

A Bayesian Compressed Sensing Kalman Filter for Direction of Arrival Estimation

Matthew Hawes^a, Lyudmila Mihaylova^a, Francois Septier^b and Simon Godsill^c

^a Department of Automatic Control and Systems Engineering, University of Sheffield, S1 3JD, UK

^b Institute Mines Telecom/Telecom Lille, CRISTAL UMR CNRS 9189, France

^c Department of Engineering, Cambridge University, CB 1PZ, UK

{m.hawes, l.s.mihaylova}@sheffield.ac.uk, francois.septier@telecom-lille.fr, sjg30@cam.ac.uk

Abstract—In this paper, we look to address the problem of estimating the dynamic direction of arrival (DOA) of a narrowband signal impinging on a sensor array from the far field. The initial estimate is made using a Bayesian compressive sensing (BCS) framework and then tracked using a Bayesian compressed sensing Kalman filter (BCSKF). The BCS framework splits the angular region into N potential DOAs and enforces a belief that only a few of the DOAs will have a non-zero valued signal present. A BCSKF can then be used to track the change in the DOA using the same framework. There can be an issue when the DOA approaches the endfire of the array. In this angular region current methods can struggle to accurately estimate and track changes in the DOAs. To tackle this problem, we propose changing the traditional sparse belief associated with BCS to a belief that the estimated signals will match the predicted signals given a known DOA change. This is done by modelling the difference between the expected sparse received signals and the estimated sparse received signals as a Gaussian distribution. Example test scenarios are provided and comparisons made with the traditional BCS based estimation method. They show that an improvement in estimation accuracy is possible without a significant increase in computational complexity.

Index Terms—DOA estimation, Bayesian compressed sensing, Kalman filter, dynamic DOA, DOA tracking

I. INTRODUCTION

Direction of arrival (DOA) estimation is the process of determining which direction a signal impinging on an array has arrived from [1]. Commonly used methods of solving this problem are: MUSIC [2], [3] and ESPRIT [4], [5]. However, these methods have two drawbacks: Firstly, we need some knowledge of the number of signals that are present. Secondly, evaluation of the covariance matrix is required, thus increasing the computational complexity required to solve the problem. This covariance matrix is estimated from the signals received by each sensor at different time snapshots. Instead, if we consider the fact that only a few of the potential DOAs will have a signal present then we can consider the problem from

the view point of compressive sensing (CS) and work directly with the received signals.

CS theory tells us that when certain conditions are met it is possible to recover some signals from fewer measurements than used by traditional methods [6], [7]. This can be applied to solve the problem of DOA estimation [8], [9], [10]. First split the angular region of interest into N potential DOAs, where signals actually impinge on the array from only L ($L \ll N$) of these directions. The problem can then be formulated as finding the minimum number of DOAs with a signal present that still gives an acceptable approximation of the array output. Those directions that have the non-zero valued signals are then used as the DOA estimates. It is also possible to convert this problem into a probabilistic form and solve using a relevance vector machine (RVM) based approach [11], [12], [13]. It has been shown in the case of static DOA estimation that methods based on this approach offer encouraging results [14].

Less work has been done on the problem of estimating a dynamic DOA. One option is to use particle filters or probability hypothesis density (PHD) filters. These filters have been used in the areas of DOA estimation and tracking of sources [15], [16], [17], [18].

Alternatively, in [19] the authors track a dynamic DOA with a Kalman filter (KF) and narrow the angular region being considered to focus in more closely on the DOA estimate from the previous iteration. However, this removes the advantage of being able to directly work 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.

Bayesian Kalman filters (BKF) have been used to track dynamic sparse signals [20], where the predicted mean of the signals at each iteration is taken as the estimate from the previous iteration and the hyper-parameters (precision) are estimated using BCS, hence the term Bayesian compressed

sensing Kalman Filter (BCSKF). There are still some issues with this method when applied to the problem of DOA estimation of a dynamic far field source using uniform linear arrays (ULAs). Namely when the DOA approaches the endfire region (i.e. the signal arrives parallel/close to parallel to the array), the estimation accuracy can degrade. This means that it is possible to initially have an accurate estimate and then not track the changes in DOA properly.

In this paper, to solve this problem, we propose modifying the BCSKF to include information about what the received array signals would be expected to be at a given time snapshot. This is done by changing the distribution used to model the received signals. Now instead of assuming a zero-mean Gaussian hierarchical prior we assume that the mean is instead centred at the value that is expected given the previous snapshot's estimate and the expected change in DOA. As a result we derive a new posterior distribution and marginal likelihood function that can be used to solve the DOA estimation problem by following a similar framework as used for the RVM. We term this framework the modified RVM, which is used to find an initial estimate of the DOA at the first snapshot and then at each subsequent snapshot to optimise the noise variance estimate as well as the hyperparameters required by the BCSKF.

The remainder of this paper is structured in the following manner: Section II gives details of the proposed estimation method. This includes the array model being used (II-A), the modified RVM framework for BCS (II-B) and the BCSKF (II-C). In Section III an evaluation of the effectiveness of the proposed method is presented and conclusions are drawn in Section IV.

II. PROPOSED DESIGN METHODS

A. Array Model

A narrowband array structure consisting of M sensors is shown in Fig. 1. The sensors are assumed to be omnidirectional with identical responses. A plane-wave signal model is assumed, i.e. the signal impinges upon the array from the far field at an angle θ as shown. In this work we assume that $0^\circ \leq \theta \leq 180^\circ$. The distance from the first sensor to subsequent sensors is denoted as d_m for $m = 1, 2, \dots, M$, with $d_1 = 0$, i.e. the distance from the first sensor to itself. Note, these values are multiples of a uniform adjacent sensor separation of Δd .

