A Bayesian Framework for Blind Adaptive Beamforming

Sarmad Malik, Jacob Benesty, and Jingdong Chen

***Abstract*—In this work, the problem of blind adaptive beamforming in the presence of steering-vector uncertainty is addressed within a Bayesian estimation framework. We express the single-input multiple-output (SIMO) observation model in the short-time-Fourier-transform (STFT) domain and employ a variational formulation to obtain iterative closed-form learning rules for inferring approximate posteriors on the steering vector and the target signal. By varying the *a priori* belief in the top-level statistical model, i.e., modeling a quantity as a random process or an unknown deterministic entity, it is shown that the considered framework yields a variety of beamforming algorithms including the celebrated minimum variance distortionless response (MVDR) beamformer. We highlight these interconnections and show by means of simulation results that the Bayesian approach alleviates signal distortion in noisy and uncertain environments as compared to the conventional MVDR beamformer by adaptively learning and incorporating uncertainty pertaining to the steering vector.**

***Index Terms*—Adaptive beamforming, Bayesian learning, steering-vector uncertainty, variational calculus.**

I. Introduction

SINCE the seminal work of J. Capon [1] regarding the spectral analysis of traveling waves by means of an array of sensors, considerable research and effort has been directed towards the development and analysis of beamforming algorithms. Application of beamformers is ubiquitous in many areas of sensor and array signal processing, e.g., wireless communications [2], speech enhancement [3], [4], source localization [5], etc.

Various forms of fixed [6]–[8] or data-independent beamformers have been considered that include delay-and-sum beamforming [9], superdirective beamformers [10]–[15], differential microphone arrays [16], as well as schemes focused on weight vectors for sidelobe control [17]. The basic motivation behind the development of data-dependent beamforming was to adaptively estimate or select weight vectors subject to relevant constraints predicated upon achieving better resolution and enhancing interference rejection capability [18], [19]. In this regard, the minimum variance distortionless response (MVDR) beamformer (also referred to as the Capon's beamformer [1]) has gained considerable importance as it does not distort the desired signal. However, precise knowledge of the steering vector is required for the distortionless characteristic to hold [20], which is usually not available due to environmental reasons or calibration errors. Works in [21] and [22], imposed additional linear constraints to arrive at adaptive algorithms that favored the signal arriving from the direction of interest, while discriminating against disturbances from all other directions in the presence of possible steering-vector errors.

Over the years, extensive research has been aimed at improving the robustness of beamforming algorithms and here we will try to present an overview of some of the approaches. Proponents in [2] utilized a kurtosis maximization approach, while Bell *et al.* in [23] adopted a Bayesian approach for robust adaptive beamforming. The uncertainty in source direction-of-arrival (DOA) was incorporated in the estimation framework via an *a priori* known probability density function (PDF) on the source DOA. This resulted in a weighted combination of MVDR filters pointed at the most probable set of DOAs. Li *et al.* in [18] showed that accounting for the steering-vector uncertainty in the Capon beamformer amounted to a diagonal loading scheme. In [24], Doclo and Moonen dealt with fixed beamformers and utilized an FIR filter-and-sum structure to put forth broadband design procedures that achieved robustness against gain and phase errors in the array characteristics. A discussion pertaining to the design of robust super-directive beamformers was reported in [15] that employed the statistics of the sensor-array characteristics.

The statistical approach presented in [25] is of vital importance. It assumed some knowledge regarding the stochastic variation of the steering vector and derived maximum-likelihood and Bayesian posterior estimators, depending on the modeling of the target signal as a deterministic quantity or as a random variable, respectively. A computationally efficient state-space beamformer was proposed in [26]. Imposition of the single convex constraint corresponding to the worst-case mismatch in conjunction with the first-order Markov model for the unknown filter weights yielded the constrained Kalman filter, which could estimate time-varying filter weights.

In [27], Chen *et al.* addressed the issue of DOA mismatch by astutely imposing two point quadratic constraint, which rendered the optimization problem solvable in closed form via Karush-Kuhn-Tucker condition. The approach was further augmented with systemic computation of a diagonal loading factor. Gaudes *et al.* in [28] aimed at attaining robustness while maintaining the ability to control side lobes. The conventional linearly constrained minimum variance (LCMV) cost function was supported with an additional regularization constraint to penalize the discrepancy between actual and target array

Manuscript received February 08, 2013; revised September 19, 2013 and December 16, 2013; accepted February 24, 2014. Date of publication March 11, 2014; date of current version April 07, 2014. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Andrzej Cichocki.

S. Malik and J. Benesty are with the National Institute of Scientific Research (INRS-EMT), University of Quebec, Montreal, QC H5A 1K6, Canada (e-mail: sarmad.malik@rub.de;benesty@emt.inrs.ca).

J. Chen is with Northwestern Polytechnical University, Xi'an 710072, China (e-mail: jingdongchen.wevoice@gmail.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2014.2310432

responses. The problem was shown to be convex and solvable by means of a support vector machine. A generalized eigen value decomposition approach was pursued in [29], where a blind acoustic beamformer was proposed. This blind approach utilized a single-channel postfilter to alleviate distortions in the estimated target signal.

The doubly constrained Capon beamformer based on a spherical uncertainty set, originally introduced in [30], was generalized by using an ellipsoidal modeling for the uncertainty in [31]. The generalized approach was shown to be efficiently solvable by means of semi-definite programming. An essential relationship between robust MVDR beamformers was highlighted in the context of probabilistic and worst-case distortionless response constraints in [32], which enabled the probabilistically constrained beamformers to be implemented using their deterministic worst-case counterparts. In a relatively recent correspondence [33], robustness was attained by focusing on the reconstruction of the interference covariance matrix rather than estimating the optimal diagonal loading factor. Moreover, steering-vector estimation was carried out without imposing a norm constraint making the estimate immune to gain perturbations.

Motivated by the works of Attias *et al.* [34], [35], we address the problem of blind adaptive beamforming using a variational Bayesian framework. Bayesian approaches have been shown to be inherently robust against outliers as the inference mechanism relies on the whole probability mass rather than just point estimates. Furthermore, they allow incorporation of *a priori* statistical belief [36]. The SIMO observation model is formulated in the STFT domain and augmented with a first-order Markov model [26], [37] on the steering vectors. Conjugate priors are described over the steering vectors as well as the target signal, leading to closed-form and efficiently implementable posterior estimators via variational optimization, which can be invoked iteratively to tighten the lower-bound on the log-likelihood function. Further, we derive model/covariance parameters by means of parametric optimization. It is shown that the variation of statistical belief in the top-level model can yield a variety of algorithms including the conventional MVDR beamformer, which in fact is a deterministic maximum-likelihood solution as shown in [5]. Owing to the ubiquitous application of signal enhancement by means of beamforming in various domains of signal processing, we evaluate the derived algorithms with respect to array again and target signal distortion, both of which have been shown to directly affect system performance, e.g., accuracy of an automatic speech recognizer [38], output signal-to-noise ratio in satellite communication systems [25], intelligibility of a processed speech signal [39], etc.

We show that our statistical modeling within the variational framework provides a built-in mechanism for adaptive learning of the *effective* steering vector in reverberant and noisy environments, which minimizes target signal distortion. Modeling of the unknown steering vector by means of the first-order Markov model will inevitably yield robust learning rules based on gradient-based adaptation with optimal step-size control. In our approach, incorporation of second order statistics of the estimated signal and observation noise prevents compromising the array gain, while still maintaining fast convergence. We analyze the performance of our formulation in varying degree of stationary and non-stationary sensor noise, and steering-vector uncertainty.

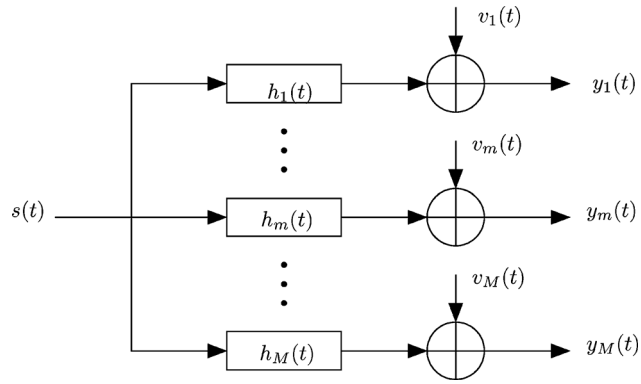

Fig. 1. Time-domain signal model depicting the source signal $s(t)$, room impulse responses $h_m(t)$, observation noise signals $v_m(t)$, and sensor observation signals $y_m(t)$.

The rest of the paper is arranged as follows. In Sections II and III, we present the STFT system model and our *a priori* belief, respectively. Section IV comprises the derivation of the variational Bayesian beamformer along with parameter learning rules. Related algorithms are derived in Section V by modifying top-level statistical modeling. Relevant instrumental measures of performance are outlined in Section VI. Sections VII and VIII present simulation results and conclusions of this work, respectively.

II. Signal Model

Consider a time-domain signal model as shown in Fig. 1, where M sensors capture a convolved source signal in the presence of additive noise. The observation signal $y_m(t)$ at the mth microphone, where $m = 1, \ldots, M$, is expressed as

$$y_m(t) = g_m(t) * s(t) + v_m(t)\,, \tag{1}$$

where $s(t)$ is the source signal, and $g_m(t)$ and $v_m(t)$ are the mth system impulse response and additive observation noise, respectively. Note that t is the sample-time index and $*$ denotes linear convolution. The short-time-Fourier-transform (STFT) representation of (1) is then given as

$$Y_m(k,n) = G_m(k,n)\,S(k,n) + V_m(k,n)\,, \tag{2}$$

such that the uppercase letters denote frequency-domain counterparts of the terms in (1) for the kth frequency bin, whereas n denotes the frame-time index. For notational convenience we drop the frequency index k and re-write the observation (2) as

$$Y_m(n) = G_m(n)\,S(n) + V_m(n)\,. \tag{3}$$

Using (3), we express the M STFT observation signals in vector notation as:

$$\begin{aligned} \mathbf{y}(n) &= \mathbf{g}(n)\,S(n) + \mathbf{v}(n)\,, &(4)\\ &= \mathbf{x}(n) + \mathbf{v}(n)\,, &(5)\\ &= \mathbf{a}(n)\,X_1(n) + \mathbf{v}(n)\,, &(6) \end{aligned}$$

where

$$\mathbf{y}(n) \triangleq [Y_1(n), \ldots, Y_m(n), \ldots, Y_M(n)]^T \,, \tag{7}$$

$$\mathbf{g}(n) \triangleq [G_1(n), \ldots, G_m(n), \ldots, G_M(n)]^T \,, \tag{8}$$

$$\mathbf{x}(n) \triangleq \mathbf{g}(n)\, S(n) \,, \tag{9}$$

$$\mathbf{v}(n) \triangleq [V_1(n), \ldots, V_m(n), \ldots, V_M(n)]^T \,, \tag{10}$$

$$\mathbf{a}(n) \triangleq \left[1, \ldots, \frac{G_m(n)}{G_1(n)}, \ldots, \frac{G_M(n)}{G_1(n)}\right]^T \,, \tag{11}$$

and

$$X_1(n) \triangleq G_1(n)\, S(n) \tag{12}$$

is then the target signal to be estimated. It can be noticed that in (6) $\mathbf{a}(n)$ is the steering vector for noise reduction [40], which we model as a first-order Markov process [26]:

$$\mathbf{a}(n) = A\, \mathbf{a}(n-1) + \mathbf{u}(n) \,, \tag{13}$$

where A and $\mathbf{u}(n)$ are the state-transition coefficient and process noise, respectively. First-order Markov modeling as expressed by (13) will enable us to recursively learn the posterior on the steering vector by means of a gradient-based adaptive estimator with optimal step-size control.

A. Statistical Modeling and a Priori Beliefs

We model the vector $\mathbf{a}(n)$ in (6) and (13) as a normally distributed complex random process with an $M \times M$ predicted error-covariance matrix [41]

$$\begin{aligned}\widehat{\boldsymbol{\Phi}}_{\mathbf{a}}^{+}(n) &= E\left\{ \left[\mathbf{a}(n) - \widehat{\mathbf{a}}^{+}(n)\right] \left[\mathbf{a}(n) - \widehat{\mathbf{a}}^{+}(n)\right]^H \right\} \\ &= A\, \widehat{\boldsymbol{\Phi}}_{\mathbf{a}}(n-1) + \boldsymbol{\Phi}_{\mathbf{u}}(n) \,, \end{aligned} \tag{14}$$

where $\widehat{\boldsymbol{\Phi}}_{\mathbf{a}}(n-1)$ is the $M \times M$ error-covariance matrix at time $n-1$, which is defined as

$$\begin{aligned}&\widehat{\boldsymbol{\Phi}}_{\mathbf{a}}(n-1) \\ &\triangleq E\left\{ [\mathbf{a}(n-1) - \widehat{\mathbf{a}}(n-1)]\, [\mathbf{a}(n-1) - \widehat{\mathbf{a}}(n-1)]^H \right\} \end{aligned} \tag{15}$$

and $\widehat{\mathbf{a}}^{+}(n)$ is the predicted mean:

$$\widehat{\mathbf{a}}^{+}(n) = A\, \widehat{\mathbf{a}}(n-1) \,, \tag{16}$$

where $\widehat{\mathbf{a}}(n-1)$ is the *a posteriori* mean. Note that

$$\boldsymbol{\Phi}_{\mathbf{u}}(n) \triangleq E\left[\mathbf{u}(n)\mathbf{u}^H(n)\right] \tag{17}$$

is the $M \times M$ process noise covariance. Here, H and $(\cdot)^*$ represent Hermitian transposition and complex conjugation, respectively. Similarly, we express the observation noise covariance and target signal variance as

$$\boldsymbol{\Phi}_{\mathbf{v}}(n) \triangleq E\left[\mathbf{v}(n)\mathbf{v}^H(n)\right] \tag{18}$$

and

$$\phi_{X_1}(n) \triangleq E\left[X_1(n) X_1^*(n)\right] \,, \tag{19}$$

respectively. Distributions corresponding to (14)–(19) are given in Appendix A.

