Blind Deconvolution of Ultrasonic Traces Accounting for Pulse Variance

Kjetil F. Kaaresen and Erik Bølviken

Abstract—The ability of pulse-echo measurements to resolve closely spaced reflectors is limited by the duration of the ultrasonic pulse. Resolution can be improved by deconvolution, but this often fails because frequency selective attenuation introduces unknown changes in the pulse shape. In this paper we propose a maximum a posteriori algorithm for simultaneous estimation of a time varying pulse and high-resolution deconvolution. A priori information is introduced to encourage estimates where the pulse only varies slowly and the reflectivity sequence is sparse. This add sufficient regularization to the problem, and no further assumptions on the pulse such as minimum phase or a particular parametric form are needed. The joint pulse and reflectivity estimate is computed iteratively by alternating steps of pulse estimation and reflectivity estimation. The first step only amounts to a linear least squares fit. The second step is a difficult combinatorial optimization problem which we solve by a sub-optimal but efficient search procedure. Due to the sparseness assumption, our approach is particularly suited for layered media containing a limited number of abrupt impedance changes. This is a situation of interest in many applications of non-destructive evaluation. Synthetic and real data results show that the algorithm works well.

Keywords— Adaptive deconvolution, MAP estimation, iterated window maximization.

I. INTRODUCTION

In ultrasonic non-destructive evaluation (NDE) a focused acoustic wave of short duration is transmitted into an object. Due to internal impedance changes, part of the incident energy is reflected and can be measured to obtain information about the internal structure of the object. Under simplifying assumptions, the recorded ultrasonic trace (A-line) can be modeled as a 1-D convolution between a pulse shape and the reflectivity of the insonified medium [1]. The convolution with the pulse "smears" out fine details in the reflectivity and make interpretation of closely spaced reflectors difficult. It is therefore desirable to (partially) remove the effect of the pulse, i.e. perform deconvolution. A major difficulty is that the incident pulse often changes considerably as it passes trough the medium. Generally, the high-frequency components will be attenuated more severely than the low-frequency components. Linear with frequency attenuation models have been shown to approx-

imate the pulse changes in some biological tissues [1], but, in general, the exact nature of the changes is poorly known. It is thus necessary to estimate both the reflectivity and the pulse from the same data (*blind* deconvolution), with the additional complication that the pulse is slowly time varying.

Many adaptive deconvolution algorithms have been designed to cope with time varying pulses. The majority are autocorrelation based solutions (cf. e.g. [2], [3]) which are incapable of estimating the phase of the pulse. These typically resort to a questionable minimum phase assumption. Adaptive deconvolution methods based on higher order statistics [4] can identify an arbitrary pulse as long the reflectivity is non-Gaussian and (locally) stationary. The price to pay for this generality, however, is that large amounts of data will be necessary for precise pulse estimates [4], [5]. In particular in the time varying case, this can be severe since only short data segments can be considered as reasonably stationary.

To improve the estimates for finite amounts of data, we shall introduce more a priori information. In NDE, the insonified material will often admit some form of layered structure. Large impedance contrasts typically exist only in a limited number of positions, corresponding to interfaces between different materials, faults etc. This implies that the reflectivity will be sparse, i.e. only at a limited number of samples have non-zero values. Sparseness is a powerful constraint that have been used extensively in seismic applications. Bernoulli-Gaussian modeling and generalized maximum likelihood solutions have been shown to provide precise identification of non-minimum phase pulses and high-resolution deconvolution results [6], [7], [8]. The solutions have mainly been limited to the time-invariant case. Natural extensions to time varying pulses require either a parametric model for the pulse variations or some sort of time gating. Parametric models have been proposed [9], but these appear quite restrictive and may easily enforce an unrealistic structure on the data. Time gating consist of dividing the trace into (possibly overlapping) intervals and carry out standard estimation in each interval. A fundamental problem is that long intervals will violate pulse invariance, whereas short intervals may not contain enough data for precise estimation. A previous approach given by Chi and Chen [10] proceeds recursively and performs time-invariant estimation in moderately long and over-

lapping intervals.

In this paper, we propose a different solution. The idea is to use short intervals, but to constrain the solution by the a priori knowledge that the pulse only varies moderately between neighboring intervals. To do so, we formulate the problem in a Bayesian framework. A priori information about the unknowns (the pulse and the reflectivity) is represented as *prior* probability distributions. To combine the information in the prior and the observed data, we use a *maximum a posteriori* (MAP) estimate. This is the maximizer of an objective function (the posterior) defined as the conditional probability density of the unknowns given the observed data. Computing this maximum is a difficult optimization problem. Instead of directly maximizing with respect to all unknowns, we use the following *block component method* (BCM) [6]:

1) Start with an initial guess for the reflectivity.
2) Compute the MAP estimate of the pulse treating the reflectivity estimate from the preceding step as known.
3) Compute the MAP estimate of the reflectivity treating the pulse estimate from the preceding step as known.
4) Repeat from 2) until convergence.

In our Bayesian setting this is a version of the iterated conditional modes procedure [11] with vector valued variables. Thus, convergence to a (possibly local) maximum of the overall objective function is ensured [11]. No guarantee for global optimality can be given, but simulations suggest that a satisfactory solution will usually be found.

The two sub-problems that must be solved are fundamentally different. With suitable Gaussian distribution assumptions, estimation of the pulse for fixed reflectivity (step 2 above) only requires a linear least squares fit. However, due to the dimensions of the involved quantities, the implied inversion may easily be too burdensome. To avoid this problem, we give an equivalent iterative solution that will be much faster in many cases. We also give an extension to the multichannel case when several traces are available. Provided the pulse is invariant across the traces, this leads to major improvements of the estimates.

Estimation of the Bernoulli Gaussian reflectivity for fixed pulse (step 3 above) is a difficult combinatorial optimization problem that is usually solved by various iterative and sub-optimal methods [8], [12], [13] [14], [15], [16]. Any of these methods are candidates for the optimization in step 3. Different choices lead to varying degrees of optimality and computational efficiency. In this paper we shall consentrate on the *iterated window maximization* (IWM) algorithm [16], which we have found to perform better than a selected sample of well-established alternatives [17]. IWM uses local updating and pre-storing of some key-quantities to achieve a very rapid iterative search. Note also a previous deconvolution algorithm devised by Powell and described by Zala [18]. This algorithm appears related to IWM and has been used as one part of a successful inversion scheme for NDE data [18].

Since IWM does not require the pulse to be constant on intervals, we interpolate the estimates from the pulse estimation step. Thus, the reflectivity estimates are based on a continuously varying and non-parametric pulse estimate.

The rest of the paper is organized as follows: In Section II the model is stated in detail and the method used for pulse estimation is derived. Section III treats the reflectivity estimation step. A short review of the IWM algorithm adapted to the present context is given. For derivation and further discussion the reader is referred to [16]. In Section IV results from synthetic and real NDE data are presented which show that the proposed method works well. Finally, differences to some previous works are discussed in Section V.