The steering vector of the array is given by

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

where $\Omega = \omega T_s$ is the normalised frequency with T_s being the sampling period, $\mu_m = \frac{d_m}{cT_s}$ for $m = 1, 2, \dots, M$, and $\{\cdot\}^T$ denotes the transpose operation. Note, the steering vector of an array gives contains information about the array geometry,

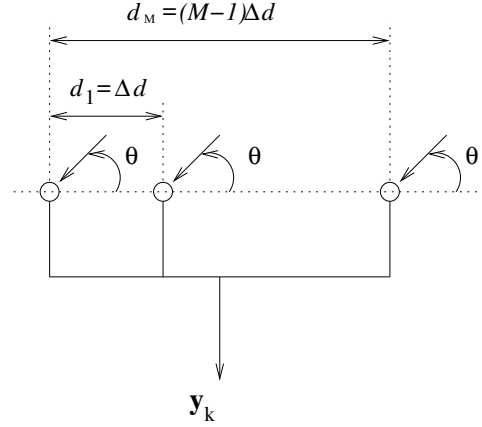


Fig. 1. Linear Array structure being considered, consisting of M sensors with a uniform adjacent sensor separation of Δd .

namely the sensor locations and the delay required for a signal to reach a given sensor.

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

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

where $\mathbf{x}_k = [x_{k,1}, x_{k,2}, \dots, x_{k,N}]^T \in \mathbb{C}^{N \times 1}$ gives the received signals, $\mathbf{n}_k = [n_{k,1}, n_{k,2}, \dots, n_{k,M}]^T \in \mathbb{C}^{M \times 1}$ the sensor noise and $\mathbf{A} = [\mathbf{a}(\Omega, \theta_1), \mathbf{a}(\Omega, \theta_2), \dots, \mathbf{a}(\Omega, \theta_N)] \in \mathbb{C}^{M \times N}$ is the matrix containing the steering vectors for each angle of interest.

B. Bayesian Compressed Sensing for DOA Estimation

First split the angular range that is being monitored into N potential DOAs. Each direction can then be considered as having a signal present. However, only $L \ll N$ of the received signals are non-zero valued, with the directions of these L signals giving the actual DOAs.

Now we split (2) into real and imaginary components in the following way

$$\tilde{\mathbf{y}}_k = \tilde{\mathbf{A}}\tilde{\mathbf{x}}_k + \tilde{\mathbf{n}}_k \quad (3)$$

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

where $\mathcal{R}(\cdot)$ and $\mathcal{I}(\cdot)$ give the real and imaginary components respectively. Note, 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 larger than for \mathbf{y}_k , but we are now only considering real valued data. Similar is true when comparing \mathbf{A} and $\tilde{\mathbf{A}}$, and \mathbf{x}_k and $\tilde{\mathbf{x}}_k$ and \mathbf{n}_k and $\tilde{\mathbf{n}}_k$.

The aim is to now find a solution for $\tilde{\mathbf{x}}_k$ that gives the minimum l_0 norm, i.e. the minimum number of non-zero valued signals. This is done by evaluating the following

$$\tilde{\mathbf{x}}_{k,opt} = \max \mathcal{P}(\tilde{\mathbf{x}}_k, \sigma^2, \mathbf{p}|\tilde{\mathbf{y}}, \mathbf{x}_e), \quad (4)$$

where σ is the variance of the Gaussian noise \mathbf{n} , $\mathbf{p} = [p_1, p_2, \dots, p_{2N}]^T$ contains the hyperparameters that are to be estimated and \mathbf{x}_e holds the expected values of $\tilde{\mathbf{x}}_k$.

To do this first obtain the following from (2):

$$\mathcal{P}(\tilde{\mathbf{y}}_k | \tilde{\mathbf{x}}_k, \sigma^2) = (2\pi\sigma^2)^{-M} \exp \left\{ -\frac{1}{2\sigma^2} \|\tilde{\mathbf{y}}_k - \tilde{\mathbf{A}}\tilde{\mathbf{x}}_k\|_2^2 \right\}. \quad (5)$$

Now exert a belief about the values of $\tilde{\mathbf{x}}_k$ that are expected by enforcing

$$\begin{aligned} \mathcal{P}(\tilde{\mathbf{x}}_k | \mathbf{p}, \mathbf{x}_e) &= (2\pi)^{-N} |\mathbf{P}|^{1/2} \\ &\times \exp \left\{ -\frac{1}{2} (\tilde{\mathbf{x}}_k - \mathbf{x}_e) \mathbf{P} (\tilde{\mathbf{x}}_k - \mathbf{x}_e)^T \right\}, \end{aligned} \quad (6)$$

where $|\mathbf{P}|$ indicates the determinant of \mathbf{P} , where $\mathbf{P} = \text{diag}(\mathbf{p})$.

It is also necessary to define the hyperparameters over \mathbf{p} and σ^2 . There are various possibilities for the structuring of the priors on \mathbf{p} , 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. [21]. One possibility would be to treat the complex components of \mathbf{x}_k as complex Student-t distributed, as detailed in [22], [23]. Here though we treat the real and imaginary components of \mathbf{x}_k as independent Student-t distributed random variables, and hence have independent Gamma priors for the mixing variables p_n over all real and imaginary components of \mathbf{x}_k :

$$\mathcal{P}(\mathbf{p}) = \prod_{n=1}^{2N} G(p_n | \beta_1, \beta_2). \quad (7)$$

A Gamma prior can also be used for σ^2

$$\mathcal{P}(\sigma^2) = G(\sigma^{-2} | \beta_3, \beta_4), \quad (8)$$

where $\beta_1, \beta_2, \beta_3$ and β_4 are scale and shape priors. Note, when $\mathbf{x}_e = [0, 0, \dots, 0]^T$ then (6) reverts to the traditional hierarchical prior used in BCS [11].

