Bayesian Compressive Sensing Approaches for Direction of Arrival Estimation with Mutual Coupling Effects

Matthew Hawes[a], Lyudmila Mihaylova[a] and François Septier[b], Simon Godsill[c]

[a] Department of Automatic Control and Systems Engineering, University of Sheffield, S1 3JD, UK
[b] IMT Lille Douai, Univ. Lille, CNRS, UMR 9189 - CRIStAL, F-59000 Lille, France
[c] Engineering Department, University of Cambridge, UK
{m.hawes, l.s.mihaylova}@sheffield.ac.uk, francois.septier@telecom-lille.fr, sjg@eng.cam.ac.uk

Abstract—The problem of estimating the dynamic direction of arrival of far field signals impinging on a uniform linear array, with mutual coupling effects, is addressed. This work proposes two novel approaches able to provide accurate solutions, including at the endfire regions of the array. Firstly, a Bayesian compressive sensing Kalman filter is developed, which accounts for the predicted estimated signals rather than using the traditional sparse prior. The posterior probability density function of the received source signals and the expression for the related marginal likelihood function are derived theoretically. Next, a Gibbs sampling based approach with indicator variables in the sparsity prior is developed. This allows sparsity to be explicitly enforced in different ways, including when an angle is too far from the previous estimate. The proposed approaches are validated and evaluated over different test scenarios and compared to the traditional relevance vector machine based method. An improved accuracy in terms of average root mean square error values is achieved (up to 73.39% for the modified relevance vector machine based approach and 86.36% for the Gibbs sampling based approach). The proposed approaches prove to be particularly useful for direction of arrival estimation when the angle of arrival moves into the endfire region of the array.

Index Terms—Dynamic DOA estimation, Bayesian compressive sensing, Kalman filter, Gibbs sampling, Relevance vector machine

I. INTRODUCTION

Direction of arrival (DOA) estimation is the process of determining which direction a signal impinging on an array has arrived from. Commonly used methods of solving this problem are: MUSIC [1], [2], ESPRIT [3]–[6] and the maximum likelihood DOA estimator [7]–[9]. However, these methods have some disadvantages, in particular they require knowledge of the number of signals present beforehand and evaluation of a covariance matrix of the array output (adding computational complexity).

Compressive Sensing (CS) theory says that when certain conditions are met it is possible to recover signals from fewer measurements than used by traditional methods [10], [11]. Hence, CS can be applied to the problem of DOA estimation [12]–[15] by splitting the angular region into N potential DOAs, where only $L << N$ of the DOAs have an impinging signal (alternatively $N - L$ of the angular directions have a

zero-valued impinging signal present). These DOAs are then estimated by finding the minimum number of DOAs with a non-zero valued impinging signal that still give an acceptable estimate of the array output.

The problem can also be converted into a probabilistic form and solved via Bayesian compressive sensing (BCS) [16], implemented with a relevance vector machine (RVM) [17]–[19]. Such a method has been used to solve the problem of static DOA estimation [20], [21], where a belief of having a sparse received signal is made and the most likely values found.

The Kalman filter (KF) can be used to track dynamic DOAs, with the angular range narrowed to focus in more closely on the DOA estimate from the previous iteration [22]. However, this prevents directly working with the measured array signals and introduces an additional stage of having to reevaluate the steering vector of the array at each iteration of the KF. Hierarchical KFs have been used to track dynamic sparse signals [23], [24], where the predicted mean of the signals at each iteration is taken as the estimate from the previous iteration and the hyperparameters are estimated using BCS, hence the term Bayesian compressive sensing Kalman Filter (BCSKF).

However, a problem remains when a BCSKF is applied to dynamic DOA estimation with a uniform linear array (ULA). The estimation accuracy can be reduced when the DOA approaches the endfire region of the array, i.e. when the impinging signal arrives parallel to or almost parallel to the array. This can be particularly problematic when there is a lot of noise present.

An additional challenge to address when considering the DOA estimation problem is that of mutual coupling. One way of modeling the mutual coupling effects is to use a mutual coupling matrix [25], [26]. In [25] the mutual coupling matrix is found using two methods: minimum mean-square matching and the mutual impedance method. The method in [26] applies a symmetric Toeplitz matrix, where only antennas within a set separation of each other can cause mutual coupling effects. In this work the method in [26] is used to ensure mutual coupling effects are included in the signal model.
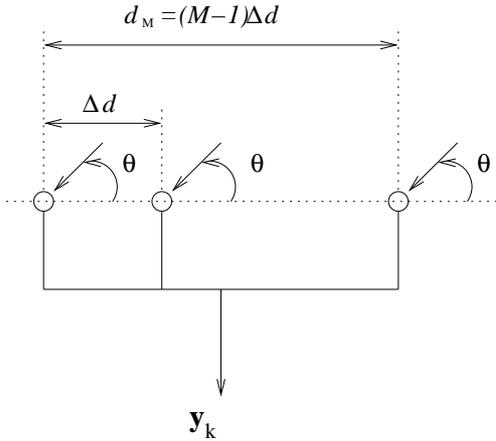
Fig. 1. Linear array structure being considered, consisting of M antennas with a uniform adjacent antenna separation of Δd.

The contributions of this paper are: i) A BCSKF with a modified RVM, where the traditional sparsity prior is replaced with a belief that the estimated signals will instead match predicted signal values, is proposed. The result of this new prior is that a new posterior distribution and marginal likelihood have been derived. Initial results for this method using a signal model without mutual coupling have been reported in [27]. ii) A Gibbs sampling approach is proposed. In this approach zero valued signals can be explicitly enforced when there is too large a change in DOA in order to alleviate the estimation accuracy problem for the endfire region of the array. iii) A comprehensive performance evaluation is provided, with the proposed methods being compared to a BCSKF using the traditional RVM approach. Significant improvements in terms of the average root mean square error ($RMSE$) values are observed (up to 73.39% for the BCSKF with modified RVM and up to 86.36% for the Gibbs sampling approach).

The remainder of this paper is structured in the following manner: Section II gives details of the proposed estimation methods, including the array model with mutual coupling effects (II-A), the modified RVM framework for BCS (II-B), the BCSKF (II-C) and the Gibbs sampling implementation (II-D). In Section III an evaluation of the effectiveness of the proposed approaches is presented and conclusions are drawn in Section IV.

II. PROPOSED ESTIMATION METHODS

A. Array Model

A narrowband ULA structure consisting of M omnidirectional antennas, with identical responses is shown in Figure 1. Here, a plane-wave signal mode is assumed, i.e. the signal impinges upon the array from the far field and the angle of arrival is limited to $0° \leq \theta \leq 180°$. The distance from the first antenna to subsequent antennas is denoted as d_m for $m = 1, 2, \ldots, M$, with $d_1 = 0$, i.e. the distance from the first antenna to itself. Note, these values are multiples of a uniform adjacent antenna separation of Δd.

The steering vector of the array is given by

$$\mathbf{a}(\Omega, \theta) = [1, e^{-j\mu_2\Omega\cos\theta}, \ldots, e^{-j\mu_M\Omega\cos\theta}]^T, \quad (1)$$

where $\Omega = \omega T_s$ is the normalised frequency with T_s being the sampling period, $\mu_m = \frac{d_m}{cT_s}$ for $m = 1, 2, \ldots, M$, c gives the wave propagation speed and $\{\cdot\}^T$ denotes the transpose operation.

The array output, $\mathbf{y}_k$, at time snapshot k is then given by

$$\mathbf{y}_k = \mathbf{A}_{st}\mathbf{x}_k + \mathbf{n}_k, \quad (2)$$

where $\mathbf{x}_k = [x_{k,1}, x_{k,2}, \ldots, x_{k,N}]^T \in \mathbb{C}^{N\times 1}$ gives the received source signals, $\mathbf{n}_k = [n_{k,1}, n_{k,2}, \ldots, n_{k,M}]^T \in \mathbb{C}^{M\times 1}$ is a noise term, given by a zero mean multivariate Gaussian random variable and $\mathbf{A}_{st} = [\mathbf{a}(\Omega, \theta_1), \mathbf{a}(\Omega, \theta_2), \ldots, \mathbf{a}(\Omega, \theta_N)] \in \mathbb{C}^{M\times N}$ is the matrix containing the steering vectors for each angle of interest. Note, N is the number of points in the grid of potential DOAs the angular region has been split into. However, only $L << N$ of these angular directions will have an impinging signal present.

In practice there will also be mutual coupling effects present, which alter the pattern of an individual antenna as compared to if it was being used on its' own. As a result (2) has to be altered to account for this fact. A mutual coupling matrix is used to achieve this [26], by giving the true steering vector matrix as

$$\mathbf{A} = \mathbf{M}_{MC}\mathbf{A}_{st}. \quad (3)$$

Here $\mathbf{M}_{MC} \in \mathbb{C}^{M\times M}$ is the mutual coupling matrix given by

$$\mathbf{M}_{MC} = \begin{pmatrix} 1 & m_2 & \ldots & m_{D-1} & \ldots & m_M \\ m_2 & 1 & m_2 & \ldots & \ddots & \vdots \\ \vdots & m_2 & 1 & m_2 & \ldots & m_{D-1} \\ m_{D-1} & \ldots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ldots & m_2 & 1 & m_2 \\ m_M & \ldots & m_{D-1} & \ldots & m_2 & 1 \end{pmatrix}. \quad (4)$$

In (4) the mutual coupling coefficients are given by $m_i = \rho_i \exp\{j\phi_i\}$ for $i = 2, \ldots, D-1, D, \ldots, M$, where ρ_i and ϕ_i give the amplitude and phase, respectively. The variable D places a limit on the separation between antennas above which there will be no mutual coupling effects. In other words when $i > D$, then $\rho_i = 0$. This then gives the following:

$$\begin{aligned} \mathbf{y}_k &= \mathbf{M}_{MC}\mathbf{A}_{st}\mathbf{x}_k + \mathbf{n}_k \\ &= \mathbf{A}\mathbf{x}_k + \mathbf{n}_k. \end{aligned} \quad (5)$$

Equation (5) can then be split into real and imaginary components (given by $\mathcal{R}(\cdot)$ and $\mathcal{I}(\cdot)$, respectively) as follows

$$\tilde{\mathbf{y}}_k = \tilde{\mathbf{A}}\tilde{\mathbf{x}}_k + \tilde{\mathbf{n}}_k$$
$$\begin{bmatrix} \mathcal{R}(\mathbf{y}_k) \\ \mathcal{I}(\mathbf{y}_k) \end{bmatrix} = \begin{bmatrix} \mathcal{R}(\mathbf{A}) & -\mathcal{I}(\mathbf{A}) \\ \mathcal{I}(\mathbf{A}) & \mathcal{R}(\mathbf{A}) \end{bmatrix} \begin{bmatrix} \mathcal{R}(\mathbf{x}_k) \\ \mathcal{I}(\mathbf{x}_k) \end{bmatrix} + \begin{bmatrix} \mathcal{R}(\mathbf{n}_k) \\ \mathcal{I}(\mathbf{n}_k) \end{bmatrix}. \quad (6)$$

The difference between $\mathbf{y}_k$ and $\tilde{\mathbf{y}}_k$ is that $\mathbf{y}_k$ has been split into its real and imaginary components in $\tilde{\mathbf{y}}_k$. As a result the

dimensions of $\tilde{\mathbf{y}}_k$ are increased. A similar relationship exists between $\mathbf{A}$ and $\tilde{\mathbf{A}}$, $\mathbf{x}_k$ and $\tilde{\mathbf{x}}_k$ and $\mathbf{n}_k$ and $\tilde{\mathbf{n}}_k$.