II. PULSE ESTIMATION

A. Model

We model the observed trace, z, by the time-variant convolutional model

$$z(n) = \sum_{k=0}^{K} x(n-k)h(k, n-k) + e(n), \quad n = 1, 2, \ldots N,$$
(1)

where x is the reflectivity of the insonified material, h is the pulse, and e is additive noise. Note that $h(\cdot, n)$ is the pulse shape resulting from a reflector at position n. Although the shape depends on n, we assume that there is a fixed K such that $h(k, n) = 0$ for $k < 0$ and $k > K$. To avoid separate treatment of some border conditions, we also assume that the observed record contain no truncated reflections, i.e. $x(n) = 0$ for $n < 0$ and $n > N - K$. The noise term accounts for measurement noise and other effects not explained by our model. We take it to be zero-mean, Gaussian, and white, with variance σ_e^2.

To achieve a computationally efficient solution, we assume that the pulse is invariant on (small) intervals, $(n_{i-1}, n_i]$, for $i = 1, 2, \ldots I$, $n_0 = 0$, and $n_I = N$. Set

$$h_i(k) = h(k, n), \quad n_{i-1} < n \le n_i$$
(2)

and

$$x_i(n) = \begin{cases} x(n), & n_{i-1} < n \le n_i, \\ 0, & \text{else.} \end{cases}$$
(3)

Using (2) and (3), model (1) can be written as

$$z(n) = \sum_{i=1}^{I} \sum_{k=0}^{K} x_i(n-k)h_i(k) + e(n).$$
(4)

In matrix form (4) becomes

$$\mathbf{z} = \sum_{i=1}^{I} \mathbf{X}_i \mathbf{h}_i + \mathbf{e} \qquad (5)$$

with obvious notation. Setting $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_I)$ and $\mathbf{h} = \mathrm{col}(\mathbf{h}_1, \ldots, \mathbf{h}_I)$, equation (5) can be more compactly be written as

$$\mathbf{z} = \mathbf{X}\mathbf{h} + \mathbf{e} \qquad (6)$$

We take the prior for $\mathbf{h}$ to be zero mean Gaussian with covariance matrix $\mathbf{\Sigma_h}$. For given $\mathbf{X}$, the MAP estimator in the linear Gaussian model (6) is then well known to be (c.f. e.g. [19])

$$\hat{\mathbf{h}} = \left(\mathbf{B} + \sigma_{\mathbf{e}}^2 \mathbf{\Sigma_h}^{-1} \right)^{-1} \mathbf{u}, \qquad (7)$$

where $\mathbf{B} = \mathbf{X}'\mathbf{X}$ and $\mathbf{u} = \mathbf{X}'\mathbf{z}$.

B. Prior

To compute (7) the covariance matrix of $\mathbf{h}$ must be specified. We want the pulse to wary only moderately between neighboring intervals. To encourage such continuity we use a first order auto-regressive vector process, defined by taking $\mathbf{h}_i$ given $\mathbf{h}_{i-1}$ to be Gaussian with mean $\alpha\mathbf{h}_{i-1}$ and variance $\mathbf{\Sigma}$. The chain is initialized by letting $\mathbf{h}_1$ be Gaussian with mean 0 and variance $(1 - \alpha^2)^{-1}\mathbf{\Sigma}$. (This makes the process stationary.) Using these definitions it is not difficult to see that $\mathbf{h}$ has covariance matrix

$$\mathbf{\Sigma_h} = \frac{1}{1 - \alpha^2} \begin{pmatrix} \mathbf{\Sigma} & \alpha\mathbf{\Sigma} & & \cdots & \alpha^{I-1}\mathbf{\Sigma} \\ & \mathbf{\Sigma} & \alpha\mathbf{\Sigma} & & \vdots \\ & & \ddots & \ddots & \\ \text{SYMMETRIC} & & \mathbf{\Sigma} & \alpha\mathbf{\Sigma} \\ & & & & \mathbf{\Sigma} \end{pmatrix}. \qquad (8)$$

The inverse, which is needed in (7), is

$$\mathbf{\Sigma_h}^{-1} = \begin{pmatrix} \mathbf{\Sigma}^{-1} & -\alpha\mathbf{\Sigma}^{-1} & & & \\ & (1+\alpha^2)\mathbf{\Sigma}^{-1} & -\alpha\mathbf{\Sigma}^{-1} & & \mathbf{0} \\ & & \ddots & \ddots & \\ \text{SYMMETRIC} & & & (1+\alpha^2)\mathbf{\Sigma}^{-1} & -\alpha\mathbf{\Sigma}^{-1} \\ & & & & \mathbf{\Sigma}^{-1} \end{pmatrix}. \qquad (9)$$

The sparse structure of the inverse results from the choice of an AR model. Higher order AR models would also give rather simple solutions. However, we think that the first order model provide an adequate description of the prior assumption of continuity and will not consider higher order models.

C. Iterative maximization

If there are many intervals, solving the linear system in (7) may be costly. Since both $\mathbf{B}$ and $\mathbf{\Sigma_h}^{-1}$ will be sparse, various standard sparse matrix techniques [20] could be used. Here, we shall use an iterative solution that is basically a vector-valued version of the familiar Gauss-Seidel procedure [21]. To gain some additional insight, we derive the solution from a statistical point of view and show that it is also an instance of iterated conditional modes with vector valued variables. Instead of directly determining the MAP estimate of $\mathbf{h}$ as in (7), we compute the conditional MAP estimate of $\mathbf{h}_i$ given all $\mathbf{h}_j$, $j \neq i$. Repeating this procedure iteratively for all i will lead at least to a local maximum of the posterior [11]. Since the posterior in this case (conditional on $\mathbf{X}$) is multivariate Gaussian and only has a global maximum, the procedure will certainly converge to the global optimum. It is thus equivalent to the one-shot inversion in (7). Note that an iterative inversion procedure is particularly well suited in the present context. Since pulse estimation is performed repeatedly as part of the overall BCM, the previous pulse estimate provides a good starting point.