Our aim is to estimate the target signal $X_1(n)$ within a Bayesian framework, which due to its modeling as a random variable, amounts to learning of a posterior distribution. In order to incorporate uncertainty regarding the steering vector $\mathbf{a}(n)$ in the estimation process, it is inevitable that we must seek learning rules for a steering-vector posterior as well. It is also essential to realize that any ensuing estimator will rely on the model parameters given by the set

$$\boldsymbol{\Theta}(n) = \{\boldsymbol{\Phi}_{\mathbf{v}}(n), \boldsymbol{\Phi}_{\mathbf{u}}(n), \phi_{X_1}(n)\} \,, \tag{20}$$

which are unknown *a priori* and need to be estimated as well. Thus, our inference tasks are summarized as:

- inference of posterior distributions on $X_1(n)$ and $\mathbf{a}(n)$, and
- estimation of point estimates of the model parameters $\boldsymbol{\Theta}(n)$.

III. The Bayesian Approach: Incorporation of *a* Priori Belief

In order to motivate the selection of a variational Bayesian framework for deriving learning rules for state-space model described by (6) and (13), an overview of different mathematical approaches with respect to optimization criteria and resulting solutions will be presented here.

Consider the target signal $X_1(n)$ in (6) as an unknown deterministic quantity. The steering vector $\mathbf{a}(n)$ for noise reduction is assumed to be known *a priori* and the observation noise vector $\mathbf{v}(n)$ is modeled as a normally distributed random variable according to (113). It is well known that the optimization of the log-likelihood function $\ln p[\mathbf{y}(n) \,|\, X_1(n)]$ with respect to the target signal $X_1(n)$, i.e.,

$$\frac{\partial}{\partial X_1^*(n)} \ln p[\mathbf{y}(n) \,|\, X_1(n)] = 0 \,, \tag{21}$$

results in the maximum-likelihood (ML)/MVDR beamformer [1].

Incorporation of statistical belief regarding $X_1(n)$ according to (114) is possible by changing the objective function to the log-posterior function, which is then given as

$$\ln p[X_1(n) \,|\, \mathbf{y}(n)] = \ln \left\{ \frac{p[\mathbf{y}(n) \,|\, X_1(n)]\, p[X_1(n)]}{p[\mathbf{y}(n)]} \right\} \,. \tag{22}$$

Again, the steering vector $\mathbf{a}(n)$ is considered to be known *a priori* and the Gaussian modeling of $\mathbf{v}(n)$ given by (113) is utilized. Optimization of the log-posterior function with respect to $X_1(n)$, i.e.,

$$\frac{\partial}{\partial X_1^*(n)} \ln p[X_1(n) \,|\, \mathbf{y}(n)] = 0 \,, \tag{23}$$

yields the maximum- *a-posteriori* (MAP)/Wiener estimator.

The reader will appreciate that in both ML and MAP beamformers steering vector is assumed to be known and the target signal is treated as an unknown deterministic quantity. As next logical steps, the aim is to

- estimate the steering vector as well, and
- model the unknown quantities, i.e., the target signal and the steering vector, as random variables rather than unknown and deterministic.

We posit that modeling the quantities of interest, especially the steering vector, as random variables will render the resulting beamformer robust in a reverberant and noisy environment.

In order to infer posteriors on more than one random variable, we revert to the variational Bayesian methodology [36], [42]. The derivation is initiated with the log-likelihood function $\ln p[\mathbf{y}(n)\,|\,\boldsymbol{\Theta}(n)]$ and $X_1(n)$ and $\mathbf{a}(n)$ are inserted using $(M+1)$–fold marginalization

$$\ln p\,[\,\mathbf{y}(n)\,|\,\boldsymbol{\Theta}(n)] = \ln \int p\,[\,\mathbf{y}(n), X_1(n), \mathbf{a}(n)\,|\,\boldsymbol{\Theta}(n)]\; dX_1(n)\,d\mathbf{a}(n)\,. \quad (24)$$

Thereafter, we consider an arbitrary function $q\,[\mathbf{a}(n), X_1(n)]$ of $X_1(n)$ and $\mathbf{a}(n)$, which in fact will behave as the approximate posterior on the random variables of interest. It is has been documented in the literature that the tractability of estimated posteriors in variational estimators can be ensured by utilizing the mean-field approximation [42], i.e.,

$$q\,[X_1(n), \mathbf{a}(n)] \approx q\,[X_1(n)]\;\, q\,[\mathbf{a}(n)]\;. \quad (25)$$

The factorized posterior given by (25) is incorporated in (24) to get

$$\ln p\,[\,\mathbf{y}(n)\,|\,\boldsymbol{\Theta}(n)] = \ln \int \frac{q[X_1(n)]\,q[\mathbf{a}(n)]}{q[X_1(n)]\,q[\mathbf{a}(n)]} \times p[\,\mathbf{y}(n), X_1(n), \mathbf{a}(n)\,|\,\boldsymbol{\Theta}(n)]\,dX_1(n)\,d\mathbf{a}(n). \quad (26)$$

Application of the Jensen's inequality [43] to (26) results in

$$\ln p[\,\mathbf{y}(n)\,|\,\boldsymbol{\Theta}(n)] \geq \int q[X_1(n)]\,q[\mathbf{a}(n)]\,dX_1(n)\,d\mathbf{a}(n) \times \ln \frac{p[\,\mathbf{y}(n), X_1(n), \mathbf{a}(n)\,|\,\boldsymbol{\Theta}(n)]}{q[X_1(n)]q[\mathbf{a}(n)]} \quad (27)$$

$$= \mathcal{J}\,\{q[\mathbf{a}(n)], q[X_1(n)], \boldsymbol{\Theta}(n)\}\;, \quad (28)$$

where $\mathcal{J}\,\{q[\mathbf{a}(n)], q[X_1(n)], \boldsymbol{\Theta}(n)\}$ is the variational lower bound (VLB).

The optimization of the VLB with respect to $q[X_1(n)]$, $q[\mathbf{a}(n)]$, and $\boldsymbol{\Theta}(n)$ can be achieved by means of the expression given in Appendix B.

IV. Variational Bayesian Algorithm

We highlight that the optimization expressions (115) and (116) require the application of functional derivative, i.e., differentiation of a *functional* with respect to a function using the Euler-Lagrange equation [42], [44] (see Appendix C). Using the fundamentals of variational calculus (see Appendix B in [45] for a detailed derivation), we summarily state the respective optimal solutions to (115) and (116) as

$$\ln q^{\star}[X_1(n)] = \kappa_1 + \int \ln\,\{p[\,\mathbf{y}(n), X_1(n), \mathbf{a}(n)\,|\,\boldsymbol{\Theta}(n)]\}\; q^{\star}[\mathbf{a}(n)]\,d\mathbf{a}(n) \quad (29)$$

$$\ln q^{\star}[\mathbf{a}(n)] = \kappa_2 + \int \ln\,\{p[\,\mathbf{y}(n), X_1(n), \mathbf{a}(n)\,|\,\boldsymbol{\Theta}(n)]\}\; q^{\star}[X_1(n)]\,dX_1(n) \quad (30)$$

which have been termed in [36] as the *theorems of variational Bayesian learning*. The constants κ_1 and κ_2 impose the normalization constraints on the resulting distributions. We use Bayes' theorem and factorize the joint distribution $p[\,\mathbf{y}(n), X_1(n), \mathbf{a}(n)\,|\,\boldsymbol{\Theta}(n)]$ in (29) and (30) as

$$p[\mathbf{y}(n), X_1(n), \mathbf{a}(n)\,|\,\boldsymbol{\Theta}(n)] = p[\,\mathbf{y}(n)\,|\,X_1(n), \mathbf{a}(n), \boldsymbol{\Theta}(n)]\;p[X_1(n), \mathbf{a}(n)\,|\,\boldsymbol{\Theta}(n)], \quad (31)$$

$$= p[\,\mathbf{y}(n)\,|\,X_1(n), \mathbf{a}(n), \boldsymbol{\Theta}(n)]\;p[X_1(n)\,|\,\boldsymbol{\Theta}(n)]\;p[\mathbf{a}(n)]. \quad (32)$$

In view of the prior distribution (114) on $X_1(n)$, we can write

$$p[X_1(n)\,|\,\boldsymbol{\Theta}(n)] = p[X_1(n)\,|\,\phi_{X_1}(n)] \quad (33)$$

and

$$p[X_1(n), \mathbf{a}(n),\,|\,\boldsymbol{\Theta}(n)] = p[X_1(n)\,|\,\boldsymbol{\Theta}(n)]\;p[\mathbf{a}(n)] \quad (34)$$

implies the assumption of mutual independence regarding $X_1(n)$ and $\mathbf{a}(n)$. Furthermore, it must be emphasized that the likelihood function $p[\,\mathbf{y}(n)\,|\,X_1(n), \mathbf{a}(n), \boldsymbol{\Theta}(n)]$ is in fact the distribution of the observation noise, i.e.,

$$p[\,\mathbf{y}(n)\,|\,X_1(n), \mathbf{a}(n), \boldsymbol{\Theta}(n)] = p[\mathbf{v}(n)\,|\,\boldsymbol{\Phi}_{\mathbf{v}}(n)] \quad (35)$$

and thus we have

$$p[\mathbf{y}(n)\,|\,X_1(n), \mathbf{a}(n), \boldsymbol{\Theta}(n)] = \frac{1}{\pi^M\,|\boldsymbol{\Phi}_{\mathbf{v}}(n)|} \times \exp\Big\{-[\mathbf{y}(n)-\mathbf{a}(n)X_1(n)]^H \boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)[\mathbf{y}(n)-\mathbf{a}(n)X_1(n)]\Big\}. \quad (36)$$

Note that the likelihood is conditioned on only one of the model parameters, which is the observation noise covariance $\boldsymbol{\Phi}_{\mathbf{v}}(n)$.

A. Target Signal Posterior

In order to derive learning rules for the target signal posterior, first- and second-order functions of $X_1(n)$ on the right-hand side of (29) are isolated. A comparison of these isolated terms with the left-hand side of (29) readily leads to the full posterior estimator.

Using the Bayesian chain rule and substituting (32) into (29) results in

$$\ln q^{\star}[X_1(n)] = \kappa_1 + \int \ln\,\{p[\,\mathbf{y}(n)\,|\,X_1(n), \mathbf{a}(n), \boldsymbol{\Theta}(n)] \times\; p[X_1(n)\,|\,\boldsymbol{\Theta}(n)]\;p[\mathbf{a}(n)]\}\; q^{\star}[\mathbf{a}(n)]\,d\mathbf{a}(n), \quad (37)$$

which simplifies to

$$\ln q^{\star}[X_1(n)] = \kappa_1' + \int \ln\,\{p[\mathbf{y}(n)\,|\,X_1(n), \mathbf{a}(n), \boldsymbol{\Theta}(n)] \times\; p[X_1(n)\,|\,\boldsymbol{\Theta}(n)]\}\; q^{\star}[\mathbf{a}(n)]\,d\mathbf{a}(n)$$

$$= \kappa_1' + E_{q_a^{\star}}\,\{\,\ln\,\{\;p[\mathbf{y}(n)\,|\,X_1(n), \mathbf{a}(n), \boldsymbol{\Theta}(n)] \times\; p[X_1(n)\,|\,\boldsymbol{\Theta}(n)]\}\;\}\;, \quad (38)$$

where the term

$$\kappa_1' = E_{q_a^{\star}}\,\{\,\ln p[\mathbf{a}(n)]\,\} + \kappa_1 \quad (39)$$

comprises all the terms that are not functions of $X_1(n)$ and thus irrelevant for deriving learning rules for the mean and variance of $q[X_1(n)]$, and

$$E_{q_a^\star}(\,\cdot\,) = \int (\,\cdot\,)\, q^\star[\mathbf{a}(n)]\, d\mathbf{a}(n) \tag{40}$$

denotes the expectation with respect to $q^\star[\mathbf{a}(n)]$.

The distribution $p[X_1(n)\,|\,\boldsymbol{\Theta}(n)] = p[X_1(n)|\phi_{X_1}]$, as given by (114), acts as a conjugate prior [42] and enforces a normal form on $q^\star[X_1(n)]$ as well, i.e.,

$$\begin{aligned}\ln q^\star[X_1(n)] = &-\ln[\pi\,\widehat{\phi}_{X_1}(n)]\\ &-\left[X_1^*(n)-\widehat{X}_1^*(n)\right]\widehat{\phi}_{X_1}^{-1}(n)\left[X_1(n)-\widehat{X}_1(n)\right],\end{aligned} \tag{41}$$

where $\widehat{X}_1(n)$ and $\widehat{\phi}_{X_1}(n)$ are the posterior mean and variance of the target signal, respectively. We carry out the following three key steps:

1) substituting (114) and (36) into (38),
2) resolving the first- and second-order expectations corresponding to $E_{q_a^\star}(\cdot) = \int(\cdot)\, q^\star[\mathbf{a}(n)]\, d\mathbf{a}$ in (38) using the identities [41]:

$$E_{q_a^\star}[\,\mathbf{a}(n)\,] \triangleq \widehat{\mathbf{a}}(n)\,, \tag{42}$$

$$E_{q_a^\star}[\,\mathbf{a}(n)\,\mathbf{a}^H(n)\,] \triangleq \widehat{\mathbf{a}}(n)\,\widehat{\mathbf{a}}^H(n) + \widehat{\boldsymbol{\Phi}}_{\mathbf{a}}(n)\,, \tag{43}$$

   where $\widehat{\mathbf{a}}(n)$ and $\widehat{\boldsymbol{\Phi}}_{\mathbf{a}}(n)$ are the estimated posterior mean and covariance of the steering vector, and
3) comparing first- and second-order terms with respect to $X_1(n)$ in (38) and (41) to extract the expressions for posterior mean and variance,

and arrive (see Appendix D) at the learning rules for the target signal posterior. The estimated covariance $\widehat{\phi}_{X_1}(n)$ of the target signal is then given as

$$\widehat{\phi}_{X_1}(n) = \left[\widehat{\mathbf{a}}^H(n)\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\widehat{\mathbf{a}}(n) + \widetilde{\phi}_{X_1}^{-1}(n)\right]^{-1}, \tag{44}$$

where

$$\widetilde{\phi}_{X_1}(n) = \left\{\mathrm{tr}\left[\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\widehat{\boldsymbol{\Phi}}_{\mathbf{a}}(n)\right] + \phi_{X_1}^{-1}(n)\right\}^{-1}, \tag{45}$$

and the mean of the target signal $\widehat{X}_1(n)$ is estimated as

$$\widehat{X}_1(n) = \mathbf{h}_{\mathrm{VB}}^H(n)\mathbf{y}(n)\,, \tag{46}$$

such that the $M\times 1$ variational Bayesian weighting vector $\mathbf{h}_{\mathrm{VB}}(n)$ turns out to be

$$\mathbf{h}_{\mathrm{VB}}(n) = \left[\widehat{\mathbf{a}}^H(n)\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\widehat{\mathbf{a}}(n) + \widetilde{\phi}_{X_1}^{-1}(n)\right]^{-1}\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\widehat{\mathbf{a}}(n)\,. \tag{47}$$

Note that the incorporation of the steering-vector uncertainty in $\mathbf{h}_{\mathrm{VB}}(n)$ is manifested via the inclusion of the steering-vector state-error covariance $\widehat{\boldsymbol{\Psi}}_{\mathbf{a}}(n)$, which is encapsulated in the composite covariance term $\widetilde{\phi}_{X_1}(n)$.

