Under review, "Circuits, Systems, and Signal Processing," special issue on Advanced Signal / Image Reconstruction

An adaptive, multiscale inverse scattering approach to photothermal depth profilometry

Eric L. Miller and Ibrahim Yavuz
Dept. of Electrical and Computer Engineering
Northeastern University,
235 Forsyth Building
Boston MA, 02115, USA
Tel: (617) 373-8386 Fax: (617) 373-8627
e-mail: elmiller@ece.neu.edu

Lena Nicolaides and Andreas Mandelis
Photothermal and Optoelectronic Diagnostics Laboratories
Dept. of Mechanical Engineering, Univ. of Toronto
5 King's College Road, Toronto M5S 3G8, CANADA
Phone/Fax: (416) 978-5106
e-mail: mandelis@mie.utoronto.ca

November 3, 1999

Abstract

Photothermal depth profilometry is formulated as a non-linear inverse scattering problem. Starting with the one dimensional heat diffusion equation, we derive a mathematical model relating arbitrary variation in the depth-dependent thermal conductivity to observed thermal wavefields at the surface of a material sample. The form of the model is particularly convenient for incorporation into a non-linear optimization framework for recovering the conductivity based on thermal wave data obtained at multiple frequencies. We develop an adaptive, multi-scale algorithm for solving this highly ill-posed inverse problem. The algorithm is designed to produce an accurate, low order representation of the thermal conductivity by automatically controlling the level of detail in the reconstruction. This control is designed to reflect both (1) the nature of the underlying physics which says scale should decrease with depth and (2) the particular structure of the conductivity profile which may require a sparse collection of fine scale components to adequately represent significant features such as a layering structure. The approach is demonstrated in a variety of synthetic examples representative of non-destructive evaluation problems seen in the steel industry.

Keywords: Adaptive signal representation, multiscale methods, B-splines, non-linear inverse scattering, photothermal depth profiling.
Running title: Multiscale adaptive inverse scattering

1 Introduction

One of the most promising applications of photothermal non-destructive evaluation (NDE) technology is in the area of *depth profilometry* (DP) [4, 8–13]. This techniques uses a modulated laser beam to induce thermal waves in a material sample with the resulting wavefield amplitude and phase measured at the surface. The experiment is repeated for a number of modulation frequencies yielding a spectrum of observations. Assuming that the structure varies only with depth, the problem is to convert the measured spectrum into a depth-varying profile of a relevant physical property such as thermal conductivity [8], thermal effusivity [9], or thermal diffusivity [10, 12, 13].

Considerable work has been done in developing physical models and algorithms based on these models for reconstructing profiles. All models reported in the literature are based on the one dimensional, heat diffusion equation (HDE). In much of this work, the material is considered to be composed of layers where the physical properties assume simple functional forms. For piecewise constant layers, the HDE can be solved in terms of a recursively defined generalized reflection coefficient [4, 9]. In the limit of infinitesimally thin layers, the recursion converges to a non-linear Ricatti equation. For finitely thick, piecewise linear layers the solution is given via a different recursively defined reflection coefficient [8].

An alternate approach based on the notion of a thermal harmonic oscillator (THO) has been pioneered by Mandelis and collaborators [11, 12]. Making a WKB approximation for variations of the thermal effusivity yields solutions to the HDE in terms of integrals of the thermal diffusivity. Using a particular functional form for this profile, analytic expressions have been reported for a semi-infinite medium and a finite depth inhomogeneity backed by a semi-infinite substrate [10].

Each of these physical models has been used as the basis for an inversion algorithm. Under suitable approximations (most notably, ignoring the non-linear term) the Ricatti equation model reduces to a Laplace transform type relationship between the derivative of the log-effusivity and

the data [4, 9]. Tikhonov regularized, linear inversion methods are then used to determine the log effusivity [9]. While the details of the inversion methods for the piecewise constant and THO models are different, they both share a common structure in which the data are processed progressively from high to low modulation frequencies. The higher frequencies are used to estimate the profile structure in slices of the material closest to the surface while the lower frequency data provide information concerning the deeper variations. For the piecewise constant profile method, the width of these slices is determined *a priori* [8]. The THO method locally fits profiles of a predefined structure to slices whose widths are determined adaptively as the algorithm progresses [10, 13].

We pursue a different approach to the DP problem. Rather than considering solutions to the HDE for specific profiles or approximations to the physics, we employ a numerical implementation of the HDE which allows for exact solutions (up to discretization error) for arbitrary profiles. In particular, it is not necessary to assume that the thermal conductivity is slowly varying. The price for this flexibility is increased computational complexity. Thus, we also introduce a highly efficient, non-linear approximation to the exact solution. For the problem of interest in this paper, we show that this method, which is similar in nature to the Extended Born Approximation [5, 18], is about 20 times faster than the exact model with a loss in accuracy close to machine precision.

The use of this physical model allows us to formulate the profile reconstruction process as a non-linear inverse scattering problem in a highly lossy medium. Problems of this type are known to be ill-posed in that small perturbations in the data can lead to large amplitude, non-physical artifacts in a reconstruction. To stabilize the inversion process, one typically uses a *regularization* procedure. Our previous efforts in this field have included the use of wavelet-based regularization techniques for two dimensional, single frequency, non-linear inverse scattering problems [14, 15]. The use of wavelets was motivated by analysis which indicated that a given level of accuracy in a reconstruction required that the resolution in the estimated profile be space varying. The

data and the physics supported fine scale estimates in regions close to the detectors while coarser scale information could only be recovered further into the material sample. Representation of the unknown image using an orthonormal wavelet basis provided a convenient means of enforcing this variable resolution. Regularization was achieved in part by concentrating the information in the data on the recovery of a low order representation of the unknown comprised of relatively few fine scale wavelet coefficients supplemented by a small number of coarse scale coefficients

For the DP problem we build on this work using an alternate multiscale representation of the unknown profile. Rather than an orthonormal basis of wavelets, we use a collection of spline functions possessing a natural multiscale structure. Specifically, coarse scale elements can be expressed as linear combinations of finer scale splines [17]. By surrendering the orthonormality property of wavelets, we have a more flexible method for modeling the unknown profile.

As in [14] our inversion approach here is designed to produce a low order reconstruction in which the distribution of fine scale detail is determined in an automatic, controlled, and rational manner. The precise method for accomplishing this, however, is substantially different from our previous wavelet related efforts. Starting with a coarse scale set of functions, we iteratively refine the reconstruction to (a) add detail and then (b) prune away unnecessary degrees of freedom to obtain a more parsimonious description of the profile. The final estimate is comprised of functions at many scales whose spatial distribution is dictated adaptively by the data and the physical model. Like the results in e.g. [10] we tend to see fine scale detail close to the surface with coarser scale elements used deeper into the material. However, our approach involves no decomposition of the material into slices (predefined or virtual). Unlike the techniques in [8, 10] where low frequency data play no role in the structure of the reconstruction near the surface, here the data from all frequencies impact the entire structure of the estimated profile.

In §4 we show that limiting the number of degrees of freedom in the reconstruction is useful

for a number of reasons. First, the low dimensionality can lead to computational efficiencies. We demonstrate via numerical experiments that the adaptive algorithm can be substantially faster than a brute-force, fine scale reconstruction with little loss in accuracy. Additionally, our approach can lead to better reconstructions. As described in greater detail in §3.1, inversion requires the solution of a non-linear optimization problem and is thus prone to convergence to local minima of an associated cost function. By constraining the reconstruction using our adaptive approach, we show that it is also possible to converge to a lower cost point in solution space than is the case for a high dimensional, fine scale approach to inversion. Currently, we have observed these advantages only through simulation. An area of work is in the development of a more rigorous theory to understand the conditions where we might expect improved computational and/or reconstruction performance.

The remainder of this paper is organized as follows. In § 2, the physical model is developed. Our approach to inversion is provided in § 3. Here we formulate the basic problem in an inverse scattering context, define the spline functions, and describe the details of the adaptive algorithm. In § 4 we present examples of this method. Conclusions and future work are the subject of § 5.

2 Formulation of the Depth Profilometry Problem

2.1 Continuous Formulation

The DP problem of interest in this paper is the reconstruction of the thermal diffusivity profile based on observations of thermal waves obtained from illumination by a modulated laser source. The starting point is the one dimensional HDE with boundary conditions stated as follows:

$$\frac{d}{dz}\kappa(z)\frac{d}{dz}T(z) - i\omega\rho(z)c(z)T(z) = 0 \tag{1}$$

$$-\kappa(z)\frac{d}{dz}T(z) = \frac{1}{2}Q \qquad z = 0 \tag{2}$$

$$T(z) = 0 \qquad z \to \infty \tag{3}$$

where κ is the thermal conductivity, ρ the density and c the specific heat. $T(z)$ is the thermal wavefield, ω the angular modulation frequency, and Q the intensity of the laser light incident at the surface which is taken to be at $z = 0$. An $e^{i\omega t}$ time dependence for the source and T is assumed.

