A Method for Reducing the Complexity in the Reconstruction of a Blurred Signal

Ibrahim Yavuz

January 26, 2000

Abstract:

The regularized least squares methods for the solution of ill-posed inverse problems are summarized, and appropriate references are stated. Additionally, an adaptive multi-scale algorithm is proposed to solve highly ill-posed inverse problems with fewer degrees of freedom and comparable performance. The algorithm controls the level of detail in the reconstruction by distributing the fine scale information to the appropriate intervals in the overall estimation interval. The method is applied to a linear inverse problem, namely the reconstruction of a signal from its blurred and noisy version. The results are stated, compared with the regular fine scale approach and the relevant properties are explained. This treatment is seen as a step for the application of the algorithm to the nonlinear inverse problems where it is forseen to provide a decrease in the complexity of the inversion as well as better convergence in the solution space than the regular approach.

1 Inroduction

Least-squares estimation methods are commonly used in solving linear inverse problems. A linear inverse problem can be expressed as the problem of estimating the vector $\mathbf{x}$ based on the knowledge of a data vector $\mathbf{b}$ which is related to $\mathbf{x}$ as $\mathbf{b} = \mathbf{A}\mathbf{x} + \mathbf{n}$. Here $\mathbf{A}$ is a known linear transformation matrix and $\mathbf{n}$ is a random noise vector. Inverse problems are typically ill-posed, meaning that small perturbations in the data can lead to large amplitude, non-physical artifacts in the reconstruction. Linear ill-posed problems arise in a variety of applications: astronomy [1], computerized tomography [2], electrocardiography [3], early vision [5] and meteorology [4] are just a few of these. Vast amount of literature on ill-posed problems exist in the setting of Hilbert spaces and other infinite dimensional spaces. See for example [6], [7],

[8], [9], [10], [11]. Also, introductions on ill-posed problems can be found in [12], [1], [13], [14], [15], [16]. The paper by Hofmann [17] contains a valuable material on infinite-dimensional as well as finite-dimensional ill-posed problems. Most numerical methods for treating discrete ill-posed problems seek to overcome the reconstruction problems by replacing the problem with a nearby well-conditioned problem whose solution approximates the required solution, moreover is a more satisfactory solution than the ordinary least squares solution. This is done by adding an extra regularization term to the cost function that is to be minimized, which itself incorporates more information about the sought solution, for example forcing the solution to be smooth.

The problem of reconstructing a signal from its blurred and noisy version is an ill-posed problem as described above and regularized estimation methods can be applied to estimate each sample of the original signal with the right choice of the regularization. When the blur is linear the solution to the problem is found by minimizing the cost function in the form of $||\mathbf{Ax} - \mathbf{b}||$. The problems in the reconstruction due to ill-posedness is reflected in the properties of the matrix $\mathbf{A}$. Namely, when the problem is ill-posed, the matrix $\mathbf{A}$ is ill conditioned and all its singular values decay to zero in such a way that there is no particular gap in the singular value spectrum (large condition number of $\mathbf{A}$) [18]. The norm to express the actual cost function is typically chosen to be 2-norm (least-squares). The norm chosen for the additional cost term can vary from 1-norm to 2-norm. When both terms are expressed with 2-norm, the solution that minimizes the total cost can be expressed by a simple linear equation.

When the blur is non-linear, the cost function is in the form of $||\mathbf{F}[\mathbf{x}] - \mathbf{b}||$ and the reason of being ill-posed is less obvious mathematically, depending on the nonlinear mapping $\mathbf{F}$. The additional regularization term again can be chosen to be anything from 1-norm to 2-norm but in any case the minimization of the cost must be done iteratively by applying numerical techniques such as Levenberg-Marquardt, Gauss-Newton etc. The additional problem associated with non-linear problems is that the iteratively converged solution is a local minima depending on the initial choice of the estimated parameters and one cannot be sure about being global of the solution. In non-linear cases, when the number of samples to be estimated is large, sample-by-sample inversion might be inappropriate because of this difficulty of approaching to the global minimum, plus the computational complexity of the iterative techniques.

In our work, instead of estimating each sample, we propose to express the signal to be inverted as a weighted sum of some set of refinable basis functions and estimate the weight of each function. This is generally known as estimation in a subspace. The basis functions we choose are refinable second order box splines (obtained by convolution of a box with itself twice) [19]. Refinability means that each basis function (parent) can be represented as a linear combination of finer scale children functions. Instead of each parent one can use its finer scale children in the basis set to get a more detailed representation of the original signal at the intervals when needed. The use of such a function tree is motivated by the fact that while inverting, one might not need the same level of refinement at every place in the estimation interval. In other words, depending on the vector to be inverted, there might be regions in the estimation interval where coarser scale representation is sufficient to give a satisfactory solution. If we can adaptively determine the level of refinement required at the proper places in the estimation interval, this allows us to have a lower order representation of the unknown signal comprised of relatively few fine scale coefficients supplemented by a small number of coarse scale coefficients. In some applications, this fact may decrease the complexity by decreasing the order of estimation and in others it can lead to better reconstructions than the ordinary regularized solution. A similar approach was used related to wavelet-based regularization techniques for two-dimensional non-linear inverse scattering problems in [20], [21], where the representation of the original signal is done using wavelet functions. Using the spline functions instead of wavelets in this work, we give up the orthonormality property of wavelets which leads to a more flexible method for modeling the unknown signal.

In order to produce a low order reconstruction based on this idea, we must determine the distribution of the fine scale detail adaptively based on the perturbated data (we don't have the original signal to determine this) in an automatic and controlled manner. The method to accomplish this starts with an estimation based on a coarse scale set of basis functions and iteratively refines the reconstruction first to add detail and then to get rid of unnecessary degrees of freedom to obtain a better description of the signal by adjusting the resolution. The details of this procedure is explained in Section 3.

As mentioned above, limiting the degrees of freedom this way in the reconstruction is useful for two reasons. First, one can achieve computational efficiencies with the help of low dimensionality. This is important especially in the non-linear reconstruction problems, because the iterative numerical techniques required in the reconstruction are computationally demanding. Second, this approach can lead to better reconstructions, which is also important, in particular, in non-linear inverse problems where the convergence to a local-minima is a problem. In other words, it is possible to converge to a lower cost point in the solution space than is the case for a high dimensional sample-by-sample inversion approach. Although both advantages are seen more likely to be abtained in non-linear problems, in our work, we present the suggested technique by applying it to a linear problem as a demonstration of its use and as a step to achieving these advantages for non-linear problems in the future applications.

Section 2 summarizes the principles of inversion using regularized least-squares estimation for linear and non-linear problems. Section 3 describes the method proposed in this work and gives the details of the algorithm that controls the level of refinement needed. In Section 4, we apply the algorithm to a linear reconstruction problem and state the associated results.

2 Regularized Least-Squares Inversion

Our problem is to reconstruct the signal $x(n)$ from its blurred and noisy (additive noise) version $y(n)$. When the blur is modeled as a linear filter:

$$y(n) = L[x(n)] + v(n) \tag{1}$$

where L represents the linear filtering operation. This can be written in the matrix form:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{v}. \tag{2}$$

$\mathbf{H}$ is the matrix representing the linear filtering operation L. In the special case where the blur is also shift invariant (Linear shift invariant filter), the system L can be represented by its unit sample response $h(n)$ and equation (1) can be written using the convolution operation as:

$$y(n) = h(n) * x(n) + v(n). \tag{3}$$

Equation (3) can also be written in matrix form as in equation (2). In this case, the matrix $\mathbf{H}$ representing the linear shift invariant filtering operation is a toeplitz matrix.