B. Steering Vector Posterior

For deriving the learning rules for the steering vector posterior, first- and second-order functions of $\mathbf{a}(n)$ on the right-hand side of (30) are isolated. A comparison of the isolated terms with the left-hand side of (30) yields the recursive posterior estimator for the steering vector.

We begin by substituting (32) into (30) to get

$$\begin{aligned}\ln q^\star[\mathbf{a}(n)] = \kappa_2 + \int &\ln\{p[\mathbf{y}(n)\,|\,X_1(n),\mathbf{a}(n),\boldsymbol{\Theta}(n)]\\ &\times p[X_1(n)\,|\,\boldsymbol{\Theta}(n)]\,p[\mathbf{a}(n)]\}\,q^\star[X_1(n)]\,dX_1(n)\,,\end{aligned} \tag{48}$$

which can be simplified to

$$\begin{aligned}\ln q^\star[\mathbf{a}(n)] = &\,\kappa_2'\\ &+E_{q_{X_1}^\star}\{\ln\{p[\mathbf{y}(n)\,|\,X_1(n),\mathbf{a}(n),\boldsymbol{\Theta}(n)]\,p[\mathbf{a}(n)]\}\}\,,\end{aligned} \tag{49}$$

where

$$\kappa_2' = E_{q_{X_1}^\star}\{\ln p[X_1(n)|\boldsymbol{\Theta}(n)]\} + \kappa_2 \tag{50}$$

encapsulates terms that are independent of $\mathbf{a}(n)$ and thus irrelevant for deriving learning rules for $q^\star[\mathbf{a}(n)]$, and

$$E_{q_{X_1}^\star}(\,\cdot\,) = \int(\,\cdot\,)q^\star[X_1(n)]\,dX_1(n) \tag{51}$$

denotes the expectation with respect to $q^\star[X_1(n)]$.

Note that $p[\mathbf{a}(n)]$, which in fact is a prediction distribution [41], is acting as a Gaussian prior according to (111). Thus, it will enforce a Gaussian form on the resulting posterior as well. Consequently, $q^\star[\mathbf{a}(n)]$ must be of the form:

$$\begin{aligned}\ln q^\star[\mathbf{a}(n)] = &-\ln\left[\pi^M\,|\widehat{\boldsymbol{\Phi}}_{\mathbf{a}}(n)|\right]\\ &-[\mathbf{a}(n)-\widehat{\mathbf{a}}(n)]^H\,\widehat{\boldsymbol{\Phi}}_{\mathbf{a}}^{-1}(n)\,[\mathbf{a}(n)-\widehat{\mathbf{a}}(n)]\,.\end{aligned} \tag{52}$$

Substitution of (111) and (36) into (49) allows us to write

$$\begin{aligned}&E_{q_{X_1}^\star}\{\ln\{p[\mathbf{y}(n)\,|\,X_1(n),\mathbf{a}(n),\boldsymbol{\Theta}(n)]\,p[\mathbf{a}(n)]\}\}\\ &=E_{q_{X_1}^\star}\Big\{-[\mathbf{y}(n)-\mathbf{a}(n)X_1(n)]^H\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)[\mathbf{y}(n)-\mathbf{a}(n)X_1(n)]\\ &\quad-\left[\mathbf{a}(n)-\widehat{\mathbf{a}}^+(n)\right]^H\widehat{\boldsymbol{\Phi}}_{\mathbf{a}}^{+^{-1}}(n)\left[\mathbf{a}(n)-\widehat{\mathbf{a}}^+(n)\right]\Big\}+\kappa_4,\end{aligned} \tag{53}$$

where

$$\kappa_4 = E_{q_{X_1}^\star}\left\{-\ln\left[\pi^M\,\left(|\boldsymbol{\Phi}_{\mathbf{v}}(n)|+|\widehat{\boldsymbol{\Phi}}_{\mathbf{a}}^+(n)|\right)\right]\right\} \tag{54}$$

consists of terms that are not functions of $\mathbf{a}(n)$ and hence dispensable for the ensuing *completion of squares*. Note that analogously to (42) and (43), the expectations with respect to $q^\star[X_1(n)]$ can be resolved using the identities [41]:

$$E_{q_{X_1}^\star}[\,X_1(n)\,] \triangleq \widehat{X}_1(n)\,, \tag{55}$$

$$E_{q_{X_1}^\star}[\,X_1(n)\,X_1^*(n)\,] \triangleq \widehat{X}_1(n)\,\widehat{X}_1^*(n) + \widehat{\phi}_{X_1}(n)\,. \tag{56}$$

After resolving expectations using (55) and (56), we compare first- and second-order terms in $\mathbf{a}(n)$ in (53) to express the learning rules for the steering vector posterior mean and error-covariance as (see Appendix E)

$$\widehat{\mathbf{a}}(n) = \widehat{\boldsymbol{\Phi}}_{\mathbf{a}}(n)\left[\widehat{X}_1^*(n)\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\mathbf{y}(n) + \widehat{\boldsymbol{\Phi}}_{\mathbf{a}}^{+^{-1}}(n)\widehat{\mathbf{a}}^+(n)\right], \tag{57}$$

and

$$\widehat{\boldsymbol{\Phi}}_{\mathbf{a}}(n) = \widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n) - \mathbf{K}(n)\widehat{X}_1(n)\widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n)\,, \tag{58}$$

respectively, where

$$\widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n) = \left[\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\widehat{\phi}_{X_1}(n) + \widehat{\boldsymbol{\Phi}}_{\mathbf{a}}^{+^{-1}}(n)\right]^{-1} \tag{59}$$

is the modified prior error-covariance and

$$\mathbf{K}(n) = \widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n)\widehat{X}_1^*(n)\left[\widehat{X}_1(n)\widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n)\widehat{X}_1^*(n) + \boldsymbol{\Phi}_{\mathbf{v}}(n)\right]^{-1} \tag{60}$$

is the Kalman gain. Substitution of (58) and (60) into (57) followed by rearrangements allow us to express the learning of the posterior mean $\widehat{\mathbf{a}}(n)$ in a form resembling gradient-based adaptation (see Appendix F):

$$\begin{aligned}\widehat{\mathbf{a}}(n) &= \left[\widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n) - \mathbf{K}(n)\widehat{X}_1(n)\widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n)\right] \\ &\quad\times\left[\widehat{X}_1^*(n)\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\mathbf{y}(n) + \widehat{\boldsymbol{\Phi}}_{\mathbf{a}}^{+^{-1}}(n)\widehat{\mathbf{a}}^+(n)\right], \quad (61)\\ &= \boldsymbol{\Psi}(n)\,\widehat{\mathbf{a}}^+(n) + \boldsymbol{\Lambda}(n)\widehat{X}_1^*(n)\mathbf{e}(n), \quad (62)\end{aligned}$$

where

$$\boldsymbol{\Psi}(n) = \widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n)\widehat{\boldsymbol{\Phi}}_{\mathbf{a}}^{+^{-1}}(n) \tag{63}$$

is a *leakage matrix*,

$$\boldsymbol{\Lambda}(n) = \widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n)\left[\widehat{X}_1(n)\widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n)\widehat{X}_1^*(n) + \boldsymbol{\Phi}_{\mathbf{v}}(n)\right]^{-1} \tag{64}$$

is the Kalman step size, and

$$\mathbf{e}(n) = \mathbf{y}(n) - \boldsymbol{\Psi}(n)\,\widehat{\mathbf{a}}^+(n)\widehat{X}_1(n) \tag{65}$$

represents the error signal.

The recursive estimator for the steering vector posterior can thus be summarized as:

$$\widehat{\mathbf{a}}^+(n) = A\,\widehat{\mathbf{a}}(n-1), \tag{66a}$$

$$\widehat{\boldsymbol{\Phi}}_{\mathbf{a}}^+(n) = A^2\,\widehat{\boldsymbol{\Phi}}_{\mathbf{a}}(n-1) + \boldsymbol{\Phi}_{\mathbf{u}}(n), \tag{66b}$$

$$\widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n) = \left[\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\widehat{\phi}_{X_1}(n) + \widehat{\boldsymbol{\Phi}}_{\mathbf{a}}^{+^{-1}}(n)\right]^{-1}, \tag{66c}$$

$$\boldsymbol{\Psi}(n) = \widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n)\widehat{\boldsymbol{\Phi}}_{\mathbf{a}}^{+^{-1}}(n), \tag{66d}$$

$$\boldsymbol{\Lambda}(n) = \widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n)\left[\widehat{X}_1(n)\widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n)\widehat{X}_1^*(n) + \boldsymbol{\Phi}_{\mathbf{v}}(n)\right]^{-1}, \tag{66e}$$

$$\mathbf{e}(n) = \mathbf{y}(n) - \boldsymbol{\Psi}(n)\,\widehat{\mathbf{a}}^+(n)\widehat{X}_1(n), \tag{66f}$$

$$\widehat{\mathbf{a}}(n) = \boldsymbol{\Psi}(n)\,\widehat{\mathbf{a}}^+ + \boldsymbol{\Lambda}(n)\widehat{X}_1^*(n)\mathbf{e}(n), \tag{66g}$$

$$\widehat{\boldsymbol{\Phi}}_{\mathbf{a}}(n) = \widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n) - \boldsymbol{\Lambda}(n)\widehat{X}_1^*(n)\widehat{X}_1(n)\widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n). \tag{66h}$$

It is interesting to observe that the recursive state-space estimator given in (66) utilizes the target signal posterior, i.e., posterior mean and variance of $X_1(n)$, which is estimated using the VB weighting vector $\mathbf{h}_{\mathrm{VB}}$ of (47), and vice versa. This highlights the exchange of the estimated first- and second-order moments among the two posterior estimators as they are iteratively executed to maximize the lowerbound on the log-likelihood function.

C. Parameter Learning Rules

In order to learn the model parameters $\boldsymbol{\Theta}(n)$, optimization equations involving the VLB are to be solved for the respective model parameters (see Appendix G). From (113), (27), and (36) it can be seen that the optimization of the VLB with respect to $\boldsymbol{\Phi}_{\mathbf{v}}(n)$ equates to the optimization of the log-likelihood function (36) subject to the expectations with respect to $q^\star[X_1(n)]$ and $q^\star[\mathbf{a}(n)]$. Thus, (154) can be simplified to

$$E_{q^\star_{X_1},q^\star_{\mathbf{a}}}\left\{\frac{\partial}{\partial\boldsymbol{\Phi}_{\mathbf{v}}(n)}p[\mathbf{y}(n)\,|\,X_1(n),\mathbf{a}(n),\boldsymbol{\Theta}(n)]\right\} = \mathbf{0}_M\,. \tag{67}$$

Note that due to the mutual independence assumption regarding $X_1(n)$ and $\mathbf{a}(n)$, the expectation operator is rendered factorizable:

$$E_{q^\star_{X_1},q^\star_{\mathbf{a}}}\{\cdot\} = E_{q^\star_{X_1}}\{\cdot\}\;E_{q^\star_{\mathbf{a}}}\{\cdot\}\,. \tag{68}$$

After applying relevant matrix calculus identities [46] to (67), we obtain the ML optimal estimate of the observation noise covariance as

$$\begin{aligned}&\bar{\boldsymbol{\Phi}}_{\mathbf{v}}(n)\\ &= E_{q^\star_{X_1},q^\star_{\mathbf{a}}}\left\{[\mathbf{y}(n) - \mathbf{a}(n)X_1(n)]\,[\mathbf{y}(n) - \mathbf{a}(n)X_1(n)]^H\right\}. \quad (69)\end{aligned}$$

Algebraic rearrangements and (42), (43), (55), and (56) can be invoked to resolve expectations in (69) to get

$$\begin{aligned}\bar{\boldsymbol{\Phi}}_{\mathbf{v}}(n) &= \mathbf{y}(n)\mathbf{y}^H(n) - \mathbf{y}(n)\widehat{\mathbf{a}}^H(n)\widehat{X}_1^*(n) - \widehat{\mathbf{a}}(n)\widehat{X}_1(n)\mathbf{y}^H(n)\\ &\quad + \left[\widehat{\mathbf{a}}(n)\widehat{\mathbf{a}}^H(n) + \widehat{\boldsymbol{\Phi}}_{\mathbf{a}}(n)\right]\left[\widehat{X}_1(n)\widehat{X}_1^*(n) + \widehat{\phi}_{X_1}(n)\right]. \quad (70)\end{aligned}$$

Utilizing the following two definitions:

$$\bar{\mathbf{e}}(n) \triangleq \mathbf{y}(n) - \widehat{\mathbf{a}}(n)\widehat{X}_1(n) \tag{71}$$

$$\begin{aligned}\boldsymbol{\Omega}(n) &\triangleq \widehat{\mathbf{a}}(n)\widehat{\mathbf{a}}^H(n)\widehat{\phi}_{X_1}(n) + \widehat{\boldsymbol{\Phi}}_{\mathbf{a}}(n)\widehat{X}_1(n)\widehat{X}_1^*(n)\\ &\quad + \widehat{\boldsymbol{\Phi}}_{\mathbf{a}}(n)\widehat{\phi}_{X_1}(n), \quad (72)\end{aligned}$$

we re-arrange (70) to express the learning rule for the observation noise covariance as:

$$\bar{\boldsymbol{\Phi}}_{\mathbf{v}}(n) = \bar{\mathbf{e}}(n)\,\bar{\mathbf{e}}^H(n) + \boldsymbol{\Omega}(n). \tag{73}$$

As the characteristics of the first-order Markov model have remained essentially unaltered, for brevity we refer to [47] for the learning rule pertaining to $\boldsymbol{\Phi}_{\mathbf{u}}(n)$.