We know that

$$\mathcal{P}(\tilde{\mathbf{x}}_k, \sigma^2, \mathbf{p} | \tilde{\mathbf{y}}_k, \mathbf{x}_e) = \mathcal{P}(\tilde{\mathbf{x}}_k | \tilde{\mathbf{y}}_k, \sigma^2, \mathbf{p}, \mathbf{x}_e) \mathcal{P}(\mathbf{p}, \sigma^2 | \tilde{\mathbf{y}}_k) \quad (9)$$

and ¹

$$\begin{aligned} \mathcal{P}(\tilde{\mathbf{x}}_k | \tilde{\mathbf{y}}_k, \mathbf{p}, \sigma^2, \mathbf{x}_e) &= \frac{\mathcal{P}(\tilde{\mathbf{y}}_k | \tilde{\mathbf{x}}_k, \sigma^2) \mathcal{P}(\tilde{\mathbf{x}}_k | \mathbf{p}, \mathbf{x}_e)}{\mathcal{P}(\tilde{\mathbf{y}}_k | \mathbf{p}, \sigma^2, \mathbf{x}_e)} \\ &= (2\pi)^{-N} |\Sigma|^{-1/2} \\ &\times \exp \left\{ -\frac{1}{2} (\tilde{\mathbf{x}}_k - \boldsymbol{\mu})^T \Sigma^{-1} (\tilde{\mathbf{x}}_k - \boldsymbol{\mu}) \right\}, \end{aligned} \quad (10)$$

where the covariance matrix is given by

$$\Sigma = (\sigma^{-2} \tilde{\mathbf{A}}^T \tilde{\mathbf{A}} + \mathbf{P})^{-1}, \quad (11)$$

and the mean given by

$$\boldsymbol{\mu} = \Sigma (\sigma^{-2} \tilde{\mathbf{A}}^T \tilde{\mathbf{y}}_k + \mathbf{P} \mathbf{x}_e). \quad (12)$$

¹See Appendix A

Note, the maximum of (10) is the posterior mean $\boldsymbol{\mu}$.

Similarly to [12], the probability $\mathcal{P}(\sigma^2, \mathbf{p} | \tilde{\mathbf{y}}_k)$ can be represented in the following form:

$$\mathcal{P}(\sigma^2, \mathbf{p} | \tilde{\mathbf{y}}_k) \approx \mathcal{P}(\tilde{\mathbf{y}}_k | \mathbf{p}, \sigma^2, \mathbf{x}_e) \mathcal{P}(\mathbf{p}) \mathcal{P}(\sigma^2) \quad (13)$$

and the second two terms on the right of (13) become constant with uniform scale priors, then maximising $\mathcal{P}(\sigma^2, \mathbf{p} | \tilde{\mathbf{y}}_k)$ is roughly equivalent to maximising $\mathcal{P}(\tilde{\mathbf{y}}_k | \mathbf{p}, \sigma^2, \mathbf{x}_e)$. This can be achieved by a type 2 maximisation of its logarithm, which is given by ²

$$\begin{aligned} \mathcal{L}(\mathbf{p}, \sigma^2) &= \log \left\{ (2\pi\sigma^2)^{-M} |\Sigma|^{\frac{1}{2}} |\mathbf{P}|^{\frac{1}{2}} \exp \left(-\frac{1}{2} \right. \right. \\ &\times \left. \left. (\tilde{\mathbf{y}}_k^T \mathbf{B} \tilde{\mathbf{y}}_k + \mathbf{x}_e^T \mathbf{C} \mathbf{x}_e - 2\sigma^2 \tilde{\mathbf{y}}_k^T \tilde{\mathbf{A}} \Sigma \mathbf{P} \mathbf{x}_e) \right) \right\} \\ &= -\frac{1}{2} \left(2M \log(2\pi) + 2M \log \sigma^2 - \log |\Sigma| - \right. \\ &\log |\mathbf{P}| + \sigma^{-2} \|\tilde{\mathbf{y}}_k - \tilde{\mathbf{A}} \boldsymbol{\mu}\|_2^2 + \boldsymbol{\mu}^T \mathbf{P} \boldsymbol{\mu} \\ &\left. + \mathbf{x}_e^T \mathbf{P} \mathbf{x}_e - \mathbf{x}_e^T \mathbf{P} \boldsymbol{\mu} \right), \end{aligned} \quad (14)$$

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

To do this (14) is differentiated with respect to p_n and σ^{-2} to obtain the update expressions ³

$$p_n^{\text{new}} = \frac{\gamma_n}{\mu_n^2 + x_{e,n}^2 - x_{e,n} \mu_n}, \quad (15)$$

where $\gamma_n = 1 - p_n \Sigma_{nn}$, Σ_{nn} is the n^{th} diagonal element of Σ and

$$\sigma_{\text{new}}^2 = \frac{\|\tilde{\mathbf{y}}_k - \tilde{\mathbf{A}} \boldsymbol{\mu}\|_2^2}{2M - \sum_n \gamma_n}. \quad (16)$$

The maximisation is then achieved by iteratively maximising (11) and (12) and (15) and (16) until a convergence criterion is met [11], [12]. Note that when $\mathbf{x}_e = [0, 0, \dots, 0]^T$ the update expressions match that of the traditional RVM. The final estimate of the received signals is then given by

$$\tilde{\mathbf{x}}_{k,\text{opt}} = \left(\frac{\tilde{\mathbf{A}}^T \tilde{\mathbf{A}}}{\sigma_{\text{opt}}^2} + \mathbf{P}_{\text{opt}} \right)^{-1} \left(\frac{\tilde{\mathbf{A}}^T \tilde{\mathbf{y}}_k}{\sigma_{\text{opt}}^2} + \mathbf{P}_{\text{opt}} \mathbf{x}_e \right) \quad (17)$$

where σ_{opt}^2 and $\mathbf{P}_{\text{opt}} = \text{diag}([p_{\text{opt},1}, p_{\text{opt},2}, \dots, p_{\text{opt},2N}]^T)$ are the result of optimising the noise estimate and hyperparameters respectively.