To produce a form of the model useful in an inverse scattering context we start by decomposing $\kappa(z)$ into the sum of a background component, $\kappa_b(z)$, and a perturbation of arbitrary magnitude, $\kappa_p(z)$. Generally the background is taken to be a nominal profile such as a halfspace as in [8, 12] or a well characterized material sample backed by air [10]. The goal of the inverse problem is the reconstruction of $\kappa_p(z)$[1]. Inserting $\kappa = \kappa_b + \kappa_p$ into (1), we write the overall model as

$$(\mathcal{F}_b + \mathcal{F}_p)T = v \tag{4}$$

where we have defined the matrices of linear operators, $\mathcal{F}_b$ and $\mathcal{F}_p$, and the source vector, v, as

$$\mathcal{F}_b = \begin{bmatrix} \frac{d}{dz}\kappa_b(z)\frac{d}{dz} - i\omega\rho(z)c(z) \\ b_0 \\ b_\infty \end{bmatrix} \quad \mathcal{F}_p = \begin{bmatrix} \frac{d}{dz}\kappa_p(z)\frac{d}{dz} \\ 0 \\ 0 \end{bmatrix} \quad v = \begin{bmatrix} 0 \\ \frac{1}{2}Q \\ 0 \end{bmatrix}. \tag{5}$$

The linear operator b_0 enforces the boundary condition, (2), and b_∞ implements (3).

We manipulate (4) to obtain a solution of the HDE in the form

$$T = (I + \mathcal{G}_b\mathcal{F}_p)^{-1}\,\mathcal{G}_b v \tag{6}$$

with $\mathcal{G}_b \equiv \mathcal{F}_b^{-1}$ the Green's operator for the boundary value problem (1)– (3) with $\kappa = \kappa_b$. Borrowing from the inverse scattering literature, the field computed with $\kappa_p = 0$ is termed the background field and is given as $T_b \equiv \mathcal{G}_b v$. The scattered field $T_s \equiv T - T_b$ is then

$$T_s = \mathcal{G}_b\mathcal{F}_p T = \mathcal{G}_b \frac{d}{dz}\kappa_p \frac{d}{dz} T \tag{7}$$

$$= \mathcal{G}_b\mathcal{F}_p\,(I + \mathcal{G}_b\mathcal{F}_p)^{-1}\,\mathcal{G}_b v. \tag{8}$$

where (7) follows from the definition of $\mathcal{F}_p$ in (5) and (6) is used to obtain (8).

[1] In theory, the specification of $\kappa_b(z)$ is somewhat arbitrary as the algorithm is designed to produce a κ_p such that $\kappa_b + \kappa_p$ is as accurate as possible. However, making κ_p "small" through as precise a description of κ_b as is possible does practically improve the convergence of the inversion.

The goal of the DP problem is to recover κ_p from observations of T_s taken at $z = 0$ for multiple modulation frequencies. In preparation for a description of the inversion algorithm in § 3, (7) indicates that $T_s(0)$ can be written as a linear functional of κ_p in which the functional explicitly depends on $\mathcal{G}_b$ and dT/dz. Because $T(z)$ itself is dependent on κ_p, the functional also implicitly depends on the thermal conductivity. Thus letting y_k be the complex value datum taken at the kth frequency ω_k, we write

$$y_k = c_k(\kappa_p)\kappa_p \tag{9}$$

where c_k is the linear functional obtained from (7). In § 3.2 we provide a explicit formula for c_k in the context of a discretized model. The data vector comprised of y_k for $k = 1, 2, \ldots, N_y$ is

$$y = \mathcal{C}(\kappa_p)\kappa_p \tag{10}$$

with $\mathcal{C}$ being the column vector of c_k's. The DP problem of interest in this work then is the recovery of κ_p from knowledge of y as well as the HDE model.

2.2 Discretization

To implement a solution to the DP problem requires a discretized form of the physical model. Here we use a standard first-order finite difference scheme for the HDE in (1)– (3). The field $T(z)$, and the physical properties $\kappa(z)$, $\rho(z)$ and $c(z)$ are sampled on a uniform grid of spacing h. First order differentiation w.r.t z then is represented using a bi-diagonal matrix, D_z, with $1/h$ on the main diagonal and $-1/h$ on the first super diagonal. For the homogeneous Neumann boundary condition, (3), we set to zero the last element in the vector of samples for the field T where we ensure that the boundary is taken far enough from the surface so that numerical artifacts are negligible.

To construct, the discrete form of the c_k in (9), consider the expression for T_s in (7). Under the discrete model we have

$$T_{s,k} = G_{b,k} D_z \mathcal{D}(\kappa_p) D_z T_k \tag{11}$$

with $G_{b,k}$ the matrix form of Green's operator for the k-th frequency, T_k the vector of samples of

the thermal field and similarly for κ_p and $T_{s,k}$. For a vector x, $\mathcal{D}(x)$ is the diagonal matrix whose (ii)th element is x_i. After some tedious but elementary algebra, it is not hard to show that from (11) we can write the surface value of the scattered wavefield under the discretized model as

$$T_s(z=0) = G_{b,k}(1,:)\mathcal{D}(T_k)D_z\mathcal{D}(D_zT_k)\kappa_p \equiv c_k(\kappa_p)\kappa_p \tag{12}$$

with $G_{b,k}(1,:)$ being the first row of the matrix G_b. From (12) we conclude that the discrete representation for the linear functional $c_k(\kappa_p)$ in (9) is the row vector $G_{b,k}(1,:)\mathcal{D}(T_k)D_z\mathcal{D}(D_zT_k)$.

The most computationally intensive part of this forward model is determination of the vector T_k. From (6), the discretized form of this this calculation requires the solution of the linear system

$$(I + G_{b,k}F_p)T_k = G_{b,k}v. \tag{13}$$

Because the inversion algorithm requires systems of this form be solved hundreds if not thousands of times, there is significant motivation for seeking accurate approximations to the determination of T_k. The similarity of the DP problem to inverse electrical conductivity problems found in the geophysical community, lead us to consider a so-called Extended Born Approximation (EBA) to (13). A detailed theoretical treatment of this method may be found in [5, 18]. The practical implementation of the EBA for our problem amount to replacing the full matrix $I + G_{b,k}F_p$ by only its diagonal elements thereby greatly reducing the number of operations need to determine T_k. In § 4, we provide experimental verification of both the accuracy and efficiency of this method.

3 A Scale-Adaptive Algorithm for the DP Problem

We start by considering the fixed-scale solution to the DP problem in which we seek to recover κ_p directly from the data. The machinery developed here then forms the basis for the adaptive method developed later in this section.

3.1 The Fixed-Scale Solution

As is typically done for an inverse problem of the type considered here [15, 16, 18], we use our discretized model to define the estimate of κ_p as the solution to the following regularized least

squares cost function:

$$\hat{\kappa}_p = \arg\min_{\kappa_p} \|y - C(\kappa_p)\kappa_p\|_2^2 + \lambda^2 \|D_z \kappa_p\|_p^p. \tag{14}$$

The first term in (14) ensures that the estimate is consistent with the data. The second term plays the role of a regularizer and is used to help combat the ill-posedness. By varying p in the range 1 to 2 one can control the smoothness in the reconstruction. For $p = 2$, one obtains a traditional smoothness regularizer. As p approaches 1, the regularizer is more encouraging of forming profiles which have edges or other sharp discontinuities as might be found in certain NDE applications. With $p = 1$, one has a total variation (TV) regularization scheme [7, 19].

The regularization parameter, λ, in (14) determines the relative importance of the two terms on $\hat{\kappa}_p$. As $\lambda \to 0$, we demand that $\hat{\kappa}_p$ just fit the data. As mentioned in § 1, the resulting estimate tends to display high frequency, large amplitude artifacts. On the other hand, as $\lambda \to \infty$, the data play a limited role in influencing $\hat{\kappa}_p$ and we obtain overly smooth estimates. Proper selection of this parameter is a non-trivial problem [1, 6]. In this paper, we set λ by trial and error.