In analogy to (67), we see that (156) effectively reduces to the optimization of the distribution that carries $\phi_{X_1}(n)$, which is the prior on $X_1(n)$. Thus, we write the optimization task as

$$E_{q^\star_{X_1}}\left\{\frac{\partial}{\partial\phi_{X_1}(n)}p[X_1(n)\,|\,\phi_{X_1}(n)]\right\} = 0\,. \tag{74}$$

Substitution of (114) into (74) followed by application of the derivative and resolution of expectations using (55) and (56), results in the learning rule for the target-signal variance as

$$\bar{\phi}_{X_1}(n) = \widehat{X}_1(n)\widehat{X}_1^*(n) + \widehat{\phi}_{X_1}(n)\,. \tag{75}$$

V. Effect of Elimination of Uncertainty on Posterior Estimators

In this section, we analyze the effect of reducing uncertainty via selection of the Dirac delta function as an alternative for prior distributions.

A. Considering the Target Signal as a Random Variable and the Steering Vector to be Known

We begin by setting the prior on $\mathbf{a}(n)$ to

$$p[\mathbf{a}(n)] = \delta\left[\mathbf{a}(n) - \mathbf{a}_0\right] , \tag{76}$$

where $\mathbf{a}_0$ indicates the known deterministic value of $\mathbf{a}(n)$. As there is no uncertainty regarding the statistics of $\mathbf{a}(n)$, the optimal posterior is the same as the prior, i.e.,

$$q^{\star}[\mathbf{a}(n)] = p[\mathbf{a}(n)] , \tag{77}$$

which when applied to (42) and (43) leads to:

$$E_{q^{\star}_{\mathbf{a}}}\left[\,\mathbf{a}(n)\right] = E_{p[\mathbf{a}(n)]}\left[\,\mathbf{a}(n)\right] = \mathbf{a}_0 , \tag{78}$$

$$E_{q^{\star}_{\mathbf{a}}}\left[\mathbf{a}(n)\mathbf{a}^H(n)\right] = E_{p[\mathbf{a}(n)]}\left[\,\mathbf{a}(n)\mathbf{a}^H(n)\right] = \mathbf{a}_0\mathbf{a}_0^H . \tag{79}$$

This implies that the VLB is then optimized under the limit

$$\frac{\partial}{\partial q[X_1(n)]}\mathcal{J}\left\{q[X_1(n)], q^{\star}[\mathbf{a}(n)], \boldsymbol{\Theta}(n)\right\}\Big|_{\widehat{\boldsymbol{\Phi}}^{+}_{\mathbf{a}}(n)\to 0} = 0 . \tag{80}$$

Using (80), the VB weighting vector of (47) reduces to the well known maximum- *a-posteriori* (MAP)/Wiener estimator:

$$\mathbf{h}_{\mathrm{MAP}}(n) = \left[\mathbf{a}_0^H\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\mathbf{a}_0 + \phi_{X_1}^{-1}(n)\right]^{-1}\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\mathbf{a}_0 . \tag{81}$$

B. Considering the Target Signal as Unknown and Deterministic Quantity and the Steering Vector as a Random Variable

We set the prior on the target signal as

$$p[X_1(n)] = \delta\left[X_1(n) - \widehat{X}_1(n)\right] , \tag{82}$$

where $\widehat{X}_1(n)$ is the deterministic estimate of the target signal. Utilization of (82) results in the modification of (45) as

$$\widetilde{\phi}_{X_1}(n) = \mathrm{tr}\left[\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\widehat{\boldsymbol{\Phi}}_{\mathbf{a}}(n)\right] , \tag{83}$$

corresponding to optimization of the VLB subject to the limit $\phi_{X_1}(n) \to 0$, i.e.,

$$\frac{\partial}{\partial q[X_1(n)]}\mathcal{J}\left\{q[X_1(n)], q^{\star}[\mathbf{a}(n)], \boldsymbol{\Theta}(n)\right\}\Big|_{\phi_{X_1}\to 0} = 0 . \tag{84}$$

Thus using (83) and (84), we can re-write the VB estimator of (47) as a maximum-likelihood (ML) weighting vector with steering-vector uncertainty (MLSU):

$$\mathbf{h}_{\mathrm{MLSU}}(n) = \left\{\widehat{\mathbf{a}}^H(n)\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\widehat{\mathbf{a}}(n) + \mathrm{tr}\left[\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\widehat{\boldsymbol{\Phi}}_{\mathbf{a}}(n)\right]\right\}^{-1}\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\widehat{\mathbf{a}}(n). \tag{85}$$

For (84), it is easy to see that the steering vector posterior estimator in (66) simplifies to

$$\widehat{\mathbf{a}}^{+}(n) = A\,\widehat{\mathbf{a}}(n-1) , \tag{86a}$$

$$\widehat{\boldsymbol{\Phi}}^{+}_{\mathbf{a}}(n) = A^2\,\widehat{\boldsymbol{\Phi}}_{\mathbf{a}}(n-1) + \boldsymbol{\Phi}_{\mathbf{u}}(n) , \tag{86b}$$

$$\widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n) = \left[\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)[\widehat{\phi}_{X_1}(n)\to 0] + \widehat{\boldsymbol{\Phi}}_{\mathbf{a}}^{+^{-1}}(n)\right]^{-1} = \widehat{\boldsymbol{\Phi}}^{+}_{\mathbf{a}}(n) , \tag{86c}$$

$$\boldsymbol{\Psi}(n) = \mathbf{I}_M , \tag{86d}$$

$$\boldsymbol{\Lambda}(n) = \widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n)\left[\widehat{X}_1(n)\widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n)\widehat{X}_1^{*}(n) + \boldsymbol{\Phi}_{\mathbf{v}}(n)\right]^{-1} , \tag{86e}$$

$$\mathbf{e}(n) = \mathbf{y}(n) - [\boldsymbol{\Psi}(n) \to \mathbf{I}_M]\,\widehat{\mathbf{a}}^{+}(n)\widehat{X}_1(n) , \tag{86f}$$

$$\widehat{\mathbf{a}}(n) = [\boldsymbol{\Psi}(n) \to \mathbf{I}_M]\,\widehat{\mathbf{a}}^{+} + \boldsymbol{\Lambda}(n)\widehat{X}_1^{*}(n)\mathbf{e}(n) , \tag{86g}$$

$$\widehat{\boldsymbol{\Phi}}_{\mathbf{a}}(n) = \widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n) - \boldsymbol{\Lambda}(n)\widehat{X}_1^{*}(n)\widehat{X}_1(n)\widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n) , \tag{86h}$$

where the leakage matrix $\boldsymbol{\Psi}(n)$ is now an identity $\mathbf{I}_M$ as highlighted by (86d). The posterior estimator in (86) corresponds to the optimization of the VLB under the limit:

$$\frac{\partial}{\partial q[\mathbf{a}(n)]}\mathcal{J}\left\{q^{\star}[X_1(n)], q[\mathbf{a}(n)], \boldsymbol{\Theta}(n)\right\}\Big|_{\phi_{X_1}\to 0} = 0 , \tag{87}$$

C. Considering the Target Signal to be Unknown and Deterministic Quantity and the Steering Vector to be Known

We set both the target signal as well as steering vector priors as

$$p[X_1(n)] = \delta\left[X_1(n) - \widehat{X}_1(n)\right] , \tag{88}$$

$$p[\mathbf{a}(n)] = \delta\left[\mathbf{a}(n) - \mathbf{a}_0\right] , \tag{89}$$

which implies that the VLB will be optimized with respect to the limits

$$\frac{\partial}{\partial q[X_1(n)]}\mathcal{J}\left\{q[X_1(n)], q^{\star}[\mathbf{a}(n)], \boldsymbol{\Theta}(n)\right\}\Big|_{\substack{\phi_{X_1}(n)\to 0 \\ \widehat{\boldsymbol{\Phi}}^{+}_{\mathbf{a}}(n)\to 0}} = 0 \tag{90}$$

causing the VB estimator in (47) to transform to the well known ML/MVDR estimator:

$$\mathbf{h}_{\mathrm{ML}}(n) = \left[\mathbf{a}_0^H\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\mathbf{a}_0\right]^{-1}\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\,\mathbf{a}_0 . \tag{91}$$

The interconnections of the VB, MLSU, MAP, and ML algorithms within the Bayesian framework are summarized in Fig. 2.

VI. Performance Measures

This section outlines instrumental measures of performance, which will be used to assess the performance of beamforming algorithms. Here, we will reintroduce the STFT frequency index k for clarity.

We consider a generic estimator $\mathbf{h}(k,n)$ and the observation model (6) to express the output of the beamformer as

$$\begin{aligned} Z(k,n) &= \mathbf{h}^H(k,n)\left[\mathbf{a}(k,n)X_1(k,n) + \mathbf{v}(k,n)\right] \\ &= X_{\mathrm{fd}}(k,n) + V_{\mathrm{rn}}(k,n) , \end{aligned} \tag{92}$$

where

$$X_{\mathrm{fd}}(k,n) = \mathbf{h}^H(k,n)\mathbf{a}(k,n)X_1(k,n) \tag{93}$$

is the filtered desired/target signal and

$$V_{\mathrm{rn}}(k,n) = \mathbf{h}^H(k,n)\mathbf{v}(k,n) \tag{94}$$

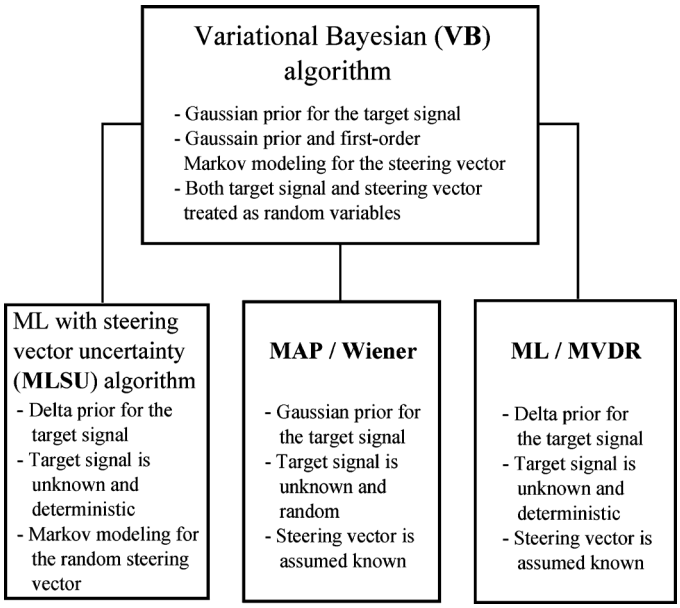

Fig. 2. Bayesian framework for deriving adaptive beamforming algorithms.

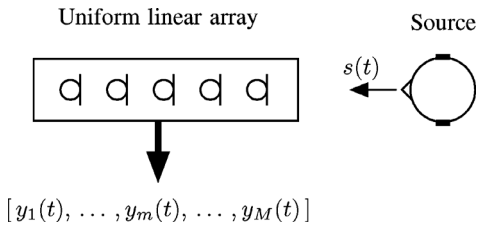

Fig. 3. Schematic diagram showing the orientation of the microphone array and the location of the sound source.

represents residual noise. In view of (93), it is possible to express the distortion in the filtered desired signal as

$$D_{\rm fd}(k,n) = X_{\rm fd}(k,n) - X_1(k,n)\,. \tag{95}$$

As the desired signal and the observation noise are modeled as uncorrelated processes, the variance $\phi_Z(k,n)$ of $Z(k,n)$ is given by [48]

$$\phi_Z(k,n) \triangleq E\left[Z(k,n)Z^*(k,n)\right] = \phi_{X_{\rm fd}}(k,n) + \phi_{V_{\rm rn}}(k,n) \tag{96}$$

where:

$$\phi_{X_{\rm fd}}(k,n) \triangleq E\left[X_{\rm fd}(k,n)X^*_{\rm fd}(k,n)\right] \tag{97}$$

$$\phi_{V_{\rm rn}}(k,n) \triangleq E\left[V_{\rm rn}(k,n)V^*_{\rm rn}(k,n)\right]\,. \tag{98}$$

Here, we define further three variances, i.e., the target signal variance

$$\phi_{X_1}(k,n) \triangleq E\left[X_1(k,n)X_1^*(k,n)\right]\,, \tag{99}$$

the observation noise variance at the first/reference sensor

$$\phi_{V_1}(k,n) \triangleq E\left[V_1(k,n)V_1^*(k,n)\right]\,, \tag{100}$$

and the signal-distortion variance

$$\phi_{D_{\rm fd}}(k,n) \triangleq E\left[D_{\rm fd}(k,n)D^*_{\rm fd}(k,n)\right]\,. \tag{101}$$

With the definitions given in (97)–(100), the frame-wise fullband input and output signal-to-noise ratios, i.e., iSNR and oSNR, can be expressed as

$$\mathrm{iSNR}(n) = \frac{\sum\limits_k \phi_{X_1}(k,n)}{\sum\limits_k \phi_{V_1}(k,n)} \tag{102}$$

and

$$\mathrm{oSNR}(n) = \frac{\sum\limits_k \phi_{X_{\rm fd}}(k,n)}{\sum\limits_k \phi_{V_{\rm rn}}(k,n)}\,, \tag{103}$$

respectively, where $\sum_k$ implies summation over all STFT bins. Expressions given in (102) and (103), enable us to write the array/SNR again as

$$\mathcal{A}_{\rm SNR}(n) = \frac{\mathrm{oSNR}(n)}{\mathrm{iSNR}(n)}\,. \tag{104}$$

Now, we consider (99) and (101), and define the fullband target signal distortion index as

$$\beta(n) = \frac{\sum\limits_k \phi_{D_{\rm fd}}(k,n)}{\sum\limits_k \phi_{X_1}(k,n)}\,. \tag{105}$$

In the ensuing results section, we will analyze the performance of the considered algorithms in terms of $\beta(n)$ and $\mathcal{A}_{\rm SNR}(n)$.