The standard linear least-squares solution to the problem stated in equation (2), for any matrix $\mathbf{H}$ (need not be toeplitz) is obtained by minimizing the mean square cost:

$$\arg\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{H}\mathbf{x}||_2^2. \tag{4}$$

The well known solution to this is:

$$\hat{\mathbf{x}} = (\mathbf{H^T H})^{-1}\mathbf{H^T y}. \tag{5}$$

In most cases, though, the problem is ill-posed, that is the solution is very sensitive to perturbations of $\mathbf{H}$ and $\mathbf{y}$. This happens when the condition number of the matrix $\mathbf{H}$ is large [22]. In such a situation the largest perturbations are associated with the smallest generalized singular value of H [18]. One of the best known regularization methods is *Tikhonov regularization* [10] (also called damped least squares [23]). The Tikhonov regularized solution $\hat{\mathbf{x}}$ is defined as the solution to the following least squares problem:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{Hx}||_2^2 + \lambda^2 ||\mathbf{Lx}||_2^2. \tag{6}$$

λ controls the weight of the regularization term on the cost function. The weight should not be so small as to not stabilize the solution and it should not be so big that it will outweigh the data term. As $\lambda \to 0$, we demand that $\hat{\mathbf{x}}$ just fit the data. On the other hand, as $\lambda \to \infty$ the data play a limited role in influencing $\hat{\mathbf{x}}$ and we obtain overly smooth estimates. Proper selection of this parameter is a non-trivial problem ([18], [24]), although in some cases it can be set by trial and error.

The choice of regularization matrix $\mathbf{L}$ in (6) depends on the problem. The choice of $\mathbf{L} = \mathbf{I}$, the identity matrix, controls the perturbations by putting extra cost proportional to the energy of the vector $\mathbf{x}$, forcing less perturbated solutions by limiting the energy. Another popular choice of $\mathbf{L}$ is the differential matrix that gives finite difference approximation to the first derrivative of $x(n)$.

$$\mathbf{L} = \begin{bmatrix} 1 & -1 & & & & \\ & 1 & -1 & & & \\ & & & \ddots & & \\ & & & & 1 & -1 \end{bmatrix}_{N-1 \times N} \tag{7}$$

where N is the length of the vector $\mathbf{x}$. This choice of $\mathbf{L}$ forces the elements of vector $\mathbf{x}$ change smoothly by putting extra cost directly proportional to the derrivative of $x(n)$.

It can be shown that Tikhonov regularization of (6) in effect dampens or filters out the the contributions to $\hat{\mathbf{x}}$ corresponding to the generalized singular values of $\mathbf{H}$ smaller than about λ, making the solution less sensitive to perturbations than the ordinary least squares solution [18]. In fact, it is shown in [22] that the condition number of the problem (6) is inversely proportional to the parameter λ. In addition, it can be shown that the contributions to $\hat{\mathbf{x}}$ corresponding to small generalized singular values are more oscillatory than large ones [25], therefore Tikhonov regularized solution is indeed smoother than the unregularized solution since the effects of small singular values are filtered out.

The regularization term in (6) could in general be $\lambda^2||\mathbf{Lx}||_p^p$, p being in the range $[1, 2]$. The $p = 2$ case is known as smoothness regularizer, based on the discussion above. As p approaches 1, the regularizer is more encouraging of forming profiles which have edges or other sharp discontinuities. With $p = 1$, one has the total variation (TV) regularization scheme [26], [27]. When p is chosen to be 2, the solution to (6) can be expressed as

$$\hat{\mathbf{x}} = \left(\mathbf{H^T H}+\lambda\mathbf{L^T L}\right)^{-1}\mathbf{H^T y}. \tag{8}$$

The discussion in Section 2 until this point has been for the case of linear blur. When the blur is non-linear, we have the model:

$$y(n) = f(x(n)) + v(n). \tag{9}$$

where $f(.)$ is a non-linear function. In matrix form:

$$\mathbf{y} = \mathbf{F}[\mathbf{x}] + \mathbf{v}. \tag{10}$$

$\mathbf{F}$ is a non-linear operator on vector $\mathbf{x}$. Least-squares solution is the vector $\mathbf{x}$ that minimizes the regularized cost function.

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{F}[\mathbf{x}]||_2^2 + \lambda^2||\mathbf{Lx}||_2^2. \tag{11}$$

The solution to (11) can be found numerically by iterative techniques such as Levenberg-Marquardt, Gauss-Newton, etc. These algorithms start by an initial choice of $\mathbf{x} = \mathbf{x}_0$, and iteratively approach to a local minima of the cost function, generally depending on $\mathbf{x}_0$.

Up to this point we applied the least-squares estimation to the samples of the signal $x(n)$. Hence the order of the problem is N, the length of the vector $\mathbf{x}$. In another approach, if we express $\mathbf{x}$ as a linear combination of M basis vectors $\mathbf{b_i}, i = 0, 1, ..., M - 1$ then

$$\mathbf{x} = \sum_{i=0}^{M-1} a_i\mathbf{b_i} \Leftrightarrow \mathbf{x} = \mathbf{Ba} \tag{12}$$

where $\mathbf{a} = [a_0 a_1 ... a_{M-1}]^T$ and $\mathbf{B} = [\mathbf{b_0 b_1 ... b_{M-1}}]$.

Our aim now is to estimate $\mathbf{a}$ and express $\hat{\mathbf{x}} = \mathbf{B}\hat{\mathbf{a}}$. Assuming a linear blur and using the equation (12) in equation (2)

$$\mathbf{y} = \mathbf{HBa} + \mathbf{v} \tag{13}$$

and regularized least squares estimate for $\mathbf{a}$ from equation (6) is

$$\hat{\mathbf{a}} = \arg\min_{\mathbf{a}} ||\mathbf{y} - \mathbf{HBa}||_2^2 + \lambda^2 ||\mathbf{LBa}||_2^2 \tag{14}$$

and using (8)

$$\hat{\mathbf{a}} = (\mathbf{B^T H^T H B} + \lambda \mathbf{B^T L^T L B})^{-1} \mathbf{B^T H^T y}. \tag{15}$$

Since the length of vector $\mathbf{a}$ is M, the order of the least squares estimation problem in this case is M. Therefore, if $M \ll N$, that is if we can satisfactorily express $\mathbf{x}$ with a number of basis vectors much less than the length of $\mathbf{x}$, then we can decrease the complexity of the solution a great deal using this kind of approach. The quality of this approach, of course, depends on how well we can approximate $\mathbf{x}$ with M number of basis vectors.

The same argument holds also for the case of non-linear blur. By expressing the vector $\mathbf{x}$ in terms of little number of basis vectors, it is possible to decrease the complexity of the solution, although the solution cannot be expressed by a simple linear eqation and must be found iteratively. The solution is the minima of the cost function similar to (14):

$$\hat{\mathbf{a}} = \arg\min_{\mathbf{a}} ||\mathbf{y} - \mathbf{F}[\mathbf{Ba}]||_2^2 + \lambda^2 ||\mathbf{LBa}||_2^2. \tag{16}$$

where $\mathbf{F}$ is the non-linear transformation on $\mathbf{x}$.

