian type radial basis function*, i.e. for $N \in \mathbb{N}$ and $n = 2^N$ we define

$$\varphi_{\ell,k}(x) := c_{\ell,k} e^{-\rho_\ell |x - x_{\ell,k}|^2}, \quad x \in [a, b] \tag{4.9}$$

for $\ell = 1, ..., N$ and $k = 1, ..., 2^{\ell-1}$ where

$$x_{\ell,k} := a + \frac{k}{2^{\ell-1}}(b - a), \qquad \rho_\ell = 2^{\ell-1}\rho_0 \tag{4.10}$$

and $\varphi_{0,0} \equiv c_{0,0}$. The normalized basis functions (4.9) are displayed in Figure 2.

For multiscale splitting we chose the lowest L modes for the space $U_1 = \tilde{U}_1$ and defined $\tilde{U}_2$ to be the space spanned by the basis functions $L + 1, ..., \tilde{n}$, where $\tilde{n}$ is the maximal number of basis functions under consideration. For the iterations we show two cases with $\ell = 100$ in Figure 3. Figure 3 displays the results for $n = 128$, $L = 16$, $\tilde{n} = 32$ and a regularization parameter $\alpha = 10^{-9}$. We show the truth in black dots, the full reconstruction without any multiscale splitting in blue, the one-step regularized multiscale in black, the first iteration of the iterative regularized multiscale method in red and the result of the iterative multiscale method in pink. As expected, the one-step multiscale has much larger errors than the full solution, arising from the non-orthogonality of the multi-scale spaces. The convergence is studied in a third example in Figure 4, where we display the result of iterations with $\ell = 1, 2, 10, 100, 500, 1000$. The iterative multiscale solution (pink line) confirms the convergence to the solution in the full subspace (blue line) according to Theorem 3.9 and Corollary 3.10.

5 Conclusions

The broad idea of a multiscale approach is to decompose your problem into different scales or levels and to use these decompositions either for constructing appropriate approximations or to solve smaller problems on each of these levels, leading to increased stability or increased efficiency.

The idea of sequential multiscale as studied in our Section 2 is to first solve the problem in a large-scale subspace and then successively move to finer scale spaces. The equivalence or non-equivalence of this approach to solving the full-scale problem at once has been studied in Section 3.1. When the spaces under consideration are orthogonal, we show equivalence for the unregularized problem. For the regularized problems equivalence can only achieved when the regularization parameters are modified appropriately.

In general the sequential multiscale approach is not equivalent to a full solution, but we have shown in Theorem 3.6 and Corollary 3.7 that under appropriate assumptions we obtain convergence of an *iterative* sequential multiscale version of the method.

For the *regularized case* we have developed a strategy to appropriately adapt the regularization when an iterative approach is taken. We have first studied a simplified situation in Lemma 3.8, for which the calculation of the modified regularization parameter could be carried out explicitly. Then, the *general operator-based case* is treated, with a core equivalence result in Theorem 3.9.

Finally, we demonstrate the validity of the iterative sequential multiscale approach by testing the method on integral equations as they appears for atmospheric temperature retrieval from infrared satellite radiances and for image denoising, here for a simplified case in one dimension. The examples show the practical validity of the results and confirm the non-equivalence and convergence theory.

The analysis and methods of this work provide basic insight into the convergence behavior of multiscale approaches to inversion and data assimilation. Here, the static case has been treated, where iteration takes place at one given point in time with fixed measurements. For data assimilation, where iteration is carried out naturally by cycled systems (compare [30]), we are interested in the extension to the case where a dynamical model is applied between different iteration steps. This is of high interest to many researchers and practitioners, the use of multiscale approaches in a cycled environment will be part of our future work.

References

[1] In Timothy J. Barth, Tony Chan, and Robert Haimes, editors, *Multiscale and Multiresolution Methods: Theory and Applications*, volume 20 of *Lecture Notes in Computational Science and Engineering*, pages X, 394. Springer-Verlag Berlin Heidelberg, 2002.

[2] In Alexander Mielke, editor, *Analysis, Modeling and Simulation of Multiscale Problems*, pages XVIII,697. Springer Berlin Heidelberg, 2006.