VII. Simulation Results

In our simulations, we considered a uniform linear array (ULA) [4] with five sensors and an inter-sensor distance of $d = 1$ cm situated in an enclosure of size 5 m $\times$ 4 m $\times$ 6 m ($x \times y \times z$). A source was placed end-fire to the ULA at 3.5 m $\times$ 2 m $\times$ 1.5 m ($x \times z \times y$), which was located 1 m away from the nearest sensor located at 2.5 m $\times$ 2 m $\times$ 1.5 m. The geometrical orientation of the ULA and the location of the sound source are shown in Fig. 3. A sampling frequency of $f_{\rm s} = 8$ kHz was selected with wave propagation velocity $c = 343$ m/s.

In order to generate the sensor signals, impulse responses were generated using the modified image method [49], [50] for reverberation times corresponding to $\mathrm{T}_{60} = 0.1, 0.2, 0.3, 0.4,$ and 0.5 s, each of duration 512 ms. Impulse responses corresponding to the nearest sensor for $\mathrm{T}_{60} = 0.1$ s and 0.5 s are shown in Fig. 4(a) and (b), respectively. Data generation via reverberant room impulse responses is a convenient way of introducing perturbations in the end-fire steering vector

$$\mathbf{a}_0(k) = \left[\,1 \quad \mathrm{e}^{-j\Delta_k} \quad \mathrm{e}^{-j2\Delta_k} \quad \mathrm{e}^{-j3\Delta_k} \quad \mathrm{e}^{-j4\Delta_k}\,\right]^T\,, \tag{106}$$

where

$$\Delta_k = 2\pi \frac{k}{N} f_{\rm s} \frac{d}{c} \tag{107}$$

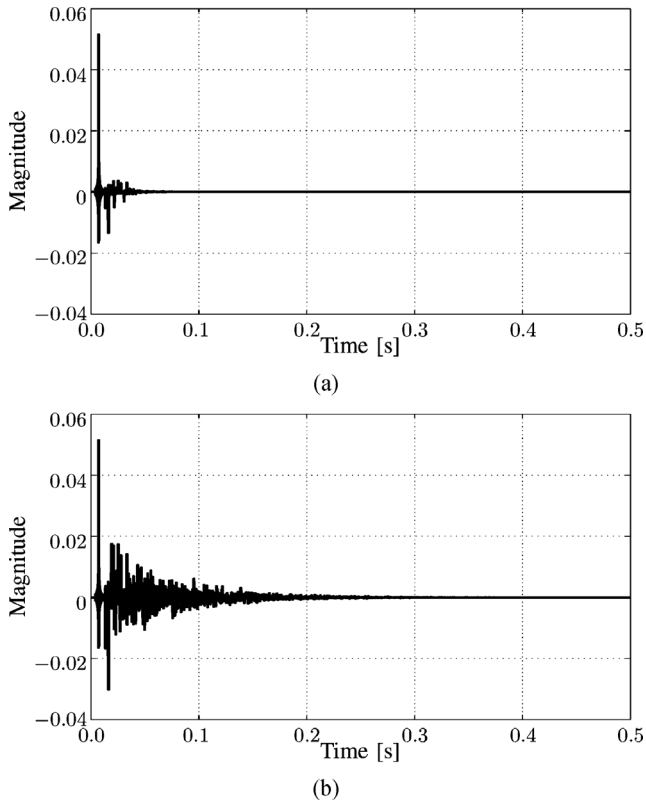

Fig. 4. Room impulse responses generated using the image method [49] corresponding to the reference microphone for the $\mathrm{T}_{60} = 0.1$ s and 0.5 s, respectively. (a) Room impulse response for $\mathrm{T}_{60} = 0.1$ s. (b) Room impulse response for $\mathrm{T}_{60} = 0.5$ s.

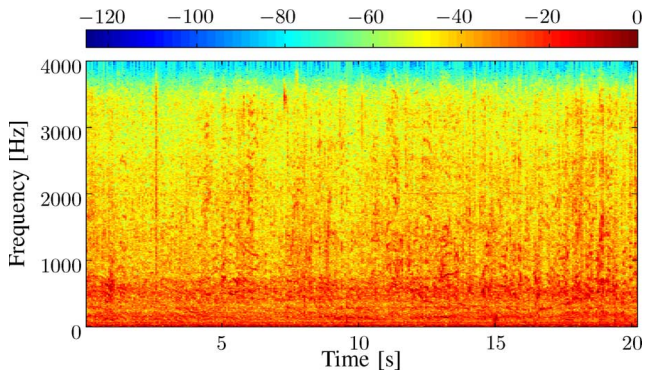

Fig. 5. Periodogram of the considered babble noise signal.

and N is the frame length. An acoustic source comprising of 20 concatenated TIMIT database sentences was convolved with the respective impulse responses and corrupted with sensor noise to yield the observation signals of length 260 s. Although experiments were conducted for a range of iSNR, representative results at $\mathrm{iSNR} = 15$ and 5 dB are presented for brevity. In order to address relevant practical cases, simulations were carried out with white Gaussian as well as babble observation noise. In Fig. 5, the periodogram of the considered babble noise signal is shown.

Processing of the observation/sensor signals was carried out in the STFT domain using the overlap-add scheme [51] with a window length of 128 ms and an overlap of 75%. The state-transition coefficient for the steering-vector posterior estimator was set to 0.9997. Here, we also highlight that the (co)variance parameter estimates $\bar{\boldsymbol{\Phi}}_{\mathbf{v}}(n)$ and $\bar{\phi}_{X_1}(n)$ in (73) and (75), respectively, and $\bar{\boldsymbol{\Phi}}_{\mathbf{u}}(n)$ in [47], are instantaneous estimates. For suitable performance of the algorithms these parameters were temporally smoothed using a first-order recursion with $\gamma = 0.9$ as the smoothing constant.

In Sections IV and V, we have derived variational Bayesian (VB), maximum *a posteriori* (MAP), maximum likelihood with steering-vector uncertainty (MLSU), and the maximum-likelihood (ML/MVDR) estimators. Of these four related algorithms, the VB and the MLSU discern themselves by incorporating the steering-vector uncertainty in the estimation process. Thus, in the following sections we will analyze the performance of VB and MLSU estimators in comparison to the ML/MVDR solution in noisy and reverberant environment. It should be noted that in order to contain computational complexity and carry out online processing we execute only one iteration per frame for the VB and MLSU algorithms.

Although other sophisticated beamforming algorithms are recognized and appreciated, e.g., [15], [24], [28], etc., the selection of the non-blind MVDR beamformer as the reference approach is principally motivated by two aspects. First, as a result of our modeling and derivation we have shown that the MVDR solution is a particular instance of the generic Bayesian beamformer derived by imposing certain statistical simplifications within the parent variational framework. Therefore, it is only logical then to compare the MVDR solution with versions that incorporate additional *a priori* belief to demonstrate direct advantages of modeling quantities of interest as random variables. Second, the MVDR beamformer is a well known and recognized algorithm which has been repeatedly used by researchers [25], [30], [52] for objective and tangible evaluation in domains ranging from digital communications to room acoustics.

A. Target Signal Distortion

In Figs. 6(a) and 7(a), we analyze the performance of three contending configurations at $\mathrm{iSNR} = 15$ dB for varying T_{60} and for two different observation noise types. It can be noticed that despite known direction-of-arrival and array geometry, the ML solution suffers from target signal distortion in the look direction due to room reverberation. As expected, it can be seen that signal distortion increases with T_{60} for all configurations. The VB and MLSU solutions iteratively learn the steering-vector posterior and incorporate the related uncertainty, which enables them to achieve lower distortion as compared to the ML configuration. The MLSU estimator consistently achieves 5 dB improvement over the ML scheme for all reverberation times and lowerbounds the VB algorithm as well. We observe a similar trend in Figs. 6(b) and 7(b), where target signal distortion was studied at $\mathrm{iSNR} = 5$ dB. Though the target signal distortion has increased as compared to Figs. 6(a) and 7(a), the MLSU and VB estimators consistently outperform the ML solution for larger reverberation times.

It is interesting to see that the VB algorithm, though it lowerbounds the ML approach, exhibits more distortion than the MLSU estimator. The understanding of this phenomenon lies in the comparison of beamformer weights $\mathbf{h}_{\mathrm{VB}}(n)$ and $\mathbf{h}_{\mathrm{MLSU}}(n)$ as expressed in (47) and (85), respectively. Both

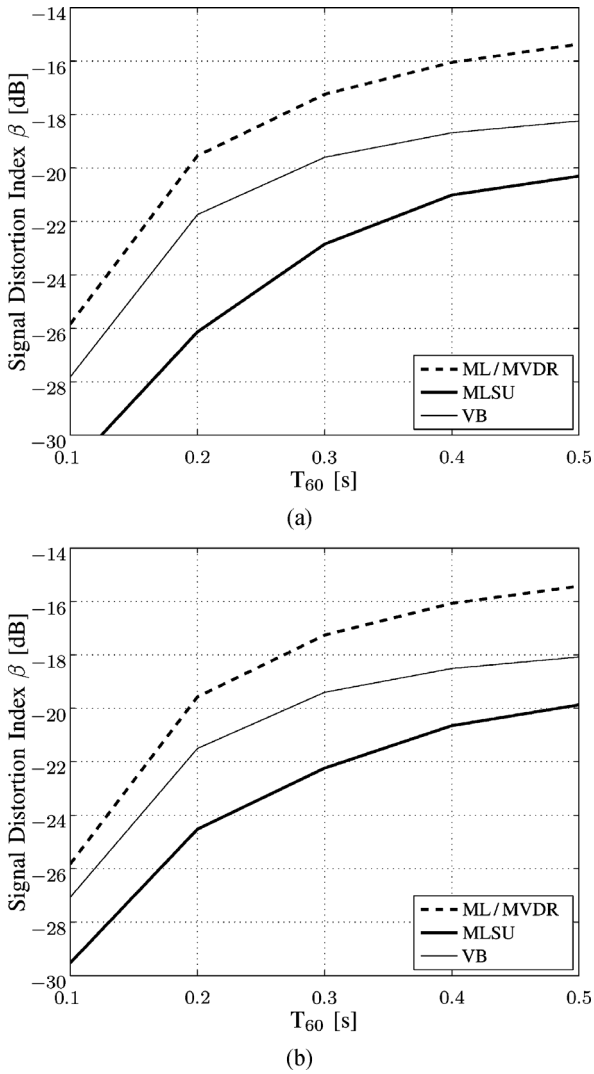

Fig. 6. Performance comparison in the presence of white observation noise for different T_{60}. (a) Signal distortion at $\mathrm{iSNR} = 15$ dB. (b) Signal distortion at $\mathrm{iSNR} = 5$ dB.

