

# Sparse Deconvolution of B-Scan Images

Tomas Olofsson and Erik Wennerström

**Abstract**—In this paper, a new computationally efficient sparse deconvolution algorithm for the use on B-scan images from objects with relatively few scattering targets is presented. It is based on a linear image formation model that has been used earlier in connection with linear minimum mean squared error (MMSE) two-dimensional (2-D) deconvolution. The MMSE deconvolution results have shown improved resolution compared to synthetic aperture focusing technique (SAFT), but at the cost of increased computation time. The proposed algorithm uses the sparsity of the image, reducing the degrees of freedom in the reconstruction problem, to reduce the computation time and to improve the resolution. The dominating task in the algorithm consists in detecting the set of active scattering targets, which is done by iterating between one up-dating pass that detects new points to include in the set, and a down-dating pass that removes redundant points. In the up-date, a spatio-temporal matched filter is used to isolate potential candidates. A subset of those are chosen using a detection criterion. The amplitudes of the detected scatterers are found by MMSE. The algorithm properties are illustrated using synthetic and real B-scan. The results show excellent resolution enhancement- and noise-suppression capabilities. The involved computation times are analyzed.

## I. INTRODUCTION

RESOLUTION enhancement of ultrasonic pulse-echo B-scan images can be achieved using synthetic aperture focusing technique (SAFT) or using estimation based approaches, such as two-dimensional (2-D) deconvolution techniques. In the estimation-based approaches, the entities of interest are the scattering strengths from points in the image plane. Although we may argue that SAFT also aims at finding this map of scattering strengths, these methods do not use any statistically-motivated method for estimation; rather they should be interpreted as mimicking the behavior of an acoustical lens using the delay-and-sum (DAS) technique. As such, they inherit the intrinsic limitations of DAS; the lateral resolution is diffraction limited, and grating lobes appear when using too coarse a spatial sampling [1].

Methods using the estimation-based approach can, for images with low noise levels, overcome the above-mentioned limitations and far outperform SAFT. For mechanically scanned B-scans, examples of this class of methods can be found in [2] in which linear minimum mean squared error (MMSE) was used to estimate the scatter map. A similar method based on singular value decomposition (SVD) regularization is found in [3]. One disadvantage

of the methods is that the computation time is typically much larger than for SAFT processing, often limiting their use only to applications without real-time constraints.

Linear methods as those in [2] and [3] are best adapted to objects containing diffuse scatterers, with a Gaussian distribution modeling the amplitudes of these. In many application, particularly in nondestructive testing (NDT), the images are better described by a collection of a few but relatively strong contribution, for instance, indicating small cracks or material inclusions. For such images, sparse models and sparse deconvolution techniques are more appropriate.

The majority of work on sparse deconvolution has concerned 1-D signals. It was originally developed in seismic exploration for detecting layered earth structures. The reflections from the layers were modeled by spiky sequences having a Bernoulli-Gaussian distribution, convolved with a wavelet describing a damped oscillation caused by an impulsive source. Sparse deconvolution can be described as the combined detection of the spikes and estimation of their amplitudes. In early approaches the spike positions were detected using a recursive Kalman filter approach [4] that was based on a state-space description of the wavelet. Later a more flexible linear algebra approach involving finite-impulse response matrices was introduced [5]. This approach has been used in ultrasonic applications for deconvolving A-scans acquired from layered structures [6], [7].

The detection of spikes in sparse deconvolution is made complicated by overlapping contributions from these and iterative search approaches have been used to find the most likely combination of spike positions. The single most likely replacement (SMLR) algorithm [4] has been widely used for this purpose in both the Kalman and linear algebra approaches.

SMLR iteratively scans every potential spike position and finds the replacement yielding the largest improvement in the chosen criterion. The replacement may consist in either including or removing a point from a set of positions found so far. This method has been further developed in two directions to improve the efficiency: In the iterated window maximization (IWM) method [8], the signals are partitioned in a number of smaller time windows, and the global detection criterion is maximized through iterative local maximizations in these windows. In [9], the multiple most likely replacement (MMLR) algorithm was proposed. At a very small increased computation cost at each iteration, the algorithm in [9] allowed several nonoverlapping spikes to be replaced at each iteration, as opposed to a single in SMLR, thus reducing the number of iterations required for convergence.

Manuscript received April 29, 2006; accepted October 18, 2006.

The authors are with the Signals & Systems Group, Uppsala University, SE-751 21 Uppsala, Sweden (e-mail: tol@signal.uu.se).

Digital Object Identifier 10.1109/TUFFC.2007.434

One of relatively few examples of sparse deconvolution in imaging applications can be found in star field astronomy [10] in which a spatially invariant point spread function (PSF) well describes the image formation. Note, however, that the linear algebra approach mentioned above allows for treating any linear measurement model, in which image generation models with spatially variant PSFs are important special cases. A recent paper [11] proposed the use of 2-D sparse deconvolution for ultrasonic B-scan images in which such spatially varying PSFs were considered. Note that the involved matrices and vectors are usually much larger in imaging applications than in 1-D deconvolution. This is particularly true for spatially variant PSFs relevant to mechanically scanned B-scans and time- and memory-efficient methods are therefore much desired.

In [11] it was shown that the sparsity of the image could be used to significantly reduce the time required for the estimation compared to linear MMSE solutions. An efficient search strategy was developed that can be regarded as a 2-D extension of the MMLR search in [9], which was based on recursions that are not applicable for images. In [11], the search strategy was combined with efficient computations of the search criterion and a simulation study verified that significant time saving was possible. However, few algorithm details were given, and no results were presented for real data. This paper gives a more thorough presentation of the algorithm and presents results for real B-scans.