B. Modified Relevance Vector machine for DOA Estimation

The aim is to now find a solution for $\tilde{\mathbf{x}}_k$ which gives the closest possible match to a predicted set of signal values. To achieve this one can follow a modified RVM framework [27], by evaluating the following

$$\tilde{\mathbf{x}}_{k,opt} = \max \mathcal{P}(\tilde{\mathbf{x}}_k, \sigma_k^2, \mathbf{p}_k | \tilde{\mathbf{y}}_k, \tilde{\mathbf{x}}_p), \qquad (7)$$

where σ_k^2 is the variance of the Gaussian noise $\mathbf{n}_k$, $\mathbf{p}_k = [p_{k,1}, p_{k,2}, ..., p_{k,2N}]^T$ contains the hyperparameters that are to be estimated and $\tilde{\mathbf{x}}_p = [\mathcal{R}(\mathbf{x}_p)^T, \mathcal{I}(\mathbf{x}_p)^T]^T = [\mathcal{R}(x_{p,1}), ..., \mathcal{R}(x_{p,N}), \mathcal{I}(x_{p,1}), ..., \mathcal{I}(x_{p,N})]^T$ holds the predicted values of $\tilde{\mathbf{x}}_k$.

From (6) it is possible to find:

$$\mathcal{P}(\tilde{\mathbf{y}}_k | \tilde{\mathbf{x}}_k, \sigma_k^2) = (2\pi\sigma_k^2)^{-M} \exp\left\{ -\frac{1}{2\sigma^2} ||\tilde{\mathbf{y}}_k - \tilde{\mathbf{A}}\tilde{\mathbf{x}}_k||_2^2 \right\}. \quad (8)$$

The traditional RVM would now apply a belief that $\tilde{\mathbf{x}}_k$ is sparse. However, here this is changed to a belief that $\tilde{\mathbf{x}}_k$ will match the predicted signals $\tilde{\mathbf{x}}_p$:

$$\begin{aligned} \mathcal{P}(\tilde{\mathbf{x}}_k | \mathbf{p}_k, \tilde{\mathbf{x}}_p) &= (2\pi)^{-N} |\mathbf{P}_k|^{\frac{1}{2}} \\ &\times \exp\left\{ -\frac{1}{2}(\tilde{\mathbf{x}}_k - \tilde{\mathbf{x}}_p)^T \mathbf{P}_k (\tilde{\mathbf{x}}_k - \tilde{\mathbf{x}}_p) \right\}. \end{aligned} (9)$$

Note, when $\tilde{\mathbf{x}}_p = [0, 0, ..., 0]^T$ then (9) reverts to the hierarchical prior used in the traditional RVM [16], [17] and $|\mathbf{P}_k|$ indicates the determinant of $\mathbf{P}_k$, where $\mathbf{P}_k = \text{diag}(\mathbf{p}_k)$.

It is also necessary to define the hyperparameters over $\mathbf{p}_k$ and σ_k^2. There are various possibilities for the structuring of the priors on $\mathbf{p}_k$, which represent mixing parameters in a scale mixture of normals representation of the marginal distribution of $\mathbf{x}_k$, which will here be in the Student-t family, see e.g. [28]. One possibility would be to treat the complex components of $\mathbf{x}_k$ as complex Student-t distributed, as detailed in [29], [30]. However, this work treats the real and imaginary components of $\mathbf{x}_k$ as independent Student-t distributed random variables, and hence there are independent Gamma priors for the mixing variables $p_{k,n}$ over all real and imaginary components of $\mathbf{x}_k$:

$$\mathcal{P}(\mathbf{p}_k) = \prod_{n=1}^{2N} G(p_{k,n} | \beta_1, \beta_2). \qquad (10)$$

A Gamma prior can also be used for σ_k^2

$$\mathcal{P}(\sigma_k^2) = G(\sigma_k^{-2} | \beta_3, \beta_4), \qquad (11)$$

where $\beta_1, \beta_2, \beta_3$ and β_4 are scale and shape priors.

It is known that

$$\mathcal{P}(\tilde{\mathbf{x}}_k, \sigma_k^2, \mathbf{p}_k | \tilde{\mathbf{y}}_k, \tilde{\mathbf{x}}_p) = \mathcal{P}(\tilde{\mathbf{x}}_k | \tilde{\mathbf{y}}_k, \sigma_k^2, \mathbf{p}_k, \tilde{\mathbf{x}}_p) \mathcal{P}(\sigma_k^2, \mathbf{p}_k | \tilde{\mathbf{y}}_k, \tilde{\mathbf{x}}_p) \qquad (12)$$

and

$$\begin{aligned} \mathcal{P}(\tilde{\mathbf{x}}_k | \tilde{\mathbf{y}}_k, \sigma_k^2, \mathbf{p}_k, \tilde{\mathbf{x}}_p) &= \frac{\mathcal{P}(\tilde{\mathbf{y}}_k | \tilde{\mathbf{x}}_k, \sigma_k^2) \mathcal{P}(\tilde{\mathbf{x}}_k | \mathbf{p}_k, \tilde{\mathbf{x}}_p)}{\mathcal{P}(\tilde{\mathbf{y}}_k | \sigma_k^2, \mathbf{p}_k, \tilde{\mathbf{x}}_p)} \\ &= (2\pi)^{-N} |\mathbf{\Sigma}|^{-1/2} \exp\left\{ -\frac{1}{2}(\tilde{\mathbf{x}}_k - \boldsymbol{\mu})^T \mathbf{\Sigma}^{-1}(\tilde{\mathbf{x}}_k - \boldsymbol{\mu}) \right\}, (13) \end{aligned}$$

where the covariance matrix and the mean are given by

$$\mathbf{\Sigma} = (\sigma_k^{-2}\tilde{\mathbf{A}}^T\tilde{\mathbf{A}} + \mathbf{P}_k)^{-1} \qquad (14)$$

and

$$\boldsymbol{\mu} = \mathbf{\Sigma}(\sigma_k^{-2}\tilde{\mathbf{A}}^T\tilde{\mathbf{y}}_k + \mathbf{P}_k\tilde{\mathbf{x}}_p), \qquad (15)$$

respectively. Note, the maximum of (13) is the posterior mean $\boldsymbol{\mu}$. For a derivation of (13) please see Appendix A.

Similarly to [17], the probability $\mathcal{P}(\sigma_k^2, \mathbf{p}_k | \tilde{\mathbf{y}}_k, \tilde{\mathbf{x}}_p)$ can be represented in the following form:

$$\mathcal{P}(\sigma_k^2, \mathbf{p}_k | \tilde{\mathbf{y}}_k, \tilde{\mathbf{x}}_p) \approx \mathcal{P}(\tilde{\mathbf{y}}_k | \sigma^2, \mathbf{p}, \tilde{\mathbf{x}}_p) \mathcal{P}(\sigma_k^2) \mathcal{P}(\mathbf{p}_k) \mathcal{P}(\tilde{\mathbf{x}}_p), \quad (16)$$

where $\mathcal{P}(\tilde{\mathbf{x}}_p)$ is constant as fixed values are used and the second two terms on the right of are constant if $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 1 \times 10^{-4}$ as in [17]. Therefore, maximising $\mathcal{P}(\sigma_k^2, \mathbf{p}_k | \tilde{\mathbf{y}}_k, \tilde{\mathbf{x}}_p)$ is roughly equivalent to maximising $\mathcal{P}(\tilde{\mathbf{y}}_k | \sigma_k^2, \mathbf{p}_k, \tilde{\mathbf{x}}_p)$. This can be achieved by a type 2 maximisation of its logarithm, which is given by (please see Appendix B):

$$\begin{aligned} \mathcal{L}(\sigma_k^2, \mathbf{p}_k) &= \log\left\{ (2\pi\sigma_k^2)^{-M} |\mathbf{\Sigma}|^{\frac{1}{2}} |\mathbf{P}_k|^{\frac{1}{2}} \exp\left(-\frac{1}{2} \right. \right. \\ &\qquad \times \left. \left. (\tilde{\mathbf{y}}_k^T \mathbf{B} \tilde{\mathbf{y}}_k + \tilde{\mathbf{x}}_p^T \mathbf{C} \tilde{\mathbf{x}}_p - 2\sigma_k^2 \tilde{\mathbf{y}}_k^T \tilde{\mathbf{A}} \mathbf{\Sigma} \mathbf{P}_k \tilde{\mathbf{x}}_p) \right) \right\} \\ &= -\frac{1}{2}\Big(2M\log(2\pi) + 2M\log\sigma_k^2 - \log|\mathbf{\Sigma}| - \\ &\qquad \log|\mathbf{P}_k| + \sigma_k^{-2}||\tilde{\mathbf{y}}_k - \tilde{\mathbf{A}}\boldsymbol{\mu}||_2^2 + \boldsymbol{\mu}^T \mathbf{P}_k \boldsymbol{\mu} \\ &\qquad + \tilde{\mathbf{x}}_p^T \mathbf{P}_k \tilde{\mathbf{x}}_p - \tilde{\mathbf{x}}_p^T \mathbf{P}_k \boldsymbol{\mu} \Big), \qquad (17) \end{aligned}$$

where $\mathbf{B} = (\sigma_k^2\mathbf{I} + \tilde{\mathbf{A}}\mathbf{P}_k^{-1}\tilde{\mathbf{A}}^T)^{-1}$ and $\mathbf{C} = \mathbf{P}_k - \mathbf{P}_k^T \mathbf{\Sigma} \mathbf{P}_k$.

This is now differentiated with respect to $p_{k,n}$ and σ_k^{-2} to obtain the update expressions

$$p_{k,n}^{new} = \frac{\gamma_n}{\mu_n^2 + \tilde{x}_{p,n}^2 - \tilde{x}_{p,n}\mu_n}, \qquad (18)$$

where $\gamma_n = 1 - p_{k,n}\Sigma_{nn}$, Σ_{nn} is the nth diagonal element of $\mathbf{\Sigma}$ and

$$\sigma_{k,new}^2 = \frac{||\tilde{\mathbf{y}}_k - \tilde{\mathbf{A}}\boldsymbol{\mu}||_2^2}{2M - \sum_n \gamma_n}. \qquad (19)$$

For the derivation of (18) and (19) please see Appendix C.