The final estimated signals are then given by

$$x_{k,\text{opt},n} = \tilde{x}_{k,\text{opt},n} + j \tilde{x}_{k,\text{opt},N+n}. \quad (18)$$

Thresholding can then be applied to keep the \tilde{L} most significant signals as in [14]. 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

²See Appendix B

³See Appendix C

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 set to be equal to 0. The remaining non-zero valued signals then give the DOA estimates and \tilde{L} as an estimate of the number of far field signals impinging on the array.

C. Bayesian Compressed Sensing Kalman Filter

In order to track the changes in the DOA estimates at each time snapshot the BCS based DOA estimation procedure detailed above is combined with a BKF, giving a BCSKF for DOA estimation. Here, the signal model described above is again used along with the prediction

$$\begin{aligned}\tilde{\mathbf{x}}_{k|k-1} &= \tilde{\mathbf{x}}_{k-1|k-1} + \Delta\mathbf{x} & \Sigma_{k|k-1} &= \Sigma_{k-1} + \mathbf{P}_k^{-1} \\ \tilde{\mathbf{y}}_{k|k-1} &= \tilde{\mathbf{A}}\tilde{\mathbf{x}}_{k|k-1} & \tilde{\mathbf{y}}_{e,k} &= \tilde{\mathbf{y}}_k - \tilde{\mathbf{y}}_{k|k-1}\end{aligned}\quad (19)$$

and update steps

$$\begin{aligned}\tilde{\mathbf{x}}_k &= \tilde{\mathbf{x}}_{k|k-1} + \mathbf{K}_k \tilde{\mathbf{y}}_{e,k} & \Sigma_{k|k} &= (\mathbf{I} - \mathbf{K}_k \tilde{\mathbf{A}}) \Sigma_{k|k-1} \\ \mathbf{K}_k &= \Sigma_{k|k-1} \tilde{\mathbf{A}}^T (\sigma^2 \mathbf{I} + \tilde{\mathbf{A}} \Sigma_{k|k-1} \tilde{\mathbf{A}}^T)^{-1}\end{aligned}\quad (20)$$

of the BKF. Here, $k|k-1$ indicates prediction at time instance k given the previous measurements and $\Delta\mathbf{x}$ is determined by the expected DOA change. For example, if we sample the angular range every 1° and the DOA increases by 2° then then $\Delta\mathbf{x}$ 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 $\tilde{\mathbf{x}}_{k|k-1}$. In this work we have assumed that there will be a constant change in the DOA.

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

$$\begin{aligned}\mathcal{L}(\mathbf{p}_k, \sigma^2) &= -\frac{1}{2} \left(2M \log(2\pi) + 2M \log \sigma^2 - \right. \\ &\quad \left. \log |\Sigma| - \log |\mathbf{P}| + \sigma^{-2} \|\tilde{\mathbf{y}}_{e,k} - \tilde{\mathbf{A}}\boldsymbol{\mu}\|_2^2 \right. \\ &\quad \left. + \boldsymbol{\mu}^T \mathbf{P} \boldsymbol{\mu} + \tilde{\mathbf{x}}_{k|k-1}^T \mathbf{P} \tilde{\mathbf{x}}_{k|k-1} - \tilde{\mathbf{x}}_{k|k-1}^T \mathbf{P} \boldsymbol{\mu} \right),\end{aligned}\quad (21)$$

which can be optimised by following the procedure described in Section II-B. Here we have used the Kalman filter prediction $\tilde{\mathbf{x}}_{k|k-1}$ as the expected estimate values \mathbf{x}_e .

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 $\mathbf{x}_e = [0, 0, \dots, 0]^T$) 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

TABLE I
PERFORMANCE SUMMARY FOR THE ENDFIRE REGION EXAMPLE.

Method	Mean <i>RMSE</i> (degrees)	Mean Computation Time (seconds)
RVM	11.03	0.35
Modified RVM	0.98	0.55

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.

III. PERFORMANCE EVALUATION

In this section a comparison in performance of the traditional RVM and the proposed modified RVM will be made. Three example scenarios will be considered. One where the initial DOA starts outside of the endfire region and then moves into it, one where the DOA remains out of the endfire region and finally one where the initial DOAs and signal values are randomly generated. 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 root mean square error (*RMSE*) in DOA estimate. This is given by

$$RMSE = \sqrt{\frac{\sum_{q=1}^Q |\theta - \hat{\theta}|^2}{Q}}, \quad (22)$$

where θ is the actual DOA, $\hat{\theta}$ is the estimated DOA and Q is the number of Monte Carlo simulations carried out, with $Q = 100$ being used in each case.

The array structure being considered is a 20 sensor ULA with an adjacent sensor separation of $\frac{\lambda}{2}$. We assume the actual noise variance is given by $\sigma_n^2 = 0.4$ and an initial estimate of the noise variance of $\sigma_{nit}^2 = 0.1$ used when initialising the RVM and proposed modified RVM.