Using the unconditional moments in (8) and basic properties of multivariate Gaussian vectors [22] the prior mean for $\mathbf{h}_i$ given $\mathbf{h}_j$, $j \neq i$ is seen to be

$$\bar{\mathbf{h}}_i = \begin{cases} \alpha\mathbf{h}_2, & i = 1, \\ \dfrac{\alpha}{1 + \alpha^2}(\mathbf{h}_{i-1} + \mathbf{h}_{i+1}), & 1 < i < I, \\ \alpha\mathbf{h}_{I-1}, & i = I. \end{cases} \qquad (10)$$

Correspondingly, the conditional covariance matrix is

$$\mathbf{\Sigma}_{\mathbf{h},i} = \begin{cases} \mathbf{\Sigma}, & i = 1, \\ \dfrac{1}{1 + \alpha^2}\mathbf{\Sigma}, & 1 < i < I, \\ \mathbf{\Sigma}, & i = I. \end{cases} \qquad (11)$$

The conditional MAP estimate for $\mathbf{h}_i$ given $\mathbf{h}_j$, $j \neq i$ is

$$\hat{\mathbf{h}}_i = \left(\mathbf{B}_{ii} + \sigma_{\mathbf{e}}^2 \mathbf{\Sigma}_{\mathbf{h},i}^{-1} \right)^{-1} \left(\mathbf{u}_i - \sum_{j \neq i} \mathbf{B}_{ij}\mathbf{h}_j - \mathbf{B}_{ii}\bar{\mathbf{h}}_i \right) + \bar{\mathbf{h}}_i, \qquad (12)$$

where $\mathbf{B}_{ij} = \mathbf{X}_i'\mathbf{X}_j$ and $\mathbf{u}_i = \mathbf{X}_i'\mathbf{z}$. This can be seen by rearranging (5) as

$$\tilde{\mathbf{z}}_i = \mathbf{X}_i\tilde{\mathbf{h}}_i + \mathbf{e}, \qquad (13)$$

where

$$\tilde{\mathbf{z}}_i = \mathbf{z} - \sum_{j \neq i} \mathbf{X}_j \mathbf{h}_j - \mathbf{X}_i \bar{\mathbf{h}}_i \quad \text{and} \quad \tilde{\mathbf{h}}_i = \mathbf{h}_i - \bar{\mathbf{h}}_i. \quad (14)$$

Model (13) is on the same form as (6). Substituting $\mathbf{z}$ by $\tilde{\mathbf{z}}_i$, $\mathbf{X}$ by $\mathbf{X}_i$, and $\boldsymbol{\Sigma}_{\mathbf{h}}$ by $\boldsymbol{\Sigma}_{\mathbf{h},i}$ in (7) gives the conditional MAP estimate of $\tilde{\mathbf{h}}_i$. Linearly transforming this estimate to an estimate for $\mathbf{h}_i$ and inserting definitions (14) gives (12).

The unconditional estimate for $\mathbf{h}$ can now be found by solving (12) iteratively for each i. Using the definition of $\mathbf{X}_i$ and the assumption that the observed record contains no truncated reflections, it is easily seen that the $\mathbf{B}_{ij}$ is Toeplitz (each diagonal band is constant). Provided that $\boldsymbol{\Sigma}^{-1}$ is chosen on a Toeplitz form to, the inversion in (12) can be carried out in order K^2 by Levinson recursion [20]. For efficient execution the $\mathbf{u}_i$ and the bands of the $\mathbf{B}_{ij}$ should be computed and stored prior to the iteration. Note also that provided the intervals are not shorter than K, $\mathbf{B}_{ij} = 0$ for $|i-j| > 1$. Thus, the sum in (12) reduces to three terms. Also since $\mathbf{B}_{ij} = \mathbf{B}'_{ji}$ it is only necessary to compute and store the bands of $\mathbf{B}_{ii}$ and $\mathbf{B}_{i,i+1}$.

After the $\mathbf{B}_{ij}$ and $\mathbf{u}_i$ have been initialized, the total operation count for the iterative pulse estimation is on the order JIK^2, where J is the number of iterations. Since direct inversion in (7) will require on the order of I^3K^3 operations it is clear that the iterative method may be much faster provided that J stays reasonable.

D. Multichannel Extension

We now consider data consisting of several traces (channels) indexed by $c = 1, 2, \ldots, C$. Assuming the pulse to be common to all traces, (5) can be generalized in an obvious way to

$$\mathbf{z}_c = \sum_{i=1}^{I} \mathbf{X}_{c,i} \mathbf{h}_i + \mathbf{e}_c. \quad (15)$$

Provided the noise from the different traces are independent and have the same variance, it can be seen that the MAP estimator for this extended model is obtained by simple averaging operations. Specifically, (12) can still be used given the new definitions

$$\mathbf{B}_{ij} = \sum_{c=1}^{C} \mathbf{B}_{c,ij}, \quad \text{and} \quad \mathbf{u}_i = \sum_{c=1}^{C} \mathbf{u}_{c,i},$$

where $\mathbf{B}_{c,ij} = \mathbf{X}'_{c,i} \mathbf{X}_{c,j}$ and $\mathbf{u}_{c,i} = \mathbf{X}'_{c,i} \mathbf{z}_c$.

E. Parameter Selection and Implementation

By varying the choice of the AR parameter α and the basic covariance matrix $\boldsymbol{\Sigma}$, it is possible to express a wide range of prior assumptions. For example, by appropriate choice of $\boldsymbol{\Sigma}$ it would be possible to express continuity of each $\mathbf{h}_i$ as well as continuity between $\mathbf{h}_i$ and $\mathbf{h}_{i+1}$. In the following we shall restrict attention to the last form of continuity which is consistent with our primary motivation as stated in the introduction. Specifically we choose $\boldsymbol{\Sigma}$ on the simple form $\sigma_{\mathbf{h}}^2 \mathbf{I}$ and $\alpha = 1$. The last choice makes the prior improper. It is non-informative [23] with respect to the overall mean level of the AR process. The posterior and resulting MAP estimate is, however, perfectly valid. Note also from (10) that this choice is intuitively pleasing since the conditional prior expectation of the pulse in any given interval reduces to the average of the pulses in the neighboring intervals.

From (11)-(12) it can be seen that the remaining parameter $\sigma_{\mathbf{h}}$ determines the degree of continuity between neighboring intervals. Its value should be chosen relative to the interval size. For short intervals the pulse will vary little between neighboring intervals and a small $\sigma_{\mathbf{h}}$ is appropriate. To obtain a truly continuous estimate, we would ideally like to use intervals containing only one sample and enforce the necessary continuity by a sufficiently small $\sigma_{\mathbf{h}}$. However, this would normally be computationally prohibitive. For the results presented here we took the interval size equal to the pulse length which was 50 samples. Based on some preliminary experiments we chose $\sigma_{\mathbf{h}} = 0.03$. This ensured reasonable continuity between neighboring intervals, while still enabling real pulse variations to be tracked. Note that the convolutional model (1) is indeterminate with respect to an inverse scaling of h and x. This ambiguity was resolved by simply normalizing each $h(\cdot, n)$ to unit energy after each step of pulse estimation. The chosen value of $\sigma_{\mathbf{h}}$ is thus relative to a unit energy pulse.

The noise standard deviation, $\sigma_{\mathbf{e}}$, was taken as the true value for the synthetic data and heuristically selected as 4% of the maximum data value for the real data. Note that (12) actually only depend on the ratio $\sigma_{\mathbf{e}}^2/\sigma_{\mathbf{h}}^2$. Thus, with the above simplifications the pulse estimation step only require specification of one free parameter, which is directly related to the continuity of the produced estimate.