The maximisation is then achieved by iteratively finding $\mathbf{\Sigma}$ and $\boldsymbol{\mu}$, followed by $p_{k,n}^{new}$ for $n = 1, ..., N$ and $\sigma_{k,new}^2$ until a convergence criterion is met [16], [17]. In other words, the new estimates for the noise variance and precision hyperparameters found from (19) and (18) are then used in (14) and (15) to find new estimates of the covariance matrix and mean of the distribution in (13). Note that when $\tilde{\mathbf{x}}_p = [0, 0, ..., 0]^T$ the update expressions match those used by the traditional RVM.

The final estimate of the received signals is then given by

$$\tilde{\mathbf{x}}_{k,opt} = \left(\frac{\tilde{\mathbf{A}}^T\tilde{\mathbf{A}}}{\sigma_{k,opt}^2} + \mathbf{P}_{k,opt} \right)^{-1} \left(\frac{\tilde{\mathbf{A}}^T\tilde{\mathbf{y}}_k}{\sigma_{k,opt}^2} + \mathbf{P}_{k,opt}\tilde{\mathbf{x}}_p \right) \quad (20)$$

where $\sigma_{k,opt}^2$ and $\mathbf{P}_{k,opt} = \text{diag}(p_{k,opt,1}, p_{k,opt,2}, ..., p_{k,opt,2N})$ are the result of optimising the noise estimate and hyperparameters, respectively. Now $\tilde{\mathbf{x}}_{k,opt}$ can be used to reconstruct

the estimated signals as

$$x_{k,opt,n} = \tilde{x}_{k,opt,n} + j\tilde{x}_{k,opt,N+n}, \qquad (21)$$

where $n = 1, 2, ..., N$.

The thresholding scheme in [20] can then be applied to keep the $\tilde{L}$ most significant signals. To do this find the total energy content of the estimated received signals and then sort them. A threshold value, η, is then defined as a percentage of the energy content that is to be retained. Starting with the most significant estimated signal, the estimated signals are summed until the threshold is reached and the remaining signals are then set to be equal to 0. The remaining non-zero valued signals then give the DOA estimates and $\tilde{L}$ is an estimate of the number of far field signals impinging on the array.

C. Bayesian Compressive Sensing Kalman Filter

In order to track the changes in the DOA estimates at each time snapshot the modified RVM based DOA estimation procedure detailed above is combined with a Bayesian KF, giving a BCSKF for DOA estimation [27]. The signal model described above is again used along with the prediction

$$\tilde{\mathbf{x}}_{k|k-1} = \tilde{\mathbf{x}}_{k-1|k-1} + \boldsymbol{\Delta}\mathbf{x} \quad \boldsymbol{\Sigma}_{k|k-1} = \boldsymbol{\Sigma}_{k-1} + \mathbf{P}_k^{-1}$$
$$\tilde{\mathbf{y}}_{k|k-1} = \tilde{\mathbf{A}}\tilde{\mathbf{x}}_{k|k-1} \qquad\qquad \tilde{\mathbf{y}}_{e,k} = \tilde{\mathbf{y}}_k - \tilde{\mathbf{y}}_{k|k-1} \quad (22)$$

and update steps

$$\tilde{\mathbf{x}}_k = \tilde{\mathbf{x}}_{k|k-1} + \mathbf{K}_k\tilde{\mathbf{y}}_{e,k} \quad \boldsymbol{\Sigma}_{k|k} = (\mathbf{I} - \mathbf{K}_k\tilde{\mathbf{A}})\boldsymbol{\Sigma}_{k|k-1}$$
$$\mathbf{K}_k = \boldsymbol{\Sigma}_{k|k-1}\tilde{\mathbf{A}}^T(\sigma_k^2\mathbf{I} + \tilde{\mathbf{A}}\boldsymbol{\Sigma}_{k|k-1}\tilde{\mathbf{A}}^T)^{-1} \quad (23)$$

of the BCSKF. Here, $k|k-1$ indicates prediction at time instance k given the previous measurements and $\boldsymbol{\Delta}\mathbf{x}$ is determined by the assumed DOA change. Note, $\boldsymbol{\Delta}\mathbf{x}$ is fixed by the predetermined constant motion rather than being a random noise term. For example, if the angular range is sampled every $1°$ and the DOA is assumed to increase by $2°$ then $\boldsymbol{\Delta}\mathbf{x}$ will be selected to increase the index of the non-zero valued entries in $\tilde{\mathbf{x}}_{k-1|k-1}$ by two to give the index of the non-zero valued entries in $\tilde{\mathbf{x}}_{k|k-1}$.

At each time snapshot it is necessary to estimate the noise variance and hyperparameters in order to evaluate the prediction and update steps of the BCSKF. This is done by considering the log likelihood function given by

$$\mathcal{L}(\sigma_k^2, \mathbf{p}_k) = -\frac{1}{2}\Big(2M\log(2\pi) + 2M\log\sigma_k^2 - \log|\boldsymbol{\Sigma}|$$
$$- \log|\mathbf{P}_k| + \sigma_k^{-2}||\tilde{\mathbf{y}}_{e,k} - \tilde{\mathbf{A}}\boldsymbol{\mu}||_2^2 + \boldsymbol{\mu}^T\mathbf{P}_k\boldsymbol{\mu}$$
$$+ \tilde{\mathbf{x}}_{k|k-1}^T\mathbf{P}_k\tilde{\mathbf{x}}_{k|k-1} - \tilde{\mathbf{x}}_{k|k-1}^T\mathbf{P}\boldsymbol{\mu}\Big), \qquad (24)$$

which can be optimised by following the procedure described in Section II-B. In other words we apply the modified RVM framework to $\tilde{\mathbf{y}}_{e,k}$, using the KF prediction $\tilde{\mathbf{x}}_{k|k-1}$ as the expected estimate values $\tilde{\mathbf{x}}_p$.

It is worth noting that the continued accuracy of the proposed BCSKF relies on the accuracy of the initial estimate and the parameter values selected. If the initial estimate (made using the framework described in Section II-B and $\tilde{\mathbf{x}}_p = [0, 0, ..., 0]^T$) of the received signals is accurate and

sparse, then the priors that are enforced will ensure this continues to be the case. However, an inaccurate initial DOA estimate or poorly matched expected DOA change can lead to the introduction of inaccuracies in subsequent estimates. Similarly, if the initial estimate of the received signals is not sparse then subsequent estimates are likely to not be sparse. As a result, care should be taken when choosing the initial parameter values and determining the likely DOA change.

D. Gibbs Sampling for DOA Estimation

The method described in the previous sections based on a BCSKF with a modified RVM required the use of prior knowledge of the predicted change in DOA. However, in practice this may not always be known, making it important to have an alternative method that can still give improved accuracy for the endfire region.

This work proposes using a sparsity prior which is given as a combination of a point mass concentrated at zero (Dirac delta function) and a zero mean Gaussian distribution, [31]–[33], giving

$$\mathcal{P}(\tilde{\mathbf{x}}_k|\mathbf{p}_k, \tilde{\mathbf{z}}_k) = \prod_{n=1}^{2N}(1 - \tilde{z}_{k,n})\delta_0 + \tilde{z}_{k,n}\mathcal{N}(\tilde{x}_{k,n}|0, p_{k,n}), \quad (25)$$

where $\tilde{\mathbf{z}}_k = [\mathbf{z}_k^T, \mathbf{z}_k^T]^T$ and $\mathbf{z}_k = [z_{k,1}, z_{k,2}, ..., z_{k,N}]^T$.

Note, $\tilde{z}_{k,n}$ is the indicator variable for $\tilde{x}_{k,n}$ and determines which of the two components in (25) is selected. When $\tilde{z}_{k,n} = 0$, the value of $\tilde{x}_{k,n}$ is determined solely by the point mass concentrated at zero. As a result, $\tilde{x}_{k,n} = 0$ and sparsity is explicitly introduced. Alternatively, when $\tilde{z}_{k,n} = 1$ the value of $\tilde{x}_{k,n}$ is determined by the Gaussian distribution allowing a non-zero valued estimate. The repetition of $\mathbf{z}_k$ in $\tilde{\mathbf{z}}_k$ means that the same indicator variable is used for both the real and imaginary parts of each entry in $\mathbf{x}_k$.

This indicator value can also be used to address the endfire accuracy problem by selecting the value of $z_{k,n} = 0$ if $|n-i| > j$. Here i is the index of the closet non-zero valued estimate from the previous time snapshot and j defines a maximum allowed change in the DOA estimate. Only $n = 1, 2, ..., N$ is considered to get the entries for $\mathbf{z}_k$, with $\tilde{\mathbf{z}}_k$ then being found as previously stated.

This leaves the following

$$z_{k,n} = \begin{cases} z_{k,n}^1 & \text{if } |n-i| \le j, \\ z_{k,n}^2 & \text{if } |n-i| > j, \end{cases} \qquad (26)$$

where $z_{k,n}^1$ and z_n^2 are defined by the following Beta distributions

$$\begin{aligned} z_{k,n}^1 &= B(z_{k,n}^1|\beta_5^1, \beta_6^1), \\ z_{k,n}^2 &= B(z_{k,n}^2|\beta_5^2, \beta_6^2). \end{aligned} \qquad (27)$$

In order to enforce zero-valued estimates when $|n-i| > j$, it is necessary to select β_5^2 and β_6^2 to ensure a zero-valued $z_{k,n}$ is preferred. However, when $|n-i| \le j$ it is necessary to choose β_5^1 and β_6^1 so that the chances of $z_{k,n} = 0$ and $z_{k,n} = 1$ are equal. This can be achieved by

$$\begin{aligned} \mathcal{P}(z_{k,n}^1|\beta_5, \beta_6) &= B(z_{k,n}^1|\beta_5, \beta_6), \\ \mathcal{P}(z_{k,n}^2|\beta_5, \beta_6) &= B\left(z_{k,n}^2|\beta_5 - \frac{1}{j}, \beta_6 + \frac{1}{j}\right), \quad (28) \end{aligned}$$

where $\beta_5 = \beta_6 = 1$.