[3] In Ronald DeVore and Angela Kunoth, editors, *Multiscale, Nonlinear and Adaptive Approximation: Dedicated to Wolfgang Dahmen on the Occasion of his 60th Birthday*, pages XXIV,660. Springer Berlin Heidelberg, 2009.

[4] In Björn Engquist, Olof Runborg, and Yen-Hsi R Tsai, editors, *Numerical Analysis of Multiscale Computations: Proceedings of a Winter Workshop at the Banff International Research Station 2009*, volume 82 of *Lecture Notes in Computational Science and Engineering*, pages X,430. Springer-Verlag Berlin Heidelberg, 2012.

[5] Habib Ammari and Hyeonbae Kang. *Inverse Problems, Multi-scale Analysis and Effective Medium Theory: Workshop in Seoul, Inverse Problems, Multi-scale Analysis, and Homogenization, June 22-24, 2005, Seoul National University, Seoul, Korea*, volume 408. American Mathematical Soc., 2006.

[6] Jeffery L Anderson. An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus A*, 59(2):210–224, 2007.

[7] Necdet Bildik and Sinan Deniz. Comparative study between optimal homotopy asymptotic method and perturbation-iteration technique for different types of nonlinear equations. *Iranian Journal of Science and Technology, Transactions A: Science*, pages 1–8, 2016.

[8] Craig H Bishop and Daniel Hodyss. Ensemble covariances adaptively localized with eco-rap. Part 1: tests on simple error models. *Tellus A*, 61(1):84–96, 2009.

[9] Martin Burger, Hendrik Dirks, and jahn Mueller. Inverse problems in imaging. In Mike Cullen, Melina A Freitag, Stefan Kindermann, and Robert Scheichl, editors, *Large Scale Inverse Problems*, volume 13 of *Radon Series on Computational and Applied Mathematics*. Walter de Gruyter, Berlin, 2013.

[10] William F Campbell, Craig H Bishop, and Daniel Hodyss. Vertical covariance localization for satellite radiances in ensemble Kalman filters. *Monthly Weather Review*, 138(1):282–290, 2010.

[11] Li Cao, Bo Han, and Wei Wang. Homotopy perturbation method for nonlinear ill-posed operator equations. *International Journal of Nonlinear Sciences and Numerical Simulation*, 10:1319–1322, 2009.

[12] E.Miller and A. Willsky. A multiscale approach to solving one dimensional inverse problems. In *Time-Frequency and Time-Scale Analysis, 1992., Proceedings of the IEEE-SP International Symposium*, pages 129–132, Oct 1992.

[13] Heinz W Engl, Martin Hankle, and Andreas Neubauer. *Regularization of Inverse Problems*. Mathematics and Its Applications. Springer Netherlands, 2000.

[14] Bjorn Engquist and Christina Frederick. Nonuniform sampling and multiscale computation. *Multiscale Modeling & Simulation*, 12(4):1890–1901, 2014.

[15] R. Escalante and M. Raydan. *Alternating Projection Methods*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2011.

[16] Christina Frederick and Bjorn Engquist. Numerical methods for multiscale inverse problems. *arXiv preprint arXiv:1401.2431*, 2014.

[17] Melina Freitag and Roland Potthast. Synergy of inverse problems and data assimilation techniques. In Mike Cullen, Melina A Freitag, Stefan Kindermann, and Robert Scheichl, editors, *Large Scale Inverse Problems*, volume 13 of *Radon Series on Computational and Applied Mathematics*. Walter de Gruyter, Berlin, 2013.

[18] Charles W Groetsch. *Inverse problems in the mathematical sciences*. Theory and Practice of Applied Geophysics Series. Vieweg, 1993.

[19] A.K. Gupta and S. Saha Ray. Comparison between homotopy perturbation method and optimal homotopy asymptotic method for the soliton solutions of boussinesqburger equations. *Computers & Fluids*, 103:34 – 41, 2014.

[20] Tijana Janjic, Lars Nerger, Alberta Albertella, Jens Schröter, and Sergey Skachko. On domain localization in ensemble-based Kalman filter algorithms. *Monthly Weather Review*, 139(7):2046–2060, 2011.

[21] B. Kaltenbacher, A Neubauer, and O Scherzer. *Iterative Regularization Methods for Nonlinear Ill-Posed Problems*. Radon Series on Computational and Applied Mathematics. De Gruyter, 2008.