To turn the piece-wise constant pulse estimate into a truly continuously varying pulse to be used as input to the reflectivity estimation step, a final step of linear interpolation was used. To ensure the validity of this interpolation, the pulse estimates from neighboring intervals were shifted (if necessary) to obtain maximal correlation. Note that the interpolation approximation makes the objective function of the pulse and reflectivity estimation steps slightly different. Theoretically this could disturb the convergence properties of the outer BCM and result in small oscillations. However, this was never observed in our experiments. A safeguard is to set an upper limit to the number of BCM iterations or to avoid interpolation by using the piece-wise constant pulse estimate directly.

III. Reflectivity Estimation

We now describe the IWM algorithm [16] which is used to compute the MAP estimate of the reflectivity, $\mathbf{x}$, for given $\mathbf{h}$. We assume that $\mathbf{x}$ is independent of $\mathbf{h}$ and has a Bernoulli-Gaussian [6] prior defined as follows: The components are independent. With probability $1 - \lambda$ they are zero and with probability λ they are sampled from a zero-mean Gaussian distribution. The variance of this Gaussian is taken to be infinite. This non-informative limit leads to some simplifications and usually works well [16]. Other assumptions are as in the previous section.

The IWM algorithm is based on an equivalent representation of $\mathbf{x}$ by two variable length vectors, $\mathbf{t}$ giving the positions of the reflectors (non-null elements) and $\mathbf{a}$ giving their amplitudes. Computing the optimal $\mathbf{a}$ for given $\mathbf{t}$ only requires a linear least squares fit. Estimation of $\mathbf{t}$, however, is a difficult combinatorial problem that is solved by a (sub-optimal) iterative search. The search proceeds through a series of local changes to $\mathbf{t}$. Only changes which increase the posterior are accepted. For each new candidate value of $\mathbf{t}$, the amplitudes must essentially be refitted. However, because the new $\mathbf{t}$ values only differ from the previous one in a few components, it is not necessary to refit all the amplitudes. Instead only those inside a small window, at least covering the changed reflectors, are updated. This reduces the dimension of the least squares fit to the number of reflectors in the window (typically in the order of 1-5). Thus, the inversion will be fast. Efficient compilation of the coefficients of the least squares problem is achieved by using a simple relationship with the two correlation functions which are computed and stored prior to the iteration. These are

$$c_{hh}(k, n) = \sum_{l=k}^{\min(K, N-n)} h(l, n) h(l-k, n+k) \quad (16)$$

and

$$c_{hz}(n) = \sum_{l=0}^{\min(K, N-n)} h(l, n) z(n+l), \quad (17)$$

where $n = 1, 2, \ldots, N$ and $k = 0, 1, \ldots, \min(K, N - n)$.

Assume now that a window w has been chosen. Let $\mathbf{t}^w$ be the components of $\mathbf{t}$ that are inside the window and $\mathbf{t}^{\overline{w}}$ be the components outside the window. Also partition $\mathbf{a}$ likewise. The optimal value of $\mathbf{a}^w$, given $\mathbf{a}^{\overline{w}}$, $\mathbf{t}$ and $\mathbf{z}$ is

$$\hat{\mathbf{a}}^w = (\mathbf{S}^w)^{-1} \mathbf{v}^w, \quad (18)$$

where

$$S_{ij}^w = c_{hh}\big(|t_i^w - t_j^w|, \min(t_i^w, t_j^w)\big) \quad (19)$$

and

$$v_i^w = c_{hz}(t_i^w) - \sum_j a_j^{\overline{w}} c_{hh}\big(|t_i^w - t_j^{\overline{w}}|, \min(t_i^w, t_j^{\overline{w}})\big). \quad (20)$$

Note from the last sum that reflectors outside the window affect the fit. However, only those producing overlapping reflections with reflectors inside the window need to be considered. Thus, the sum should be limited to j such that $|t_i^w - t_j^{\overline{w}}| \leq K$. The $(\mathbf{t}^w, \hat{\mathbf{a}}^w)$ pair corresponding to the largest value of the posterior is the one which yields the larger value of the criterion

$$l(\mathbf{t}) = (\mathbf{v}^w)' \hat{\mathbf{a}}^w - \theta M^w. \quad (21)$$

Here M^w is the number of reflectors inside the window and θ is a parameter to be discussed later.

The window positions are systematically cycled trough the entire data record. Within each window new candidate values for $\mathbf{t}^w$ are obtained by trying all changes in a predetermined *transition set*. The changes considered here were: inserting a reflector, deleting a reflector, and moving a reflector with one or more samples. In addition, some changes affecting two reflectors simultaneously were also considered (c.f. the specification of IWM 2 in [16]). Equations (18)-(21) are computed for the original value of $\mathbf{t}^w$ and for each new candidate value. If an improving candidate is found, both $\mathbf{a}$ and $\mathbf{t}$ is updated. Otherwise only $\mathbf{a}$ is updated. Then, a new window is chosen, and the procedure is repeated until neither $\mathbf{t}$ nor $\mathbf{a}$ change any more. This guarantees that the final value of $\mathbf{a}$ will be globally optimal given the final value of $\mathbf{t}$, but the final value of $\mathbf{t}$ may still be suboptimal. The degree of sub-optimality depend on the richness of the transition set. For the transition set used here, the IWM algorithm usually work well and provide higher optimality than well-established alternatives [12], [13] which are based on a much more limited exploration strategy [17]. For the applications presented here, the windows were chosen to include all reflectors with distance 15 samples or less from those that were updated.

The parameter θ govern the sparsity of the produced estimate by determining the "penalty" for each extra reflector in (21). Although the MAP estimator defines it as a function of σ_e and the parameters of the Bernoulli-Gaussian prior, this value can generally not be used [16]. Instead, we used $\theta = 50\sigma_e^2$ in all our experiments. (An argument for choosing θ proportional to σ_e^2 for known pulse is given in [16]. Here we use a somewhat larger proportionality constant to account for estimation uncertainty in the pulse.)

For the first step of reflectivity estimation, no pulse estimate is available. In this case a simple estimate formed by inserting reflectors in all local maxima of the data record was used. (A point was considered as a local maxima if it had larger absolute value than the 40 nearest samples.) In spite of the crudeness of this initial estimate, all presented reconstructions converged reasonably fast to sensible solutions.
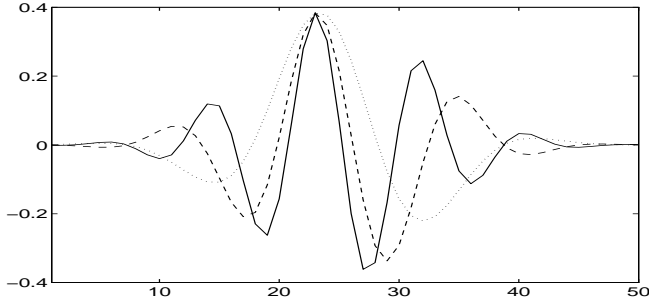
Fig. 1.   The variant pulse used for data generation: Solid line $n = 1$; dashed line $n = 500$, dotted line $n = 1000$.