The posterior distribution of $\tilde{\mathbf{x}}_k$ can be written as [33]

$$\mathcal{P}(\tilde{\mathbf{x}}_k|\tilde{\mathbf{y}}_k, \sigma_k^2, \mathbf{p}_k, \tilde{\mathbf{z}}_k) \propto \left\{ \prod_{n=1}^{2N} [(1 - \tilde{z}_{k,n})\delta_0 + \tilde{z}_{k,n} \right.$$
$$\left. \times \mathcal{N}(\tilde{x}_{k,n}|0, p_{k,n})] \right\} \mathcal{N}(\tilde{\mathbf{y}}_k|\tilde{\mathbf{A}}\tilde{\mathbf{x}}_k, \sigma_k^{-2}). \quad (29)$$

Now also define $\tilde{\mathbf{A}}_n$ as being the entries in $\tilde{\mathbf{A}}$ relating to the index n and $\tilde{\mathbf{A}}_{-n}$ are the entries of $\tilde{\mathbf{A}}$ excluding the entries relating to index n (and similarly for $\tilde{\mathbf{x}}_k$). Then as per Appendix D this gives

$$\mathcal{P}(\tilde{x}_{k,n}|\tilde{\mathbf{y}}_k, \tilde{\mathbf{x}}_{k,-n}, \sigma_k^2, \mathbf{p}_k, \tilde{\mathbf{z}}_k) = (1 - \hat{z}_{k,n})\delta_0 + \hat{z}_{k,n}$$
$$\times \mathcal{N}(\tilde{x}_{k,n}|\hat{\mu}_{k,n}, \hat{p}_{k,n}) \quad (30)$$
$$\hat{p}_{k,n} = p_{k,n} + p_0 \tilde{\mathbf{A}}_n^T \tilde{\mathbf{A}}_n, \quad (31)$$
$$\hat{\mu}_{k,n} = \hat{p}_{k,n}^{-1} p_0 \tilde{\mathbf{A}}_n^T \tilde{\mathbf{y}}_{k,n}, \quad (32)$$
$$\frac{\hat{z}_{k,n}}{1 - \hat{z}_{k,n}} = \frac{\tilde{z}_{k,n}}{1 - \tilde{z}_{k,n}}$$
$$\times \frac{\mathcal{N}(0|0, p_{k,n})}{\mathcal{N}(0|\hat{\mu}_{k,n}, \hat{p}_{k,n})}, \quad (33)$$

where $p_0 = 1/\sigma_k^2$ and $\tilde{\mathbf{y}}_{k,n} = \tilde{\mathbf{y}}_k - \tilde{\mathbf{A}}_{-n}\tilde{\mathbf{x}}_{k,-n}$.

There are two further posterior distributions that have to be considered. That is the distributions for $p_{k,n}$ and p_0 which are given by

$$\mathcal{P}(p_{k,n}|\mathbf{x}_{k-1}) = G(\beta_1 + ||\mathbf{x}_{k-1,nj}||_0, \beta_2 + ||\mathbf{x}_{k-1,nj}||_2^2) \quad (34)$$

and

$$\mathcal{P}(p_0|\tilde{\mathbf{y}}_k, \tilde{\mathbf{x}}_k) = G(\beta_3 + M, \beta_4 + \frac{1}{2}||\tilde{\mathbf{y}}_k - \tilde{\mathbf{A}}\tilde{\mathbf{x}}_k||_2^2), \quad (35)$$

respectively. Note, in (34) $\mathbf{x}_{k-1,nj}$ gives the entries within $\mathbf{x}_{k-1}$ that have an index within the distance j of index n. By using $\mathbf{x}$ rather than $\tilde{\mathbf{x}}$ to find $\mathbf{x}_{k-1,nj}$ it guarantees the same value of $||\mathbf{x}_{k-1,nj}||_2^2$ and $||\mathbf{x}_{k-1,nj}||_0$ for both the real and imaginary components.

As a result the Gibbs sampling steps are as detailed below:
1) Sample $\tilde{x}_{k,n}$ from $\mathcal{P}(\tilde{x}_{k,n}|\tilde{\mathbf{y}}_k, \tilde{\mathbf{x}}_{k,-n}, \sigma_k^2, \mathbf{p}_k, \tilde{\mathbf{z}}_k)$.
2) Sample $p_{k,n}$ from $\mathcal{P}(p_{k,n}|\mathbf{x}_{k-1})$.
3) **if** $n \leq N$ then Sample $z_{k,n}^1$ from $\mathcal{P}(z_{k,n}^1|\beta_5, \beta_6)$, **else** $z_{k,n}^1 = z_{k,n-N}^1$.
4) **if** $n \leq N$ then Sample $z_{k,n}^2$ from $\mathcal{P}(z_{k,n}^2|\beta_5, \beta_6)$, **else** $z_{k,n}^2 = z_{k,n-N}^2$.
5) Sample p_0 from $\mathcal{P}(p_0|\tilde{\mathbf{y}}_k, \tilde{\mathbf{x}}_k)$.

These steps are done for each of the T iterations of the Gibbs sampler, where the first T_{BI} iterations are the burn-in iterations. The final estimate of the received array signals is then given by the mean values of the final $T - T_{BI}$ iterations [32]. The DOA estimate can then be found using the previously described thresholding scheme (see II-B), with the remaining non-zero valued estimates corresponding to the DOA estimates.

Note, the performance of this method will again heavily depend on the accuracy of the first estimate. As a result, it is possible to use the traditional BCS DOA estimation method (Section II-B with $\tilde{\mathbf{x}}_p = [0, 0, ..., 0]^T$) to ensure an as accurate as possible initial estimate at the first time snapshot. The proposed Gibbs sampling based method can then be used at the subsequent time snapshots to get the next DOA estimate.

III. PERFORMANCE EVALUATION

In this section a comparison in performance of the proposed methods and the traditional RVM based BCSKF method will be made over five example scenarios, under the same test conditions. Firstly, an example is considered where the initial DOA starts outside of the endfire region and then moves into it. Secondly, an example is given where the DOA remains out of the endfire region. In the third scenario the initial DOAs and the signal values are randomly generated. Then the evaluation will also consider the scenario where there is a mismatch between the actual and assumed change in DOA. Finally, the evaluation will consider a random change in DOA at each time snapshot. This means that $\Delta \mathbf{x}$ which is selected for the modified RVM based BCSKF will not be a true reflection of how the DOA actually changes for the last two examples.

Note, the term traditional RVM based BCSKF method means the entries of $\mathbf{P}_k$ in the prediction step of the BCSKF are found using the RVM optimisation method as detailed in [16], [17]. In other words this is the method detailed in Section II-B with $\tilde{\mathbf{x}}_p = \mathbf{0}$. All of the examples are implemented in Matlab on a computer with an Intel Xeon CPU E3-1271 (3.60GHz) and 16GB of RAM.

The performance of each method will be measured using the $RMSE$ in the DOA estimate. This is given by

$$RMSE = \sqrt{\frac{\sum_{q=1}^{Q} \sum_{l=1}^{\tilde{L}} |\theta_l - \hat{\theta}_l|^2}{Q\tilde{L}}}, \quad (36)$$

where θ_l is the actual DOA, $\hat{\theta}_l$ is the estimated DOA and Q is the number of Monte Carlo simulations carried out, with $Q = 100$ being used in each case. This gives a measure of the estimation accuracy and the computation time will be used as a measure of the complexity of each method.

For the Gibbs sampling method a burn-in period of 250 iterations is used and then 50 further iterations used to find the final estimate of the received array signals. When a distance of $j = 5°$ is exceeded a zero-valued estimate of the received signals is enforced in order to alleviate the endfire accuracy problem.

For all the design examples considered the selection of $\sigma_k^2 = 0.4$ as the noise variance is used, with an initial estimate of the noise variance given by $\sigma_{k,0}^2 = 0.1$. The array geometry being used is that of a ULA with $M = 20$ antennas and an adjacent antenna separation of $\frac{\lambda}{2}$, where λ is the wavelength of the signal of interest. This gives an array aperture of 9.5λ. For the mutual coupling matrix a value of $D = 3$ is selected, meaning that a separation of 1.5λ or greater gives negligible mutual coupling effects. The values $\rho_1 = 0.65$, $\rho_2 = 0.25$, $\phi_1 = \pi/7$ and $\phi_2 = \pi/10$ are then also used. Finally, in each example a single narrowband signal impinging on the array is considered, meaning $L = 1$.

TABLE I
PERFORMANCE SUMMARY FOR THE ENDFIRE REGION EXAMPLE.

Method	Average $RMSE$ (degrees)	Average Computation Time (seconds)
RVM	19.88	0.76
Modified RVM	6.46	0.98
Gibbs	3.82	17.83 excluding burn-in
		107.84 including burn-in

Fig. 2. $RMSE$ values for the endfire region example.

TABLE II
PERFORMANCE SUMMARY FOR THE ENDFIRE REGION EXAMPLE WITH REDUCED ADJACENT ANTENNA SEPARATIONS.

Method	Average $RMSE$ (degrees)	Average Computation Time (seconds)
RVM	2.21	0.98
Modified RVM	0.86	1.25
Gibbs	3.02	16.81 excluding burn-in
		101.65 including burn-in

Fig. 3. $RMSE$ values for the endfire region example with reduced antenna separation.

A. Endfire Region

For this example the initial DOA of the signal is $\theta = 20°$, which then decreases by $1°$ at each time snapshot. The signal value at each snapshot is set to be 1. Table I summarises the performance of the three methods for this example, with the $RMSE$ values at each time snapshot being shown in Figure 2.

Here it can be seen that there has been a significant decrease in the average $RMSE$ values for both the modified RVM method (67.51% improvement) and the Gibbs sampling based method (80.78% improvement). Overall this suggests that a more accurate estimate of the DOA is possible. It is worth noting that there has still been an increase in the RMSE for the modified RVM based approach in the endfire region of the angular range. However, this has come much later on than for the traditional RVM based approach (indicating a degradation in performance for a smaller angular range) and the maximum $RMSE$ value reached is lower (indicating the degradation is less severe).

These improvements have come at the cost of an increased computation time in both instances. For the modified RVM method this increase is insignificant as the average computation time is still less than one second. The increase for the Gibbs sampling based method is larger, illustrating an increase in computational complexity. However, it is worth remembering that this increase has resulted in a more accurate DOA estimate being achieved without prior knowledge about what the change in DOA will be.

In this instance the results suggest that one of the two proposed methods should be used when the estimated DOA approaches the endfire region of the array. If the change in DOA is known in advance and computational complexity is a primary concern then the modified RVM based method is the most suitable (a more accurate estimate can be achieved without a large increase in computation time). However, when this information is not available, or computational complexity is not a concern, it is possible to get a significant improvement in accuracy (at the cost of computation time) using the Gibbs sampling based method.

In the previous simulation an adjacent antenna separation of $\lambda/2$ is used as it is known that this is the largest separation that can be used while still avoiding a degraded performance due to the introduction of grating globes [34]. However, an example of what the relative performance of the methods is when a smaller adjacent antenna separation will now be considered.

In this instance an adjacent antenna separation of $\lambda/4$ is selected. As the array aperture is kept constant (to allow a fair comparison between adjacent antenna separation sizes) this means the number of antennas is given by $M = 39$. This also means a value of $D = 9$ is required to keep the same distance limits on mutual coupling occurring. The values of ρ_i and ϕ_i are then selected to be uniformly spread over the range of 0.65 to 0.25 and $\pi/7$ to $\pi/10$, respectively. The remaining parameters are kept constant and the same test scenario as for the previous example is used.

The performances of each of the methods in this instance are summarised in Table II and Figure 3, respectively. Here it can be seen that the larger number of antennas used has resulted in a lower average $RMSE$ values for all three of the methods. In this instance only the modified RVM method has performed better that the traditional RVM based method when comparing average $RMSE$ values (decrease in average $RMSE$ of 61.09%). However, by looking at the maximum $RMSE$ values it can be seen that the largest estimation error possible with the traditional RVM based method is larger than that for the Gibbs sampling based method ($16.42°$ as compared

TABLE III
PERFORMANCE SUMMARY FOR THE NON-ENDFIRE REGION EXAMPLE.