[22] Andreas Kirsch. *An Introduction to the Mathematical Theory of Inverse Problems*. Applied Mathematical Sciences. Springer-Verlag New York, 2011.

[23] Yoonsang Lee and Andrew J. Majda. Multiscale methods for data assimilation in turbulent systems. *Multiscale Modeling & Simulation*, 13(2):691–713, 2015.

[24] Zhiyong Meng and Fuqing Zhang. Limited-area ensemble-based data assimilation. *Monthly Weather Review*, 139(7):2025–2045, 2011.

[25] T. Miyoshi and K. Kondo. A multi-scale localization approach to an ensemble kalman filter. *SOLA*, 9:170–173, 2013.

[26] Nicola Montaldo and John D Albertson. Multi-scale assimilation of surface soil moisture data for robust root zone moisture predictions. *Advances in water resources*, 26(1):33–44, 2003.

[27] Carsten Montzka, Valentijn R.N.Pauwels, Harrie-Jan Hendricks Franssen, Xujun Han, and Harry Vereecken. Multivariate and multiscale data assimilation in terrestrial systems: A review. *Sensors*, 12(12):16291–16333, 2012.

[28] Siegfried Mller. *Adaptive Multiscale Schemes for Conservation Laws*, volume 27 of *Lecture Notes in Computational Science and Engineering*. Springer Berlin Heidelberg, 2003.

[29] Aamir Nadeem and Roland Potthast. Transformed and generalized localization for ensemble methods in data assimilation. *Mathematical Methods in the Applied Sciences*, 39(4):619–634, 2016.

[30] Gen Nakamura and Roland Potthast. *Inverse Modeling*. 2053-2563. IOP Publishing, 2015.

[31] James Nolen, Grigorios A. Pavliotis, and Andrew M. Stuart. Multiscale modelling and inverse problems. In Ivan G. Graham, Thomas Y. Hou, Omar Lakkis, and Robert Scheichl, editors, *Numerical Analysis of Multiscale Problems*, volume 83 of

Lecture Notes in Computational Science and Engineering, pages 1–34. Springer Berlin Heidelberg, 2012.

[32] Ming Pan and Eric F Wood. A multiscale ensemble filtering system for hydrologic data assimilation. part ii: Application to land surface modeling with satellite rainfall forcing. *Journal of Hydrometeorology*, 10(6):1493–1506, 2009.

[33] Ming Pan, Eric F. Wood, Dennis B.McLaughlin, Dara Entekhabi, and Lifeng Luo. A multiscale ensemble filtering system for hydrologic data assimilation. part i: Implementation and synthetic experiment. *Journal of Hydrometeorology*, 10(3):794–806, 2009.

[34] Clemens Pechstein. *Finite and Boundary Element Tearing and Interconnecting Solvers for Multiscale Problems*, volume 90 of *Lecture Notes in Computational Science and Engineering*. Springer-Verlag Berlin Heidelberg, 2013.

[35] A Perianez, H Reich, and R Potthast. Optimal localization for ensemble Kalman filter systems. *Journal of the Meteorological Society of Japan*, 92(6):585–597, 2014.

[36] Santha R.Akella et al. Reservoir multiscale data assimilation using the ensemble kalman filter. *Applied Mathematics*, 2(02):165–180, 2011.

[37] David R.Stauffer and Nelson L.Seaman. Multiscale four-dimensional data assimilation. *Journal of Applied Meteorology*, 33(3):416–434, 1994.

[38] Jiang Wan and Nicholas Zabaras. A bayesian approach to multiscale inverse problems using the sequential monte carlo method. *Inverse Problems*, 27(10):105004, 2011.

[39] Y.Xie, S.Koch, J.McGinley, S.Albers, P.E.Bieringer, M.Wolfson, and M.Chan. A space-time multiscale analysis system: A sequential variational analysis approach. *Monthly Weather Review*, 139(4):1224–1240, 2011.

[40] Yuhua Zhou, Dennis McLaughlin, Dara Entekhabi, and Gene-Hua Crystal Ng. An ensemble multiscale filter for large nonlinear data assimilation problems. *Monthly Weather Review*, 136(2):678–698, 2008.

[41] Yu Zou and Roger Ghanem. A multiscale data assimilation with the ensemble kalman filter. *Multiscale Modeling & Simulation*, 3(1):131–150, 2005.