This paper is organized as follows: In Section II, the image formation model and the search criterion are presented. In Section III, the algorithm is presented along with some of the related computational issues. Simulation and experimental results are presented in Section IV. In Section V, conclusions and comments are given.

## II. THEORY

### A. Image Model

Consider the measurement setup depicted in Fig. 1 in which a transducer is scanned in the  $x$ -direction and transmits in the  $z$ -direction. At each of the  $L$  scanning positions, a pulse-echo measurement is performed insonifying the region of interest (ROI). The ROI, which lies in the  $zx$ -plane, consists of a rectangular grid of  $M \times N$  equally spaced potential scatterers. The spacing is  $\Delta x$  and  $\Delta z$  in  $x$  and  $z$  directions, respectively. Furthermore,  $\Delta x$  is set equal to the distance between the scanning positions. Although nothing prevents one from choosing a quite arbitrary grid, this particular choice results in simplified algorithm implementations.

The received A-scans are sampled at  $K$  time instants with a sampling period  $T_s$ . We have chosen  $\Delta z = T_s c_p / 2$ , where  $c_p$  is the sound speed. This corresponds to one grid point per time sample.

Let  $y(k, l)$  denote the  $l$ th A-scan, sampled at time  $k$  and let  $\mathbf{y}_l = \text{col}(y(1, l), y(2, l), \dots, y(K, l))$  represent this A-scan as a  $K \times 1$  vector.  $L$  such column vectors are acquired

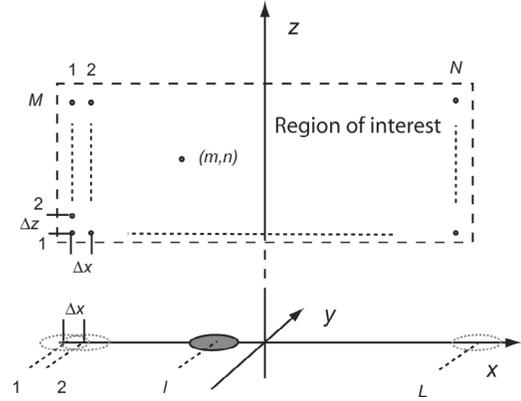


Fig. 1. Illustration of the measurement setup in the text. The ROI is divided in an  $M \times N$  grid with the center points in the grid separated with  $\Delta x$  and  $\Delta z$  in the  $x$ - and  $z$ -directions, respectively. Pulse-echo measurements are performed at  $L$  transducer positions with a scan increment of  $\Delta x$ , resulting in a B-scan of size  $K \times L$ .

and the B-scan can be represented by a column vector  $\mathbf{y} = (\mathbf{y}_1^T \dots \mathbf{y}_L^T)^T$ , i.e., a column vector of size  $KL \times 1$  consisting of the A-scans placed on top of each other.

Let further the column vector  $\mathbf{p}_{(m,n)}$  denote the vectorized B-scan that a hypothetical (noise free) response from a unit strength scatterer at grid position  $(m, n)$  would result in. This template is defined by the electromechanical impulse response of the transducer and the double-path spatial impulse responses (SIRs) associated with the scatterer position (see [2] for details).

We consider the B-scan to be the result of backscattering from a subset,  $\mathcal{I}$ , of the points in the ROI. In the following we term this subset the set of active scatterers. Let  $n_a$  denote the number of points in this set.

If we neglect multiple scattering, we can model the B-scan as a noise-corrupted superposition of contributions from active scatterers:

$$\mathbf{y} = \sum_{(m,n) \in \mathcal{I}} \mathbf{P}_{(m,n)} o(m,n) + \mathbf{e} = \mathbf{P}_{\mathcal{I}} \mathbf{o}_{\mathcal{I}} + \mathbf{e}, \quad (1)$$

where the sum runs over only the positions listed in  $\mathcal{I}$ . The matrix  $\mathbf{P}_{\mathcal{I}}$  consists of  $n_a$  columns  $\mathbf{p}_{(m,n)}$  with  $(m, n) \in \mathcal{I}$  and the  $n_a \times 1$  vector  $\mathbf{o}_{\mathcal{I}}$  consists of the corresponding amplitudes  $o(m, n)$ . The vector  $\mathbf{e}$  represents measurement noise. In practice, we let this vector model anything that cannot be explained by the linear relation in (1). This could include model discrepancies, such as errors in the template  $\mathbf{p}_{(m,n)}$  predicting the response from a scatterer at position  $(m, n)$ . This can, for instance, occur when using an inappropriate method for the SIR calculation, errors in the measurement of the transducer impulse response, or because the true scatterer positions are not found exactly on the grid points.

### B. Statistical Assumptions

The image is modeled using a Bernoulli-Gaussian distribution. A priori, a scatterer is independent of the other

scatterers assumed to belong to the active set with a probability  $\lambda$ . This number roughly describes how large a fraction of the grid positions that can be expected to contain an active scatterer.

We model the noise as zero mean white Gaussian,  $\mathbf{e} \sim N(0, \sigma_e^2 \mathbf{I}_{KL})$ , where  $\mathbf{I}_{KL}$  denotes a  $KL \times KL$  identity matrix. Furthermore, we model the amplitudes of the active scatterers to be independent and identically distributed (iid) as  $o_m \sim N(0, \sigma_o^2)$ , yielding the joint distribution  $\mathbf{o}_{\mathcal{I}} \sim N(0, \sigma_o^2 \mathbf{I}_{n_a})$ . If necessary, these assumptions can be relaxed to include colored noise as well as correlations between elements in  $\mathbf{o}_{\mathcal{I}}$ .