Method	Average $RMSE$ (degrees)	Average Computation Time (seconds)
RVM	2.91	0.73
Modified RVM	1.85	0.78
Gibbs	1.46	13.08 excluding burn-in
		79.57 including burn-in

TABLE IV
PERFORMANCE SUMMARY FOR THE RANDOM INITIAL DOA EXAMPLE.

Method	Average $RMSE$ (degrees)	Average Computation Time (seconds)
RVM	12.10	0.74
Modified RVM	3.22	0.86
Gibbs	2.89	13.88 excluding burn-in
		84.19 including burn-in

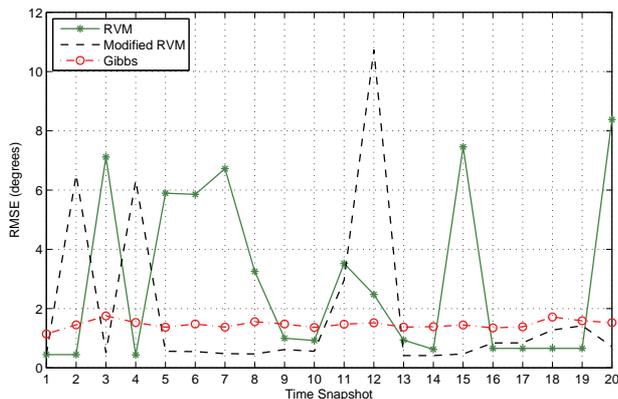

Fig. 4. $RMSE$ values for the non-endfire region example.
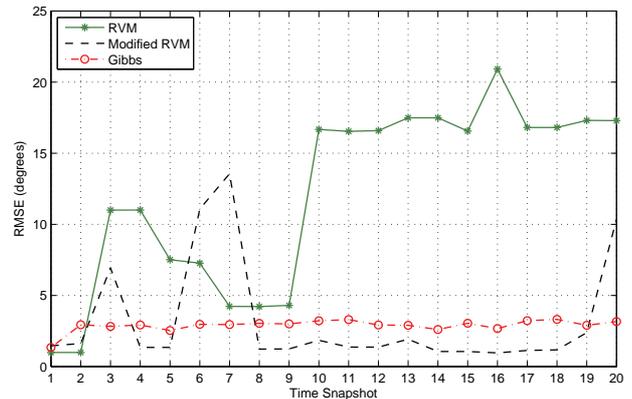
Fig. 5. $RMSE$ values for the random initial DOA example.

to $3.71°$).

It is worth noting that such an array configuration is unlikely to be used in practice. This is due to the costs associated with the number of antennas required. As a result, the remaining examples will stick to the adjacent antenna separation of $\lambda/2$ and associated parameters previously defined.

B. Non-Endfire Region

For this example the initial DOA is $\theta = 100°$ with the DOA increasing by $1°$ at each time snapshot, with the signal value remaining constant at -1. The performance of the three methods is summarised in Table III, with the $RMSE$ values illustrated in Figure 4.

In this instance it can be seen that there has not been as large an increase in $RMSE$ for the traditional RVM method, as the DOA does not enter the endfire region. However, both the modified RVM and Gibbs sampling based methods have managed to achieve improvements in average $RMSE$ values of 36.42% and 50.00%, respectively. For the Gibbs sampling based method this comes at the expenses of an increase in computation time, whereas the time for the modified RVM based method is comparable to the traditional RVM based method. As with the previous test scenario this would suggest that the modified RVM based method should be used when the expected DOA change information is available and the Gibbs sampling based method when this is not the case, or when computational complexity is not a major concern.

C. Random Initial DOA

Next consider the case where the initial DOA is randomly chosen from the entire angular range and increased by $1°$ at each time snapshot. The signal value is randomly selected

as ± 1 for each simulation and remains constant as the DOA changes.

Table IV and Figure 5 summarise the performance of the various methods in this instance. Again, it can be seen that the modified RVM based method has offered improvements in terms of $RMSE(73.39\%)$, without a significant increase in computation time. The Gibbs sampling based method has also given an estimation accuracy improvement (76.12%) and has even outperformed the modified RVM based method, without prior knowledge of how the DOA was going to change. However, this has come at the expense of an increased computation time.

D. Mismatched Actual and Assumed DOA Change

This subsection compares the performances of the estimation methods for two situations where the actual change in DOA is not known. First, consider the case where there is an initial DOA of $\theta = 100°$ which increases by $1°$ for 9 time snapshots before decreasing by $1°$ for the remaining time snapshots. In this instance, assume a constant signal value of 1 throughout.

The performance comparison is now made between the traditional RVM based method, the modified RVM based method with the assumed DOA change set to a constant increase of $1°$, the modified RVM based method with the assumed DOA change set to a constant decrease of $1°$ and the Gibbs sampling based method. The performances for each are summarised in Figure 6 and Table V, respectively.

In this instance the average $RMSE$ values suggest a comparable performance in terms of estimation accuracy between the traditional RVM based methods and the two modified RVM based examples. This can be explained by the fact that for both of the modified RVM based examples, the assumed
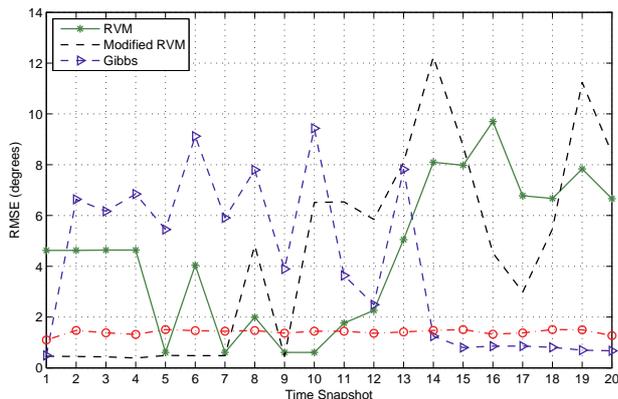
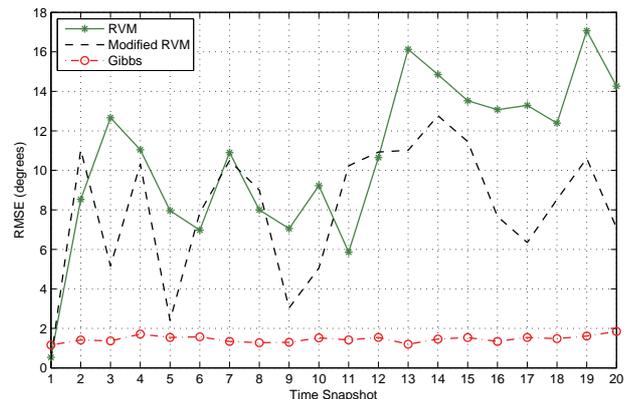Fig. 6. $RMSE$ values for the increasing DOA followed by decreasing DOA example.

Fig. 7. $RMSE$ values for the increasing DOA with an assumed decrease in DOA of $3°$ example.

TABLE V
PERFORMANCE SUMMARY FOR THE INCREASING DOA FOLLOWED BY DECREASING DOA EXAMPLE.

Method	Average $RMSE$ (degrees)	Average Computation Time (seconds)
RVM	4.49	0.69
Modified RVM with constant $+1°$ DOA change	4.45	0.77
Modified RVM with constant $-1°$ DOA change	4.08	0.79
Gibbs	1.41	16.12 excluding burn-in
		97.59 including burn-in

TABLE VI
PERFORMANCE SUMMARY FOR THE INCREASING DOA WITH AN ASSUMED DECREASE IN DOA OF $3°$ EXAMPLE.

Method	Average $RMSE$ (degrees)	Average Computation Time (seconds)
RVM	10.70	0.66
Modified RVM	8.07	0.77
Gibbs	1.46	13.99 excluding burn-in
		84.72 including burn-in

DOA change does not match the actual DOA changes for the entire time range which means the same improvements as for the previous scenario can no longer be guaranteed. Figure 6 highlights this in the results for the two modified RVM examples. It demonstrates that with an assumed increasing DOA the modified RVM offers some initial improvements, while the performance is significantly degraded when the DOA starts to decrease again. On the other hand, the example with an assumed decreasing DOA performs worse than the traditional RVM based method initially and then offers significant improvements when the actual DOA also starts to decrease.

It can also be seen that for the Gibbs sampling based method there has been an improvement in DOA estimation accuracy. In terms of average $RMSE$ values this is a decrease of 68.60%, which has been achieved without any knowledge of how the DOA was going to change. However, there is again an increase in the computation time.

To illustrate how a larger mismatch between actual and assumed DOA changes effects the performance of the modified RVM based method now consider an example where the actual DOA is increasing by $1°$ in each snapshot, while the assumed change is a decrease of $3°$. Here, the initial DOA is $100°$, with a constant signal value of -1. The $RMSE$ values for the methods are shown in Figure 7 and summarised in Table VI along with the computation times.

Here it can be seen that the Gibbs sampling based method has offered an 86.36% improvement in average $RMSE$ as compared to the traditional RVM based method. There has again been a significant increase in the computational com-

plexity.

For the modified RVM based method the average $RMSE$ values suggests that there has been an improvement in estimation accuracy. However, this is smaller than when the actual and assumed DOA changes match. It is also unlikely that this improvement would be obtained in every scenario the the method could be applied to. From looking at Figure 7 we can see that the traditional and modified RVM based methods are showing comparable performance for the just over half of the time frame considered. This is the relative performance that would be expected in the majority of cases.

E. Random Changes in Direction of Arrival

Finally, consider the example where the initial signal value is assumed to be equal to 1 and the initial DOA is chosen to be $100°$. The actual DOA is then allowed to randomly change by up to $\pm3°$ for each time snapshot. For the modified RVM method assume that the actual DOA change is an increase of $3°$. This gives the results as summarised in Table VII and Figure 8.

It can be seen that the Gibbs sampling based method has again outperformed the modified RVM based method in terms of estimation accuracy, due to the fact that no prior knowledge of how the DOA will change is required. As compared to the traditional RVM based method there has been an improvement in $RMSE$ of 78.81%. However, as is expected this is at the cost of computation time.

It is also worth noting that the average $RMSE$ values suggest that the modified RVM and traditional RVM have offered a comparable performance. This is due to the fact that the assumption of how the DOA will change is not valid,

TABLE VII
PERFORMANCE SUMMARY FOR THE RANDOM CHANGES IN DOA WITH AN ASSUMED INCREASE IN DOA OF $3°$ EXAMPLE.