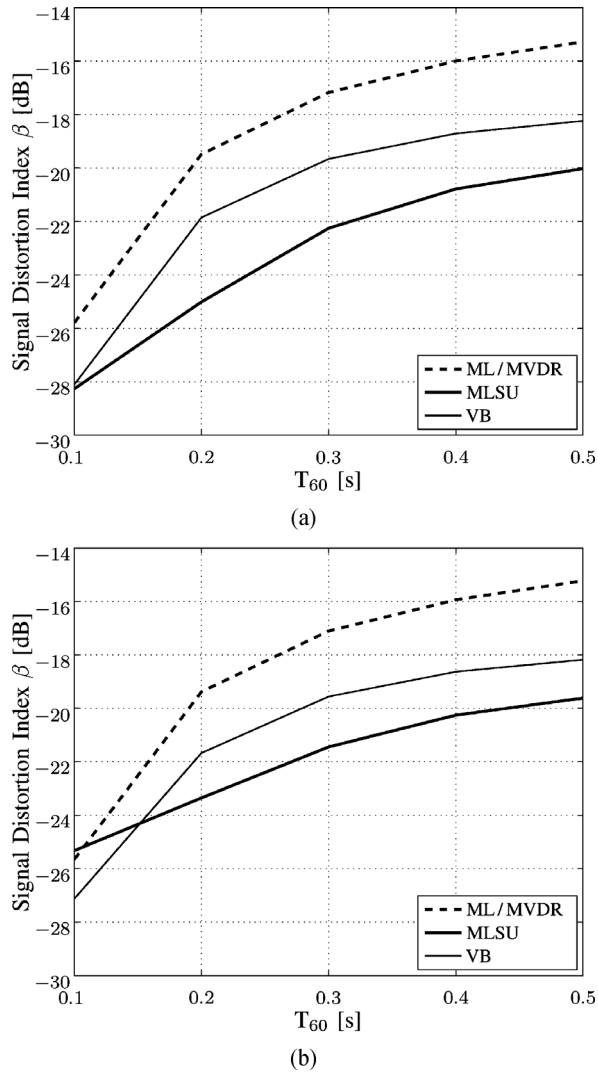

Fig. 7. Performance comparison in the presence of babble observation noise for different T_{60}. (a) Signal distortion at $\mathrm{iSNR} = 15$ dB. (b) Signal distortion at $\mathrm{iSNR} = 5$ dB.

of these weights are functions of the posteriors on the steering vector $\mathbf{a}(n)$ that are learned by the state-space estimators given by (66) and (86), respectively. The inherent and common aspect of the respective state-space posterior estimators is that the state-error covariance $\widehat{\boldsymbol{\Psi}}_{\mathbf{a}}(n)$ decays monotonically with time [42]. This asymptotically reduces the VB and MLSU weighting functions to

$$\mathbf{h}_{\mathrm{VB}}(n)\,|_{\widehat{\boldsymbol{\Psi}}_{\mathbf{a}}(n)\to\mathbf{0}} \approx \left[\widehat{\mathbf{a}}^{H}(n)\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\widehat{\mathbf{a}}(n) + \phi_{X_1}^{-1}(n)\right]^{-1}\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\widehat{\mathbf{a}}(n) \quad (108)$$

and

$$\mathbf{h}_{\mathrm{MLSU}}(n)\,|_{\widehat{\boldsymbol{\Psi}}_{\mathbf{a}}(n)\to\mathbf{0}} \approx \left[\widehat{\mathbf{a}}^{H}(n)\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\widehat{\mathbf{a}}(n)\right]^{-1}\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\widehat{\mathbf{a}}(n), \quad (109)$$

respectively. We can deduce easily from (108) and (109) that due to the presence of the prior target signal variance $\phi_{X_1}(n)$, $\mathbf{h}_{\mathrm{VB}}(n)$ will cause relatively more distortion as compared to $\mathbf{h}_{\mathrm{MLSU}}(n)$ even if ideal estimate of the steering vector $\widehat{\mathbf{a}}(n)$ were available.

B. Array Gain

Now, we consider performance with respect to the array gain as depicted in Figs. 8 and 9 for $\mathrm{iSNR} = 15$ dB and $\mathrm{iSNR} = 5$ dB, respectively. For both iSNR cases, it can be seen that the SNR gains remain more or less constant with T_{60}. It is interesting to see in the white observation noise case that despite exhibiting more signal distortion, as evident from Fig. 8(a) and (b), the VB solution carries out more noise suppression to attain an SNR gain higher than the MLSU algorithm. The very structural attribute of the VB weighting function $\mathbf{h}_{\mathrm{VB}}(n)$ as highlighted in Section VII-A via (108), which made it more prone to distortion, makes it now more suitable for noise reduction. This is a manifestation of the classical trade-off between SNR gain and target signal distortion. Generally, we can conclude that reduced target signal distortion achieved by the state-space algorithms does not particularly compromise their array gains, as they lie in very close vicinity of the ML/MVDR anchor.

C. Convergence Analysis

The rate of convergence is surely a non-trivial matter in any blind iterative system. In order to reduce the convergence time,

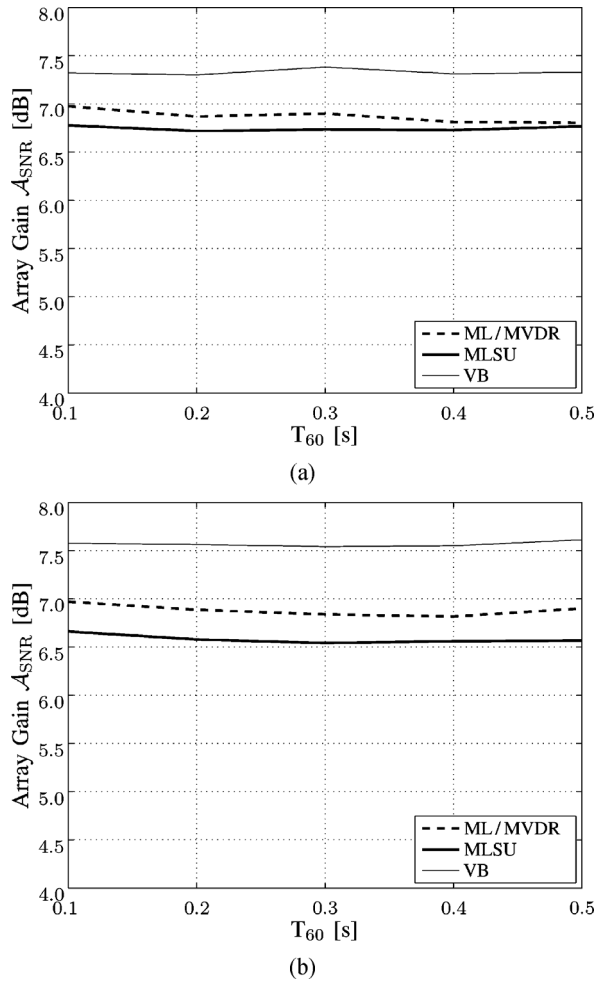

Fig. 8. Array gain with white observation noise for different T_{60}. (a) Array gain at $\mathrm{iSNR} = 15$ dB. (b) Array gain at $\mathrm{iSNR} = 5$ dB.

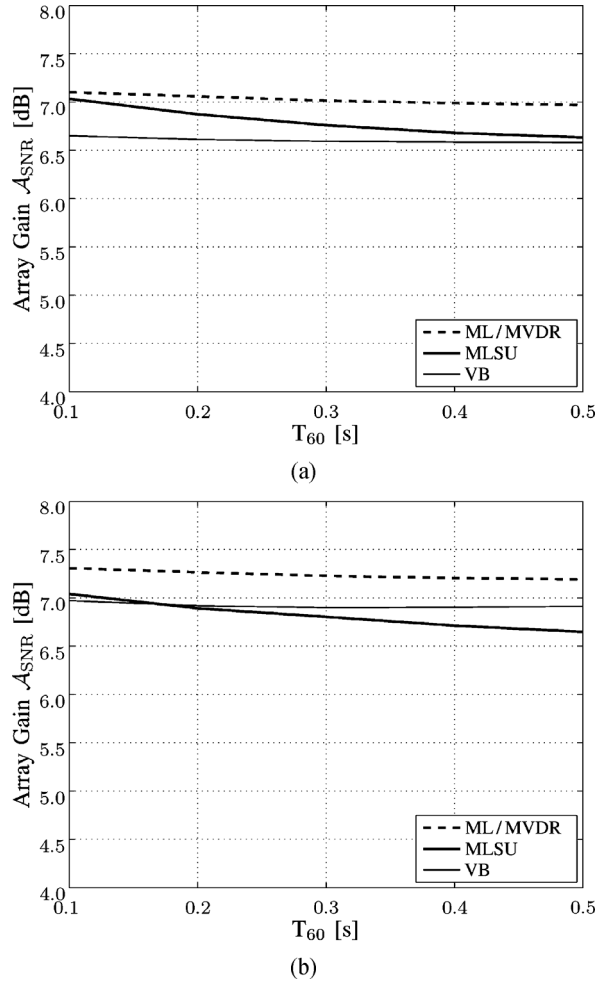

Fig. 9. Array gain with babble observation noise for different T_{60}. (a) Array gain at $\mathrm{iSNR} = 15$ dB. (b) Array gain at $\mathrm{iSNR} = 5$ dB.

we have reduced the *degree of blindness* in our problem by setting the steering-vector posterior as

$$\begin{aligned} \widehat{a}_1(n) &= 1 \\ \widehat{\Phi}_{a_{1,1}}(n) &= \delta \end{aligned} \tag{110}$$

in (66) and (86) in each iteration, where $\widehat{a}_1(n)$ and $\widehat{\Phi}_{a_{1,1}}$ are the mean and state-error covariance of the steering vector corresponding to the 1st sensor, respectively. Configuring $\delta = 10^{-16}$, which is a very small value, is indicative of a high degree of *certainty* regarding the belief pertaining to $\widehat{a}_1(n)$.

The plots in Figs. 10 and 11, outline the convergence of the MLSU algorithm at $\mathrm{iSNR} = 15$ dB for different T_{60} in the presence of white Gaussian observation noise. It can be seen that the algorithm convergences swiftly and does not exhibit any long term divergence. The convergence floor reduces from almost -20 dB to below -32 dB as T_{60} varies from 0.5 s to 0.1 s. The MLSU algorithm exhibits similar characteristics at $\mathrm{iSNR} = 5$ dB, as well. As the underlying state-space estimator is naturally affected by the observation noise, the respective convergence floors now lie approximately between -18 and -28 dB for the considered range of reverberation times.

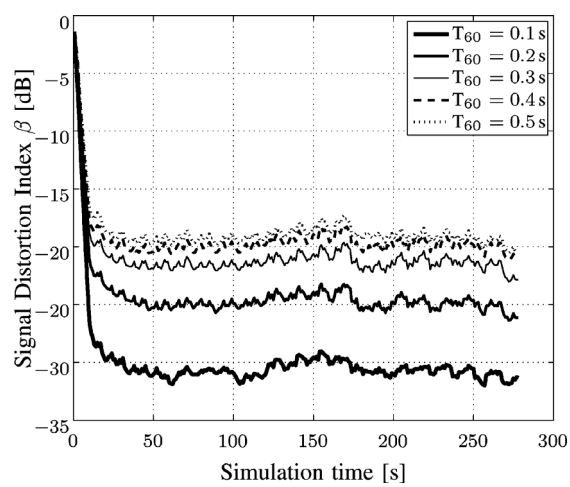

Fig. 10. MLSU convergence characteristics for different T_{60} at $\mathrm{iSNR} = 15$ dB.

VIII. Conclusions

A rigorous description of a variational Bayesian framework for blind adaptive beamforming was presented. We initiated our discussion with a single-input multiple-output observation model in the STFT domain. *A priori* stochastic belief was expressed regarding the elements of the observation model, which was then incorporated within the variational framework to yield an iterative STFT-domain beamformer in closed-form comprising target signal and steering vector posterior estimators.

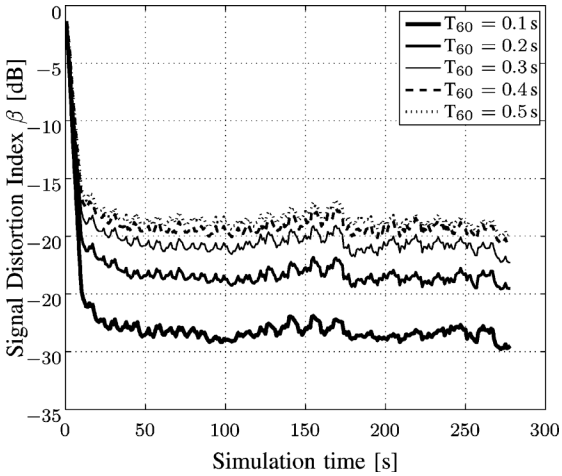

Fig. 11. MLSU convergence characteristics for different T_{60} at iSNR $=$ 5 dB.

It was shown that the systematic elimination of uncertainty pertaining to the elements in the observation model, results in different variants of the state-space variational Bayesian (VB) algorithm. These variants included the maximum- *a-posteriori* (MAP) estimator, a maximum-likelihood estimator operating with steering-vector uncertainty (MLSU), and the conventional maximum-likelihood (ML) estimator or the MVDR beamformer. Thus, we showed the interconnections between these variants using a unifying Bayesian framework. Finally, performance analysis of the VB and MLSU algorithms was presented in noisy and uncertain environments with respect to target signal distortion, array gain, and convergence characteristics. In future work, it would also be interesting to evaluate the performance of the algorithms under test in real-time applications, e.g., automatic speech recognition, and with different array geometries.

Appendix A
Complex Normal Distributions

Distributions corresponding to (14)–(19) are written as [45], [47], [53]:

$$p[\mathbf{a}(n)] = \frac{1}{\pi^M \left|\widehat{\boldsymbol{\Phi}}_{\mathbf{a}}^{+}(n)\right|} \times \exp\left\{-\left[\mathbf{a}(n)-\widehat{\mathbf{a}}^{+}(n)\right]^H \widehat{\boldsymbol{\Phi}}_{\mathbf{a}}^{+^{-1}}(n)\left[\mathbf{a}(n)-\widehat{\mathbf{a}}^{+}(n)\right]\right\}, \quad (111)$$

$$p[\mathbf{u}(n)\,|\,\boldsymbol{\Phi}_{\mathbf{u}}(n)] = \frac{1}{\pi^M \left|\boldsymbol{\Phi}_{\mathbf{u}}(n)\right|} \times \exp\left\{-\mathbf{u}^H(n)\boldsymbol{\Phi}_{\mathbf{u}}^{-1}(n)\mathbf{u}(n)\right\}, \quad (112)$$

$$p[\mathbf{v}(n)\,|\,\boldsymbol{\Phi}_{\mathbf{v}}(n)] = \frac{1}{\pi^M \left|\boldsymbol{\Phi}_{\mathbf{v}}(n)\right|} \times \exp\left\{-\mathbf{v}^H(n)\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\mathbf{v}(n)\right\}, \quad (113)$$

and

$$p\left[X_1(n)\,|\,\phi_{X_1}(n)\right] = \frac{1}{\pi\,\phi_{X_1}(n)} \times \exp\left\{-X_1^*(n)\phi_{X_1}^{-1}(n)X_1(n)\right\}, \quad (114)$$

respectively, and $|\cdot|$ denotes the determinant of a matrix. Note that only for notation convenience and brevity we have shortened the notation $p[\mathbf{a}(n)\,|\,\widehat{\boldsymbol{\Phi}}_{\mathbf{a}}(n-1)]$ in (111) to $p[\mathbf{a}(n)]$.