### III. SEARCH CRITERION

We define sparse deconvolution as the combination of joint detection [i.e., finding the set  $\mathcal{I}$  that maximizes the probability  $P(\mathcal{I}|\mathbf{y})$ ], which is equivalent to maximizing the product  $p(\mathbf{y}|\mathcal{I})P(\mathcal{I})$ , then, as a final stage, estimating the amplitudes indicated by the so found  $\mathcal{I}$ . The first step is the dominating task and is solved iteratively by successively improving a criterion function that is obtained as follows.

We take as the criterion the logarithm of the above product, with constant terms removed. The logarithm of  $P(\mathcal{I})$  is given by:

$$\ln P(\mathcal{I}) = \ln(\lambda^{n_a} (1-\lambda)^{NM-n_a}) = n_a \ln \frac{\lambda}{1-\lambda} + c, \quad (2)$$

with  $c$  being a constant. To find  $\ln p(\mathbf{y}|\mathcal{I})$ , the marginalization integral is solved yielding:

$$p(\mathbf{y}|\mathcal{I}) = \int p(\mathbf{y}, \mathbf{o}_{\mathcal{I}}|\mathcal{I}) d\mathbf{o}_{\mathcal{I}} = \frac{\exp(-\frac{1}{2}\mathbf{y}^T \mathbf{D}_{\mathcal{I}}^{-1} \mathbf{y})}{(2\pi)^{NM/2} |\mathbf{D}_{\mathcal{I}}|^{1/2}}, \quad (3)$$

where  $\mathbf{D}_{\mathcal{I}}$  is given by:

$$\mathbf{D}_{\mathcal{I}} = \sigma_e^2 \mathbf{I}_{KL} + \sigma_o^2 \mathbf{P}_{\mathcal{I}} \mathbf{P}_{\mathcal{I}}^T. \quad (4)$$

This matrix is of size  $KL \times KL$ , which makes the quadratic form and determinant in (3) computationally awkward. However, both these factors can be simplified to more attractive forms that result in the log marginal likelihood:

$$\ln p(\mathbf{y}|\mathcal{I}) = n_a \ln \beta - \frac{1}{2} \ln |\mathbf{B}_{\mathcal{I}} + \beta^2 \mathbf{I}_{n_a}| - \frac{1}{\sigma_e^2} \|\mathbf{y}\|^2 + \frac{1}{2\sigma_e^2} \mathbf{c}_{\mathcal{I}}^T (\mathbf{B}_{\mathcal{I}} + \beta^2 \mathbf{I}_{n_a})^{-1} \mathbf{c}_{\mathcal{I}}, \quad (5)$$

where  $\beta = \sigma_e / \sigma_o$ , where the  $n_a \times 1$  vector  $\mathbf{c}_{\mathcal{I}}$  is defined as:

$$\mathbf{c}_{\mathcal{I}} = \mathbf{P}_{\mathcal{I}}^T \mathbf{y}, \quad (6)$$

and with the symmetric  $n_a \times n_a$  matrix  $\mathbf{B}_{\mathcal{I}}$  defined as:

$$\mathbf{B}_{\mathcal{I}} = \mathbf{P}_{\mathcal{I}}^T \mathbf{P}_{\mathcal{I}}. \quad (7)$$

Note that the elements in  $\mathbf{B}_{\mathcal{I}}$  are scalar products between the different  $\mathbf{p}_{(m,n)}$ -vectors in the active set. These can be extracted from a matrix (look-up-table)  $\mathbf{B} = \mathbf{P}^T \mathbf{P}$ , where  $\mathbf{P}$  is the  $KL \times MN$  matrix consisting of the  $\mathbf{p}$ -vectors for all points in the ROI. Note also that  $\mathbf{B}$  does not depend on the received data, and it can be computed and stored off-line.

Similarly, the vector  $\mathbf{c}_{\mathcal{I}}$  for a certain  $\mathcal{I}$  can be extracted from a  $MN \times 1$  vector  $\mathbf{c} = \mathbf{P}^T \mathbf{y}$ , that we can compute before we begin the search for the optimal  $\mathcal{I}$ . The vector  $\mathbf{c}$  is the output from a spatiotemporal matched filter and large entries in  $\mathbf{c}$  indicate the likely presence of an active scatterer. This is used in the search strategy described in Section III-B.

By combining (2) and (5) and removing terms that are constant with respect to  $\mathcal{I}$ , we obtain the search criterion:

$$J(\mathcal{I}) = n_a \alpha - \frac{1}{2} \ln |\mathbf{B}_{\mathcal{I}} + \beta^2 \mathbf{I}_{n_a}| + \frac{1}{2\sigma_e^2} \mathbf{c}_{\mathcal{I}}^T (\mathbf{B}_{\mathcal{I}} + \beta^2 \mathbf{I}_{n_a})^{-1} \mathbf{c}_{\mathcal{I}}. \quad (8)$$

with the scalar  $\alpha = \ln(\lambda/(1-\lambda)) + \ln \beta$ .

Note also that, for a given  $\mathcal{I}$ , the linear MMSE estimate of the amplitudes is given by:

$$\hat{\mathbf{o}}_{\mathcal{I}} = (\mathbf{B}_{\mathcal{I}} + \beta^2 \mathbf{I}_{n_a})^{-1} \mathbf{c}_{\mathcal{I}}. \quad (9)$$

The linear (nonsparse) MMSE solution used in [2] is obtained as the special case when  $\mathcal{I}$  consists of all points in the ROI:

$$\hat{\mathbf{o}}_{\text{MMSE}} = (\mathbf{P}^T \mathbf{P} + \beta^2 \mathbf{I}_{MN})^{-1} \mathbf{P}^T \mathbf{y}. \quad (10)$$

## IV. ALGORITHM

### A. Initialization

The initialization consists of calculating the look-up-tables  $\mathbf{B}$  and  $\mathbf{c}$ , and for this we first need to calculate the  $\mathbf{p}$ -vectors. These are defined by the transducer's electrical impulse response, which is measured, and the SIRs associated with the points in the ROI. In this work, we have used the DREAM (Discrete REpresentation Array Modeling) toolbox [12], [13] for calculating the SIRs.