To solve (14) we start by writing $\|D_z \kappa_p\|_p^p = g(\kappa_p)^T g(\kappa_p)$ with g the vector whose ith element is $[D_z \kappa_p]_i^{p/2}$. Thus (14) is formulated as a non-linear least squares optimization problem:

$$\hat{\kappa}_p = \arg\min_{\kappa_p} e^T(\kappa_p) e(\kappa_p) \tag{15}$$

$$e(\kappa_p) = \begin{bmatrix} y - C(\kappa_p)\kappa_p \\ \lambda g(\kappa_p) \end{bmatrix} \tag{16}$$

Eq. (15) is solved using the `leastsq` routine provided in Matlab's Optimization Toolbox. For a given κ_p this program requires the ability to evaluate the RHS in (15) and a method for computing the Jacobian of the vector e in (16) with respect to each element of κ_p. The cost evaluation is done using the equations developed in the previous section. We approximate the Jacobian, J, as

$$J = \begin{bmatrix} -C(\kappa_p) \\ \frac{\partial g}{\partial \kappa_p} \end{bmatrix}. \tag{17}$$

After some straightforward calculus, it is not hard to show that

$$\frac{\partial g}{\partial \kappa_p} = \frac{p}{2}\mathcal{D}\left(g\right)^{-p/2}\mathcal{D}\left(\text{sign}(g)\right)D_z$$

with sign(g) the vector whose ith element is the signum of g_i. Note that in (17) we ignore the dependence of C on κ_p. This is the primary motivation for formulating the physical problem in the manner of §2. This approximation is closely related to the Born Iterative Method [2] for problems in which the ρc product in (1) is to be recovered rather than κ.

3.2 Adaptive Multiscale Inversion

As discussed in §1, a key issue of interest in this paper is the determination of a low order, multiscale representation for the vector κ_p as a means of reducing the complexity of the inversion process and improving the quality of the ultimate reconstruction by "appropriately" distributing fine scale information. To achieve this, we constrain κ_p to live in the linear span of a set of vectors where we adaptively determine both the vectors and the required expansion coefficients. Formally, we consider representation of κ_p of the form

$$\kappa_p = \sum_{i=1}^{N_b} b_i a_i = Ba \tag{18}$$

where b_i is the ith vector, a_i the expansion coefficient, B the matrix whose ith column is b_i, and a the vector of a_i. Note that a is of length N_b.

Using (18) and (15), the determination of $\hat{\kappa}_p$ now reduces to the estimation of the a_i via

$$\hat{a} = \arg\min_a e^T(Ba)e(Ba) = \arg\min_a \|e(Ba)\|_2^2 \tag{19}$$

where $e(Ba)$ is (16) evaluated at the vector $\kappa_p = Ba$. Letting N be the length of κ_p, then typically we have $N_b << N$ so that (19) is a far smaller optimization problem than (15). In the remainder of this section, we describe the family of vectors used in our approach as well as the methods we have developed to adaptively determine those b_i to include in $\hat{\kappa}_p$.

The b_i vectors are sampled versions of a family of continuous functions which satisfy the following

two scale dilation equation [17]

$$b(z) = \sum_{k=0}^{N_k-1} \alpha_k b(2z-k). \tag{20}$$

In words, (20) says that the function b can be expressed as a linear combination of dilated and translated versions, $b(2z-k)$, of itself. While specification of the coefficients α_i can lead to many functions which satisfy (20), including wavelets, here we are concerned with B-splines [17, Chap. 7.4]. For example, with $\alpha_0 = \alpha_1 = 1$, the $b(z)$ satisfying (20) is nothing more than the box function from $z = 0$ to $z = 1$. Such zeroth order splines form a basis for piecewise constant functions. With $\alpha_0 = \alpha_2 = 1$ and $\alpha_1 = 2$, we obtain a hat function, or first order B-spline, which is piecewise linear.

Here we employ piecewise quadratic B-splines generated using $\alpha_0 = \alpha_3 = 1$ and $\alpha_1 = \alpha_2 = 3$. In Figure 1, we illustrate these functions on two spatial scales: a coarse scale at the top and a finer scale at the bottom. Note that the relative spacing of the spline functions is proportional to their width. Additionally, because these functions satisfy (20), wide functions on one scale can be expressed as linear combinations of narrow functions on the next scale. For example, the highlighted element shown at the top of Figure 1 can be decomposed via (20) in terms of the fine scale functions in the bottom panel. We represent the relationship among these functions using a graph as shown in Figure 2. The nodes in the graph represent the functions (or associated expansion coefficients) and the edges indicate the links from one scale to the next.

We make use of the non-linear least squares formulation of § 3.1 and the functions defined above to obtain a reconstruction of κ_p using the algorithm outlined in Fig. 3. We start with a low order, coarse scale collection of quadratic B-splines. With these b_i, we estimate a set of expansion coefficients to get a rough $\hat{\kappa}_p$. The remainder of the algorithm is a loop where the current collection of vectors is alternately refined and then unneeded detail is pruned away. The goal is to provide sufficient flexibility to add arbitrary detail to $\hat{\kappa}_p$ followed by a stage in which we determine which, if any, degrees of freedom associated with this detail were warranted. Those not needed are removed.

While one might consider many methods to accomplish this objective, in the following paragraphs, we detail an approach which we have found to work well at least for the DP problem.

In the refinement stage, we replace all b_i in our representation of $\hat{\kappa}_p$ by their finer scale children[2]. This allows for the recovery of finer scale information at the expense of a higher order representation. Once the new collection of vectors is constructed, we solve for a new vector of expansion coefficients to produce a finer scale estimate of κ_p. While this requires the solution of a new non-linear least squares problem, the coarse scale estimate is used to initialize the finer scale reconstruction procedure to speed convergence. Specifically, the a_i for each child is initialized using the average of the previously generated parent estimates. From Figure 2, assume that vectors 3-8 had just been replaced by 9-22. Then, e.g., the initial values for a_{13} and a_{14} would both be $\frac{a_4+a_5}{2}$.

After refining the estimate, we next remove unnecessary detail. The goal here is to reduce the complexity of the model by

- **Coarsening** Replacing fine scale children with their parents in regions where such detail is not warranted
- **Pruning** Removing from the representation vectors which contribute little to $\hat{\kappa}_p$

Coarsening is done in a sequence of steps in which limited detail is removed at each stage until we can no longer represent the current estimate of κ_p to within some predetermined tolerance. A similar procedure is carried out for pruning except that we remove from the estimate in a one-by-one manner vectors whose $|a_i|$ are small. The result is a lower order model which is a "small" perturbation of the optimal estimate computed from the previous optimization step.

As illustrated in Fig. 4, we begin coarsening by looking for all collections vectors in our current B matrix which represent a complete set of children for a given parent. For example, in Figure 2, vectors 13-16 are a complete set of children for 5. For each such children-parent set, we consider

[2]Obviously, those b_i already at the finest scale are not refined.

a new collection of vectors obtained by replacing those fine scale children with *all* of their coarse scale parents keeping the remaining members of the current collection the same. Again, referring to Figure 2 suppose that the current B matrix was comprised of vectors 9-22. One set of vectors considered by the pruning process would be 9–12, 4–6, and 17–22. Here 13–16 are removed and replaced by all relevant parents, in this case 4, 5, and 6. Another possible configuration is 9, 10, 3–5, and 15–22 where 11–14 are replaced by 3–5.

This replacement strategy yields some finite number of possible sets of vectors for representing κ_p with each set having fewer members than the current one. To determine which, if any collection is selected, we project $\hat{\kappa}_p$ onto the linear span of the vectors in each set. Let B_i be the matrix constructed from one such coarse set of vectors. The projection, $\hat{\kappa}_{p,i}$, is computed as $\hat{\kappa}_{p,i} = B_i B_i^\dagger \hat{\kappa}_p$ where $B_i^\dagger$ is the pseudo-inverse of B_i [3, p. 243][3]. As explained in Appendix 6, to evaluate the utility of $\hat{\kappa}_{p,i}$ it is useful to compute a weighted error, $\|J(\hat{\kappa}_p - \hat{\kappa}_{p,i})\|$ where J is given in (17) and is evaluated at $\hat{\kappa}_p$. If the minimum weighted error over all i is less than some threshold, we replace the fine scale collection with the elements from the new, coarsened set, otherwise we stop. If we do not stop, then the pruning process is repeated. Specifically, if B_j^* denotes the selected set, we replace children in B_j^* by parents again using the relative norm criteria just described. If the minimum error still is not too small, B_j^* is replaced by a new, smaller set and the process continues.

To do pruning, we start by removing from the final set of vectors produced by coarsening that vector whose $|a_i|$ is smallest. As is done in pruning, we look at the weighted error between $\hat{\kappa}_p$ and its projection onto the remaining vectors. If this error exceeds a threshold, then we stop, otherwise, the procedure continues using the expansion coefficients associated with this projection.