A. Endfire Region

For this example the initial DOA of the signal is $\theta = 20^\circ$, 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 two methods for this example. Here we can see that there is an improved accuracy in the DOA estimates, as the mean *RMSE* has decreased for the proposed modified RVM method. This is also supported by the overall *RMSE* values as illustrated in Fig. 2. It is also worth noting that the mean computation times show that this improvement has not been at the expense of a significant increase in computational complexity.

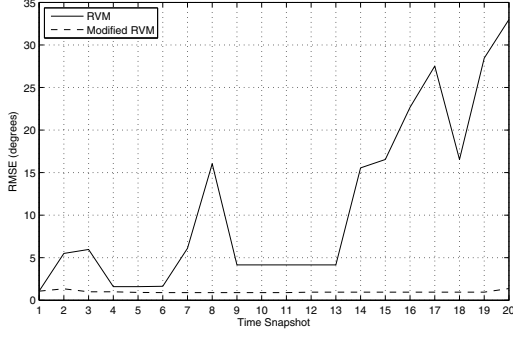


Fig. 2. *RMSE* values for the endfire region example.

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

Method	Mean <i>RMSE</i> (degrees)	Mean Computation Time (seconds)
RVM	5.59	0.33
Modified RVM	0.36	0.41

B. Non-Endfire Region

In this instance the initial DOA is $\theta = 100^\circ$ with the DOA increasing by 1° at each time snapshot, with the signal value remaining constant at -1. The performance of the two methods is summarised in Table II, with the *RMSE* values illustrated in Fig. 3. Again this illustrates the improved performance offered by the modified RVM has not been at the expense of a significant increase in computation time.

C. Random Initial DOA

Finally, we 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. As for the previous two examples Table. III and Fig. 4 indicate that the proposed modified RVM has

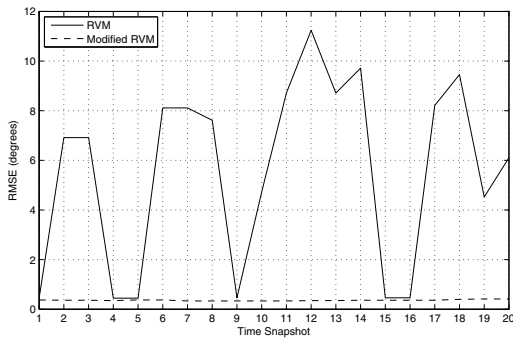


Fig. 3. *RMSE* values for the non-endfire region example.

TABLE III
PERFORMANCE SUMMARY FOR THE RANDOM INITIAL DOA EXAMPLE.

Method	Mean <i>RMSE</i> (degrees)	Mean Computation Time (seconds)
RVM	10.98	0.34
Modified RVM	3.52	0.43

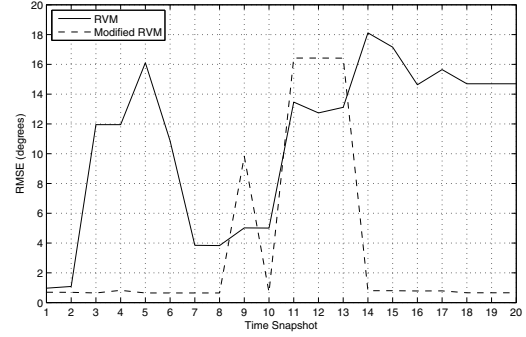


Fig. 4. *RMSE* values for the random initial DOA example.

obtained an improved accuracy without a significant increase in computational complexity.

IV. CONCLUSIONS

This paper proposes a BCSKF to estimate the DOA of a single signal impinging on a ULA from the far field. A new posterior distribution and marginal likelihood has been found and unlike traditional BCS the expected values of the estimates are accounted for. This is done to combat the problem of inaccurate DOA estimates when the actual DOA approaches the endfire region of the angular range. Then a similar optimisation framework to what is used in the RVM is applied to find the optimal hyperparameters and noise variance estimate, which are then used to estimate the received array signals. Example test scenarios have shown the proposed modified RVM is more accurate in not only the endfire region, but also in the entire angular region as a whole. This is also without a significant increase in computational complexity.

APPENDIX

A. Derivation of Posterior Distribution

From Bayes' rule we know that

$$\mathcal{P}(\tilde{\mathbf{x}}_k | \tilde{\mathbf{y}}_k, \mathbf{p}, \sigma^2, \mathbf{x}_e) \mathcal{P}(\tilde{\mathbf{y}}_k | \mathbf{p}, \sigma^2, \mathbf{x}_e) = \mathcal{P}(\tilde{\mathbf{y}}_k | \tilde{\mathbf{x}}_k, \sigma^2) \mathcal{P}(\tilde{\mathbf{x}}_k | \mathbf{p}, \mathbf{x}_e), \quad (23)$$

where $\mathcal{P}(\tilde{\mathbf{y}}_k | \tilde{\mathbf{x}}_k, \sigma^2)$ and $\mathcal{P}(\tilde{\mathbf{x}}_k | \mathbf{p}, \mathbf{x}_e)$ are known from (5) and (6), respectively.