Note that choosing the grid point distance and scanning step to be identical results in that the response from a scatterer  $(m, n)$  at scan positions  $l$  is identical to a response from  $(m+1, n)$  at  $l+1$ . As a result from this shift invariance, the matrix  $\mathbf{P}$  is block Toeplitz containing  $L \times N$  blocks, each of size  $K \times M$ . Only  $M+N$  blocks need to be calculated to fully determine  $\mathbf{P}$ .

We have developed an algorithm that uses this block Toeplitz structure in the computation of  $\mathbf{B}$ . Matrix products of the smaller blocks occur at several instants in  $\mathbf{B}$  and by avoiding repeated calculations of these blocks we reduce the computation time from being of order  $(KM)^2 \times (LN)^2$  to  $(KM)^2 \times \max(L, N)^2$ .

## B. Search

We aim at, as efficiently as possible, finding the set of  $\mathcal{I}$  that maximizes the criterion  $J(\mathcal{I})$ . In the search we generate a sequence of sets that yield monotonically increasing  $J(\mathcal{I})$ , and the search ends when no further increase is achieved. We start at iteration  $i = 0$  with the empty set  $\mathcal{I}^0 = \emptyset$ .

At iteration  $i$ , the search has produced the set  $\mathcal{I}^i$ . The next iteration then consists of the following steps:

- Adding elements (up-dating). The elements to add are chosen as follows:
  - Compute the output from the matched filter applied to a residual image,  $\mathbf{c}_{res} = \mathbf{P}^T \mathbf{y}_{res}$  where  $\mathbf{y}_{res} = \mathbf{y} - \mathbf{P}_{\mathcal{I}^i} \hat{\mathbf{o}}_{\mathcal{I}^i}$ .  $\hat{\mathbf{o}}_{\mathcal{I}^i}$  is the vector of estimated amplitudes for the scatterers indicated in  $\mathcal{I}^i$ . Intuitively,  $\mathbf{y}_{res}$  is the remaining B-scan after having removed the effects of the so far found scatterers. By applying the matched filter, we try to detect scatterers that best help to explain the residual B-scan.
  - Find a set of candidates,  $\mathcal{C}$ , by extracting peaks in the pointwise, squared matched filter output  $|\mathbf{c}_{res}|^2$  and including also their neighboring points in  $\mathcal{C}$ . These neighboring points are found by a user-defined mask centered at each peak.
  - Compute  $J(\mathcal{I}^i \cup (m, n))$  for all points  $(m, n) \in \mathcal{C}$ . Remove all points from  $\mathcal{C}$  that do not yield improvements compared to  $J(\mathcal{I}^i)$ .
  - Iterate until  $\mathcal{C}$  is empty: pick the currently best point  $(m^*, n^*) = \arg \max_{(m, n) \in \mathcal{C}} J(\mathcal{I}^i \cup (m, n))$ ; remove from  $\mathcal{C}$  all elements that are not orthogonal to  $\mathbf{p}_{(m^*, n^*)}$ . Let  $\mathcal{U}$  denote the set of points found in this iteration.
- Removing redundant elements (down-dating). For all  $(m, n) \in (\mathcal{I}^i \cup \mathcal{U})$ , calculate  $J((\mathcal{I}^i \cup \mathcal{U}) \setminus (m, n))$ , where  $\setminus$  denotes set difference. Remove those elements that yield improvements compared to  $J(\mathcal{I}^i \cup \mathcal{U})$ .

Step 4 relies on the following. Suppose we have several candidates that all yield increases in  $J$  if they are included in  $\mathcal{I}^i$ . Unfortunately, there is no guarantee that these elements will yield an increase in  $J$  if they are simultaneously included in  $\mathcal{I}^i$ . However, it can be shown that, if their associated  $\mathbf{p}$ -vectors are all mutually orthogonal, the increase in  $J$  that is obtained by including all candidates can be written as a sum of the individual increases. This sum is guaranteed to be positive because all these terms are positive.

The  $\mathbf{p}$ -vectors describe contributions from different scatterers. Even for well separated points, there will be a slight overlap of these contributions in the B-scans; therefore, the corresponding vectors typically will not be perfectly orthogonal. Because for all practical purposes, it is usually sufficient that the scalar product is small, but not necessarily zero, we say that the vectors  $\mathbf{p}_1$  and  $\mathbf{p}_2$  are orthogonal if  $|\mathbf{p}_1^T \mathbf{p}_2| < tol$ , where  $tol$  is a user-defined tolerance.

## C. Computational Aspects

For a fast search, we need efficient methods to evaluate  $J$ . The majority of the evaluations occur in up-date (step 3) and down-date (step 6) and different computation strategies are suitable in these cases. In the up-date, we base the computations on lower triangular Cholesky factors of  $\mathbf{B}_{\mathcal{I}} + \beta^2 \mathbf{I}_{n_a}$ , satisfying:

$$\mathbf{G}_{\mathcal{I}} \mathbf{G}_{\mathcal{I}}^T = \mathbf{B}_{\mathcal{I}} + \beta^2 \mathbf{I}_{n_a}. \quad (11)$$