Depending on the application, one can choose a fixed number of basis vectors ($M \ll N$) that are appropriate to approximate $\mathbf{x}$ with some satisfaction (estimation in a subspace). However, in most cases we do not have much prior information about the properties of the original signal $\mathbf{x}$, hence we do not know the basis functions that would be appropriate to express it. Solution to this, as we showed in our work, is obtained by choosing the number of the refinable basis vectors variable and adapting the refinement in the basis set by iteratively updating the reconstructions based on the available perturbated vector $\mathbf{y}$. Refinability discussed here means that each basis vector (parent) can be represented as a linear combination of finer scale children vectors. Instead of each parent one can use its finer scale children in the basis set to get a more detailed representation of the original vector $\mathbf{x}$, which improves the reconstruction at the price of increase in the inversion complexity. Choosing the basis vectors to be refinable, we can represent the original vector with any combination of fine and coarse scale basis vectors and have a control of adaptively concentrating the reconstruction quality to the intervals of the original signal where more information is carried. This control can be done by using the vector $\mathbf{y}$, without knowing $\mathbf{x}$, as will be shown later.

The trade-off involved in the control of refinement process is the following: We do not want a lot of refinements everywhere, because refinement in the basis increases the number of the vectors used to represent $\mathbf{x}$, hence the number of parameters that have to be estimated in equation (15) and this increases the complexity of the reconstruction process. The parent vector might just give a satisfactory representation of the original vector $\mathbf{x}$ in some intervals without the need for its children and it is a good idea to stick with the parents in these intervals to limit the complexity. So, the trade-off, while doing refinement is between quality of the reconstruction (better with more refinement) and the complexity (less with less refinement).

A good way of implementing a strategy based on this idea is to construct a basis vector tree (Figure 1) that is composed of splitting each vector to its children, further splitting the children to their children and so on. The vectors at the top of the tree cover a broader range in the interval in which we want to estimate our original vector $\mathbf{x}$, hence provide a coarser approximation to $\mathbf{x}$. As we come down the tree, the children cover a narrower range in the estimation interval and provide a finer approximation to $\mathbf{x}$. Figure 1 shows the graphical representation of such a tree, where each parent has four children. By defining such a tree, the problem of choosing a good basis that balances the reconstructon quality and the solution complexity by adaptively controlling the degrees of freedom involved, reduces to the problem of starting from a coarse scale basis vectors and determining the intervals where we need to further refine the initial basis vectors and the intervals we stick to the coarser scale vectors. The more splitting we do, the better is the reconstruction, however, the more complex we make the inversion.

One important point is worth mentioning once again. The vectors to give a good representation of the original vector $\mathbf{x}$ must be determined without knowing $\mathbf{x}$. Therefore, the algorithm that makes the search in the refinable vector tree to control the level of refinement needed, should find the good combination of vectors utilizing the available perturbated vector $\mathbf{y}$. The following section describes the details of this algorithm and the basis vector tree we used in our work, namely the second order box splines.

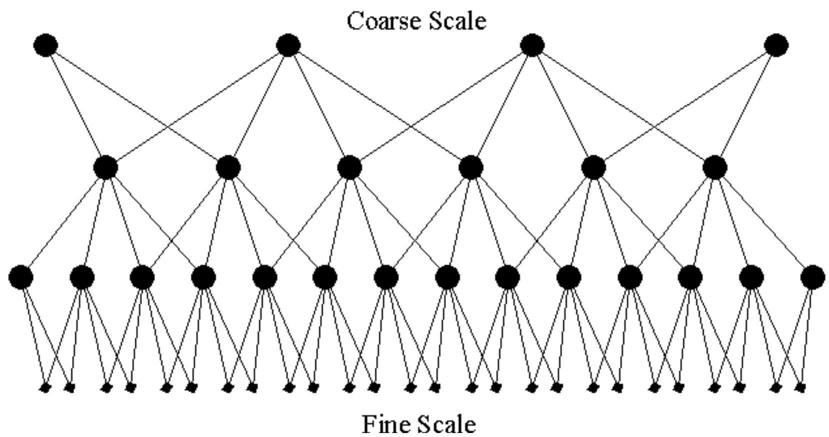

Figure 1: The graphical representation of the vector tree

3 Basis Function Tree and Refinement Strategy

The basis vectors that we use to represent the original signal $x(n)$ are the second order box splines which are obtained by convoluton of a box with itself twice [19]. As explained above, each vector can be split into its finer scale children and we have a tree of vectors that are obtained this way. In our case each vector has four children and the graphical representation of the tree is similar to the one shown in Figure 1. The spline vectors are such that they reach a maximum and they attenuate fast after a certain point and become zero (Figure 2). The vectors at the same level of the tree have the same width, which means they cover same length in the estimation interval (Figure 3a). Each parent is a linear combination of its children and the children give a more detailed representation in the interval that the parent covers. Figure 2 shows this relation between a parent and its children. The parent is obtained by the sum of the children at both ends and three times the children in the middle.

As explained above, we want to express vector $\mathbf{x}$ as a linear combination of a set of basis vectors that are to be chosen from this tree, which will provide that the fine scale information is appropriately distributed in the estimation region. To demonstrate what we mean by "appropriate distribution of fine scale information", assume for a moment that we know vector $\mathbf{x}$ a-priori and we are seeking a good set of basis vectors (little in number,

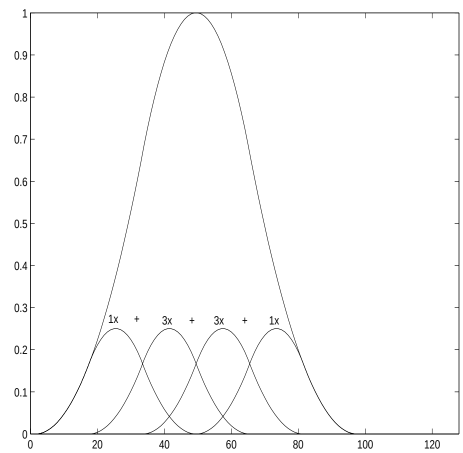

Figure 2: A parent and its children

good quality in the MSE sense) to represent $\mathbf{x}$. Figure 3b shows how we can express a square signal with 18 evenly distributed basis vectors (least square fit, projection) that are at the same level of the vector tree as shown in Figure 3a. We express the same square signal in Figure 4b with another set of 16 basis vectors shown in Figure 4a, this time finer scale vectors are also used and concentrated around the discontinuity. The approximation in Figure 4b is much better than the approximation in Figure 3b both subjectively and in the MSE sense, MSE in 4b is 0.5234, MSE in 3b is 1.3783. The reason for this improvement with less functions is that the basis vectors chosen in Figure 4a are concentrated wisely in the intervals where more information about $\mathbf{x}$ exist (the discontinuities) and coarser scale vectors are used at the intervals where they are satisfactory (the flat intervals). Also, vectors whose contributions to the overall function are small are discarded. Such adaptation of refined basis vectors can be done for any signal and we will show that it is possible to do it based on perturbated vector $\mathbf{y}$, without knowing the original vector $\mathbf{x}$.