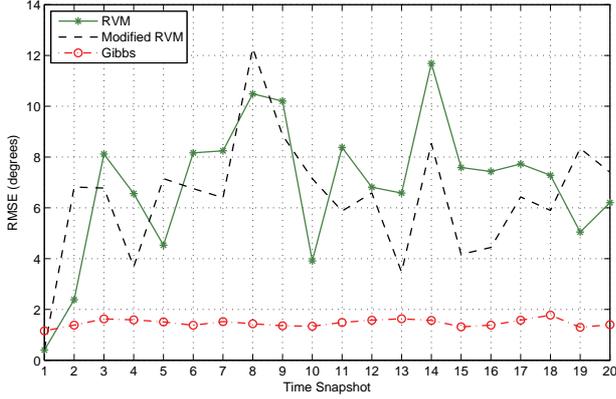| Method	Average $RMSE$ (degrees)	Average Computation Time (seconds)
RVM	6.89	0.64
Modified RVM	6.37	0.86
Gibbs	1.46	22.26 excluding burn-in
		134.56 including burn-in

Fig. 8. $RMSE$ values for the random DOA change example.

meaning the modified RVM no longer offers any improvements. Therefore, in this situation the Gibbs sampling based method would be the best to use, assuming computational complexity is not the main motivating factor.

IV. CONCLUSIONS

This paper has proposed two novel approaches for the estimation of a dynamic direction of arrival using uniform linear arrays with mutual coupling. The first approach is a Bayesian compressed Kalman filter with a modified relevance vector machine, where the traditional sparsity assumption is replaced by an assumption that the estimated signals will instead match predicted signal values. This results in the derivation of a new posterior probability density function of the received signals and the expression for the related marginal likelihood function. The second proposed approach is a Gibbs sampling approach, where sparsity is explicitly enforced if there is a large difference between the previous direction of arrival estimate and the angle currently being considered. The proposed approaches will be particularly useful when applied to the problem of dynamic direction of arrival estimation in the endfire region of antenna arrays. Such problems can arise in numerous application areas such as in communications and surveillance.

An extensive performance evaluation is provided and shows that both of the proposed approaches outperform the traditional relevance vector machine based Bayesian compressive sensing Kalman filter in terms of mean root mean square error values, by up to 73.39% for the modified relevance vector machine based method and 86.36% for the Gibbs sampling based method.

APPENDIX

A. Derivation of Posterior Distribution

Bayes' rule gives

$$\mathcal{P}(\tilde{\mathbf{x}}_k|\tilde{\mathbf{y}}_k, \sigma_k^2, \mathbf{p}_k, \tilde{\mathbf{x}}_p)\mathcal{P}(\tilde{\mathbf{y}}_k|\sigma_k^2, \mathbf{p}_k, \tilde{\mathbf{x}}_p) = \mathcal{P}(\tilde{\mathbf{y}}_k|\tilde{\mathbf{x}}_k, \sigma_k^2)\mathcal{P}(\tilde{\mathbf{x}}_k|\mathbf{p}_k, \tilde{\mathbf{x}}_p), \quad (37)$$

where $\mathcal{P}(\tilde{\mathbf{y}}_k|\tilde{\mathbf{x}}_k, \sigma_k^2)$ and $\mathcal{P}(\tilde{\mathbf{x}}_k|\mathbf{p}_k, \tilde{\mathbf{x}}_p)$ are known from (8) and (9), respectively.

Now following the method suggested in [17] carry out the multiplication on the right hand side of (37), collect terms in $\tilde{\mathbf{x}}_k$ in the exponential and complete the square.

$$-\frac{1}{2}\Big[\sigma_k^{-2}(\tilde{\mathbf{y}}_k - \tilde{\mathbf{A}}\tilde{\mathbf{x}}_k)^T(\tilde{\mathbf{y}}_k - \tilde{\mathbf{A}}\tilde{\mathbf{x}}_k) + (\tilde{\mathbf{x}}_k - \tilde{\mathbf{x}}_p)^T\mathbf{P}_k(\tilde{\mathbf{x}}_k - \tilde{\mathbf{x}}_p)\Big]$$

$$= -\frac{1}{2}\Big[\sigma^{-2}\tilde{\mathbf{y}}_k^T\tilde{\mathbf{y}}_k - \sigma_k^{-2}\tilde{\mathbf{y}}_k^T\tilde{\mathbf{A}}\tilde{\mathbf{x}}_k - \sigma^{-2}\tilde{\mathbf{x}}_k^T\tilde{\mathbf{A}}^T\tilde{\mathbf{y}}_k + \sigma_k^{-2}\tilde{\mathbf{x}}_k^T\tilde{\mathbf{A}}^T\tilde{\mathbf{A}}\tilde{\mathbf{x}}_k + \tilde{\mathbf{x}}_k^T\mathbf{P}_k\tilde{\mathbf{x}}_k - \tilde{\mathbf{x}}_k^T\mathbf{P}_k\tilde{\mathbf{x}}_p - \tilde{\mathbf{x}}_p^T\mathbf{P}_k\tilde{\mathbf{x}}_k + \tilde{\mathbf{x}}_p^T\mathbf{P}_k\tilde{\mathbf{x}}_p\Big]$$

$$= -\frac{1}{2}\Big[(\tilde{\mathbf{x}}_k - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\tilde{\mathbf{x}}_k - \boldsymbol{\mu}) - \boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \sigma^{-2}\tilde{\mathbf{y}}_k^T\tilde{\mathbf{y}}_k + \tilde{\mathbf{x}}_p^T\mathbf{P}_k\tilde{\mathbf{x}}_p\Big] \quad (38)$$

where $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$ are given by (14) and (15), respectively. This then gives the posterior distribution as (13), with the remaining exponential terms

$$-\frac{1}{2}\Big[\sigma_k^{-2}\tilde{\mathbf{y}}_k^T\tilde{\mathbf{y}}_k + \tilde{\mathbf{x}}_p^T\mathbf{P}_k\tilde{\mathbf{x}}_p - \boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\Big]. \quad (39)$$

B. Derivation of Marginal Likelihood

From (37) the following is known:

$$\mathcal{P}(\tilde{\mathbf{y}}_k|\sigma_k^2, \mathbf{p}_k, \tilde{\mathbf{x}}_p) = \frac{\mathcal{P}(\tilde{\mathbf{y}}_k|\tilde{\mathbf{x}}_k, \sigma_k^2), \mathcal{P}(\tilde{\mathbf{x}}_k|\mathbf{p}_k, \tilde{\mathbf{x}}_p)}{\mathcal{P}(\tilde{\mathbf{x}}_k|\tilde{\mathbf{y}}_k, \sigma_k^2, \mathbf{p}_k, \tilde{\mathbf{x}}_p)}, \quad (40)$$

meaning the term in the exponential will be (39) where

$$\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} = (\sigma_k^{-2}\tilde{\mathbf{A}}^T\tilde{\mathbf{y}}_k + \mathbf{P}_k\tilde{\mathbf{x}}_p)^T\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}^{-1}$$
$$\times \boldsymbol{\Sigma}(\sigma_k^{-2}\tilde{\mathbf{A}}^T\tilde{\mathbf{y}}_k + \mathbf{P}_k\tilde{\mathbf{x}}_p)$$
$$= (\sigma_k^{-2}\tilde{\mathbf{A}}^T\tilde{\mathbf{y}}_k + \mathbf{P}_k\tilde{\mathbf{x}}_p)^T(\sigma_k^{-2}\boldsymbol{\Sigma}\tilde{\mathbf{A}}^T\tilde{\mathbf{y}}_k + \boldsymbol{\Sigma}\mathbf{P}_k\tilde{\mathbf{x}}_p)$$
$$= \sigma_k^{-4}\tilde{\mathbf{y}}_k^T\tilde{\mathbf{A}}\boldsymbol{\Sigma}\tilde{\mathbf{A}}^T\tilde{\mathbf{y}}_k + \sigma^{-2}\tilde{\mathbf{y}}_k^T\tilde{\mathbf{A}}\boldsymbol{\Sigma}\mathbf{P}_k\tilde{\mathbf{x}}_p + \sigma_k^{-2}\tilde{\mathbf{x}}_p^T\mathbf{P}_k^T\tilde{\mathbf{A}}^T\tilde{\mathbf{y}}_k + \tilde{\mathbf{x}}_p^T\mathbf{P}_k^T\boldsymbol{\Sigma}\mathbf{P}_k\tilde{\mathbf{x}}_p. \quad (41)$$

Therefore the exponential term is given by

$$-\frac{1}{2}\Big[\sigma_k^{-2}\tilde{\mathbf{y}}_k^T\tilde{\mathbf{y}}_k + \tilde{\mathbf{x}}_p^T\mathbf{P}_k\tilde{\mathbf{x}}_p - \sigma_k^{-4}\tilde{\mathbf{y}}_k^T\tilde{\mathbf{A}}\boldsymbol{\Sigma}\tilde{\mathbf{A}}^T\tilde{\mathbf{y}}_k - \sigma_k^{-2}\tilde{\mathbf{y}}_k^T\tilde{\mathbf{A}}\boldsymbol{\Sigma}\mathbf{P}_k\tilde{\mathbf{x}}_p - \sigma_k^{-2}\tilde{\mathbf{x}}_p^T\mathbf{P}_k^T\tilde{\mathbf{A}}^T\tilde{\mathbf{y}}_k - \tilde{\mathbf{x}}_p^T\mathbf{P}_k^T\boldsymbol{\Sigma}\mathbf{P}_k\tilde{\mathbf{x}}_p\Big]$$

$$= -\frac{1}{2}\Big[\tilde{\mathbf{y}}_k^T[\sigma_k^{-2} - \sigma_k^{-4}\tilde{\mathbf{A}}\boldsymbol{\Sigma}\tilde{\mathbf{A}}^T]\tilde{\mathbf{y}}_k + \tilde{\mathbf{x}}_p^T[\mathbf{P}_k - \mathbf{P}_k^T\boldsymbol{\Sigma}\mathbf{P}_k]\tilde{\mathbf{x}}_p - \sigma_k^{-2}\tilde{\mathbf{y}}_k^T\tilde{\mathbf{A}}\boldsymbol{\Sigma}\mathbf{P}_k\tilde{\mathbf{x}}_p - \sigma_k^{-2}\tilde{\mathbf{x}}_p^T\mathbf{P}_k^T\tilde{\mathbf{A}}^T\tilde{\mathbf{y}}_k\Big] \quad (42)$$

The term outside of the exponential is given by

$$\frac{(2\pi\sigma_k^2)^{-M}(2\pi)^{-N}|\mathbf{P}_k|^{1/2}}{(2\pi)^{-N}|\boldsymbol{\Sigma}|^{-\frac{1}{2}}} = (2\pi\sigma_k^2)^{-M}|\boldsymbol{\Sigma}|^{\frac{1}{2}}|\mathbf{P}_k|^{\frac{1}{2}}. \quad (43)$$