This factorization requires approximately  $n_a^3/3$  floating point operations (flops). Having  $\mathbf{G}_{\mathcal{I}}$ , we then can compute  $J$  as:

$$J(\mathcal{I}) = n_a \alpha - \sum_{k=1}^{n_a} \ln(\mathbf{G}_{\mathcal{I}})_{k,k} + \frac{1}{2\sigma_e^2} \|\mathbf{v}_{\mathcal{I}}\|^2, \quad (12)$$

where  $(\mathbf{G}_{\mathcal{I}})_{k,k}$  is the  $k$ th diagonal element of  $\mathbf{G}_{\mathcal{I}}$  and  $\mathbf{v}_{\mathcal{I}}$  is the solution to the equation  $\mathbf{G}_{\mathcal{I}} \mathbf{v}_{\mathcal{I}} = \mathbf{c}_{\mathcal{I}}$ . For moderate-to-large  $n_a$ , the computations in (12) require approximately  $2n_a + n_a^2/2$  flops. Note further that  $\hat{\mathbf{o}}_{\mathcal{I}}$  in (9) can be obtained by solving  $\mathbf{G}_{\mathcal{I}}^T \hat{\mathbf{o}}_{\mathcal{I}} = \mathbf{v}_{\mathcal{I}}$ . Thus, having computed  $\mathbf{v}_{\mathcal{I}}$ , approximately  $n_a^2/2$  flops are required to estimate the amplitudes.

These estimated amplitudes are used in step 2 for evaluating  $\mathbf{c}_{res} = \mathbf{P}^T \mathbf{y}_{res}$ . It can be written as  $\mathbf{c}_{res} = \mathbf{c} - \mathbf{P}^T \mathbf{P}_{\mathcal{I}^i} \hat{\mathbf{o}}_{\mathcal{I}^i}$  where  $\mathbf{c}$  has been precomputed and all elements in the product  $\mathbf{P}^T \mathbf{P}_{\mathcal{I}^i}$  can be directly extracted from the matrix  $\mathbf{B}$ .

In the up-date, in step 3 we evaluate  $J$  for a number of related sets. We have already computed  $\mathbf{G}_{\mathcal{I}}$  and  $\mathbf{v}_{\mathcal{I}}$  in an earlier iteration and we then can find the values for these sets using a recursive up-date of the Cholesky factor that results in the following up-date equations for obtaining  $J(\mathcal{I}^i \cup (m, n))$ :

$$J(\mathcal{I}^i \cup (m, n)) = J(\mathcal{I}^i) + \alpha - \ln q + \left( \mathbf{P}_{(m,n)}^T \mathbf{y} - \mathbf{z}^T \mathbf{v}_{\mathcal{I}} \right)^2 / (2q^2 \sigma_e^2), \quad (13)$$

where  $\mathbf{z}$  is the solution to  $\mathbf{G}_{\mathcal{I}} \mathbf{z} = \mathbf{b}_{(m,n)}$ , with  $\mathbf{b}_{(m,n)}$  being a vector of scalar products between  $\mathbf{p}_{(m,n)}$  and those  $\mathbf{p}$ -vectors corresponding to the elements in  $\mathcal{I}^i$ . All these are available in  $\mathbf{B}$ . The scalar product  $\mathbf{p}_{(m,n)}^T \mathbf{y}$  is found in  $\mathbf{c}_{\mathcal{I}}$  and finally,  $q = \|\mathbf{p}_{(m,n)}\|^2 + \beta^2 - \|\mathbf{z}\|^2$ , where the first term, again, is available in  $\mathbf{B}$ . The computations in (13) involve approximately  $n_a + n_a^2/2$  flops.

In the down-dating, we calculate  $\hat{\mathbf{o}}_{\mathcal{I}}$  as explained earlier and explicitly compute the inverse  $\mathbf{D}_{\mathcal{I}} = (\mathbf{B}_{\mathcal{I}} + \beta^2 \mathbf{I}_{n_a})^{-1}$ , which can be efficiently done using  $\mathbf{G}_{\mathcal{I}}$  available from an earlier up-date step. Let  $d$  denote the element on the diagonal in  $\mathbf{D}_{\mathcal{I}}$  corresponding to point  $(m, n)$ . The following equations then can be used:

$$J(\mathcal{I} \setminus (m, n)) = J(\mathcal{I}) - \alpha - \frac{1}{2} \ln d - \frac{\hat{\delta}(m, n)^2}{2\sigma_e^2 d}, \quad (14)$$

where  $\hat{\delta}(m, n)$  is the estimated amplitude at point  $(m, n)$ .

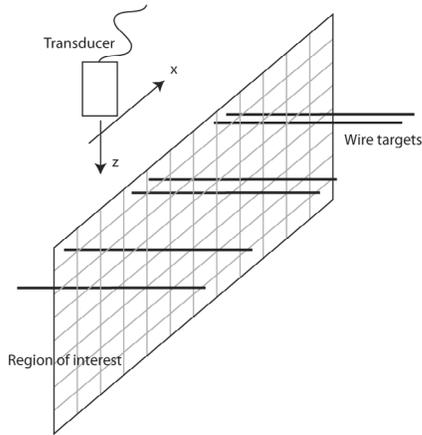


Fig. 2. Measurement setup. Wire targets in immersion.

## V. SIMULATIONS AND EXPERIMENTAL RESULTS

### A. Experimental Setup

The performance of the proposed algorithm was evaluated using simulated and measured data. The measurements were performed in immersion, and 0.3 mm thin steel wires were used to simulate point-like targets. A water sound speed of 1480 m/s was assumed in the experiment. The data were acquired using a circular 2.25 MHz piston transducer with 6 mm radius. The spatial sampling distance was  $\Delta x = 1$  mm, and temporal sampling frequency was 25 MHz. The targets were located at a distance of 120 mm from the transducer and were grouped in three pairs, in which the targets were separated by a distance of 5, 3, and 1 mm, respectively. This separation was in the scanning direction  $x$  as illustrated in Fig. 2.

The considered ROI had the size 5 mm  $\times$  70 mm yielding  $160 \times 70 = 11200$  grid points. The corresponding B-scan had the same size.