After pruning is finished, we pass the new collection of expansion vectors into the non-linear

[3]Because we allow B_i to contain B-spline vectors at multiple scales and arbitrary dyadic shifts, it is possible that elements of B_j will be linearly dependent hence the need to use a pseudo-inverse. Both theoretically and practically this linear dependence causes no difficulty for the algorithm in this paper.

least squares solver to obtain the best estimate of the conductivity using this coarsened set. Again, another optimization is performed. As described in the previous paragraphs however, the pruning process yields a low order collection of vector which still supports an accurate approximation of $\hat{\kappa}_p$. Thus, it is reasonable to assume that the set of optimal expansion coefficients should be close to the coefficients obtained by projecting $\hat{\kappa}_p$ onto the coarse set. Throughout our numerical experiments, we have found that initializing in this way does in fact lead to rapid convergence of the non-linear least squares routine for this stage of the algorithm.

This refinement-pruning loop is continued until the relative difference of the two-norm between estimates produced at the end of two successive pruning stages is below some threshold. We note that this approach does not guarantee a monotonic decrease in the cost function as we change the vectors in B. Indeed, we have seen in our experiments that the optimal κ_p obtained after a coarsening/pruning step can have a higher cost than the previous $\hat{\kappa}_p$ comprised of a larger number of expansion vectors. As examined in Appendix 6 however, the choice of weighted error criteria for accepting a coarsened or pruned collection of vectors is constructed in a way to limit the resulting rise in cost. Thus, while the cost may increase, we guarantee that this increase is bounded.

4 Examples

Here we consider the performance of our inversion approach on a variety of profiles. The background is meant to model case hardened steel as in [8] so $\kappa_b = 0.45$ W cm^{-1} K^{-1}, $\rho = 7.7$ g cm^{-3}, and $c = 0.48$ J g^{-1} K^{-1}. In all cases depth profiles are constructed for z between 0 and 400 μm from the surface based on thermal wave data from 16 logarithmically spaced frequencies in the range 10 to 10000 s^{-1}. The signal to noise ratio on T_s is taken to be 30 dB.

Over the 400 μm range of interest, we discretize the forward model on a grid of 256 equally spaced points corresponding to an interpoint spacing of about 1.57 μm. All data vectors are

generated using the exact, discretized forward model while the inversion algorithm makes use of the EBA-type model described at the end of §2.2. For all examples, ϵ,the tolerance parameters used in the coarsening and pruning stages was $1e-4$ The threshold used to terminate the refinement-pruning loop was 5%. In all cases, $\lambda = 1$ (this value was determined by trial and error) and the reconstruction algorithms were initialized with zero for the perturbation. Finally, there were a total of five scales worth of B-spline functions that could be used in the construction of $\hat{\kappa}_p$.

A Monte-Carlo analysis was used to evaluate the Extended Born Approximation to the true scattering model. We started by generating 100 random κ_p profiles via

$$\kappa_p = \sum_{i=1}^{17} a_i b_i$$

where the b_i are a collection "mid-scale" B-splines shown in Figure 5(a) and the a_i were taken to be zero mean, independent Gaussian random variables with standard deviation 20[4]. The goal here was to produce profiles which were not merely small perturbations on the background. Some samples are shown in Figure 5(b). For each profile, we solved the forward problem using the exact method and the EBA. Over the 100 trials, the average relative error between the exact and approximate data vectors was about 10^{-13} while the EBA required about 18 times fewer floating point operations than the exact model. Thus, for the problems of interest here, the EBA represents a highly efficient, essentially exact solver for the HDE.

We start be examining the performance of our approach for a profile with a layer-like inclusion shown as the solid line in Figure 6(a). To gauge the average performance in the presence of noise, in this and the next example, we performed the inversion ten times for ten difference realizations of the additive disturbance. In Figure 6(a) the dashed line is a plot of the average reconstruction for the adaptive method while the results of inverting using the finest scale grid of 256 points are given by the dashed-dot line. Here we set $p = 1.1$ in (14) to obtain an edge-preserving type of regularization.

[4]Too keep κ positive, we set to zero any values of $\kappa_p < 45$, where 45 is the background thermal conductivity.

In practice it would be necessary to also develop a method for choosing this quantity; however for the sake of simplicity, as with λ, we assume that p is known *a priori*.

Figure 6(a) shows that, on average, the results of using the adaptive approach are comparable to those of a fixed, fine scale inversion. For the adaptive method, the peak value of the reconstruction is slightly closer to that of the true profile. However, for the fixed scale method, the average reconstruction is slightly flatter over the region of the layer. The average mean square error between the two methods differed by about 1% with the fine scale case having smaller average error.

As stated in the Introduction, one of the potential advantages of the adaptive method over a fixed scale inversion is computational. To explore this issue, we "instrumented" the `leastsq` procedure to collect statistics of the inversion process. For one of the ten runs in Figure 6(b) and (c), we plot the value of the cost function and the value of the error against the number of times the inversion routine made a call to evaluate the cost. Since cost evaluation requires the evaluation of the HDE model which is the most computationally intensive step in the process, the number of times the cost function is called is a valid measure of the complexity of the inversion procedure. Also, for the adaptive approach, the plot displays the cost and error for *all* refinement-pruning stages of the algorithm and thus is an accurate portrayal of the complexity of the entire routine.

We observe from Figure 6(b) and (c) that the adaptive method converges at a faster rate than the fine scale method. Thus, while the two method converge essentially to the same profile, by constraining the inversion, we get closer to the final result faster. We note that this behavior was true in general for all ten runs. Moreover, because the adaptive approach required fewer unknowns, for this problem, significantly fewer floating point operations were needed. The average number of unknowns in the final reconstruction over the ten runs was about 19 for the adaptive approach versus 256 for the fine scale method. Reduction in the size of the inverse problem translated into a floating point count that was about 2.86 times smaller for the adaptive method than the fixed scale

inversion. Thus, for this problem, the adaptive method produces essentially the same reconstruction as a fixed scale technique at a significantly reduced level of complexity.

In addition to this quantitative advantage, the adaptive inversion also yields interesting qualitative results. By examining *which* vectors are used in the reconstruction, we see that the adaptive approach placed limited detail precisely where it belongs in a reconstruction using fewer, coarser scale vectors where κ_p varies more slowly. Indeed, Figure 6(a) indicates that for both fixed and adaptive inversions, the edge closer to the surface is recovered more accurately than the edge deeper into the medium. For the same run used to produce the graphs in Figure 6, we show in Figure 7(a) the individual vectors which contribute to the final adaptive reconstruction. We see in particular that most of the fine scale detail is distributed near the leading edge with wider, coarser scale vectors used for much of the remainder of the reconstruction. This point is further highlighted in Figure 7(b) where we plot as a function of depth, the finest scale in the reconstruction. Specifically, at each point, z, there may be a number of vectors whose support includes this point. Figure 7(b) then shows the scale of the finest vector having support at each position. Larger numbers imply finer scale and for these examples, there are 5 scales of B-splines which may be used.

Typically, for a diffusive inverse problem of the type considered here, physical intuition dictates that the level of detail decreases as a function of increasing depth due to the heavy damping of the thermal waves. While this may be true in general, it also ignores the fact that one may want limited fine scale detail to support the reconstruction of important sharp features (e.g. edges) which may be present in the profile. The results in Figure 7 demonstrate that the algorithm described in this paper is capable of adapting the detail in the reconstruction in a way which reflects both the underlying physics as well as the structure of the particular profile. The general trend is a decrease in resolution as a function of depth; however, we see an increase in the level of detail specifically in the neighborhood of the edge. Thus, to obtain the shaper profile, the algorithm automatically

chooses to distribute some fine scale degrees of freedom near the feature of import.

In Figs. 8 and 9, we examine the results of the method represented in this paper for a smoother profile. For this analysis, we set $p = 2$ in (14) to obtain a more standard smoothness type of regularizer. The results here indicate a different advantage the adaptive method may have with respect to a fixed, fine scale inversion. Figure 8 shows that over the ten runs used to evaluate the performance of the algorithm, on average the fixed scale approach converges to a far from global minimum of the cost function. Alternatively, by constraining the search space, the adaptive inversion routine essentially settles into a lower cost local minimum which is also of much lower error. The convergence curves for one of the runs shown in Figure 8(b) and (c) provide further evidence of this behavior. For this problem, despite the fact that the adaptive approach uses on average about 11 vectors, approximately 16% more floating point operations are required versus the finest scale inversion. Thus, in contrast to the first example, here we see a slightly higher computational cost producing a significantly improved reconstruction.