IV. Results

A. Computer Simulations

A record of 1000 samples were generated from model (1) using a Bernoulli-Gaussian reflectivity with reflector density $\lambda = 0.05$. In order to provide a difficult test for the algorithm, a severely time variant pulse was used (Fig. 1). It was generated according to

$$h(k, n) = \cos\left[\omega_n (k - 22)\right] \exp\left[-(k - 24)^2/100\right],$$

for $\omega_n = 0.7 - 0.0004n$, $n = 1, 2, \ldots 1000$, and $k = 0, 1, \ldots, 49$. High frequency attenuation is evident. Otherwise, we attach no particular physical reality to the selected shape. The additive noise was taken to be Gaussian, white, and to provide signal-to-noise ratio 20 dB (defined by mean power of noiseless signal divided by noise variance).

The resulting reconstruction recovered most of the large reflectors, but there was also several missed or false detections (Fig. 2b). The basic pulse shape was tracked well, but some small deviations are clearly seen (Fig. 3a-b).

In the next example, 20 independent traces were used. This required the multichannel modification given in Section II-D. Otherwise, conditions were unchanged. In this case the pulse estimates were almost indistinguishable from the truth (Fig. 3c-d). These improved pulse estimates also lead to a clear improvement in the corresponding reflectivity estimate (Fig. 2c). We conclude that severely varying pulses can be accurately tracked provided enough data are available. It is also interesting that the quite moderate estimation error of the pulses in Fig. 3a-b worsen the corresponding reflectivity estimate considerably. This makes it clear that failing to account for even moderate pulse variations will significantly reduce the ability to resolve closely spaced reflectors. This can be expected to be true not only for the present algorithm, but also for other high resolution deconvolution techniques.
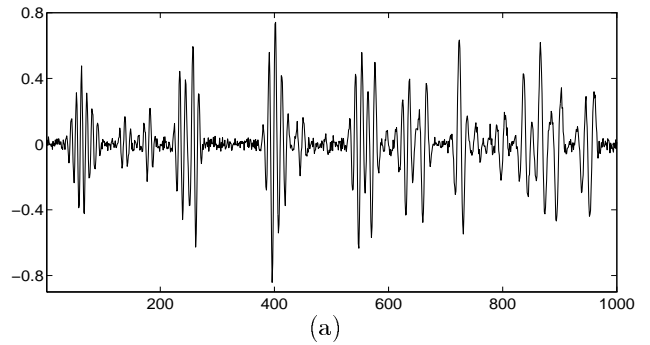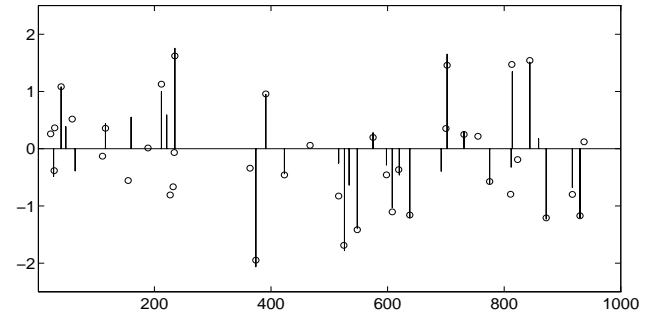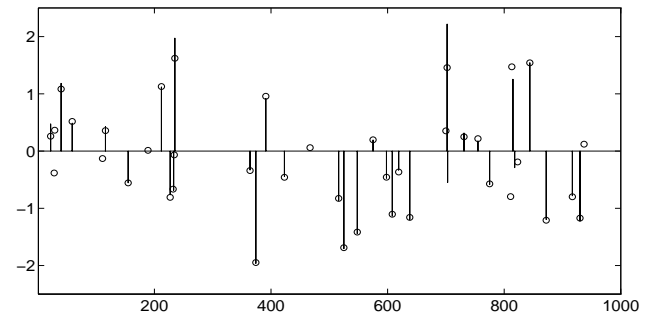
Fig. 2.   Deconvolution of synthetic data: (a) Non-stationary data; (b) deconvolution based on only one trace; (c) deconvolution based on 20 traces (only the first shown). Bars denote estimates and circles denote true values.

B. Application NDE data

The next data set is from an NDE experiment. A sample of a fiber composite material used in the production of a Swedish airplane had been exposed to a falling object. The resulting damage was investigated by scanning an ultrasound transducer along the surface of the sample. Before deconvolution, the data were exponentially weighted to strengthen week reflections at high times and a non-zero mean level was subtracted. The pre-processed data is displayed in Fig. 4a. The reflection seen in all traces around $0.2\mu s$ is the front-wall echo. The reflection around $1.7\mu s$ is the back-wall echo, while the remaining reflections are from inner delaminations cased by the falling object. Note that the back-wall echo
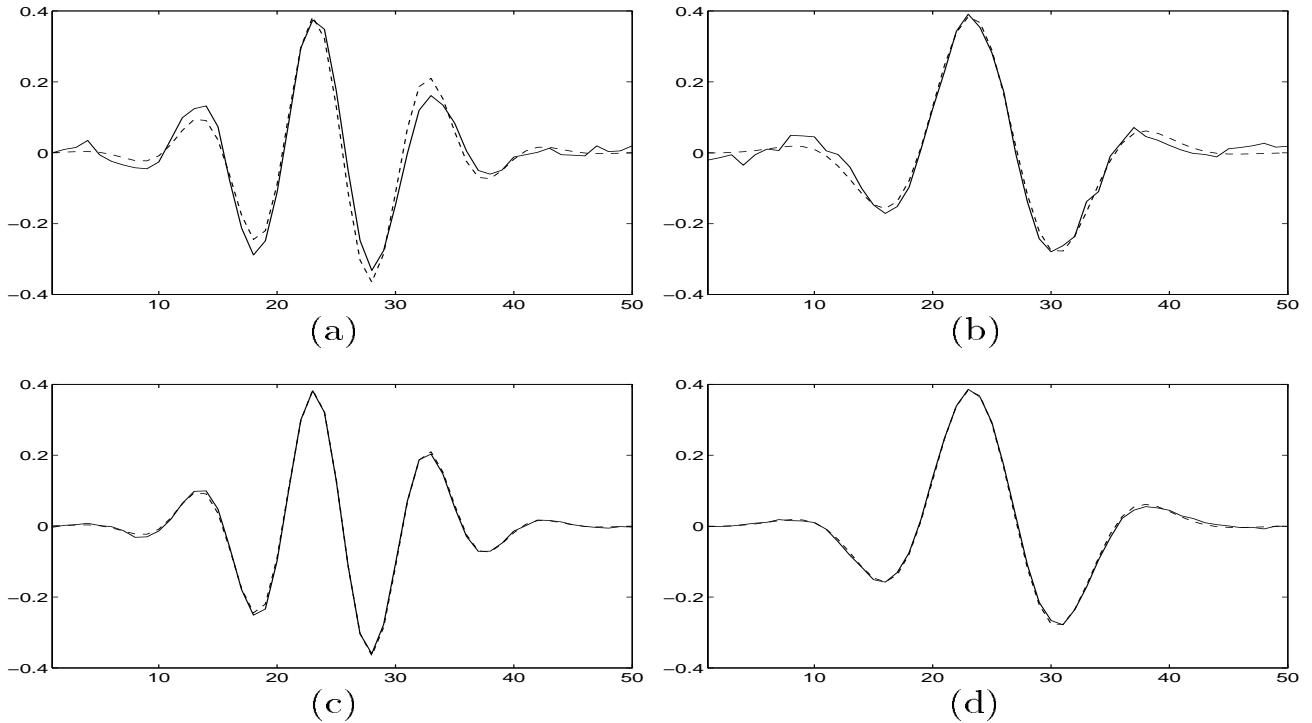
Fig. 3.   Pulse estimates for synthetic data: (a) One trace, $n = 200$; (b) one trace, $n = 800$; (c) 20 traces, $n = 200$; (d) 20 traces, $n = 800$. Solid lines denote estimates and dashed lines denote true values. Accuracy for other n was comparable to the displayed results.