Now following the method suggested in [12] carry out the multiplication on the right hand side, collect terms in $\tilde{\mathbf{x}}_k$ in the exponential and complete the square.

$$\begin{aligned}
& -\frac{1}{2} \left[\sigma^{-2} (\tilde{\mathbf{y}}_k - \tilde{\mathbf{A}} \tilde{\mathbf{x}}_k)^T (\tilde{\mathbf{y}}_k - \tilde{\mathbf{A}} \tilde{\mathbf{x}}_k) + \right. \\
& \quad \left. (\tilde{\mathbf{x}}_k - \mathbf{x}_e)^T \mathbf{P} (\tilde{\mathbf{x}}_k - \mathbf{x}_e) \right] \\
& = -\frac{1}{2} \left[\sigma^{-2} \tilde{\mathbf{y}}_k^T \tilde{\mathbf{y}}_k - \sigma^{-2} \tilde{\mathbf{y}}_k^T \tilde{\mathbf{A}} \tilde{\mathbf{x}}_k - \sigma^{-2} \tilde{\mathbf{x}}_k^T \tilde{\mathbf{A}}^T \tilde{\mathbf{y}}_k + \right. \\
& \quad \left. \sigma^{-2} \tilde{\mathbf{x}}_k^T \tilde{\mathbf{A}}^T \tilde{\mathbf{A}} \tilde{\mathbf{x}}_k + \tilde{\mathbf{x}}_k^T \mathbf{P} \tilde{\mathbf{x}}_k - \tilde{\mathbf{x}}_k^T \mathbf{P} \mathbf{x}_e - \mathbf{x}_e^T \mathbf{P} \tilde{\mathbf{x}}_k \right. \\
& \quad \left. + \mathbf{x}_e^T \mathbf{P} \mathbf{x}_e \right] \\
& = -\frac{1}{2} \left[(\tilde{\mathbf{x}}_k - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\tilde{\mathbf{x}}_k - \boldsymbol{\mu}) - \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} + \right. \\
& \quad \left. \sigma^{-2} \tilde{\mathbf{y}}_k^T \tilde{\mathbf{y}}_k + \mathbf{x}_e^T \mathbf{P} \mathbf{x}_e \right]
\end{aligned} \quad (24)$$

where $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$ are given by (11) and (12), respectively. This then gives the posterior distribution as (10), with the remaining exponential terms

$$-\frac{1}{2} \left[\sigma^{-2} \tilde{\mathbf{y}}_k^T \tilde{\mathbf{y}}_k + \mathbf{x}_e^T \mathbf{P} \mathbf{x}_e - \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \right]. \quad (25)$$

B. Derivation of Marginal Likelihood

From (23) we know that

$$\mathcal{P}(\tilde{\mathbf{y}}_k | \mathbf{p}, \sigma^2, \mathbf{x}_e) = \frac{\mathcal{P}(\tilde{\mathbf{y}}_k | \tilde{\mathbf{x}}_k, \sigma^2), \mathcal{P}(\tilde{\mathbf{x}}_k | \mathbf{p}, \mathbf{x}_e)}{\mathcal{P}(\tilde{\mathbf{x}}_k | \tilde{\mathbf{y}}_k, \mathbf{p}, \sigma^2, \mathbf{x}_e)}, \quad (26)$$

meaning the term in the exponential will be (25) where

$$\begin{aligned}
\boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} & = (\sigma^{-2} \tilde{\mathbf{A}}^T \tilde{\mathbf{y}}_k + \mathbf{P} \tilde{\mathbf{x}}_e)^T \boldsymbol{\Sigma}^T \boldsymbol{\Sigma}^{-1} \\
& \times \boldsymbol{\Sigma} (\sigma^{-2} \tilde{\mathbf{A}}^T \tilde{\mathbf{y}}_k + \mathbf{P} \tilde{\mathbf{x}}_e) \\
& = (\sigma^{-2} \tilde{\mathbf{A}}^T \tilde{\mathbf{y}}_k + \mathbf{P} \tilde{\mathbf{x}}_e)^T (\sigma^{-2} \boldsymbol{\Sigma} \tilde{\mathbf{A}}^T \tilde{\mathbf{y}}_k + \boldsymbol{\Sigma} \mathbf{P} \tilde{\mathbf{x}}_e) \\
& = \sigma^{-4} \tilde{\mathbf{y}}_k^T \tilde{\mathbf{A}} \boldsymbol{\Sigma} \tilde{\mathbf{A}}^T \tilde{\mathbf{y}}_k + \sigma^{-2} \tilde{\mathbf{y}}_k^T \tilde{\mathbf{A}} \boldsymbol{\Sigma} \mathbf{P} \tilde{\mathbf{x}}_e + \\
& \quad \sigma^{-2} \mathbf{x}_e^T \mathbf{P}^T \tilde{\mathbf{A}}^T \tilde{\mathbf{y}}_k + \mathbf{x}_e^T \mathbf{P}^T \boldsymbol{\Sigma} \mathbf{P} \tilde{\mathbf{x}}_e.
\end{aligned} \quad (27)$$

Therefore the exponential term is given by

$$\begin{aligned}
& -\frac{1}{2} \left[\sigma^{-2} \tilde{\mathbf{y}}_k^T \tilde{\mathbf{y}}_k + \mathbf{x}_e^T \mathbf{P} \mathbf{x}_e - \sigma^{-4} \tilde{\mathbf{y}}_k^T \tilde{\mathbf{A}} \boldsymbol{\Sigma} \tilde{\mathbf{A}}^T \tilde{\mathbf{y}}_k - \right. \\
& \quad \left. \sigma^{-2} \tilde{\mathbf{y}}_k^T \tilde{\mathbf{A}} \boldsymbol{\Sigma} \mathbf{P} \tilde{\mathbf{x}}_e - \sigma^{-2} \mathbf{x}_e^T \mathbf{P}^T \tilde{\mathbf{A}}^T \tilde{\mathbf{y}}_k - \right. \\
& \quad \left. \mathbf{x}_e^T \mathbf{P}^T \boldsymbol{\Sigma} \mathbf{P} \tilde{\mathbf{x}}_e \right] \\
& = -\frac{1}{2} \left[\tilde{\mathbf{y}}_k^T [\sigma^{-2} - \sigma^{-4} \tilde{\mathbf{A}} \boldsymbol{\Sigma} \tilde{\mathbf{A}}^T] \tilde{\mathbf{y}}_k + \mathbf{x}_e^T [\mathbf{P} - \mathbf{P}^T \boldsymbol{\Sigma} \mathbf{P}] \mathbf{x}_e \right. \\
& \quad \left. - \sigma^{-2} \tilde{\mathbf{y}}_k^T \tilde{\mathbf{A}} \boldsymbol{\Sigma} \mathbf{P} \tilde{\mathbf{x}}_e - \sigma^{-2} \mathbf{x}_e^T \mathbf{P}^T \tilde{\mathbf{A}}^T \tilde{\mathbf{y}}_k \right]
\end{aligned} \quad (28)$$