Finally, in Figure 9, the constituent components of the one of the ten final reconstructions are displayed. We see here that the absence of an edge or other sharp feature in κ_p has resulted in coarser scale vector being used here relative to those in Figure 7. Generally, the trend is one of decreasing resolution as a function of depth.

5 Conclusions and Future Work

We have described an inverse scattering type of approach to the photothermal depth profilometry problem. Using the heat diffusion equation as the basic physical model, a measurement model is derived based on a decomposition of the thermal conductivity into the sum of a background component and an arbitrary sized perturbation. The physical model is cast in a form appropriate for use in a regularized, iterative inversion scheme. We have presented an accurate, low complexity

approximation to the exact HDE model having the form of an Extended Born Approximation.

We have introduced and explored the performance of an inversion technique designed to focus the information in the data on a small number of carefully selected degrees of freedom. Specifically, we have made use of multiscale, B-spline functions to construct an adaptive algorithm in which the presence of fine scale detail in the reconstruction is carefully controlled. The algorithm is based on an iterative loop in which the current set of vectors is first refined, to allow more detail, and subsequently pruned, to remove unwarranted degrees of freedom. The utility of this approach has been demonstrated on a number of simulated examples representative of applications of this technology to the non-destructive evaluation of case hardened steel. We plan on testing this method on experimental data in the near future.

In comparing the adaptive method to a fixed, fine scale inversion we have seen two types of behavior. For one problem the two approaches produce essentially the same reconstruction but the computational cost of the adaptive method is far smaller than that of the fixed scale technique. Alternatively, for a second problem, the adaptive method was shown to converge to a far lower cost of the optimization function; however the price paid for this improved performance was about a 20% increase in complexity. One important objective in the future is to develop a theoretical understanding of the circumstances (classes of profiles, noise conditions, sensing system parameters, etc.)under which one might expect one of these two types of performance.

6 Appendix 1

Because the basis coarsening and pruning steps remove degrees of freedom from the representation of $\hat{\kappa}_p$, it is possible value of the cost function will rise after these procedures. The objective of this appendix is to derive a method for accepting a coarsened set of vectors which guarantees that the relative rise in cost is no larger than some predefined threshold.

Referring to (15) we define $F(\kappa_p)$ as $e^T(\kappa_p)e(\kappa_p)$. Let us assume that we have a fine scale estimate, $\hat{\kappa}_p$ which we wish to coarsen. Because $\hat{\kappa}_p$ is by definition a solution to the non-linear least squares optimization problem for some B, a second order Taylor expansion of F about $\hat{\kappa}_p$ yields []

$$F(\hat{\kappa}_p + \delta) \approx F(\hat{\kappa}_p) + 1/2\delta^T H\delta \tag{21}$$

where H is the Hessian of the cost function. Since δ is small, the Hessian is well approximated by $J^T J$ [] with J the Jacobian of e evaluated at $\hat{\kappa}_p$. Thus, after some algebra we write the relative change in cost due to a perturbation δ from $\hat{\kappa}_p$ as

$$\frac{F(\hat{\kappa}_p + \delta) - F(\hat{\kappa}_p)}{F(\hat{\kappa}_p)} \approx \frac{\|J\delta\|_2^2}{2F(\hat{\kappa}_p)}. \tag{22}$$

From § 3.2, for the coarsening problem, $\delta = \hat{\kappa}_p - \hat{\kappa}_{p,i}$. Thus, to ensure that the relative change in the cost rises by at most ϵ, we require that $\|J(\hat{\kappa}_p - \hat{\kappa}_{p,i})\|_2^2 < 2\epsilon F(\hat{\kappa}_p)$.

7 Acknowledgments

The work of the first and second authors was supported by a CAREER Award from the National Science Foundation MIP-9623721, an ODDR&E MURI under Air Force Office of Scientific Research contract F49620-96-1-0028, and the Army Research Office Demining MURI under Grant DAAG55-97-1-0013. The work of the other authors was supported by a research contract from Material and Manufacturing Ontario (MMO). The first author would like to thank Dr. Oliver Dorn of Northeastern University and Prof. Misha Kilmer of Tufts University for some very helpful discussions which greatly improved the material in this manuscript.

References

[1] Murat Belge, Eric Miller, and Misha Kilmer. Simultaneous multiple regularization parameter selection by means of the l-hypersurface with applications to linear inverse problems posed in the wavelet transform domain. In *SPIE International Symposium on Optical Science, Engineering, and Instrumentation: Bayesian Inference for Inverse Problems*. SPIE, July 1998.

[2] Weng Cho Chew. *Waves and Fields in Inhomogeneous Media*. Van Nostrand Reinhold, New York, 1990.

[3] Gene H. Golub and Charles F. Van Loan. *Matrix computations*. Johns Hopkins Univ. Press,

2nd edition, 1989.

[4] P. Grossel and F. Depasse. Alternating heat diffusion in thermophysical depth profile: multilayer and continuous descriptions. *J. Phys. D: Appl. Phys.*, 31:216–223, 1998.

[5] Tarek M. Habashy, Ross W. Groom, and Brian R. Spies. Beyond the Born and Rytov approximations: A nonlinear approach to electromagnetic scattering. *J. Geophys. Res.*, 98(B2):1759–1775, February 1993.

[6] Per Christian Hansen. Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Review*, 34(4):561–580, December 1992.

[7] S. Osher L. I. Rudin and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D*, 60:259–268, 1992.

[8] T. T. N. Lan, U. Seidel, and H. G. Walther. Theory of microstructure depth profiling by photothermal measurements. *J. Appl. Phys.*, 77(9):4739–4745, 1 May 1995.

[9] Roberto LiVoti, Mario Bertolotti, and Concita Sibilia. Thermal conductivity and diffusivity depth profiles by phtothermal technqiue: The direct and inverse problem. In Xavier Maldague, editor, *III International Workshop -Advances in Signal Processing for NDE of Materials*, volume 3 of *Topics in Nondestructive Evaluation Series*, pages 379–386. The Americal Society for Nondestructive Testing, Inc., 1998.

[10] Andeas Mandelis, Frank Funak, and Mahendra Munidasa. Generalized methodology for thermal diffusivity depth profile reconstruction in semi-infinite and finitely thick inhomogenous solids. *J. Appl. Phys.*, 80(10):5570–5578, 15 Nov. 1996.

[11] Andreas Mandelis. Hamilton-Jacobi formulation and quantum theory of thermal wave propagation in the solid state. *J. Appl. Phys.*, 26(10):2676–2683, October 1985.

[12] Andreas Mandelis and Samuel B. peralta. Photoacoustic frequency domain depth profiling of continuously inhomogeneous condensed phases: Theory and simululations for the inverse problem. *J. Appl. Phys.*, 70(3):1761–1770, 1 August 1991.

[13] Andreas Mandelis, Els Schoubs, Samuel B. Peralta, and Jan Thoen. Quantitative photoacoustic depth profilometry of magnetic field-induced thermal diffusivity inhomogeneity in the liquid crystal octylcyanobiphenyl. *J. Appl. Phys.*, 70(3):1771–1777, 1 August 1991.

[14] Eric L. Miller, Lena Nicolaides, and Andreas Mandelis. Nonlinear inverse scattering methods for thermal wave slice tomography: A wavelet domain approach. *Journal of the Optical Society of America (A)*, 15(6):1545–1556, June 1998.

[15] Eric L. Miller and Alan S. Willsky. Wavelet-based methods for the nonlinear inverse scattering problem using the Extended Born Approximation. *Radio Sci.*, 31(1):51–67, Jan–Feb 1996.

[16] M. Moghaddam, W.C. Chew, and M. Oristaglio. Comparison of the Born iterative method and Tarantola's method for an electromagnetic time-domain inverse problem. *International Journal of Imagin Systems and Technology*, 3:318–333, 1991.

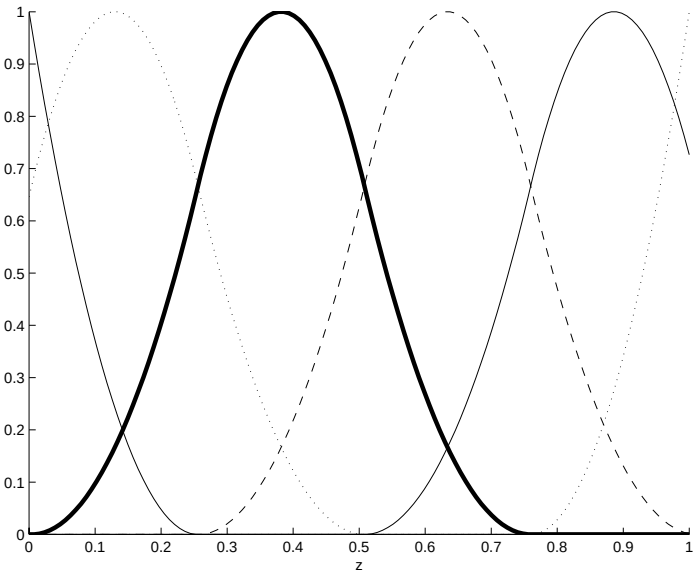
[17] Gilbert Strang and Truong Nguyen. *Wavelets and Filter Banks.* Wellesley-Cambridge Press, Wellesley, MA, 1996.