Appendix B
Optimization Expressions for the VLB

The optimization of the VLB with respect to $q[X_1(n)]$, $q[\mathbf{a}(n)]$, and $\boldsymbol{\Theta}(n)$, i.e.,

$$\frac{\partial}{\partial q[X_1(n)]}\mathcal{J}\left\{q[\mathbf{a}(n)], q[X_1(n)], \boldsymbol{\Theta}(n)\right\} = 0, \quad (115)$$

$$\frac{\partial}{\partial q[\mathbf{a}(n)]}\mathcal{J}\left\{q[\mathbf{a}(n)], q[X_1(n)], \boldsymbol{\Theta}(n)\right\} = 0, \quad (116)$$

$$\frac{\partial}{\partial \boldsymbol{\Theta}(n)}\mathcal{J}\left\{q[\mathbf{a}(n)], q[X_1(n)], \boldsymbol{\Theta}(n)\right\} = \mathbf{0}, \quad (117)$$

subject to the following normalization constraints:

$$\int q[X_1(n)]\,dX_1(n) = 1, \quad (118)$$

$$\int q[\mathbf{a}(n)]\,d\mathbf{a}(n) = 1, \quad (119)$$

will yield the required equations of learning for the posterior distributions and unknown model parameters.

Appendix C
Application of Euler-Lagrange Equation

In the context of Bayesian learning, the stationary point of the VLB $\mathcal{J}\left\{q[\mathbf{a}(n)], q[X_1(n)], \boldsymbol{\Theta}(n)\right\}$ with respect to, e.g., $q[X_1(n)]$, is found by solving

$$\frac{\partial}{\partial q[X_1(n)]}\mathcal{L}_{X_1}\left\{q[\mathbf{a}(n)], q[X_1(n)], \boldsymbol{\Theta}(n)\right\} + \lambda_{X_1}\left\{\int q[X_1(n)]dX_1(n) - 1\right\} = 0 \quad (120)$$

for $q[X_1(n)]$, where λ_{X_1} is a Lagrangian multiplier that imposes the normalization constraint and $\mathcal{L}_{X_1}\left\{q[\mathbf{a}(n)], q[X_1(n)], \boldsymbol{\Theta}(n)\right\}$ satisfies:

$$\mathcal{J}\left\{q[\mathbf{a}(n)], q[X_1(n)], \boldsymbol{\Theta}(n)\right\} = \int \mathcal{L}_{X_1}\left\{q[\mathbf{a}(n)], q[X_1(n)], \boldsymbol{\Theta}(n)\right\} dX_1(n)\,. \quad (121)$$

Appendix D
Target Signal Posterior

The substitution of (114) and (36) into (38) followed by expansion enables us to write:

$$E_{q_a^\star}\left\{\ln\left\{p[\mathbf{y}(n)\,|\,X_1(n), \mathbf{a}(n), \boldsymbol{\Theta}(n)]\,p[X_1(n)\,|\,\boldsymbol{\Theta}(n)]\right\}\right\} = E_{q_a^\star}\Big\{-[\mathbf{y}(n)-\mathbf{a}(n)X_1(n)]^H\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)[\mathbf{y}(n)-\mathbf{a}(n)X_1(n)] - X_1^*(n)\phi_{X_1}^{-1}(n)X_1(n)\Big\} + \kappa_3, \quad (122)$$

where

$$\kappa_3 = E_{q_a^\star}\left\{-\ln\left[\pi^M\left|\boldsymbol{\Phi}_{\mathbf{v}}(n)\right| + \pi\,\phi_{X_1}(n)\right]\right\} \quad (123)$$

comprises terms that do not depend upon $X_1(n)$ and hence do not play a role in the *completion of squares*. After comparing

(122) and (41) and using (42), we can write the terms involving first-order expectation as

$$\begin{aligned} X_1^*(n)\widehat{\phi}_{X_1}^{-1}(n)\widehat{X}_1(n) &= E_{q_a^\star}\left[\, X_1^*(n)\mathbf{a}^H(n)\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\mathbf{y}(n)\,\right], \\ &= X_1^*(n)E_{q_a^\star}\left[\,\mathbf{a}^H(n)\,\right]\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\mathbf{y}(n), \\ &= X_1^*(n)\widehat{\mathbf{a}}^H(n)\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\mathbf{y}(n), \end{aligned} \tag{124}$$

from which it is straightforward to express the posterior mean of the target signal as

$$\widehat{X}_1(n) = \widehat{\phi}_{X_1}(n)\widehat{\mathbf{a}}^H(n)\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\mathbf{y}(n). \tag{125}$$

For terms involving the second-order functions of $X_1(n)$, we again compare (122) and (41) and write

$$\begin{aligned} &X_1^*(n)\widehat{\phi}_{X_1}^{-1}(n)X_1(n) \\ &= X_1^*(n)\left\{E_{q_a^\star}\left[\,\mathbf{a}^H(n)\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\mathbf{a}(n)\,\right] + \phi_{X_1}^{-1}(n)\right\}X_1(n), \end{aligned} \tag{126}$$

which allows us to express the inverse posterior variance of the target signal as

$$\widehat{\phi}_{X_1}^{-1}(n) = E_{q_a^\star}\left[\,\mathbf{a}^H(n)\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\mathbf{a}(n)\,\right] + \phi_{X_1}^{-1}(n). \tag{127}$$

In order to resolve the second-order expectation in (127) by means of (43), we exploit the matrix-vector identity [46]:

$$E_{q_a^\star}\left[\,\mathbf{a}^H(n)\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\mathbf{a}(n)\,\right] \triangleq E_{q_a^\star}\left\{\,\mathrm{tr}\left[\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\mathbf{a}(n)\,\mathbf{a}^H(n)\right]\,\right\} \tag{128}$$

where $\mathrm{tr}[\cdot]$ is the trace operator. This rearrangement allows us to take the expectation operator inside the trace operator, i.e.,

$$\begin{aligned} &E_{q_a^\star}\left[\,\mathbf{a}^H(n)\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\mathbf{a}(n)\,\right] \\ &\qquad = \mathrm{tr}\left\{\,\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)E_{q_a^\star}\left[\mathbf{a}(n)\,\mathbf{a}^H(n)\right]\,\right\} \end{aligned} \tag{129}$$

and directly apply (43) followed by the reversal of the identity (128) to obtain

$$\begin{aligned} &E_{q_a^\star}\left[\,\mathbf{a}^H(n)\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\mathbf{a}(n)\,\right] \\ &\qquad = \widehat{\mathbf{a}}^H(n)\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\widehat{\mathbf{a}}(n) + \mathrm{tr}\left[\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\widehat{\boldsymbol{\Phi}}_{\mathbf{a}}(n)\right]. \end{aligned} \tag{130}$$

Substitution of (130) into (127) makes it possible to express the posterior variance of the target signal as

$$\widehat{\phi}_{X_1}(n) = \left[\widehat{\mathbf{a}}^H(n)\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\widehat{\mathbf{a}}(n) + \widetilde{\phi}_{X_1}^{-1}(n)\right]^{-1}, \tag{131}$$

where

$$\widetilde{\phi}_{X_1}(n) = \left\{\mathrm{tr}\left[\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\widehat{\boldsymbol{\Phi}}_{\mathbf{a}}(n)\right] + \phi_{X_1}^{-1}(n)\right\}^{-1}. \tag{132}$$

Substitution of (131) into (125) results in

$$\widehat{X}_1(n) = \mathbf{h}_{\mathrm{VB}}^H(n)\mathbf{y}(n), \tag{133}$$

where

$$\mathbf{h}_{\mathrm{VB}}(n) = \left[\widehat{\mathbf{a}}^H(n)\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\widehat{\mathbf{a}}(n) + \widetilde{\phi}_{X_1}^{-1}(n)\right]^{-1}\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\widehat{\mathbf{a}}(n). \tag{134}$$

Appendix E
Steering Vector Posterior

We isolate first-order terms in $\mathbf{a}(n)$ in (53) and compare with (52) to write:

$$\widehat{\boldsymbol{\Phi}}_{\mathbf{a}}^{-1}(n)\widehat{\mathbf{a}}(n) = E_{q_{X_1}^\star}\left[X_1^*(n)\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\mathbf{y}(n) + \widehat{\boldsymbol{\Phi}}_{\mathbf{a}}^{+^{-1}}(n)\widehat{\mathbf{a}}^+(n)\right]. \tag{135}$$

Application of (55) to (135) results in

$$\widehat{\boldsymbol{\Phi}}_{\mathbf{a}}^{-1}(n)\widehat{\mathbf{a}}(n) = \widehat{X}_1^*(n)\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\mathbf{y}(n) + \widehat{\boldsymbol{\Phi}}_{\mathbf{a}}^{+^{-1}}(n)\widehat{\mathbf{a}}^+(n), \tag{136}$$

which enables the expressing of the posterior mean of the steering vector as

$$\widehat{\mathbf{a}}(n) = \widehat{\boldsymbol{\Phi}}_{\mathbf{a}}(n)\left[\widehat{X}_1^*(n)\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\mathbf{y}(n) + \widehat{\boldsymbol{\Phi}}_{\mathbf{a}}^{+^{-1}}(n)\widehat{\mathbf{a}}^+(n)\right]. \tag{137}$$

Now, we isolate second-order terms in $\mathbf{a}(n)$ in (53) and compare with (52) to write:

$$\widehat{\boldsymbol{\Phi}}_{\mathbf{a}}^{-1}(n) = E_{q_{X_1}^\star}\left[\, X_1^*(n)\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)X_1(n) + \widehat{\boldsymbol{\Phi}}_{\mathbf{a}}^{+^{-1}}(n)\right]. \tag{138}$$

We resolve the expectation in (138) using (56) to get

$$\begin{aligned} \widehat{\boldsymbol{\Phi}}_{\mathbf{a}}^{-1}(n) &= E_{q_{X_1}^\star}\left[\,\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)X_1(n)\,X_1^*(n)\,\right] + \widehat{\boldsymbol{\Phi}}_{\mathbf{a}}^{+^{-1}}(n), && (139) \\ &= \bar{\boldsymbol{\Phi}}_{\mathbf{v}}^{-1}(n)\left[\widehat{X}_1(n)\,\widehat{X}_1^*(n) + \widehat{\phi}_{X_1}(n)\right] + \widehat{\boldsymbol{\Phi}}_{\mathbf{a}}^{+^{-1}}(n), && (140) \\ &= \widehat{X}_1^*(n)\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\widehat{X}_1(n) + \boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\widehat{\phi}_{X_1}(n) + \widehat{\boldsymbol{\Phi}}_{\mathbf{a}}^{+^{-1}}(n). && (141) \end{aligned}$$

From (141), we conveniently write the expression for the steering-vector posterior error-covariance as

$$\widehat{\boldsymbol{\Phi}}_{\mathbf{a}}(n) = \left[\widehat{X}_1^*(n)\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\widehat{X}_1(n) + \widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}^{-1}(n)\right]^{-1}, \tag{142}$$

where

$$\widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n) = \left[\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\widehat{\phi}_{X_1}(n) + \widehat{\boldsymbol{\Phi}}_{\mathbf{a}}^{+^{-1}}(n)\right]^{-1} \tag{143}$$

is the modified prior error-covariance. Application of the matrix inversion lemma [37] to (143) leads to

$$\widehat{\boldsymbol{\Phi}}_{\mathbf{a}}(n) = \widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n) - \mathbf{K}(n)\widehat{X}_1(n)\widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n), \tag{144}$$

such that the adaptation controller

$$\mathbf{K}(n) = \widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n)\widehat{X}_1^*(n)\left[\widehat{X}_1(n)\widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n)\widehat{X}_1^*(n) + \boldsymbol{\Phi}_{\mathbf{v}}(n)\right]^{-1} \tag{145}$$

is structurally similar to the Kalman gain.

Appendix F
Gradient-Based Adaptation

In order to attain a gradient-based adaptation rule from

$$\begin{aligned} \widehat{\mathbf{a}}(n) = &\left[\widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n) - \mathbf{K}(n)\widehat{X}_1(n)\widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n)\right] \\ &\times \left[\widehat{X}_1^*(n)\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\mathbf{y}(n) + \widehat{\boldsymbol{\Phi}}_{\mathbf{a}}^{+^{-1}}(n)\widehat{\mathbf{a}}^+(n)\right] \end{aligned} \tag{146}$$

we rearrange to get

$$\begin{aligned} \widehat{\mathbf{a}}(n) = &\boldsymbol{\Psi}(n)\,\widehat{\mathbf{a}}(n) + \widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n)\widehat{X}_1^*(n)\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\mathbf{y}(n) \\ &- \mathbf{K}(n)\widehat{X}_1(n)\widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n)\widehat{X}_1^*(n)\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\mathbf{y}(n) \\ &- \mathbf{K}(n)\boldsymbol{\Psi}(n)\,\widehat{\mathbf{a}}(n)\widehat{X}_1(n), \end{aligned} \tag{147}$$

where

$$\boldsymbol{\Psi}(n) = \widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n)\widehat{\boldsymbol{\Phi}}_{\mathbf{a}}^{+^{-1}}(n). \tag{148}$$