The term outside of the exponential is given by

$$\frac{(2\pi\sigma^2)^{-M} (2\pi)^{-N} |\mathbf{P}|^{1/2}}{(2\pi)^{-N} |\boldsymbol{\Sigma}|^{-\frac{1}{2}}} = (2\pi\sigma^2)^{-M} |\boldsymbol{\Sigma}|^{\frac{1}{2}} |\mathbf{P}|^{\frac{1}{2}}. \quad (29)$$

This gives the marginal likelihood as

$$\begin{aligned}
\mathcal{P}(\tilde{\mathbf{y}}_k | \mathbf{p}, \sigma^2, \mathbf{x}_e) & = (2\pi\sigma^2)^{-M} |\boldsymbol{\Sigma}|^{\frac{1}{2}} |\mathbf{P}|^{\frac{1}{2}} \\
& \times \exp \left\{ -\frac{1}{2} [\tilde{\mathbf{y}}_k^T \mathbf{B} \tilde{\mathbf{y}}_k + \mathbf{x}_e^T \mathbf{C} \tilde{\mathbf{x}}_e - \right. \\
& \quad \left. 2\sigma^2 \tilde{\mathbf{y}}_k^T \tilde{\mathbf{A}} \boldsymbol{\Sigma} \mathbf{P} \tilde{\mathbf{x}}_e] \right\},
\end{aligned} \quad (30)$$

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

$$\begin{aligned}
\mathcal{L}(\mathbf{p}, \sigma^2) & = \log \left\{ (2\pi\sigma^2)^{-M} |\boldsymbol{\Sigma}|^{\frac{1}{2}} |\mathbf{P}|^{\frac{1}{2}} \right. \\
& \times \exp \left\{ -\frac{1}{2} [\tilde{\mathbf{y}}_k^T \mathbf{B} \tilde{\mathbf{y}}_k + \mathbf{x}_e^T \mathbf{C} \tilde{\mathbf{x}}_e - \right. \\
& \quad \left. 2\sigma^2 \tilde{\mathbf{y}}_k^T \tilde{\mathbf{A}} \boldsymbol{\Sigma} \mathbf{P} \tilde{\mathbf{x}}_e] \right\} \left. \right\} \\
& = -M \log(2\pi) - M \log \sigma^2 + \frac{1}{2} \log |\boldsymbol{\Sigma}| + \\
& \quad \frac{1}{2} \log |\mathbf{P}| - \frac{1}{2} [\tilde{\mathbf{y}}_k^T \mathbf{B} \tilde{\mathbf{y}}_k + \mathbf{x}_e^T \mathbf{C} \tilde{\mathbf{x}}_e - \\
& \quad 2\sigma^2 \tilde{\mathbf{y}}_k^T \tilde{\mathbf{A}} \boldsymbol{\Sigma} \mathbf{P} \tilde{\mathbf{x}}_e].
\end{aligned} \quad (31)$$

Using the Woodbury matrix inversion identity we have

$$\mathbf{B} = \sigma^{-2} \mathbf{I} - \sigma^{-2} \tilde{\mathbf{A}} (\mathbf{P} + \sigma^{-2} \tilde{\mathbf{A}}^T \tilde{\mathbf{A}})^{-1} \tilde{\mathbf{A}}^T \sigma^{-2}, \quad (32)$$

which means we have

$$\begin{aligned}
\tilde{\mathbf{y}}_k^T \mathbf{B} \tilde{\mathbf{y}}_k & = \tilde{\mathbf{y}}_k^T \sigma^{-2} \tilde{\mathbf{y}}_k - \tilde{\mathbf{y}}_k^T (\sigma^{-2} \mathbf{I} - \sigma^{-2} \tilde{\mathbf{A}} \\
& \times (\mathbf{P} + \sigma^{-2} \tilde{\mathbf{A}}^T \tilde{\mathbf{A}})^{-1} \tilde{\mathbf{A}}^T \sigma^{-2}) \tilde{\mathbf{y}}_k \\
& = \tilde{\mathbf{y}}_k^T \sigma^{-2} \tilde{\mathbf{y}}_k - \tilde{\mathbf{y}}_k^T \sigma^{-2} \tilde{\mathbf{A}} \boldsymbol{\Sigma} \tilde{\mathbf{A}}^T \sigma^{-2} \tilde{\mathbf{y}}_k \\
& = \sigma^{-2} \tilde{\mathbf{y}}_k^T (\tilde{\mathbf{y}}_k - \tilde{\mathbf{A}} \boldsymbol{\mu}) + \sigma^{-2} \tilde{\mathbf{y}}_k^T \tilde{\mathbf{A}} \boldsymbol{\Sigma} \mathbf{P} \tilde{\mathbf{x}}_e \\
& = \sigma^{-2} \|\tilde{\mathbf{y}}_k - \tilde{\mathbf{A}} \boldsymbol{\mu}\|_2^2 + \boldsymbol{\mu}^T \mathbf{P} \boldsymbol{\mu} + \sigma^{-2} \tilde{\mathbf{y}}_k^T \tilde{\mathbf{A}} \boldsymbol{\Sigma} \mathbf{P} \tilde{\mathbf{x}}_e.
\end{aligned} \quad (33)$$