The algorithm we use in order to find a good set of basis vectors that represent vector $\mathbf{x}$ based on the vector $\mathbf{y}$, starts with a low order, coarse scale collection of basis vectors. It is obvious that when we split each vector in the basis to their children, we will get a better representation of $\mathbf{x}$. After splitting, though, we can search for those children vectors that don't contribute more than their parents to the quality of $\mathbf{x}$ and merge them back to their parents. By repeating splitting and merging back, we only keep the finer scale children which improve the representation of $\mathbf{x}$ and keep the

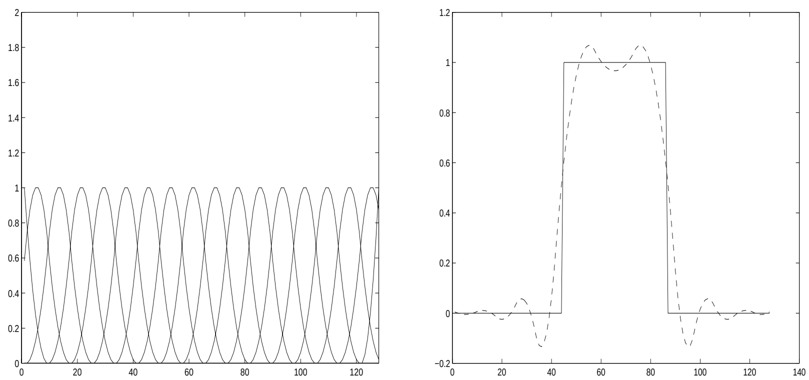

Figure 3: a)The basis vectors at the same level of the tree (b)The representation of a square signal with this set

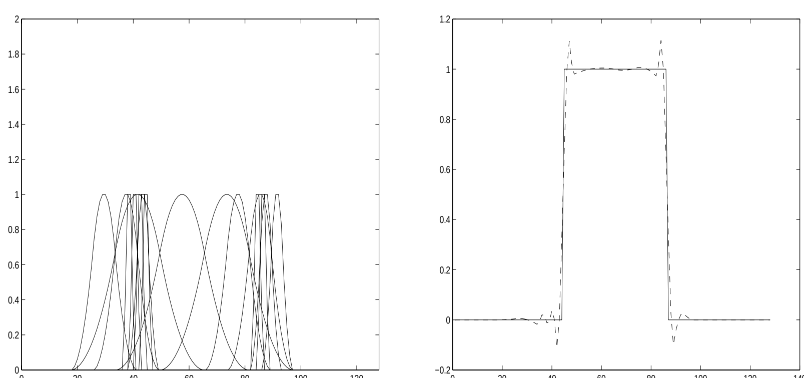

Figure 4: (a)The basis vectors refined more at the discontinuity (b)The representation of the same square signal with this set

parents whereever they are good enough. The goal is to provide arbitrary detail to $\mathbf{x}$, followed by a stage in which we determine which, if any, degrees of freedom associated with this detail warranted. Those not needed are removed. While one might consider many methods to decide for how to do the merging of the children to the parents to accomplish the stated goal, in Figure 5 and in the paragraph following it, we explain an approach which we have found to work well in our case.

After splitting all the vectors in the current basis set, we search for the collections of nodes which represent a complete set of children for a given parent. This collection normally is composed of four vectors while it has only two vectors at the end points of the tree. Figure 5 shows these 4-vector and 2-vector collections in rectangles. For each of these collections we consider a new basis obtained by replacing the fine scale children with all of their coarse scale parents. In the 4-vector case, there are three relevant parents, while in the 2-vector case there are two relevant parents. Figure 5 shows these parent vectors that are considered to replace the children in ellipses.

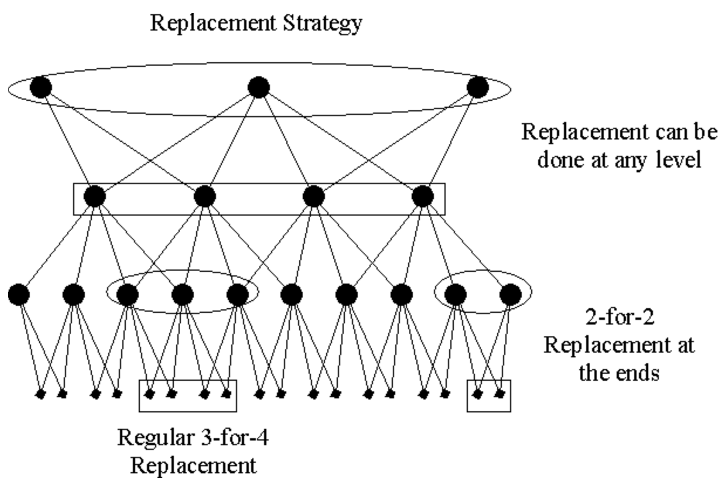

Figure 5: 3-for-4 and 2-for-2 replacements for merging of children to parents

After splitting all the vectors in the current basis set, in order to prune away the unnecessary detail we want to try these kind of replacements and keep those that do not effect the reconstruction quality much. In 3-for-4 replacements, whenever the four vectors as shown in Figure 5 are detected in the basis set, they can be replaced by their three-vector equivalent if this replacement does not effect the reconstruction more than a certain treshold.

Same is true for 2-for-2 replacements. Therefore, after splitting the basis, the reconstruction for each 3-for-4 and 2-for-2 replacement is done and for those that do not effect the reconstruction too much, we keep the replacement in the updated basis to reduce the complexity (less vectors in the basis). At each step, after splitting, this kind of search and merging continues until there is a significant difference between the reconstruction qualities.

After finishing the merging of fine scale vectors this way, one can further decrease the number of associated degrees of freedom by searching for and discarding the vectors with small coefficients, that is whose contribution to $\mathbf{x}$ is small. We also incorporate this idea to our algorithm following a very similar approach. After splitting and merging the basis vectors enough times, we determine the vector with minimum coefficient in the resulting basis set. This vector is discarded if, again, doing so does not affect the reconstruction significantly. This finding the minimum-coefficient vector and comparison of reconstructions continues until the effect of discarding the minimum-coefficient vector is above a certain treshold.

The algorithm repeats splitting, merging and discarding stages (merging and discarding phases are loops themselves) satisfactory number of times and the resulting basis vectors distribute the information detail appropriately as explained in Figure 4, based on the perturbated vector $\mathbf{y}$. We formalize and summarize the details of the steps of the algorithm next using the formulas we stated in Section 2.