This gives the marginal likelihood as

$$\mathcal{P}(\tilde{\mathbf{y}}_k|\sigma_k^2, \mathbf{p}_k, \tilde{\mathbf{x}}_p) = (2\pi\sigma_k^2)^{-M}|\boldsymbol{\Sigma}|^{\frac{1}{2}}|\mathbf{P}_k|^{\frac{1}{2}}$$
$$\times \exp\left\{ -\frac{1}{2}[\tilde{\mathbf{y}}_k^T\mathbf{B}\tilde{\mathbf{y}}_k + \tilde{\mathbf{x}}_p^T\mathbf{C}\tilde{\mathbf{x}}_p - 2\sigma_k^2\tilde{\mathbf{y}}_k^T\tilde{\mathbf{A}}\boldsymbol{\Sigma}\mathbf{P}_k\tilde{\mathbf{x}}_p] \right\}, \quad (44)$$

where $\mathbf{B}$ and $\mathbf{C}$ are defined as in Section II-B. The log likelihood is then given by

$$\mathcal{L}(\sigma_k^2, \mathbf{p}_k) = \log\left\{ (2\pi\sigma_k^2)^{-M}|\boldsymbol{\Sigma}|^{\frac{1}{2}}|\mathbf{P}_k|^{\frac{1}{2}} \right.$$
$$\left. \times \exp\left\{ -\frac{1}{2}[\tilde{\mathbf{y}}_k^T\mathbf{B}\tilde{\mathbf{y}}_k + \tilde{\mathbf{x}}_p^T\mathbf{C}\tilde{\mathbf{x}}_p - 2\sigma_k^2\tilde{\mathbf{y}}_k^T\tilde{\mathbf{A}}\boldsymbol{\Sigma}\mathbf{P}_k\tilde{\mathbf{x}}_p] \right\} \right\}$$
$$= -M\log(2\pi) - M\log\sigma_k^2 + \frac{1}{2}\log|\boldsymbol{\Sigma}| +$$
$$\frac{1}{2}\log|\mathbf{P}_k| - \frac{1}{2}[\tilde{\mathbf{y}}_k^T\mathbf{B}\tilde{\mathbf{y}}_k + \tilde{\mathbf{x}}_p^T\mathbf{C}\tilde{\mathbf{x}}_p - 2\sigma_k^2\tilde{\mathbf{y}}_k^T\tilde{\mathbf{A}}\boldsymbol{\Sigma}\mathbf{P}_k\tilde{\mathbf{x}}_p]. \quad (45)$$

Using the Woodbury matrix inversion identity gives

$$\mathbf{B} = \sigma_k^{-2}\mathbf{I} - \sigma_k^{-2}\tilde{\mathbf{A}}(\mathbf{P}_k + \sigma_k^{-2}\tilde{\mathbf{A}}^T\tilde{\mathbf{A}})^{-1}\tilde{\mathbf{A}}^T\sigma_k^{-2}, \quad (46)$$

which means

$$\begin{aligned}
\tilde{\mathbf{y}}_k^T\mathbf{B}\tilde{\mathbf{y}}_k &= \tilde{\mathbf{y}}_k^T\sigma_k^{-2}\tilde{\mathbf{y}}_k - \tilde{\mathbf{y}}_k^T(\sigma_k^{-2}\mathbf{I} - \sigma_k^{-2}\tilde{\mathbf{A}} \\
&\times (\mathbf{P} + \sigma_k^{-2}\tilde{\mathbf{A}}^T\tilde{\mathbf{A}})^{-1}\tilde{\mathbf{A}}^T\sigma_k^{-2})\tilde{\mathbf{y}}_k \\
&= \tilde{\mathbf{y}}_k^T\sigma_k^{-2}\tilde{\mathbf{y}}_k - \tilde{\mathbf{y}}_k^T\sigma_k^{-2}\tilde{\mathbf{A}}\boldsymbol{\Sigma}\tilde{\mathbf{A}}^T\sigma_k^{-2}\tilde{\mathbf{y}}_k \\
&= \sigma_k^{-2}\tilde{\mathbf{y}}_k^T(\tilde{\mathbf{y}}_k - \tilde{\mathbf{A}}\boldsymbol{\mu}) + \sigma_k^{-2}\tilde{\mathbf{y}}_k^T\tilde{\mathbf{A}}\boldsymbol{\Sigma}\mathbf{P}_k\tilde{\mathbf{x}}_p \\
&= \sigma_k^{-2}||\tilde{\mathbf{y}}_k^T - \tilde{\mathbf{A}}\boldsymbol{\mu}||_2^2 + \boldsymbol{\mu}^T\mathbf{P}_k\boldsymbol{\mu} + \sigma^{-2}\tilde{\mathbf{y}}_k^T\tilde{\mathbf{A}}\boldsymbol{\Sigma}\mathbf{P}_k\tilde{\mathbf{x}}_p.
\end{aligned}$$
$$(47)$$

Also, we know that $\mathbf{P}_k^T = \mathbf{P}_k$ as $\mathbf{P}_k$ is a real valued diagonal matrix. This means

$$\begin{aligned}
\tilde{\mathbf{x}}_p^T\mathbf{C}\tilde{\mathbf{x}}_p &= \tilde{\mathbf{x}}_p^T[\mathbf{P}_k - \mathbf{P}_k\boldsymbol{\Sigma}\mathbf{P}_k]\tilde{\mathbf{x}}_p \\
&= \tilde{\mathbf{x}}_p^T\mathbf{P}_k\tilde{\mathbf{x}}_p - \tilde{\mathbf{x}}_p^T\mathbf{P}_k\boldsymbol{\Sigma}\mathbf{P}_k\tilde{\mathbf{x}}_p \\
&= \tilde{\mathbf{x}}_p^T\mathbf{P}_k\tilde{\mathbf{x}}_p - \tilde{\mathbf{x}}_p^T\mathbf{P}_k\boldsymbol{\mu} + \sigma_k^{-2}\tilde{\mathbf{y}}_k^T\tilde{\mathbf{A}}\boldsymbol{\Sigma}\mathbf{P}_k\tilde{\mathbf{x}}_p, \quad (48)
\end{aligned}$$

which then gives the log likelihood function in (17).

C. Derivation of Update Expressions for Modified RVM

Firstly, differentiating (17) with respect to $p_{k,n}$ gives

$$-\frac{1}{2}\left[\Sigma_{nn} - \frac{1}{p_{k,n}} + \mu_n^2 + \tilde{x}_{e,n}^2 - \tilde{x}_{e,n}\mu_n \right] \quad (49)$$

and equating to zero gives

$$\Sigma_{nn} - \frac{1}{p_{k,n}} + \mu_n^2 + \tilde{x}_{e,n}^2 - \tilde{x}_{e,n}\mu_n = 0$$
$$1 - p_{k,n}\Sigma_{nn} - p_{k,n}\mu_n^2 - p_{k,n}\tilde{x}_{e,n}^2 + p_{k,n}\tilde{x}_{e,n}\mu_n = 0$$
$$\gamma_n - p_{k,n}[\mu_n^2 + \tilde{x}_{e,n}^2 - \tilde{x}_{e,n}\mu_n] = 0 \quad (50)$$

which leads to (18).

Now collect the terms with σ_k^2 in to give

$$-\frac{1}{2}\left[2M\log\sigma_k^2 - \log|\boldsymbol{\Sigma}| + \sigma_k^{-2}||\tilde{\mathbf{y}}_k - \tilde{\mathbf{A}}\boldsymbol{\mu}||_2^2 \right] \quad (51)$$

and then define $\tau = \sigma_k^{-2}$ giving

$$-\frac{1}{2}\left[2M\log\tau^{-1} - \log|\boldsymbol{\Sigma}| + \tau||\tilde{\mathbf{y}}_k - \tilde{\mathbf{A}}\boldsymbol{\mu}||_2^2 \right]$$
$$= -\frac{1}{2}\left[-2M\log\tau - \log|\boldsymbol{\Sigma}| + \tau||\tilde{\mathbf{y}}_k - \tilde{\mathbf{A}}\boldsymbol{\mu}||_2^2 \right]. (52)$$

Now differentiate (52) with respect to τ and equate to zero to give

$$-\frac{2M}{\tau} + \text{tr}(\boldsymbol{\Sigma}\tilde{\mathbf{A}}^T\tilde{\mathbf{A}}) + ||\tilde{\mathbf{y}}_k - \tilde{\mathbf{A}}\boldsymbol{\mu}||_2^2 = 0, \quad (53)$$

where $\text{tr}(\cdot)$ indicates the trace. As $\text{tr}(\boldsymbol{\Sigma}\tilde{\mathbf{A}}^T\tilde{\mathbf{A}})$ can be written as $\tau^{-1}\sum_n\gamma_n$ giving

$$\tau^{-1}\left(2M - \sum_n\gamma_n\right) = ||\tilde{\mathbf{y}}_k - \tilde{\mathbf{A}}\boldsymbol{\mu}||_2^2, \quad (54)$$

which in turn gives (19).

D. Derivation of (30), (31), (32) and (33)

From (29) it is known that

$$\mathcal{P}(\tilde{x}_{k,n}|\tilde{y}_{k,n}, \sigma_k^2, p_{k,n}, \tilde{z}_{k,n}) \propto (1 - \tilde{z}_{k,n})\delta_0$$
$$\times\mathcal{N}(\tilde{\mathbf{y}}_{k,n}|\tilde{\mathbf{A}}_n\tilde{x}_{k,n}, \sigma_k^2)$$
$$+\tilde{z}_{k,n}\mathcal{N}(\tilde{x}_{k,n}|0, p_{k,n})\mathcal{N}(\tilde{\mathbf{y}}_{k,n}|\tilde{\mathbf{A}}_n\tilde{x}_{k,n}, \sigma_k^2). \quad (55)$$

If we then combine the exponential terms in the second term in (55) we get

$$\begin{aligned}
&-\frac{1}{2}\left[\tilde{x}_{k,n}^Tp_{k,n}\tilde{x}_{k,n} + p_0(\tilde{\mathbf{y}}_{k,n} - \tilde{\mathbf{A}}_n\tilde{x}_{k,n})^T \right. \\
&\left. \times(\tilde{\mathbf{y}}_{k,n} - \tilde{\mathbf{A}}_n\tilde{x}_{k,n}) \right] \\
&= -\frac{1}{2}\left[\tilde{x}_{k,n}^Tp_{k,n}\tilde{x}_{k,n} + p_0(\tilde{\mathbf{y}}_{k,n}^T\tilde{\mathbf{y}}_{k,n} - \tilde{\mathbf{y}}_{k,n}^T\tilde{\mathbf{A}}_n\tilde{x}_{k,n} \right. \\
&\left. -\tilde{x}_{k,n}^T\tilde{\mathbf{A}}_n^T\tilde{\mathbf{y}}_{k,n} + \tilde{x}_{k,n}^T\tilde{\mathbf{A}}_n^T\tilde{\mathbf{A}}_n\tilde{x}_{k,n}) \right] \\
&= -\frac{1}{2}\left[\tilde{x}_{k,n}^T[p_{k,n} + p_0\tilde{\mathbf{A}}_n^T\tilde{\mathbf{A}}_n]\tilde{x}_{k,n} + p_0\tilde{\mathbf{y}}_{k,n}^T\tilde{\mathbf{y}}_{k,n} - \right. \\
&\left. p_0\tilde{\mathbf{y}}_{k,n}^T\tilde{\mathbf{A}}_n\tilde{x}_{k,n} - p_0\tilde{x}_{k,n}^T\tilde{\mathbf{A}}_n^T\tilde{\mathbf{y}}_{k,n} \right]. \quad (56)
\end{aligned}$$

Completing the square gives

$$-\frac{1}{2}\left[(\tilde{x}_{k,n} - \hat{\mu}_n)^T\hat{p}_{k,n}(\tilde{x}_{k,n} - \hat{\mu}_n) - \hat{\mu}_n\hat{p}_{k,n}\hat{\mu}_n \right.$$
$$\left. +p_0\tilde{\mathbf{y}}_{k,n}^T\tilde{\mathbf{y}}_{k,n} \right], \quad (57)$$

where $\hat{p}_{k,n}$ and $\hat{\mu}_{k,n}$ are given by (31) and (32), respectively.