[18] Carlos Torres-Verdín and Tarek M. Habashy. Rapid 2.5-D forward modeling and inversion via a new nonlinear scattering approximation. *Radio Sci.*, 29(4):1051–1079, July-August 1994.

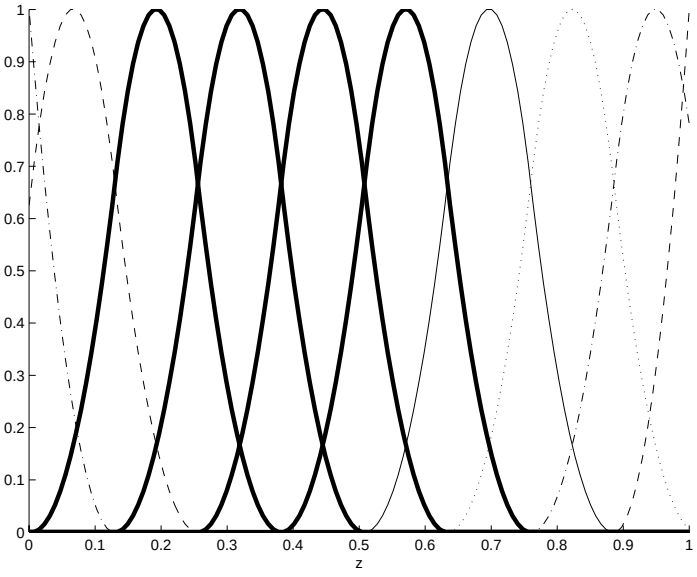
[19] C. R. Vogel and M. E. Oman. Fast, robust total variation-based reconstruction of noisy, blurred images. *IEEE Trans. Image Process.*, 7(7):813–824, July 1998.

8 Figure Captions

- **Fig. 1:** Examples of quadratic B-spline functions on two spatial scales
- **Fig. 2:** Graphical representation of multiscale B-spline vectors. Nodes on one horizontal level represent vectors or their associated expansion coefficients at a given scale. Links from one node to the next indicate cross-scale dependencies induced by the underlying two scale dilation equation, (20). Coarse scales are at the top with finer scales at the bottom.
- **Fig. 5:** In (a) we plot the B-spline vectors used to generate profiles for the testing of the EBA. A few samples of the resulting profiles are shown in (b)
- **Fig. 6:** Reconstruction performance and convergence behavior for adaptive method and fixed, finest scale approach for a layer-like profile. In (a), the average reconstructions taken over 10 noise realizations at 30 dB is compared to the true profile. For one of these runs, (b) and (c) demonstrate the improved convergence behavior typically exhibited by the adaptive method.
- **Fig. 7:** Reconstruction (a) and finest scale distribution (b) for the adaptive inversion method for layer-like profile.
- **Fig. 8:** Reconstruction performance and convergence behavior for adaptive method and fixed, finest scale approach for a smoothly varying profile. In (a), the average reconstruction taken over 10 noise realizations at 30 dB is compared to the true profile. For one of these runs, (b) and (c) demonstrate the convergence behavior of the two methods.
- **Fig. 9:** Reconstruction (a) and finest scale distribution (b) for the adaptive inversion method for smooth profile.



(a) Coarse scale vectors



(b) Fine scale vectors

Figure 1: Examples of quadratic B-spline vectors on two spatial scales

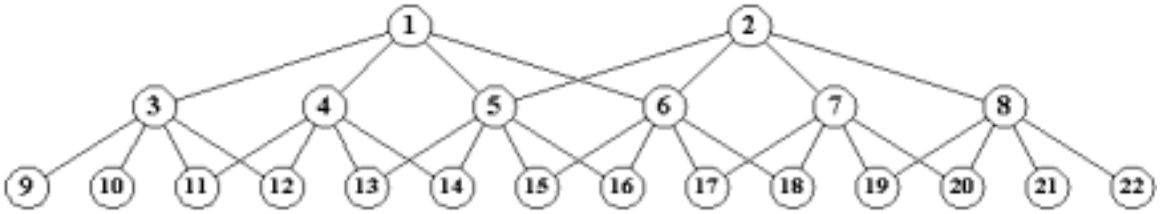


Figure 2: Graphical representation of multiscale B-spline vectors. Nodes on one horizontal level represent vectors or their associated expansion coefficients at a given scale. Links from one node to the next indicate cross-scale dependencies induced by the underlying two scale dilation equation, (20). Coarse scales are at the top with finer scales at the bottom.

Basic Inversion Algorithm

INITIALIZE

$B :=$ coarse set of B spline functions

$\hat{a} := \arg\min_a \|e(Ba)\|_2^2$

$\hat{\kappa}_p := B\hat{a}$

done := false

LOOP

while (done == false)

FIRST REFINE

$B_{fine} :=$ refine(B)

$\hat{a}_{fine} := \arg\min_a \|e(B_{fine}a)\|_2^2$

$\hat{\kappa}_{p,fine} := B_{fine}\hat{a}_{fine}$

$F := \|e(B_{fine}\hat{a}_{fine})\|_2^2$

$J :=$ Jacobian of e evaluated at $\hat{\kappa}_p$

NEXT COARSEN

$B_{coarse} :=$ coarsen$(B_{fine}, \hat{\kappa}_{p,fine}, F, J,$ threshold$)$

FINALLY PRUNE

$B_{prune} :=$ prune$(B_{coarse}, \hat{\kappa}_{p,fine}, F, J,$ threshold$)$

$\hat{a}_{prune} := \arg\min_a \|e(B_{prune}a)\|_2^2$

$\hat{\kappa}_{p,prune} := B_{prune}\hat{a}_{prune}$

CHECK TO SEE IF WE ARE DONE

if $\|\hat{\kappa}_p - \hat{\kappa}_{p,prune}\|_2^2 / \|\hat{\kappa}_p\|_2^2 >$ threshold

$B := B_{prune}$

$\hat{\kappa}_p := \hat{\kappa}_{p,prune}$

else

done := true

end if

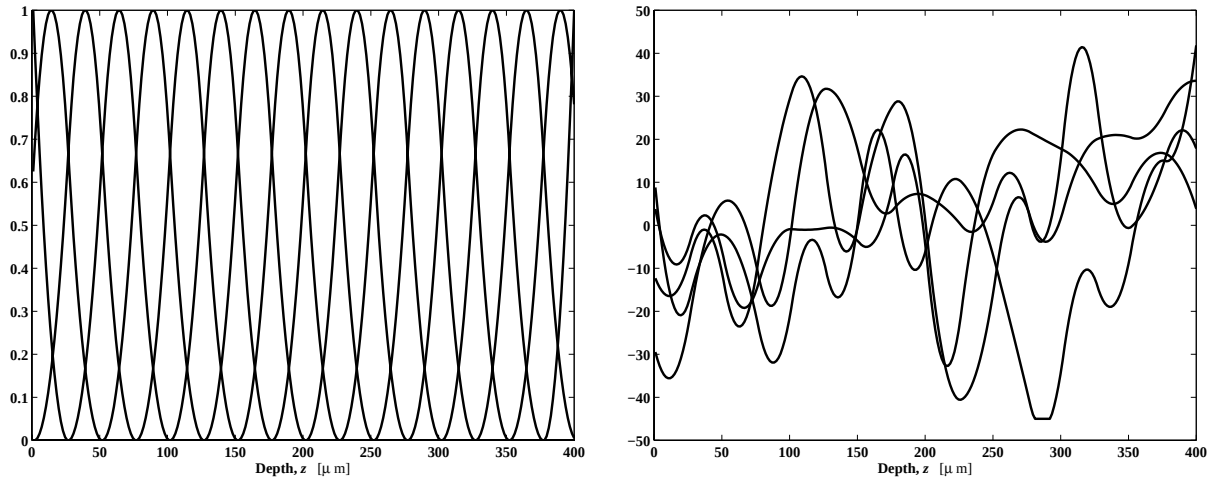
end while

Figure 3: Pseudo-code for Inversion Algorithm

Coarsening Algorithm

$B := \mathbf{prune}(B_{in}, \hat{\kappa}_p, F, J, \text{threshold})$
INITIALIZE
$B := B_{in}$
done := false
LOOP while (done == false)
    for $i = 1, 2, \ldots$ number of ways to generate coarsened versions of B
        $B_i :=$ ith valid coarsened version of B
        $\hat{\kappa}_{p,i} := B_i B_i^{\dagger} \hat{\kappa}_p$
        $\text{error}_i := \|J(\hat{\kappa}_{p,i} - \hat{\kappa}_p)\|_2^2$
        if $\min_i$ $\text{error}_i < 2F$ threshold
            $i^* := \arg\min_i$ error_i
            $B := B_{i^*}$
        else
            done := true
        end if
    end for
end while