The wire targets were chosen to avoid problems with too weak signals when using small point-like reflectors. However, the use of these violates the point target assumption; the response from the wires can be approximated by an integral over the points on a line that corresponds to a temporal smearing effect of the signals compared to smaller targets. Although we can redefine our model to include this effect, we chose in this work the approximate but simpler method of including the smearing directly in the transducer impulse response. This impulse response was measured as the pulse-echo signal obtained from a wire target in the far field, deconvolved with the SIR associated with the point at which the target was located.

### B. Simulation Results

Measurements from the above setup were simulated. The SIRs used in the simulations were obtained using the DREAM toolbox [12], and the transducer impulse response was the same as in the real data experiments. The scattering amplitudes at the wire positions were set to  $3 \times 10^{-8}$ ,

which yielded a peak-to-peak amplitude of 300 quantization levels.<sup>1</sup> White Gaussian noise of variance  $\sigma_e^2 = 100$  was added to the simulated data to obtain an SNR of approximately 10 dB.

The results from the simulations are presented in Figs. 3(a) to (d). Fig. 3(a) shows the simulated RF data. Fig. 3(b) shows the sparse solution obtained with parameters  $\sigma_e^2 = 100$ ,  $\sigma_o^2 = 10^{-15}$ , and  $\lambda = 0.01$ . In Fig. 3(c), a corresponding linear MMSE solution [2] obtained using (10) is shown. To obtain comparable results in the sense that the models predict approximately the same energy in the images, the prior variance of amplitudes was set to,  $\sigma_o^2 \lambda = 10^{-17}$  in the linear MMSE solution. The residual for the sparse solution,  $\mathbf{y}_{res} = \mathbf{y} - \mathbf{P}_{\mathcal{I}} \hat{\mathbf{o}}_{\mathcal{I}}$ , is displayed in Fig. 3(d).

We see in Fig. 3 that the sparse algorithm can distinguish the point target pairs that are separated by 5 mm and 3 mm. However, the pair of targets that are separated with only 1 mm are estimated as a single, but stronger scatterer. As a comparison, the linear MMSE can well resolve the pair separated with 5 mm, and a separation of the next pair of targets is only faintly indicated. No separation of the 1 mm targets can be seen.

Except for the last error, the detection of the positions in the sparse solution is perfect, and the amplitude estimation errors are less than 10% of the true values.

The residual of the sparse solution shows mostly white noise. Most of the energy in the original data has been accounted for, which could be expected as no model errors are present; the simulated data are generated from the same model used for reconstruction.

The simulation results serve as an example of the performance that can be achieved using sparse deconvolution in an ideal situation with no model errors present. We note the excellent noise resilience and detection capabilities that can be achieved under such circumstances.

### C. Experimental Results

The results from processing of the measured data are displayed in Figs. 4(a) to (c). The wires can be seen as the white spots, corresponding to high positive estimated scattering amplitudes. Weaker, dark indications<sup>2</sup> near the main echoes are most probably caused by model errors that, for instance, may stem from slightly inaccurate SIRs or targets that are not residing exactly at the grid positions.

As can be seen in Fig. 4, the wires were not positioned on a perfectly horizontal line. The leftmost wire in the first pair, seen at positions 14–19 mm, was slightly closer to the transducer than the others.

<sup>1</sup>The choice of scatterer amplitudes are arbitrary as long as they, in combination with the SIRs and the transducer impulse response, yield signals of realistic magnitudes. We may choose to up-scale the amplitudes and down-scale the impulse response the same amount without any effect on the detection results.

<sup>2</sup>The dark spots have negative amplitudes. Their magnitudes are approximately 1/5 of the main (bright) echoes.

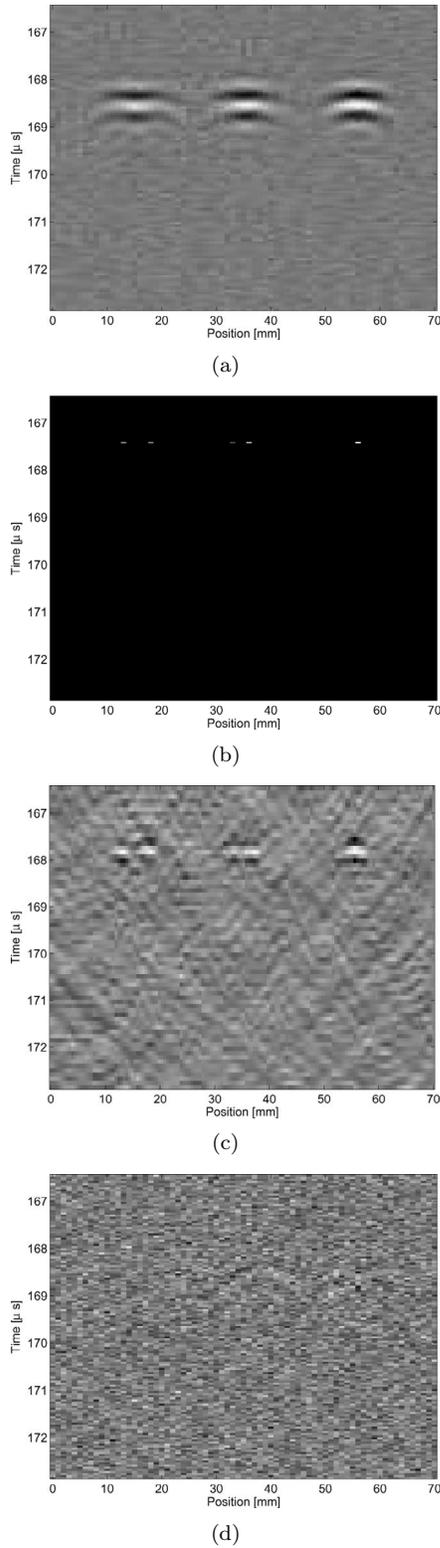


Fig. 3. Results from simulations. (a) Original simulated data. (b) Sparse solution. (c) Linear MMSE solution. (d) Residual for the sparse solution.