In order to complete the expression given in (30) it is now necessary to get a new indicator variable, $\hat{z}_{k,n}$ for the new posterior distribution for $\tilde{x}_{k,n}$. To do this, assume that

$$\frac{\tilde{z}_{k,n}}{1 - \tilde{z}_{k,n}}\mathcal{N}(0|0, p_{k,n}) = \frac{\hat{z}_{k,n}}{1 - \hat{z}_{k,n}}\mathcal{N}(0|\hat{\mu}_n, \hat{p}_{k,n}). \quad (58)$$

Thus giving us (33), allowing us to write the posterior distribution for $x_{k,n}$ in the form given in (30).

ACKNOWLEDGMENTS

REFERENCES

[1] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[2] A. Swindlehurst and T. Kailath, "A performance analysis of subspace-based methods in the presence of model errors. I. the MUSIC algorithm," *IEEE Transactions on Signal Processing*, vol. 40, no. 7, pp. 1758–1774, 1992.

[3] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984–995, 1989.

[4] M. Zoltowski, M. Haardt, and C. P. Mathews, "Closed-form 2-D angle estimation with rectangular arrays in element space or beamspace via unitary ESPRIT," *IEEE Transactions on Signal Processing*, vol. 44, no. 2, pp. 316–328, 1996.

[5] N. Tayem and H. Kwon, "Conjugate ESPRIT (C-SPRIT)," *IEEE Transactions on Antennas and Propagation*, vol. 52, no. 10, pp. 2618–2624, 2004.

[6] F. Gao and A. Gershman, "A generalized ESPRIT approach to direction-of-arrival estimation," *IEEE Signal Processing Letters*, vol. 12, no. 3, pp. 254–257, 2005.

[7] I. Ziskind and M. Wax, "Maximum likelihood localization of multiple sources by alternating projection," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 10, pp. 1553–1560, 1988.

[8] Y.-D. Huang and M. Barkat, "A dynamic programming algorithm for the maximum likelihood localization of multiple sources," *IEEE Transactions on Antennas and Propagation*, vol. 40, no. 9, pp. 1023–1030, 1992.

[9] P. Stoica and A. Gershman, "Maximum-likelihood DOA estimation by data-supported grid search," *IEEE Signal Processing Letters*, vol. 6, no. 10, pp. 273–275, 1999.

[10] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489 – 509, 2006.

[11] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[12] D. Malioutov, M. Cetin, and A. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3010–3022, 2005.

[13] M. Hyder and K. Mahata, "A robust algorithm for joint-sparse recovery," *IEEE Signal Processing Letters*, vol. 16, no. 12, pp. 1091–1094, 2009.

[14] I. Bilik, T. Northardt, and Y. Abramovich, "Expected likelihood for compressive sensing-based DOA estimation," in *Proc. IET International Conference on Radar Systems*, 2012, pp. 1–4.

[15] Q. Shen, W. Liu, W. Cui, S. Wu, Y. Zhang, and M. Amin, "Group sparsity based wideband DOA estimation for co-prime arrays," in *Proc. IEEE China Summit International Conference on Signal and Information Processing*, 2014, pp. 252–256.

[16] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2346–2356, 2008.

[17] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, pp. 211–244, 2001.

[18] M. E. Tipping and A. Faul, "Fast marginal likelihood maximisation for sparse Bayesian models," in *Proc. of the International Workshop on Artificial Intelligence and Statistics*, 2003, pp. 3–6.

[19] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, USA: Springer, 2006.

[20] M. Carlin, P. Rocca, G. Oliveri, F. Viani, and A. Massa, "Directions-of-arrival estimation through Bayesian compressive sensing strategies," *IEEE Transactions on Antennas and Propagation*, vol. 61, no. 7, pp. 3828–3838, 2013.

[21] Z. Yang, L. Xie, and C. Zhang, "Off-grid direction of arrival estimation using sparse Bayesian inference," *IEEE Transactions on Signal Processing*, vol. 61, no. 1, pp. 38–43, 2013.

[22] P. Khomchuk and I. Bilik, "Dynamic direction-of-arrival estimation via spatial compressive sensing," in *Proc. IEEE Radar Conference*, 2010, pp. 1191–1196.

[23] E. Karseras, K. Leung, and W. Dai, "Tracking dynamic sparse signals using hierarchical Bayesian Kalman filters," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2013, pp. 6546–6550.

[24] J. Filos, E. Karseras, W. Dai, and S. Yan, "Tracking dynamic sparse signals with hierarchical Kalman filters: A case study," in *Proc. International Conference on Digital Signal Processing*, 2013, pp. 1–6.

[25] T. Su and H. Ling, "On modeling mutual coupling in antenna arrays using the coupling matrix," *Microwave and Optical Technology Letters*, vol. 28, no. 4, pp. 231–237, 2001.

[26] B. Liao, Z. G. Zhang, and S. C. Chan, "DOA estimation and tracking of ULAs with mutual coupling," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 1, pp. 891–905, 2012.

[27] M. Hawes, L. Mihaylova, F. Septier, and S. Godsill, "A Bayesian compressed sensing Kalman filter for direction of arrival estimation," in *Proc. International Conference on Information Fusion*, 2015, pp. 969–975.

[28] D. F. Andrews and C. L. Mallows, "Scale mixtures of normal distributions," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 36, no. 1, pp. 99–102, 1974.

[29] P. J. Wolfe and S. J. Godsill, *Bayesian modelling of time-frequency coefficients for audio signal enhancement, in Advances in Neural Information Processing Systems 15*. Cambridge, MA: MIT press, 2002.

[30] P. J. Wolfe, S. J. Godsill, and W.-J. Ng, "Bayesian variable selection and regularization for timefrequency surface estimation," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 66, no. 3, pp. 575–589, 2004.

[31] C. Fevotte and S. Godsill, "Sparse linear regression in unions of bases via Bayesian variable selection," *IEEE Signal Processing Letters*, vol. 13, no. 7, pp. 441–444, 2006.

[32] L. He and L. Carin, "Exploiting structure in wavelet-based Bayesian compressive sensing," *IEEE Transactions on Signal Processing*, vol. 57, no. 9, pp. 3488–3497, 2009.

[33] L. Yu, H. Sun, J. Barbot, and G. Zheng, "Bayesian compressive sensing for cluster structured sparse signals," *Signal Processing*, vol. 92, no. 1, pp. 259 – 269, 2012.

[34] H. L. Van Trees, *Optimum Array Processing, Part IV of Detection, Estimation, and Modulation Theory*. New York, U.S.A.: John Wiley & Sons, Inc., 2002.

Matthew Hawes received his MEng and PhD degree from the Department of Electronic and Electrical Engineering at the University of Sheffield in 2010 and 2014, respectively. Since then he has been employed as a research associate in the Department of Automatic Control and Systems Engineering at the same university. He is currently working on the EU funded SETA project, the main scope of which is the development of models, methods and a platform for mobility prediction, congestion avoidance and sensor data fusion for smart cities. His research interests include array signal processing, localisation and tracking, big data, modelling complex systems, intelligent transportation systems, mobility, data fusion, sequential Monte Carlo methods and Markov chain Monte Carlo methods.

Lyudmila Mihaylova (M'98, SM'2008) is Professor of Signal Processing and Control at the Department of Automatic Control and Systems Engineering at the University of Sheffield, United Kingdom. Her research is in the areas of machine learning and autonomous systems with various applications such as navigation, surveillance and sensor network systems. She has given a number of talks and tutorials, including the plenary talk for the IEEE Sensor Data Fusion 2015 (Germany), invited talks University of California, Los Angeles, IPAMI Traffic Workshop 2016 (USA), IET ICWMMN 2013 in Beijing, China. Dr. Mihaylova is an Associate Editor of the IEEE Transactions on Aerospace and Electronic Systems and of the Elsevier Signal Processing Journal. She was elected in March 2016 as a president of the International Society of Information Fusion (ISIF). She is on the board of Directors of ISIF and a Senior IEEE member. She was the general co-chair IET Data Fusion & Target Tracking 2014 and 2012 Conferences, Program co-chair for the 19th International Conference on Information Fusion, Heidelberg, Germany, 2016, academic chair of Fusion 2010 conference.

François Septier received the Engineer Degree in electrical engineering and signal processing in 2004 from Télécom Lille France, and a Ph.D in Electrical Engineering from the University of Valenciennes France in 2008. From March 2008 to August 2009, he was a Research Associate in the Signal Processing and Communications Laboratory, Cambridge University, Engineering Department, UK. From August 2009, he is an Associate Professor with the IMT Lille Douai / CRIStAL UMR CNRS 9189, France. His research focuses on Bayesian computational methodology with a particular emphasis on the development of Monte Carlo based approaches for complex and high-dimensional problems.

Simon Godsill is Professor of Statistical Signal Processing in the Engineering Department at Cambridge University. He is also a Professorial Fellow and tutor at Corpus Christi College Cambridge. He coordinates an active research group in Signal Inference and its Applications within the Signal Processing and Communications Laboratory at Cambridge, specializing in Bayesian computational methodology, multiple object tracking, audio and music processing, and financial time series modeling. A particular methodological theme over recent years has been the development of novel techniques for optimal Bayesian filtering, using Sequential Monte Carlo or Particle Filtering methods. Prof. Godsill has published extensively in journals, books and international conference proceedings, and has given a number of high profile invited and plenary addresses at conferences such as the Valencia conference on Bayesian Statistics and the IEEE Statistical Signal Processing Workshop. He was technical chair of the successful IEEE NSSPW workshop in 2006 on sequential and nonlinear filtering methods, and has been on the conference panel for numerous other conferences/workshops. Prof. Godsill has served as Associate Editor for IEEE Tr. Signal Processing and the journal Bayesian Analysis. He was Theme Leader in Tracking and Reasoning over Time for the UK's Data and Information Fusion Defence Technology Centre (DIF-DTC) and Principal Investigator on many grants funded by the EU, EPSRC, QinetiQ , MOD, Microsoft UK, Citibank and Mastercard.