STEPS OF THE ALGORITHM

1 Start with initial set of basis vectors $\mathbf{B_0} = [\mathbf{b_0 b_1 ... b_{M-1}}]$

2 Repeat sufficient times

2.1 Split each vector $\mathbf{b_i}$ in $\mathbf{B_0}$ to get $\mathbf{B_1}$

2.2 Find current estimate using $\mathbf{B_1}$

$$\mathbf{\hat{x}_{curr}} = \mathbf{B_1 \hat{a}} = \mathbf{B_1}(\mathbf{B_1^T H^T H B_1} + \lambda \mathbf{B_1^T L^T L B_1})^{-1} \mathbf{B_1^T H^T y}$$

2.3 Start merging

2.3.1 Find all three for four replacement possibilities in $\mathbf{B_1}$. Let $\mathbf{B_{1,1}}, \mathbf{B_{1,2}}, ..., \mathbf{B_{1,r}}$ are obtained by doing only one replacement for each possibility

2.3.2 Find the r reconstructions corresponding to each $\mathbf{B_{1,i}}$ by projecting the current estimate onto span of each $\mathbf{B_{1,i}}$

$$\mathbf{\hat{x}_i} = \mathbf{B_{1,i}}(\mathbf{B_{1,i}^T B_{1,i}})^{-1} \mathbf{B_{1,i}^T \hat{x}_{curr}}, i = 1, 2, ..., r$$

2.3.3 Find the replacement that produces minimum error compared to the current estimate

$$\mathbf{\hat{x}_p} = \underset{\mathbf{\hat{x}_i}}{argmin} \quad ||\mathbf{\hat{x}_{curr}} - \mathbf{\hat{x}_i}||_2^2$$

2.3.4 If relative error between $\mathbf{\hat{x}_p}$ and $\mathbf{\hat{x}_{curr}}$ is below a certain treshold, update $\mathbf{B_1}$ to $\mathbf{B_{1,p}}$ and continue merging (go to step **2.3**)

$$\text{relative error} = \frac{||\mathbf{\hat{x}_{curr}} - \mathbf{\hat{x}_p}||_2}{||\mathbf{\hat{x}_{curr}}||_2}$$

otherwise start discarding small-coefficient vectors (go to step **2.4**)

2.4 Start discarding vectors that have little coefficients

2.4.1 Find the vector with minimum coefficient by looking at the vector of weights, $\mathbf{\hat{a}} = (\mathbf{B_1^T B_1})^{-1} \mathbf{B_1^T \hat{x}_{curr}}$. Discard it from $\mathbf{B_1}$ to get a candidate basis $\mathbf{B_{cand}}$.

2.4.2 If the relative error between the estimate based on $\mathbf{B_{cand}}$ and the current estimate $\mathbf{\hat{x}_{curr}}$ is below the *threshold* update $\mathbf{B_1}$ to $\mathbf{B_{cand}}$ and continue discarding (go to **2.4**)

$$\text{relative error} = \frac{||\mathbf{\hat{x}_{curr}} - (\mathbf{B_{cand}^T B_{cand}})^{-1} \mathbf{B_{cand}^T \hat{x}_{curr}}||_2}{2}$$

otherwise update $\mathbf{B_0}$ to $\mathbf{B_1}$ and go to step **2**.

The *threshold* parameter in the merging and discarding phases can be different from each other, however in our implementation we chose the same *threshold* in both phases. In the merging phase, we do the merging when the relative error is smaller than the *threshold*. In the discarding phase we discard the minimum-coefficient vector if the relative error is smaller than the *threshold*. Therefore, larger the threshold, the more merging and discarding we do. When we choose a larger *threshold*, we are willing to get rid of more vectors from the basis set by merging and discarding (less complexity) compared to a smaller threshold. The result is that the *threshold* parameter can be used to control the complexity of the inversion process and the quality of the reconstruction which are inversely proportional to each other as explained before.

In the next section, we apply the algorithm proposed here to verify the the results and the comments made in this section.

4 Application of the Algorithm and the Results

Let's choose $x(n)$ in equation (1) to be a square signal followed by a ramp as shown in Figure 6a. The length of the signal is chosen to be $N = 128$. Let's choose the blur in the same equation to be a linear blur modeled with a bell shaped impulse response as shown in Figure 6b.

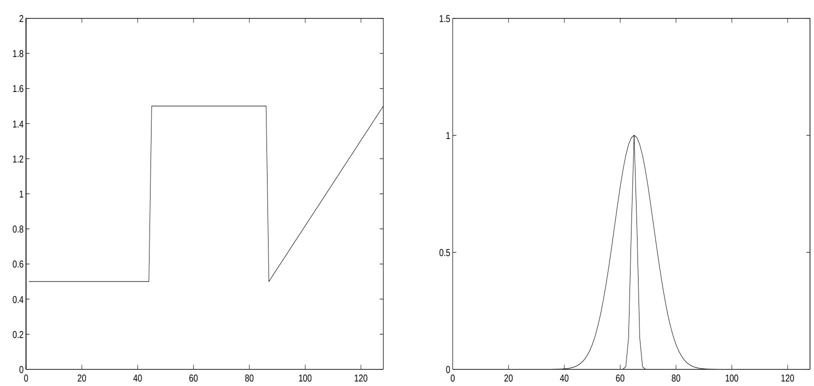

Figure 6: (a) The signal to be recovered (b) Two possible blurs modelled by narrow and wide impulse response $h(n)$

The impulse response of the blur can be narrow or wide. Also choosing the additive noise to be gaussian and independent from signal $x(n)$, our aim is to recover $x(n)$ from the blurred and noisy signal $y(n)$. First choice is to apply regularized least-squares inversion to each sample of $x(n)$ as described in equation (6) with L being the differential matrix shown in (**??**) to force smooth reconstructions. Figure 7 shows the result for the combination of the cases of narrow/wide blur and 10/40 dB noise realizations. The regularization parameter λ is set by trial and error to some value that balances the regularization term and the actual cost term. For more rigorous discussion of how to choose λ see [18], [24].

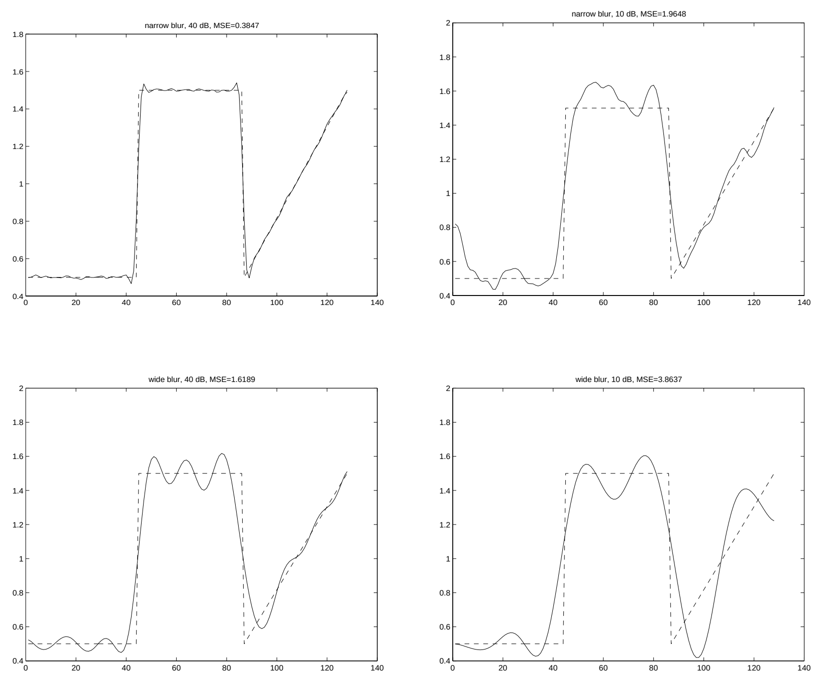

Figure 7: Regularized least-squares reconstruction (a) narrow blur, SNR=40dB, $\lambda = 1$, MSE=0.3847 (b) narrow blur, SNR=10dB, $\lambda = 50$, MSE=1.9648 (c) wide blur, SNR=40dB, $\lambda = 0.005$, MSE=1.6189 (d) wide blur, SNR=10dB, $\lambda = 5$, MSE=3.8637

Looking at Figure 7, we see that wider blur and more noise make the reconstructions worse in the MSE sense as expected, apparantly they effect the data that we base our reconstruction on in a negative way. Also it is very important to choose a right λ, especially for low SNR to stabilize the reconstruction, while not getting too much smoothing effect.