is partly obscured by the strong reflections from the inner delaminations and that some of the events are clearly due to multiple reflections (in particular the event shown after the back-wall echo around $2.2\mu s$).

The obtained reflectivity estimate is shown in Fig. 4b. Note that most of the reflections from the delaminations have the same sign. Based on the conjecture that the few delaminations with opposite sign were artifacts, we computed a new estimate using a slight modification of (21). The penalty for reflectors with positive sign was doubled while the penalty for reflectors with negative sign was kept unchanged. This did not change the front-wall echo which was clearly positive, but the inner reflections now turned out to be almost purely negative (Fig. 4c). The true structure is not known, but the good lateral continuity of this reconstruction strengthen our confidence in it. (Note that such continuity was not exploited in the algorithm, though it would be possible to do so [24], [25].) The corresponding pulse estimates show slight, but clearly evident, high frequency attenuation (Fig. 5). For comparison, an estimate using an invariant pulse was also computed (Fig. 4d). This estimate appear clearly inferior. Note in particular the severe splitting of the strong reflector around $1.2\mu s$ and of the back-wall echo. This particular data set was obtained with a probe with nominal center frequency 10 MHz (in practice measured

to be somewhat higher). For probes with higher center frequency, the frequency selective attenuation would be more severe. In such case it would be even more important to account for pulse variations.

An interesting application of the given algorithm is to resolve reflections from deep (and thus serious) delaminations that overlap with multiple reflections from delaminations close to the surface. As an illustration, trace 28 from Fig. 4c is selected for detailed display in Fig. 6. The reflector labeled F is the front wall. Label A denotes a strong primary reflection. Sound energy ringing back and forth between reflector A and the front wall gives rise to the multiple reflections B and C. The point of interest is the deeper primary reflection X, that has successfully been detected by the algorithm in spite of severe overlap with multiple B. From inspection of this trace alone, it is not clear that the estimate has captured the true structure, but continuity with the estimates from neighboring traces in Fig. 4c support the obtained solution.

C. Computational Load

Each of the real data reconstructions in Fig. 4 required 10-15 iteration of the outer BCM, resulting in a total execution time of less than one minute on a 60 MHz Pentium PC. Most of the computer time were spent in the reflectivity estimation step. The computational load of this
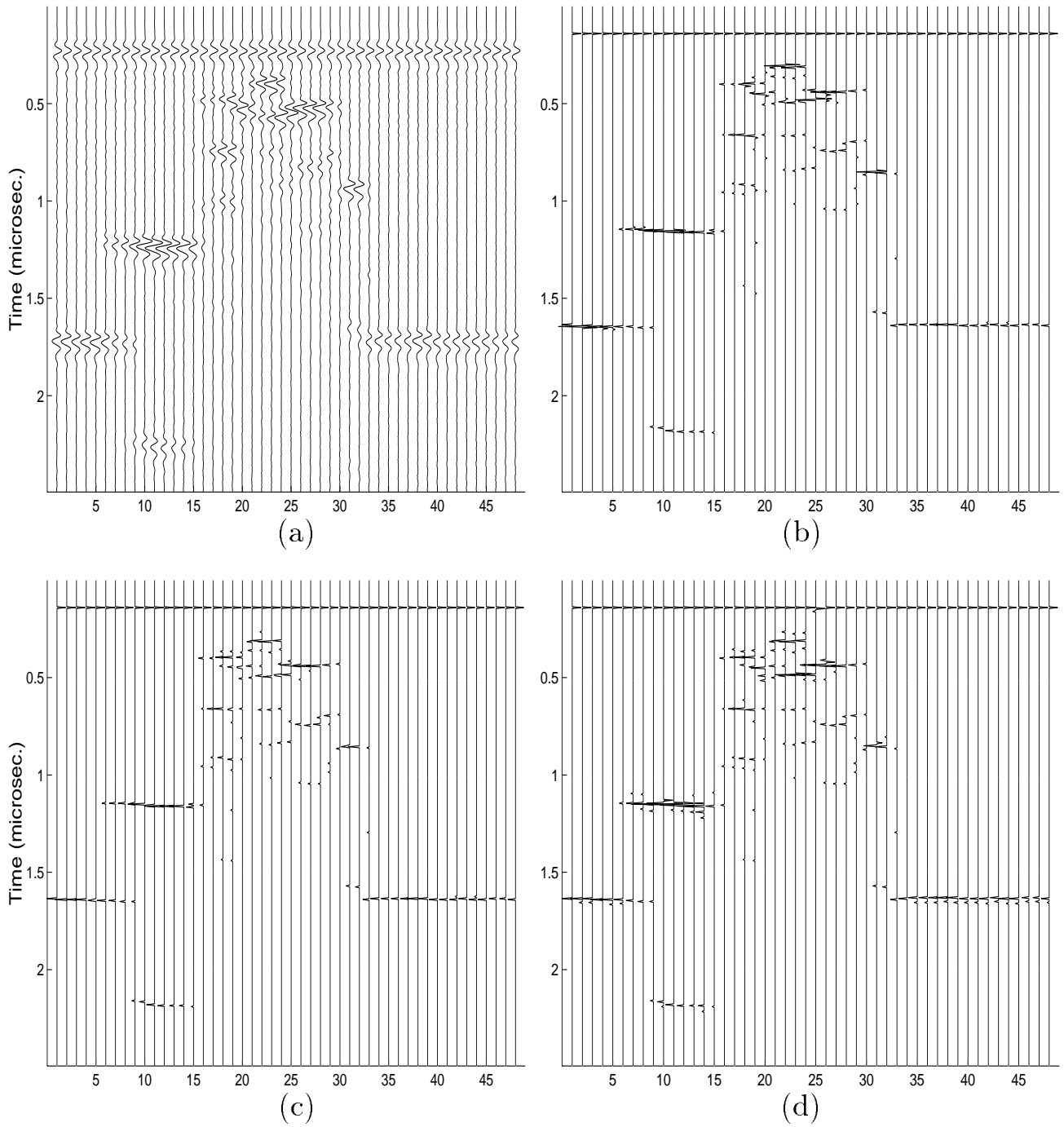
Fig. 4.   Deconvolution of NDE data: (a) Data (sampling rate 200 MHz); (b) deconvolution with variant pulse; (c) deconvolution with variant pulse and extra penalty for positive reflections; (d) deconvolution with invariant pulse and extra penalty for positive reflections.