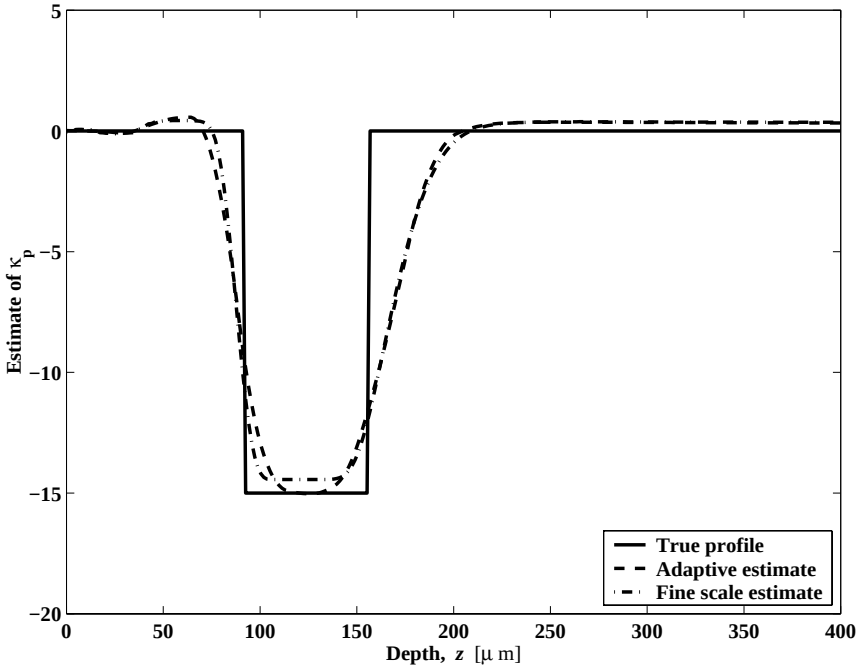
Figure 4: Pseudo-code for Coarsening Algorithm



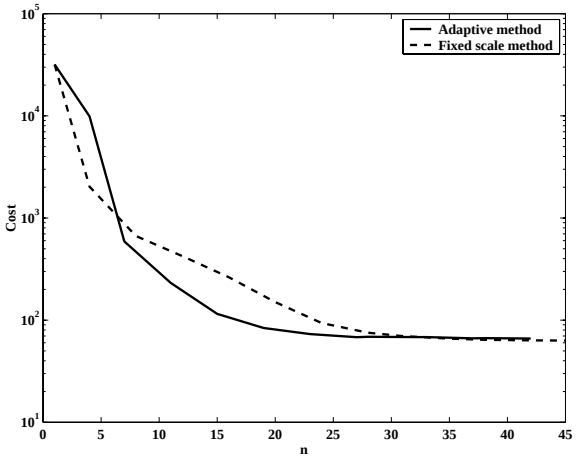
(a) B-splines used to evaluate EBA

(b) Sample κ_p profiles

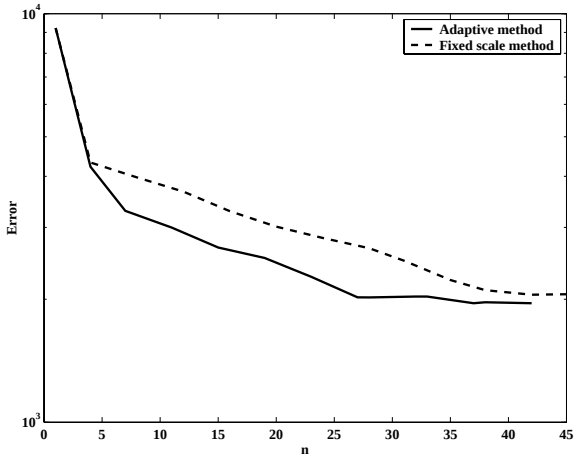
Figure 5: In (a) we plot the B-spline vectors used to generate profiles for the testing of the EBA. A few samples of the resulting profiles are shown in (b)



(a) Average reconstruction

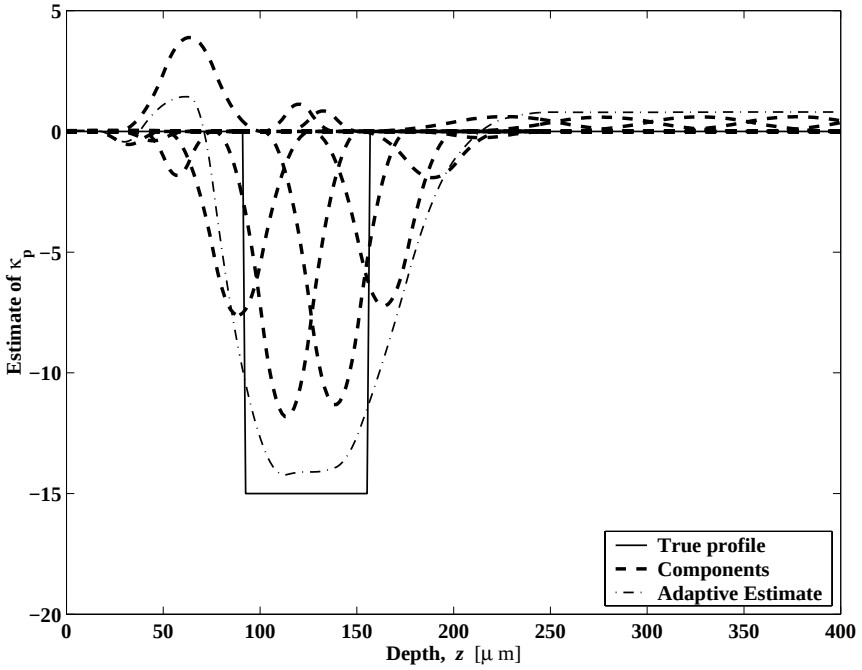


(b) Value of cost vector vs. iteration for one run

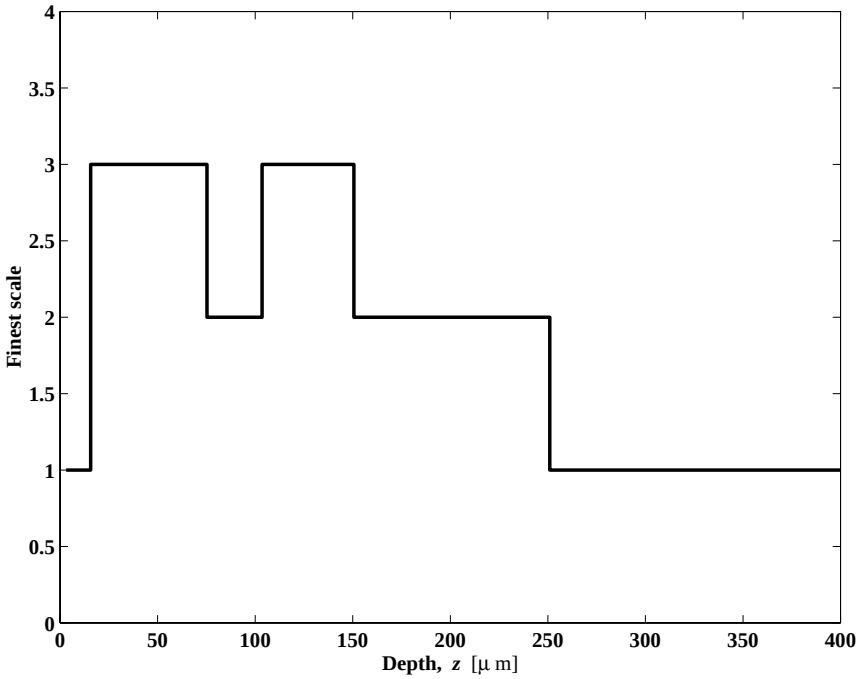


(c) Error vs. iteration for one run

Figure 6: Reconstruction performance and convergence behavior for adaptive method and fixed, finest scale approach for a layer-like profile. In (a), the average reconstructions taken over 10 noise realizations at 30 dB is compared to the true profile. For one of these runs, (b) and (c) demonstrate the improved convergence behavior typically exhibited by the adaptive method.