As explained in section 2 and 3, let's now express $x(n)$ as a linear combination of second order box splines and invert for the weights of each vector. To decide for the vectors to be in the basis with the knowledge of $y(n)$, that distributes the resolution appropriately in the estimation region and thereby reducing the inversion complexity, we use the algorithm described in section

3, which determines the appropriate refinement in the basis vector tree described before. The reconstruction with a good choice of λ (set by trial and error) and a *threshold* parameter (that balances the reconstruction quality (MSE) and the inversion complexity) is shown in Figure 8 (narrow blur) and Figure 9 (wide blur). The decided basis vectors for each case are also shown. The reconstructions in figures 8 and 9 are done based on the same data (same noise realizations) that is used to obtain the reconstructions in figure 7.

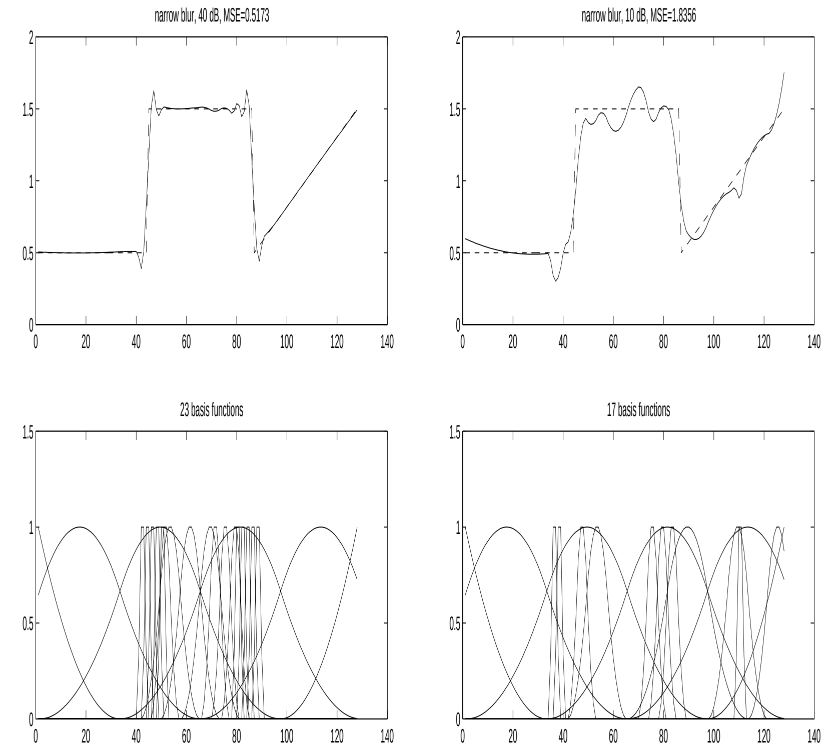

Figure 8: Reconstruction and the basis vectors decided by the algorithm (a) narrow blur, 40dB (b) narrow blur, 10 dB

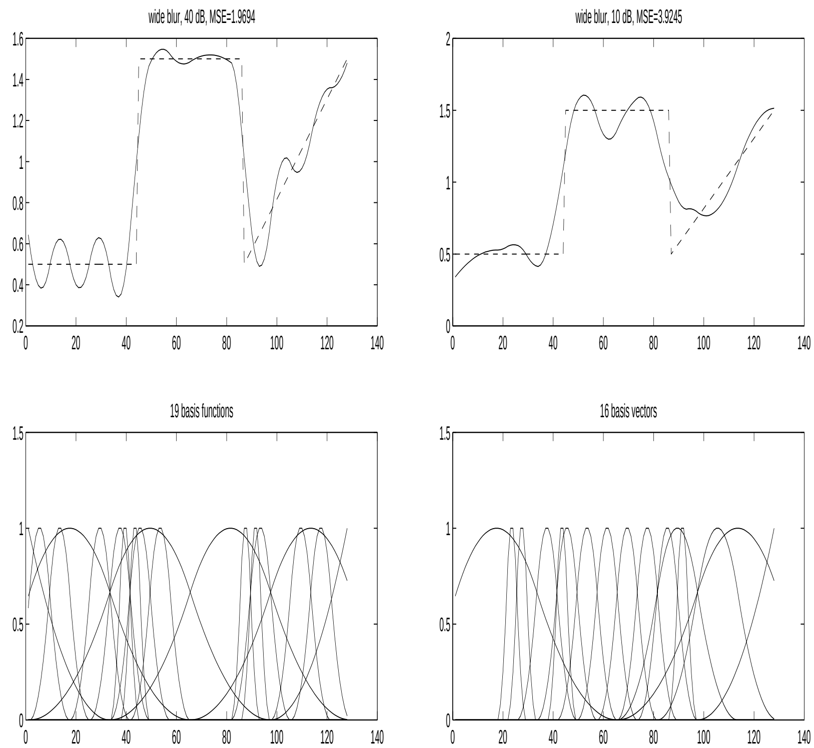

Figure 9: Reconstruction and the basis vectors decided by the algorithm (a) wide blur, 40dB (b) wide blur, 10 dB

A close look at the Figures 8 and 9 shows that the basis vectors that are decided by the algorithm that we described in Section 3 are such that the finer scale vectors are used around the discontinuities of the original signal and coarser scale vectors are used in the flat regions as expected and explained in Section 3. The observation is that the narrower the blur and the higher the SNR, the better this decision can be done, and the decision is not as good otherwise. The better reconstruction of the signal observed for less noise and narrower blur is due to the quality of this decision, as well as due to that we can estimate the weights of each vector better under these conditions. The important point is that in Figures 8 and 9 we obtain the reconstruction qualities comparable to the sample-by-sample inversion shown in Figure 7 in the MSE sense, although the order of the inversion is much less (order in Figures 8 and 9 is around 20, whereas in Figure 7 it is 128). This means that parameters estimated in Figures 8 and 9 (around 20) have about the same information as the parameters estimated in Figure 7 (the 128 samples), which is the result of concentrating the fine scale information to the intervals where it is needed.

Figure 10 compares the average reconstructions obtained by using the adaptive approach and that is obtained by fixed, fine scale inversion where SNR=40 dB and blur is chosen to be narrow. The reconstructions are averaged over 50 noise realizations. The fixed, fine scale inversion was done by using 2-norm and differential matrix in the regularization term in equation (6) ($p = 2$) and finding the solution using the closed form equation (8) of Section 2. The adaptive inversion was done by running the step 2 of the algorithm explained in section 3 ten times. It is seen that, on the average the reconstruction at the flat parts of the signal is better for the adaptive approach, because of its ripple-free nature, which is an advantage gained by using coarse scale functions at those flat areas. However, the average mean square error that is obtained by fixed, fine scale inversion in this case is slightly better than the adaptive approach, being 0.5059, whereas it is 0.5634 for adaptive method (10.2% better).