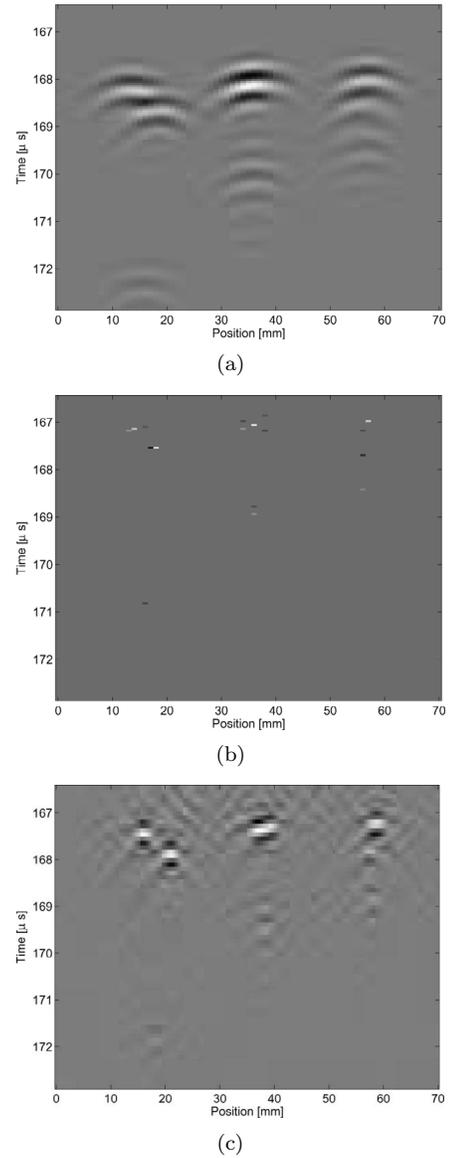


Fig. 4. Results from experimental data. (a) Original measured data. (b) Scatter amplitude estimates. (c) MMSE result.

Similar to the simulations, the wires that were separated with 5 mm and 3 mm can be distinguished by the sparse algorithm, but the two wires that were 1 mm apart are estimated as a single, stronger scatterer. The corresponding linear MMSE solution now can clearly separate only the wires in the first pair.

In the measured data, we can observe additional contributions that can be explained by the sound paths; transducer-wire-wire-transducer. These contribution appear approximately  $3.5 \mu\text{s}$ ,  $1.8 \mu\text{s}$ , and  $0.8 \mu\text{s}$  after for main echoes, for the wires separated by 5 mm, 3 mm, and 1 mm, respectively. This agrees fairly well with the values  $3.4 \mu\text{s}$ ,  $2 \mu\text{s}$ , and  $0.7 \mu\text{s}$  predicted by a sound speed of  $c_p = 1480 \text{ m/s}$ .

To illustrate the updating process, Figs. 5(a) to (c) show the residual B-scan after iteration one, three, and six when processing the measured data. At each iteration, the esti-

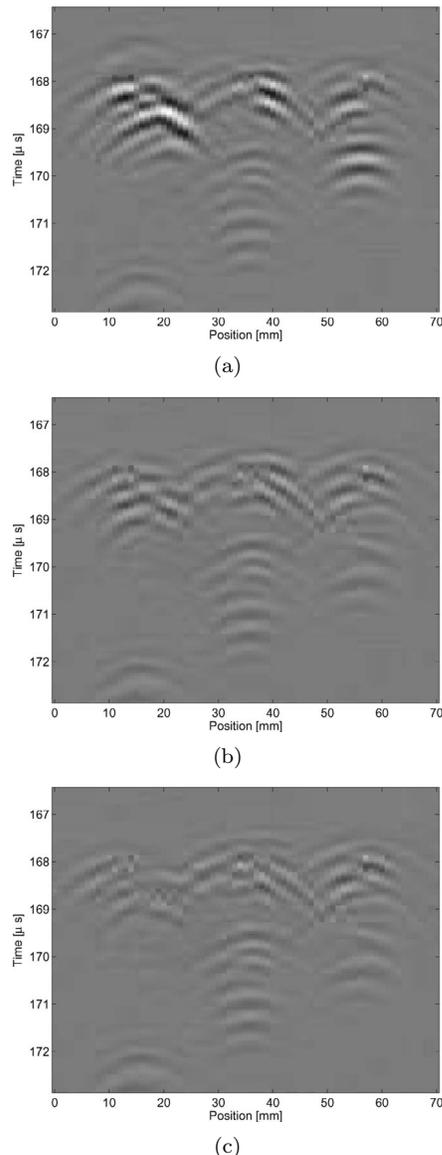


Fig. 5. Residual B-scans after a few of the iterations. The algorithm converged in six iterations. (a) First iteration. (b) Third iteration. (c) Sixth and last iteration.

mate is refined and the residuals are reduced in magnitude. We can note that the residuals for the measured data contain patterns that are more regular than the residuals for the simulated data. A probable cause of this is model errors. For a nonperfect model, the strongest contributions to the residual are expected close to strong scatterers.

#### D. Computation Time

To illustrate the computation times involved when using the algorithm, we here give the times required in the processing of the real data. A study showing how the computation time scales with the number of true scatterers, and the image size can be found in [11].

The calculations were performed on a Pentium III (Intel Inc., Santa Clara, CA) with clock frequency of 2 GHz and 4 GB of RAM memory, enough to hold all the data re-

TABLE I  
COMPUTATION TIME FOR DIFFERENT PARTS OF THE SEARCH  
ALGORITHM.