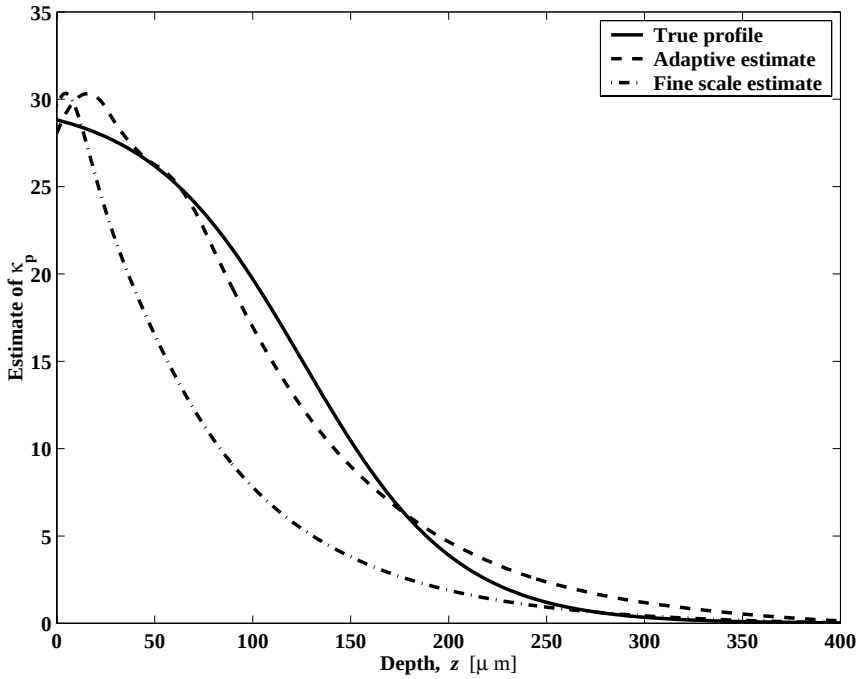


(a) Final estimate and constituent components

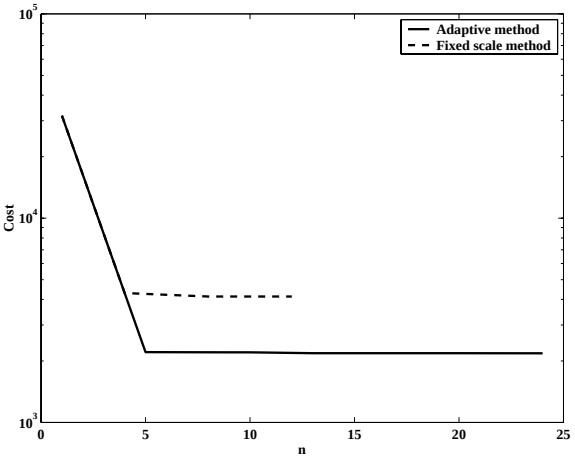


(b) Finest scale distribution

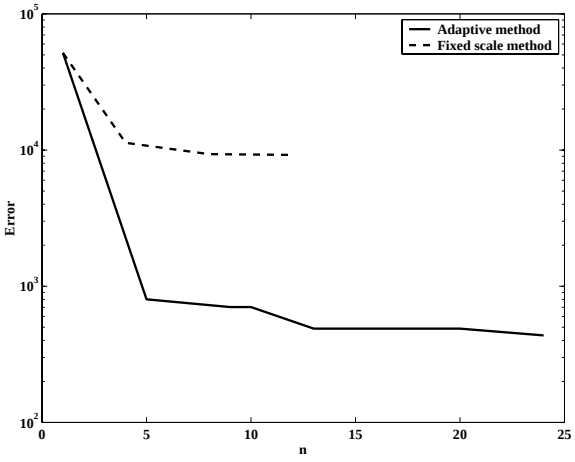
Figure 7: Reconstruction (a) and finest scale distribution (b) for the adaptive inversion method.



(a) Average reconstruction

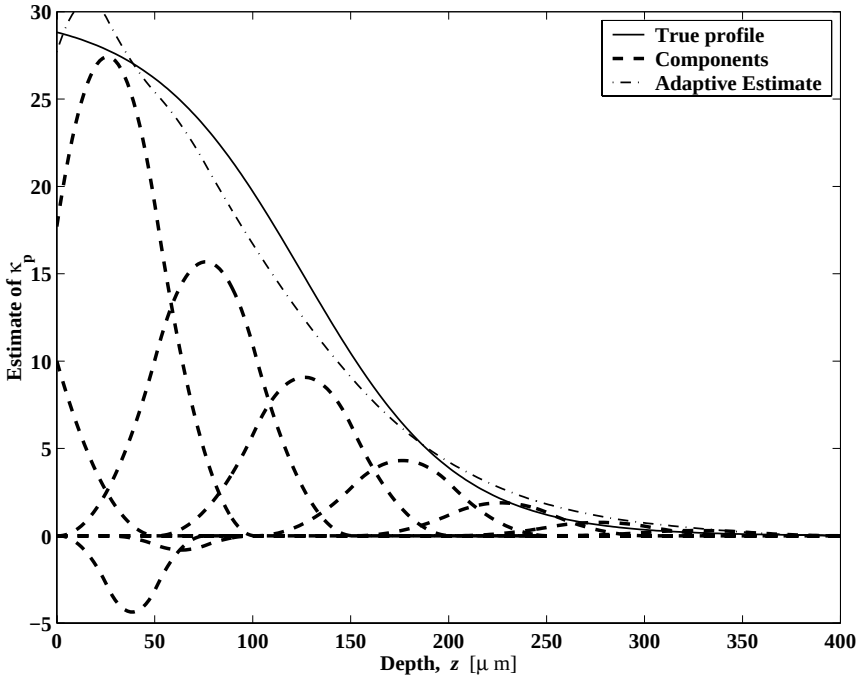


(b) Value of cost function vs. iteration for one run

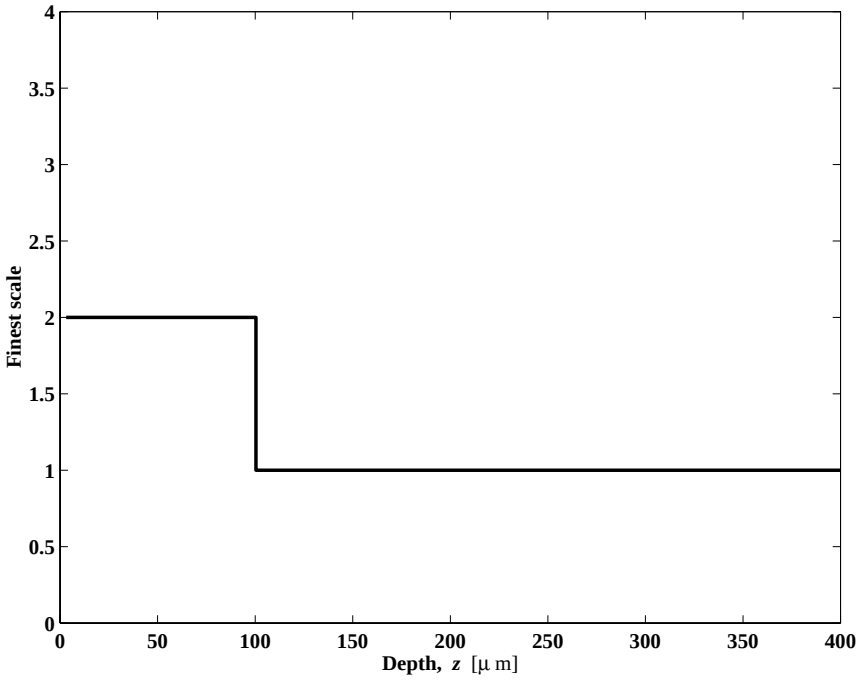


(c) Error vs. iteration for one run

Figure 8: Reconstruction performance and convergence behavior for adaptive method and fixed, finest scale approach for a smoothly varying profile. In (a), the average reconstruction taken over 10 noise realizations at 30 dB is compared to the true profile. For one of these runs, (b) and (c) demonstrate the convergence behavior of the two methods.



(a) Final estimate and constituent components



(b) Finest scale distribution

Figure 9: Reconstruction (a) and finest scale distribution (b) in adaptive inversion method.