step is quite sensitive to the number of estimated reflectors. Denser estimates will increase the average number of reflectors in each window and thereby slow down computation of (18)-(21). Reconstruction of the synthetic example with 20 traces required close to 10 minutes, even though the number of samples were approximately the same as for the real data (20 000 and 24 000 respectively). The slower execution for the synthetic data was mainly related to the much higher reflector density, but convergence was also slower (30 BCM iterations).
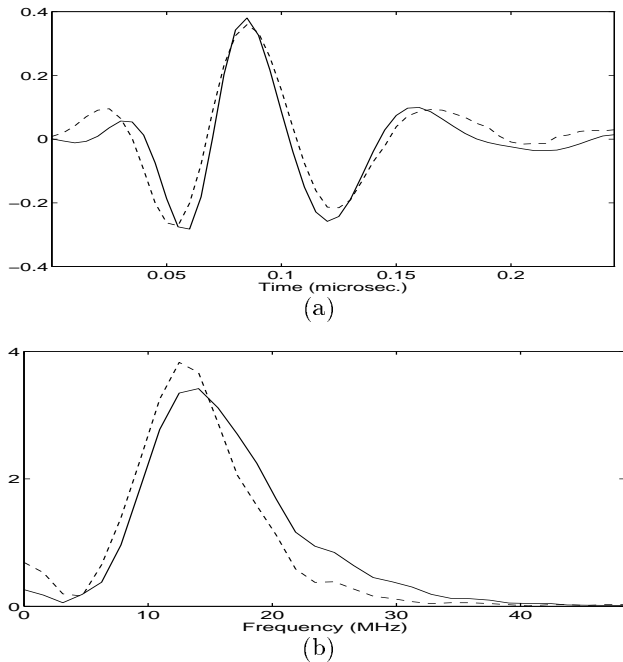
Fig. 5. Pulse estimates for NDE data: (a) Pulses; (b) amplitude spectra. Solid lines denote estimate at $0.5\,\mu s$ and dashed lines denote estimate at $2\,\mu s$. (These estimates correspond to the reflectivity estimate in Fig. 4b. Those corresponding to Fig. 4c were almost identical.)
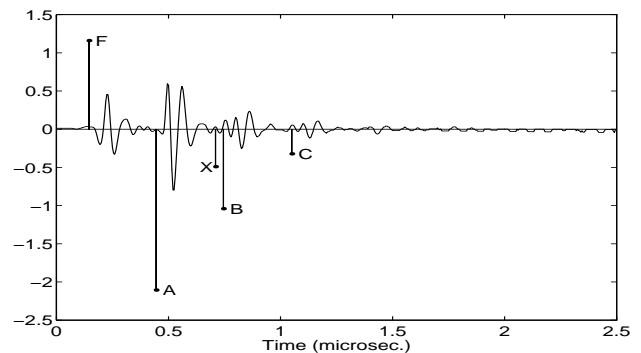

Fig. 6. Trace 28 selected from Fig. 4(c) for detailed display. Estimates are shown with bars and superposed on the data. F indicates the front-wall echo. A indicates a primary reflections with multiples B and C. X indicates a second primary reflection detected by the algorithm in spite of severe overlap with multiple B.

In all reconstructions, the iterative solution from Section II-C was used for pulse estimation. For the real data, this required 20-30 iterations for the first step of pulse estimation, but quickly reduced to 3-5 in later BCM iterations. (The reduction is due to the previous pulse estimate providing a good starting point.) We have observed, however, that convergence may be very slow in some cases. This can happen either when there is very little data in each interval or when particularly strong continuity is assumed (by using a small $\sigma_{\mathbf{h}}$). In these cases, other sparse matrix inversion techniques may be preferable. In the synthetic example with only one trace, the average number of reflectors in each interval were only 2. Here each step of pulse estimation varied from 60 to 10 iterations taking from 10 to 2 seconds. However, the dimension of the linear system was 1000×1000. Attempting direct (non-sparse) inversion required approximately 2 minutes, thus making the iterative solution preferable, even though convergence was rather slow. Adding more data improved the convergence speed. In the example with 20 traces, only 8 to 3 iterations were required taking approximately 1.5 to 0.5 seconds.

V. Conclusion and Discussion

We have presented an iterative maximum a posteriori algorithm that can simultaneously estimate a time variant pulse and perform high resolution deconvolu-

tion. The given simulations show that acceptable results can be obtained even for small data sets. For moderately large data sets, rapidly varying pulses are tracked with high accuracy, thus enabling successful resolution of severely overlapping reflections. The presented NDE example show that the algorithm also behave well for real data that will inevitably deviate slightly from the idealized model.

Sparsity of the reflectivity is fundamental to the approach. This makes the algorithm well suited for many NDE applications and some special applications of medical ultrasound [26]. In typical use of medical ultrasound, however, most of the information in the data is due to scattering caused by numerous small inhomogenities (diffuse reflections). For such data other methods should be used.

Our approach require specification of two basic parameters. One controls the continuity of the pulse estimate and the other controls the sparsity of the reflectivity estimate. Block component methods used in seismic applications have traditionally been extended by an additional step where all unknown parameters are estimated [6]. This approach has been criticized [27] for leading to possibly unstable estimates. In accordance with this, we consider these two parameters as the necessary a priori input to stabilize the estimates. They may be viewed as "tuning parameters" or possibly be estimated by other means.