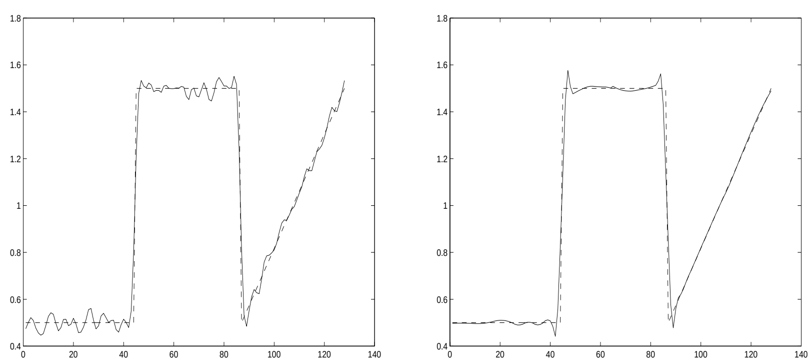

Figure 10: Comparison of the average reconstructions over 50 noise realizations, narrow blur, SNR=40 dB (a)fixed, fine scale inversion (b) adaptive approach

Looking at Figure 10, it can be commented that the adaptive approach leads to more stable solutions on the average. This is more obvious when we make the relevant perturbation worse. Figure 11 shows the average reconstructions (based on 50 noise realizations) for both methods where SNR=10 dB and blur is wide. The regularization parameter λ is chosen to be same for both cases. The adaptive approach (10 runs of step 2), in this case is better than the fixed, fine scale approach in the MSE sense too, MSE being 3.8534, whereas it is 16.8712 for the fine scale approach. 16.8712 is obtained again by choosing $p = 2$ and using the differential matrix of equation (7). Choosing p closer to 1 in the regularization term, one may get better solutions in the MSE sense for the discontinuous function that we try to invert here. In any case, the adaptive approach successfully makes the reconstruction while keeping the order of inversion much smaller.

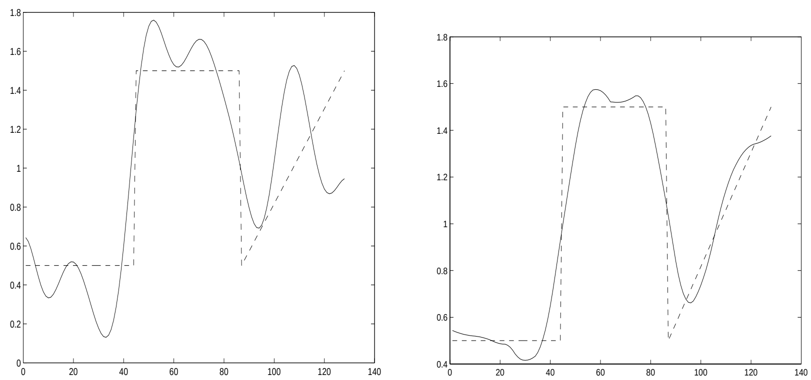

Figure 11: Comparison of the average reconstructions over 50 noise realizations, wide blur, SNR=10 dB (a)fixed, fine scale inversion (b) adaptive approach

This smaller order inversion in the adaptive approach, however, does not directly mean less complex solution than the regular approach in the linear blur cases. Because we have to go through an additional complexity in order to obtain the appropriate basis functions adapted to the signal to be converted. However, in the non-linear cases, this property of being low order might be exploited to obtain less complex solutions, since the iterative numerical techniques used in such problems are computationally intense.

Further results are presented below about the appropriate way of choosing the *threshold* parameter to balance the reconstruction quality and the order of the inversion. As explained before, the larger the *threshold* value, the less number of basis vectors are decided by the adaptive algorithm to represent the original signal, which, in turn, results in less quality in reconstruction in the MSE sense. As a study of this issue, we made the plots of average MSE and average order of inversion resulted by the adaptive approach based on 50 noise realizations in Figure 12 (narrow blur) and Figure 13 (wide blur).

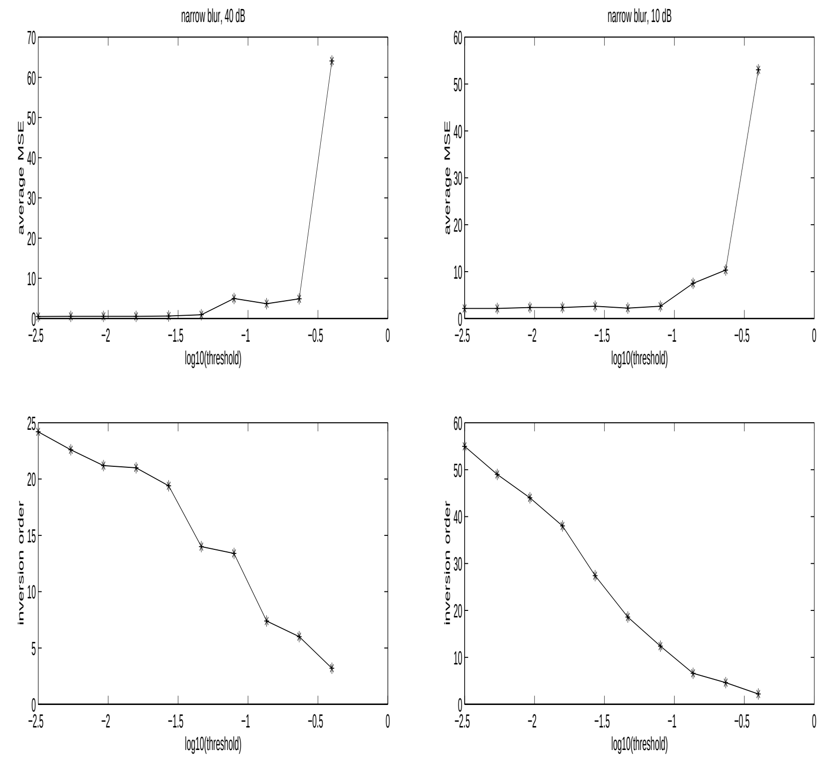

Figure 12: Average MSE and average inversion order based on 50 noise realizations (a) narrow blur, 40dB (b) narrow blur, 10 dB

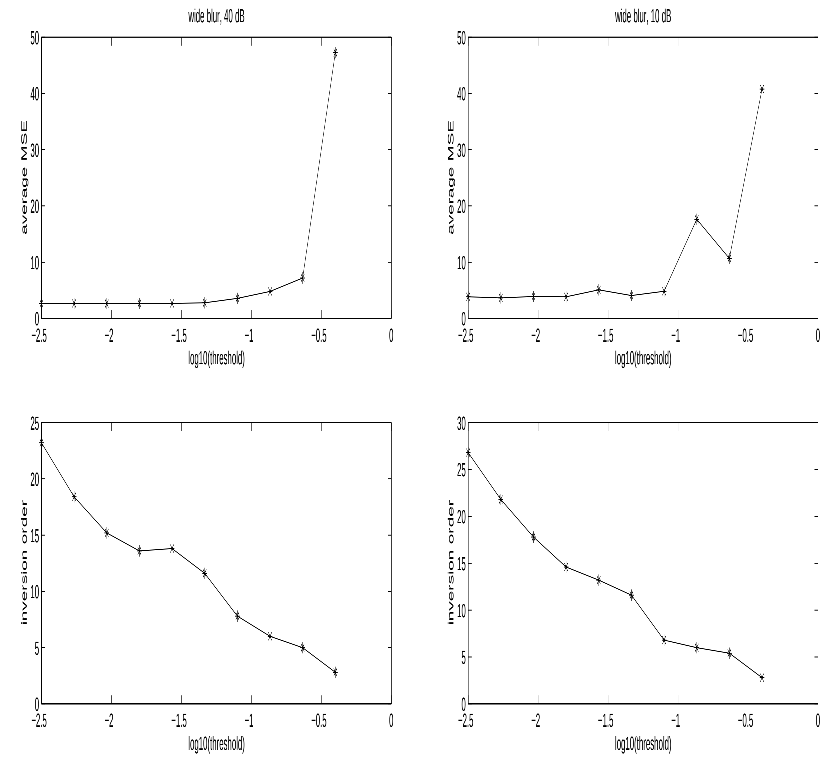

Figure 13: Average MSE and average inversion order based on 50 noise realizations (a) wide blur, 40dB (b) wide blur, 10 dB