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

$$\begin{aligned}
\mathbf{x}_e^T \mathbf{C} \mathbf{x}_e & = \mathbf{x}_e^T [\mathbf{P} - \mathbf{P} \boldsymbol{\Sigma} \mathbf{P}] \mathbf{x}_e \\
& = \mathbf{x}_e^T \mathbf{P} \mathbf{x}_e - \mathbf{x}_e^T \mathbf{P} \boldsymbol{\Sigma} \mathbf{P} \mathbf{x}_e \\
& = \mathbf{x}_e^T \mathbf{P} \mathbf{x}_e - \mathbf{x}_e^T \mathbf{P} \boldsymbol{\mu} + \sigma^{-2} \tilde{\mathbf{y}}_k^T \tilde{\mathbf{A}} \boldsymbol{\Sigma} \mathbf{P} \tilde{\mathbf{x}}_e,
\end{aligned} \quad (34)$$

which then gives the log likelihood function in (14).

C. Derivation of Update Expressions for Modified RVM

Firstly, differentiating with respect to p_i gives

$$-\frac{1}{2} \left[\Sigma_{nn} - \frac{1}{p_n} + \mu_n^2 + x_{e,n}^2 - x_{e,n} \mu_n \right] \quad (35)$$

and equating to zero gives

$$\begin{aligned}
\Sigma_{nn} - \frac{1}{p_n} + \mu_n^2 + x_{e,n}^2 - x_{e,n} \mu_n & = 0 \quad (36) \\
1 - p_n \Sigma_{nn} - p_n \mu_n^2 - p_n x_{e,n}^2 + p_n x_{e,n} \mu_n & = 0 \\
\gamma_n - p_n [\mu_n^2 + x_{e,n}^2 - x_{e,n} \mu_n] & = 0
\end{aligned}$$

which leads to (15).

Now collect the terms with σ in to give

$$-\frac{1}{2} \left[2M \log \sigma^2 - \log |\Sigma| + \sigma^{-2} \|\tilde{\mathbf{y}}_k - \tilde{\mathbf{A}}\boldsymbol{\mu}\|_2^2 \right] \quad (37)$$

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

$$\begin{aligned} & -\frac{1}{2} \left[2M \log \tau^{-1} - \log |\Sigma| + \tau \|\tilde{\mathbf{y}}_k - \tilde{\mathbf{A}}\boldsymbol{\mu}\|_2^2 \right] \quad (38) \\ & = -\frac{1}{2} \left[-2M \log \tau - \log |\Sigma| + \tau \|\tilde{\mathbf{y}}_k - \tilde{\mathbf{A}}\boldsymbol{\mu}\|_2^2 \right]. \end{aligned}$$

Now differentiate (38) with respect to τ and equate to zero to give

$$-\frac{2M}{\tau} + \text{tr}(\Sigma \tilde{\mathbf{A}}^T \tilde{\mathbf{A}}) + \|\tilde{\mathbf{y}}_k - \tilde{\mathbf{A}}\boldsymbol{\mu}\|_2^2 = 0, \quad (39)$$

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

$$\tau^{-1} (2M - \sum_n \gamma_n) = \|\tilde{\mathbf{y}}_k - \tilde{\mathbf{A}}\boldsymbol{\mu}\|_2^2, \quad (40)$$

which in turn gives (16).

ACKNOWLEDGMENTS

We appreciate the support of the UK Engineering and Physical Sciences Research Council (EPSRC) via the project Bayesian Tracking and Reasoning over Time (BTaRoT) grant EP/K021516/1.

REFERENCES

- [1] 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.
- [2] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [8] 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.
- [9] M. Hyder and K. Mahata, "A robust algorithm for joint-sparse recovery," *IEEE Signal Processing Letters*, vol. 16, no. 12, pp. 1091–1094, 2009.
- [10] Q. Shen, W. Liu, W. Cui, S. Wu, Y. Zhang, and M. Amin, "Group sparsity based wideband DOA estimation for co-prime arrays," in *In Proc. IEEE China Summit International Conference on Signal and Information Processing*, 2014, pp. 252–256.
- [11] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2346–2356, 2008.
- [12] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, pp. 211–244, 2001.
- [13] M. E. Tipping and A. Faul, "Fast marginal likelihood maximisation for sparse Bayesian models," in *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics*, 2003, pp. 3–6.
- [14] 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.
- [15] D. B. Ward, E. Lehmann, and R. Williamson, "Particle filtering algorithms for tracking an acoustic source in a reverberant environment," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 6, pp. 826–836, 2003.
- [16] B. Balakumar, A. Sinha, T. Kirubarajan, and J. Reilly, "PHD filtering for tracking an unknown number of sources using an array of sensors," in *Proc. IEEE Workshop on Statistical Signal Processing*, 2005, pp. 43–48.
- [17] X. Zhong, A. Premkumar, and W. Wang, "Direction of arrival tracking of an underwater acoustic source using particle filtering: Real data experiments," in *In Proc. IEEE TENCON Spring Conference*, 2013, pp. 420–424.
- [18] M. Baum, P. Willett, and U. Hanebeck, "MMOSPA-based direction-of-arrival estimation for planar antenna arrays," in *Proc. IEEE Workshop on Sensor Array and Multichannel Signal Processing*, 2014, pp. 209–212.
- [19] P. Khomchuk and I. Bilik, "Dynamic direction-of-arrival estimation via spatial compressive sensing," in *Proc. IEEE Radar Conference*, 2010, pp. 1191–1196.
- [20] E. Karsenas, 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.
- [21] 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.
- [22] 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.
- [23] 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.