In contrast to the recursive approach of Chi and Chen [10], our solution require the entire trace to be available at the time of processing. On the other hand, better estimates can be expected since they are based on more data (all future as well as past samples). In particular, the recursive approach use an initial part of the trace before it converges to the true pulse. As a result, the first part of the estimate will often be in gross er-

ror. This problem is not shared by the iterative solution given here. A further difference is the way continuity of the pulse is ensured. Here we use short non-overlapping intervals combined by a prior that encourage continuity. This is not equivalent to the smoothing obtained by using longer and overlapping intervals as in [10]. In particular when the data (the number and size of reflectors) are unevenly distributed, the difference can be large. As an illustration, consider application of the algorithm given here to a (multi-trace) data set containing a large amount of data in the first and last interval, say a front-wall and a back-wall echo, and little or no data in between. In the first and last interval, the influence of the prior will be overwhelmed by the data, and the estimate will be close to the true value. If there is no data in the intermediate intervals, the estimate in these will simply be the average of the two neighboring intervals, thus providing a smooth transition between the front- and back-wall echo. If there is some data in the intermediate intervals deviations from the linear transition may result. The degree of deviation will be consistent with the amount of data relative to the prior strength.

Consider instead using a longer sliding window for smoothing. In this case the estimate will be almost identical to the front-wall echo as long as this reflection is included in the window. Then, the estimate will abruptly change to become highly variable or even undefined when there is little or no data inside the window. Finally, a new abrupt change will occur when the back-wall echo enters the window. This sort of unsmooth behavior appears clearly undesirable. Furthermore, suppose the pulse changes significantly shortly after the front-wall echo. Say, in the second of the small intervals. If there is enough data in this interval relative to the strength of the prior, the change can be picked up by our approach. With a longer window, on the other hand, the change will be overwhelmed by the much larger amount of data provided by the front-wall echo. Thus, even if there is enough data to indicate that a statistically significant change has occurred, the long window approach will not account correctly for this in the presence of unevenly distributed data. In fact, the solution given here smooths in an optimal fashion relative to the prior belief in continuity as expressed by $\sigma_{\mathbf{h}}$ and the amount of data in each interval.

The present approach makes no distinction between primary and multiple reflections, thus relying on some post-processing or manual identification of the multiples. The inversion method of Zala [18] performs automatic detection and removal of multiples and has successfully been applied to NDE data. On the other hand, his work requires the pulse to be invariant and known a priori. Combination of his work with the present would give a complete solution that could simultaneously remove unwanted multiples and cope with unknown pulse variations.

REFERENCES

[1] J.A. Jensen and S. Leeman, "Nonparametric estimation of ultrasound pulses," *IEEE Trans. Biomed. Eng.*, vol. 41, no. 10, pp. 929–936, 1994.

[2] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, Englewood Cliffs, NJ: Prentice-Hall, 1985.

[3] S. Haykin, *Adaptive Filter Theory*, Englewood Cliffs, NJ: Prentice-Hall, 1986.

[4] H.-H. Chiang and C. L. Nikias, "Adaptive deconvolution and identification of nonminimum phase FIR systems based on cumulants," *IEEE Trans. Automatic Contr.*, vol. 35, no. 1, pp. 36–47, 1990.

[5] G. D. Lazear, "Mixed-phase wavelet estimation using fourth-order cumulants," *Geophysics*, vol. 58, no. 7, pp. 1042–1051, 1993.

[6] J. M. Mendel, *Optimal Seismic Deconvolution: An Estimation-Based Approach*, New York: Academic Press, 1983.

[7] C.-Y. Chi, J. M. Mendel, and D. Hampson, "A computationally fast approach to maximum-likelihood deconvolution," *Geophysics*, vol. 49, no. 5, pp. 550–565, 1984.

[8] J. M. Mendel, *Maximum-Likelihood Deconvolution: A Journey into Model-Based Signal Processing*, New York: Springer-Verlag, 1990.

[9] K. B. Rasmussen, "Maximum likelihood estimation of the attenuated ultrasound pulse," *IEEE Trans. Signal Processing*, vol. 42, no. 1, pp. 220– 222, 1994.

[10] C.-Y. Chi and W.-T. Chen, "An adaptive maximum-likelihood deconvolution algorithm," *Signal Processing, Elsevier*, vol. 24, pp. 149–163, 1991.

[11] J. Besag, "On the statistical analysis of dirty pictures," *Journ. Royal Statist. Soc. Ser. B*, vol. 48, no. 3, pp. 259–302, 1986.

[12] J. J. Kormylo and J. M. Mendel, "Maximum likelihood detection and estimation of Bernoulli-Gaussian processes," *IEEE Trans. Inform. Theory*, vol. 28, no. 3, pp. 482–488, 1982.

[13] Y. Goussard, G. Demoment, and J. Idier, "A new algorithm for iterative deconvolution of sparse spike trains," *Proc. IEEE, Int. Conf. Acoust., Speech, Signal Processing*, pp. 1547–1550, 1990.

[14] M. Lavielle, "Bayesian deconvolution of Bernoulli-Gaussian processes," *Signal Processing, Elsevier*, vol. 33, no. 1, pp. 67–79, 1993.

[15] F. Champagnat, Y. Goussard, and J. Idier, "Unsupervised deconvolution of sparse spike trains using stochastic approximation," *IEEE Trans. Signal Processing*, vol. 44, no. 12, pp. 2988– 2997, 1996.

[16] K. F. Kaaresen, "Deconvolution of sparse spike trains by iterated window maximization," *IEEE Trans. Signal Processing*, vol. 45, no. 5, pp. 1173– 1183, 1997.

[17] K. F. Kaaresen, "Evaluation and applications of the iterated window maximization method for sparse deconvolution," *IEEE Trans. Signal Processing, to appear*.

[18] C. A. Zala, "High-resolution inversion of ultrasonic traces," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 39, no. 4, pp. 458–463, 1992.

[19] B. P. Carlin and T. A. Louis, *Bayes and Empirical Bayes Methods for Data Analysis*, London: Chapman & Hall, 1996.

[20] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, New York: Cambridge University Press, 2 edition, 1992.

[21] H. Anton, *Elementary Linear Algebra*, New York: Wiley, 5 edition, 1987.

[22] K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis*, London: Academic Press, 1995.

[23] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*, New York: Springer-Verlag, 2 edition, 1988.

[24] J. Idier and Y. Goussard, "Multichannel seismic deconvolution," *IEEE Trans. Geosci. Remote Sensing*, vol. 31, no. 5, pp. 961–979, 1993.

[25] K. F. Kaaresen and T. Taxt, "Multichannel blind deconvolution of seismic signals," Tech. Rep. 5, Dept. of Mathematics, University of Oslo, Norway, 1997.

[26] J. Bai, W. Qi, and L. Yu, "A system for the reconstruction of acoustic impedance with high resolution," *Ultrasonics*, vol. 31, pp. 125–134, 1993.

[27] E. Gassiat, F. Monfront, and Y. Goussard, "On simultaneous signal estimation and parameter identification using a generalized likelihood approach," *IEEE Trans. Inform. Theory*, vol. 38, no. 1, pp. 157–162, 1992.