Step	Time [s]	Time [%]
2	0.16	4.6
3	0.2	5.7
4	2.24	63.8
5	0.7	19.9
6	0.21	6.0

quired for each set of calculations. No disc access or virtual memory operations were needed. Calculations were implemented in MATLAB (The MathWorks Inc., Natick, MA) m-script. The initialization of the matrix  $\mathbf{B}$  took approximately 10 minutes. Note that this is a one-time calculation. When scanning a volume, several B-scans are collected under similar conditions, and they all share the same matrix  $\mathbf{B}$ . The computation of  $\mathbf{c}$ , see Section III-A, required 3.3 s.

The search required in total 3.5 s, which is approximately the same as the time for calculating  $\mathbf{c}$ . The time to perform each of the steps of the search algorithm is detailed in Table I. The five different steps in the method is explained in detail in Section III-B.

It is clear that the updating part of the search dominates the search time. The single, most time-consuming operation is calculating the criterion  $J(I \cup (m, n))$  for all potential candidates in step 4. Calls to this operation consume a total of 1.94 seconds, or 55% of the total search time. In this case, approximately 7500 candidate points were evaluated in step 4 over the six iterations in total.

## VI. CONCLUSIONS

We have presented a new 2-D sparse deconvolution algorithm for deconvolution of ultrasonic B-scans. The simulation results show that the method under ideal conditions combines excellent resolution and noise suppression properties. The reason is that the available information is optimally used; we know that only a few significant scatterers are present. Therefore, we search for only a few significant scatterers. This also gives the advantage of a reduced computation time compared to the linear MMSE solution, despite the fact that an iterative search is performed. The computations in each iteration are far less demanding than those found in linear MMSE. For truly sparse images, the iterative search converges in only a few iterations.

One disadvantage with this method compared to the simpler SAFT methods is the small size of the images that can be processed. In the present form, this method should be seen mainly as a complement to SAFT-based system. Small ROIs can be chosen manually, or possibly automatically using some simple method based on SAFT results, so that only small fractions of the data need to be processed.

Note, however, that the overall time for processing the B-scan used in this paper is approximately 7 s; this is less than the approximately 10 s that was used to collect the

data. Thus it can for images of this size be implemented for real-time processing.

An evaluation of the algorithm under less ideal premises than those given here should be performed in the future. Because we are approaching the limits of how much the linear model can be used, we need to take more robust methods under consideration. The challenge is to develop methods in which model uncertainties are taken into account and that, at the same time, can be implemented using computationally efficient methods.

## REFERENCES

- [1] J. Shen and E. Ebbini, "A new coded-excitation ultrasound imaging system—Part I: Basic principles," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 43, pp. 131–140, Jan. 1996.
- [2] F. Lingvall, T. Olofsson, and T. Stepinski, "Synthetic aperture imaging using sources with finite aperture: Deconvolution of the spatial impulse response," *J. Acoust. Soc. Amer.*, vol. 114, pp. 225–234, July 2003.
- [3] H. Desoky, M. Abou-Bakr, and Y. Kadah, "Reconstruction using optimal spatially variant kernel for B-mode ultrasonic imaging," in *Proc. SPIE*, 2003, pp. 147–153.
- [4] J. Kormylo and J. Mendel, "Maximum likelihood detection and estimation of Bernoulli-Gaussian processes," *IEEE Trans. Inform. Theory*, vol. 28, pp. 482–488, May 1982.
- [5] Y. Goussard, G. Demoment, and J. Idier, "A new algorithm for iterative deconvolution of sparse spike trains," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Processing*, 1990, pp. 1547–1550.
- [6] T. Olofsson and T. Stepinski, "Maximum a posteriori deconvolution of sparse ultrasonic signals using genetic optimization," *Ultrasonics*, vol. 37, pp. 423–432, Sep. 1999.
- [7] K. Kaarensen and E. Bølviken, "Blind deconvolution of ultrasonic traces accounting for pulse variance," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 46, pp. 564–573, May 1999.
- [8] K. Kaarensen, "Deconvolution of sparse spike trains by iterated window maximization," *IEEE Trans. Signal Processing*, vol. 45, pp. 1173–1183, May 1997.
- [9] C. Chi, J. Goutsias, and J. Mendel, "A fast maximum-likelihood estimation and detection algorithm for Bernoulli-Gaussian processes," in *Proc. Int. Conf. Acoust. Speech Signal Processing*, 1985, pp. 1297–1300.
- [10] K. Kaarensen, "Evaluation and applications of the iterated window maximization method for sparse deconvolution," *IEEE Trans. Signal Processing*, vol. 46, pp. 609–624, Mar. 1998.
- [11] T. Olofsson, "Computationally efficient sparse deconvolution of B-scan images," in *Proc. IEEE Ultrason. Symp.*, 2005, pp. 540–543.
- [12] F. Lingvall, B. Piwakowski, and G. Zhang, The DREAM (Discrete REpresentation Array Modeling) toolbox, freeware toolbox available at <http://www.signal.uu.se/Toolbox/dream/>.
- [13] B. Piwakowski and K. Sbai, "A new approach to calculate the field radiated from arbitrarily structured transducer arrays," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 46, pp. 422–440, Mar. 1999.



**Tomas Olofsson** was born 1968 in Sandviken, Sweden. He received his M.Sc. degree in engineering physics in 1994 and the Ph.D. degree in signal processing in 2000, both from Uppsala University. He is currently working as an associate professor at the Signals and Systems group, Uppsala University, Sweden. His research concerns inference problems, in particular, inverse problems in ultrasonics.



**Erik Wennerström** was born in 1977 in Sweden. He received an M.Sc. degree in engineering from Uppsala University, Uppsala, Sweden. He has been a Ph.D. student at the Signals and Systems group at Uppsala University since 2003.