Figures 12 and 13 verify our discussion about how *threshold* parameter effect MSE and the inversion order. One important observation is that, as the order of inversion decreases with increasing *threshold*, the MSE level stays steady at the first, then rises very fast for the large values of the *threshold*. This suggests that actually there is not much difference in the MSE sense between the reconstruction done by 15 basis functions and the one done by 30 basis functions. This observation, graphically gives us an idea of how to set the *threshold* parameter. The wise way of doing this is to set the *threshold* to a value that provides a satisfactory reconstruction in the MSE sense with the least number of basis functions. These conditions are satisfied in the interval of the *threshold* values right before the MSE steeps up.

5 Conclusion and Future Work

We have summarized the regular least squares methods in order to reconstruct a signal from its blurred and noisy version as well as proposed an alternative adaptive method to do the reconstruction with a lower order inversion. First the conditions on making these kinds of problems ill-posed are described, then the regularization methods to overcome ill-posedness are summarized. References from the literature has been stated for the reader when appropriate. The summary was concentrated on the solutions for the linear blur models, although the basics of the techniques to do the inversion in non-linear cases were also explained, and the difficulties are expressed.

The main subject has been an alternative adaptive method to do the inversion in the ill-posed cases. The method relies on expressing the original signal as a weighted sum of refinable basis functions and estimating the weight of each function in the inversion process. The appropriate refinement in the basis vectors is determined in a controlled manner based on the perturbated signal at hand and this way the resolution of information is distributed appropriately on the estimation interval. Basically the result of this decision is that the finer scale basis vectors are used in the steeper portions of the original signal and the coarse scale vectors are used in the flat regions. This result is obtained based on the perturbated vector without an a-priori information about the original signal.

By using this kind of approach instead of doing fine scale inversion evenly distributed in the estimation interval, we concentrate the information appropriately on much smaller number of parameters and estimate them, thereby decreasing the order of inversion. These properties of the method was tested by applying to a linear inverse problem, its performance was investigated and the relevant properties were presented. It was shown that the method gives comparable and sometimes better performance in the MSE sense than the fixed fine-scale approach with much less order of inversion. The issues about choosing the threshold parameter to balance the MSE and the inversion order were also disscussed.

The smaller order of inversion of the adaptive approach does not necessarily mean less complexity, because of the additional complexity required in the decision of the appropriate refinement in the basis vectors. However, less complexity solutions can be found in treating nonlinear inverse problems based on this smaller order inversion idea since in such problems, one has to use iterative numerical methods which are computationally intensive. The method so far is applied to a linear inverse problem as a demonstration of its use and its application to nonlinear problems is left as a future work, which is foreseen to lead to less complex solutions of these problems.

Another area to be studied about the proposed adaptive method is whether it can lead to more global solutions in the treatment of nonlinear inverse problems. It is important to understand this and the circumstances (classes of profiles, noise conditions, etc.) under which distributing the refinement in a controlled manner this way can converge to a more global minima in the solution space.

6 References

[**1**] I. J. D. Craig and J. C. Brown, *Inverse Problems in Astronomy*, Adam Hilger, Bristol, UK, 1986.

[**2**] F. Natterer, *The Mathematics of Computerized Tomography*, John Wiley, New York, 1986.

[**3**] *Calculating the isochromes of ventricular depolarization*, SIAM J. Sci. Statist. Comput., 5 (1984), pp. 105-120.

[**4**] G. Wahba, *Three topics in ill-posed problems* in Inverse and Ill-Posed Problems, H. W. Engl and C. W. Groetsch, eds., Academic Press, New York, 1987.

[**5**] M. Bertero, T. A. Poggio and V. Torre, *Ill-posed problems in early vision*, Proc. IEEE, 76 (1988), pp. 869-889.

[**6**] H. W. Engl and C. W. Groetsch, Eds., *Inverse and Ill-Posed Problems*, Academic Press, New York, 1987.

[**7**] V. B. Glasko, *Inverse Problems of Mathematical Physics*, Amer. Inst. Phys. Transl. Ser., New York, 1988.

[**8**] C. W. Groetsch and C. R. Vogel, *Asymptotic theory of filtering for linear operator equations with discrete noisy data*, Math. Comp., 49 (1987), pp. 499-506.

[**9**] V. A. Mozorov, *Methods for Solving Incorrectly Posed Problems*, Springer-Verlag, New York, 1984.

[**10**] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*, John Wiley, New York, 1977.

[**11**] *Spline Models for Observational Data*, CBMS-NSF Regional Conference Series in Applied Mathematics, Vol. 59, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1990.

[**12**] A. Björk and L. Eldén, *Methods in numerical algebra for ill-posed problems*, Report LiTH-MAT-R33-1979, Dept of Mathematics, Linköping University, Linköping, Sweden, 1979.

[**13**] *Numerical treatment of ill-posed problems*, in Inverse Problems, A. Dold and E. Eckmann, eds., Lecture Notes in Math. 1225, Springer-Verlag, New York, 1986.

[**14**] B. W. Rust and W. R. Burrus, *Mathematical Programming and the Numerical Solution of Linear Equations*, Elsevier, New York, 1972.

[**15**]J. M. Varah, *A practical examination of some numerical methods for linear discrete ill-posed problems*, SIAM Rev., 21 (1979), pp. 100-111.

[**16**] *Pitfalls in the numerical solutions of the ill-posed problems*, SIAM J. Sci. Statist. Comput., 4 (1983), pp. 164-176.

[**17**] B. Hofmann, *Regularization for Appliead Inverse and Ill-Posed Problems*, Teubner-Texte Mathe., 85, Teubner, Leipzig, 1986.

[**18**] Per Christian Hansen, *Analysis of Discrete Ill-Posed Problems by Means of the L-curve*, SIAM Rev., Vol. 34, No.4 , pp. 561-580, 1992.

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

[**20**] 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), pp. 1545-1556, June 1998.

[**21**] Eric L. Miller, Alan Willsky, *Wavelet based methods for the nonlinear inverse scattering problem using the Extended Born Approximation*, Radio Sci., 31(1), pp. 51-67, 1996.

[**22**] *Perturbation bounds for discrete Tikhonov regularization*, Inverse Problems, 5 (1989), pp. L41-L45.

[**23**] A. Björk, *Least Squares Methods*, in Handbook of Numerical Analysis, Vol. I: Finite Difference Methods-Solution of Equations in R^n, P. G. Ciarlet and J. L. Lions, eds., Elsevier, New York, 1990.

[**24**] 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.

[**25**] *Regularization, GSVD and truncated GSVD*, BIT, 29 (1989), pp. 491-504.

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

[**27**] S. Osher L. I. Rudin and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys., D, 60, pp. 561-580, 1992.