Now, we insert the expression

$$\mathbf{K}(n) = \widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n)\widehat{X}_1^*(n)\left[\widehat{X}_1(n)\widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n)\widehat{X}_1^*(n) + \boldsymbol{\Phi}_{\mathbf{v}}(n)\right]^{-1} \tag{149}$$

for Kalman gain in (147) to obtain

$$\begin{aligned}\widehat{\mathbf{a}}(n) =& \boldsymbol{\Psi}(n)\,\widehat{\mathbf{a}}(n) + \widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n)\widehat{X}_1^*(n)\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\mathbf{y}(n)\\ &- \mathbf{K}(n)\boldsymbol{\Psi}(n)\,\widehat{\mathbf{a}}(n)\widehat{X}_1(n)\\ &- \widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n)\widehat{X}_1^*(n)\left[\widehat{X}_1(n)\widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n)\widehat{X}_1^*(n) + \boldsymbol{\Phi}_{\mathbf{v}}(n)\right]^{-1}\\ &\times\left[\widehat{X}_1(n)\widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n)\widehat{X}_1^*(n) + \boldsymbol{\Phi}_{\mathbf{v}}(n) - \boldsymbol{\Phi}_{\mathbf{v}}(n)\right]\\ &\times\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(n)\mathbf{y}(n)\,, \qquad (150)\\ =& \boldsymbol{\Psi}(n)\,\widehat{\mathbf{a}}^{+}(n) + \boldsymbol{\Lambda}(n)\widehat{X}_1^*(n)\mathbf{e}(n)\,, \qquad (151)\end{aligned}$$

where

$$\boldsymbol{\Lambda}(n) \triangleq \widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n)\left[\widehat{X}_1(n)\widetilde{\boldsymbol{\Phi}}_{\mathbf{a}}(n)\widehat{X}_1^*(n) + \boldsymbol{\Phi}_{\mathbf{v}}(n)\right]^{-1} \qquad (152)$$

and

$$\mathbf{e}(n) \triangleq \mathbf{y}(n) - \boldsymbol{\Psi}(n)\,\widehat{\mathbf{a}}^{+}(n)\widehat{X}_1(n)\,. \qquad (153)$$

APPENDIX G
OPTIMIZATION EXPRESSIONS FOR OBTAINING PARAMETER LEARNING RULES

$$\frac{\partial}{\partial\boldsymbol{\Phi}_{\mathbf{v}}(n)}\mathcal{J}\left\{q^{\star}[X_1(n)], q^{\star}[\mathbf{a}(n)], \boldsymbol{\Theta}(n) = \boldsymbol{\Phi}_{\mathbf{v}}(n)\right\} = \mathbf{0}_M, \qquad (154)$$

$$\frac{\partial}{\partial\boldsymbol{\Phi}_{\mathbf{u}}(n)}\mathcal{J}\left\{q^{\star}[X_1(n)], q^{\star}[\mathbf{a}(n)], \boldsymbol{\Theta}(n) = \boldsymbol{\Phi}_{\mathbf{u}}(n)\right\} = \mathbf{0}_M, \qquad (155)$$

$$\frac{\partial}{\partial\phi_{X_1}(n)}\mathcal{J}\left\{q^{\star}[X_1(n)], q^{\star}[\mathbf{a}(n)], \boldsymbol{\Theta}(n) = \phi_{X_1}(n)\right\} = 0. \qquad (156)$$

REFERENCES

[1] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, pp. 1408–1418, Aug. 1969.

[2] Z. Ding and T. Nguyen, "Stationary points of a kurtosis maximization algorithm for blind signal separation and antenna beamforming," *IEEE Trans. Signal Process.*, vol. 48, no. 6, pp. 1587–1596, Jun. 2000.

[3] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*. Berlin, Germany: Springer-Verlag, 2008.

[4] E. A. P. Habets, J. Benesty, and P. A. Naylor, "A speech distortion and interference rejection constraint beamformer," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 3, pp. 854–867, Mar. 2012.

[5] K. Harmanci, J. Tabrikian, and J. L. Krolik, "Relationships between adaptive minimum variance beamforming and optimal source localization," *IEEE Trans. Signal Process.*, vol. 48, no. 1, pp. 1–12, Jan. 2000.

[6] J. M. Kates, "Superdirective arrays for hearing aids," *J. Acoust. Soc. Amer.*, vol. 94, no. 4, pp. 1930–1933, Oct. 1993.

[7] W. Soede, A. J. Berkhout, and F. A. Bilsen, "Development of a directional hearing instrument based on array technology," *J. Acoust. Soc. Amer.*, vol. 94, no. 2, pp. 785–798, Aug. 1993.

[8] R. W. Stadler and W. M. Rabinowitz, "On the potential of fixed arrays for hearing aids," *J. Acoust. Soc. Amer.*, vol. 94, no. 3, pp. 1332–1342, Sep. 1993.

[9] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1997.

[10] E. N. Gilbert and S. P. Morgan, "Optimum design of directive antenna arrays subject to random deviations," *Bell Syst. Tech. J.*, vol. 34, pp. 637–663, May 1955.

[11] H. Cox, R. M. Zeskind, and T. Kooij, "Practical supergain," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 3, pp. 393–398, Jun. 1986.

[12] J. G. Ryan and R. A. Goubran, "Array optimization applied in the near field of a microphone array," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 8, no. 2, pp. 173–176, Mar. 2000.

[13] J. Bitzer and K. U. Simmer, "Superdirective microphone arrays," in *Microphone Arrays: Signal Processing Techniques and Applications*, M. S. Brstein and D. B. Ward, Eds. New York, NY, USA: Springer-Verlag, 2001, ch. 2, pp. 19–38.

[14] S. Yan and Y. Ma, "Robust supergain beamforming for circular array via second-order cone programming," *Appl. Acoust.*, vol. 66, no. 9, pp. 1018–1032, Sep. 2005.

[15] S. Doclo and M. Moonen, "Superdirective beamforming robust against microphone mismatch," *IEEE Trans. Signal Process.*, vol. 15, no. 2, pp. 617–631, Feb. 2007.

[16] G. Elko, "Superdirectional microphone arrays," in *Acoustic Signal Processing for Telecommunication*, S. L. Gay and J. Benesty, Eds. Boston, MA, USA: Kluwer, 2000, ch. 10, pp. 181–237.

[17] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part IV, Optimum Array Processing*. New York: Wiley, 2002.

[18] J. Li, P. Stoica, and Z. Wang, "On robust Capon beamforming and diagonal loading," *IEEE Trans. Signal Process.*, vol. 51, no. 7, pp. 1702–1715, Jul. 2003.

[19] W. Herbordt, H. Buchner, S. Nakamura, and W. Kellermann, "Multichannel bin-wise robust frequency-domain adaptive filtering and its application to adaptive beamforming," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 4, pp. 1340–1351, May 2007.

[20] O. Besson and F. Vincent, "Performance analysis of beamformers using generalized loading of the covariance matrix in the presence of random steering vector errors," *IEEE Trans. Signal Process.*, vol. 53, no. 2, pp. 452–459, Feb. 2005.

[21] O. L. Frost, III, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, no. 8, pp. 926–935, Aug. 1972.

[22] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas Propag.*, vol. 30, no. 1, pp. 27–34, Oct. 1982.

[23] K. L. Bell, Y. Ephraim, and H. L. Van Trees, "A Bayesian approach to robust adaptive beamforming," *IEEE Trans. Signal Process.*, vol. 48, no. 2, pp. 386–398, Feb. 2000.

[24] S. Doclo and M. Moonen, "Design of broadband beamformers robust against gain and phase errors in the microphone array characteristics," *IEEE Trans. Signal Process.*, vol. 51, no. 10, pp. 2511–2526, Oct. 2003.

[25] O. Besson, A. A. Monakov, and C. Chalus, "Signal waveform estimation in the presence of uncertainties about the steering vector," *IEEE Trans. Signal Process.*, vol. 52, no. 9, pp. 2432–2440, Sep. 2004.

[26] A. El-Keyi, T. Kirubarajan, and A. B. Gershman, "Robust adaptive beamforming based on the Kalman filter," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3032–3041, Aug. 2005.

[27] C.-Y. Chen and P. P. Vaidyanathan, "Quadratically constrained beamforming robust against direction-of-arrival mismatch," *IEEE Trans. Signal Process.*, vol. 55, no. 8, pp. 4139–4150, Aug. 2007.

[28] C. C. Gaudes, I. Santamaria, J. Via, E. M. M. Gomez, and T. S. Paules, "Robust array beamforming with sidelobe control using support vector machines," *IEEE Trans. Signal Process.*, vol. 55, no. 2, pp. 574–584, Feb. 2007.

[29] E. Warsitz and M. R. Haeb-Umbach, "Blind acoustic beamforming based on generalized eigenvalue decomposition," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 5, pp. 1529–1539, Jul. 2007.

[30] J. Li, P. Stoica, and Z. Wang, "Doubly constrained robust Capon beamformer," *IEEE Trans. Signal Process.*, vol. 52, no. 9, pp. 2407–2423, Sep. 2004.

[31] A. Beck and Y. C. Eldar, "Doubly constrained robust Capon beamformer with ellipsoidal uncertainty sets," *IEEE Trans. Signal Process.*, vol. 55, no. 2, pp. 753–758, Jan. 2007.

[32] S. A. Vorobyov, H. Chen, and A. B. Gershman, "On the relationship between robust minimum variance beamformers with probabilistic and worst-case distortionless response constraints," *IEEE Trans. Signal Process.*, vol. 56, no. 11, pp. 5719–5724, Nov. 2008.

[33] Y. Gu and A. Leshem, "Robust adaptive beamforming based on interference covariance matrix reconstruction and steering vector estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3881–3885, Jul. 2012.

[34] H. Attias, "A variational Bayesian framework for graphical models," in *In Advances in Neural Information Processing Systems 12*. : , 2000, pp. 209–215.

[35] H. Attias, J. C. Platt, A. Acero, and L. Deng, "Speech denoising and dereverberation using probabilistic models," in *In Advances in Neural Information Processing Systems 13*. Cambridge, MA, USA: MIT Press, 2001.

[36] M. J. Beal, "variational algorithms for approximate Bayesian Inference," Ph.D. dissertation, University College London, Gatsby Computational Neurosci. Unit, London, U.K., 2003.

[37] S. Haykin, *Adaptive Filter Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 2002.

[38] *Robust Speech Recognition of Uncertain or Missing Data—Theory and Applications*, D. Kolossa and R. Haeb-Umbach, Eds. New York, NY, USA: Springer-Verlag, 2011.
[39] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An evaluation of objective measures for intelligibility prediction of time-frequency weighted noisy speech," *J. Acoust. Soc. Amer.*, vol. 130, no. 5, pp. 3013–3027, Nov. 2011.
[40] J. P. Dmochowski and J. Benesty, "Microphone arrays: Fundamental concepts," in *Speech Processing in Modern Communication*, I. Cohen, J. Benesty, and S. Gannot, Eds. Berlin, Germany: Springer, 2010.
[41] L. L. Scharf, *Statistical Signal Processing*. Reading, MA, USA: Addison-Wesley, 1991.
[42] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.
[43] M. Beal and Z. Ghahramani, The Variational Kalman Smoother Gatsby Computational Neuroscience Unit, London, U.K., Tech. Rep. GCNU TR 2001-003, 2001.
[44] F. Scheck, *Mechanics: From Newton's Laws to Deterministic Chaos*. Berlin, Germany: Springer, 2005.
[45] S. Malik, "Bayesian learning of linear and nonlinear acoustic system models in hands-free communication," Ph.D. dissertation, Ruhr-Universität Bochum, Ins. of Communication Acoustics, Bochum, Germany, 2012.
[46] K. B. Petersen and M. S. Pedersen, *The Matrix Cookbook,* 2008.
[47] S. Malik and G. Enzner, "Online maximum-likelihood learning of time-varying dynamical models in block-frequency domain," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Dallas, TX, USA, Mar. 2010, pp. 3822–3825.
[48] J. Benesty, J. Chen, and E. A. P. Habets, *Speech Enhancement in the STFT Domain*. Heidelberg, Germany: Springer, 2012.
[49] E. Lehmann and A. Johansson, "Prediction of energy decay in room impulse responses simulated with an image-source model," *J. Acoust. Soc. Amer.*, vol. 124, no. 1, pp. 269–277, Jul. 2008.
[50] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.
[51] J. J. Shynk, "Frequency-domain and multirate adaptive filtering," *IEEE Signal Process. Mag.*, vol. 9, no. 1, pp. 14–37, Jan. 1992.
[52] E. A. P. Habets and J. Benesty, "A two-stage beamforming approach for noise reduction and dereverberation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 5, pp. 945–958, May 2013.
[53] N. R. Goodman, "Statistical analysis based on a certain multivariate complex Gaussian distribution," *Ann. Math. Statist.*, vol. 34, no. 1, pp. 152–177, Mar. 1963.

Sarmad Malik received the B.Sc. degree in Electrical Engineering from the University of Engineering and Technology Lahore, Pakistan and the M.Sc. degree in INFOTECH from Universität Stuttgart, Germany, in 2003 and 2007, respectively. In 2012, he graduated from Ruhr-Universität Bochum, Germany with the Dr.-Ing. degree in Electrical Engineering and Information Technology.

Dr. Malik worked at And-Or Logic, Inc. in Islamabad, Pakistan as Design Engineer from January 2003 to February 2005 on projects pertaining to ASIC and FPGA design. From October 2006 to March 2007, he interned at the Stuttgart Technology Center of Sony Corporation in Stuttgart, Germany, where he was working on deblurring algorithms for digital still cameras. From November 2012 to October 2013, he was a postdoctoral researcher at INRS-EMT, University of Quebec, Montreal, Canada. His research interests include Bayesian learning, linear and nonlinear acoustic echo control, multichannel adaptive filtering, blind channel identification, and speech dereverberation.

Jacob Benesty was born in 1963. He received a Master degree in microwaves from Pierre & Marie Curie University, France, in 1987, and a Ph.D. degree in control and signal processing from Orsay University, France, in April 1991. During his Ph.D. (from Nov. 1989 to Apr. 1991), he worked on adaptive filters and fast algorithms at the Centre National d'Etudes des Telecomunications (CNET), Paris, France. From January 1994 to July 1995, he worked at Telecom Paris University on multichannel adaptive filters and acoustic echo cancellation. From October 1995 to May 2003, he was first a Consultant and then a Member of the Technical Staff at Bell Laboratories, Murray Hill, NJ, USA. In May 2003, he joined the University of Quebec, INRS-EMT, in Montreal, Quebec, Canada, as a Professor. He is also an Adjunct Professor at Aalborg University, Denmark.

His research interests are in signal processing, acoustic signal processing, and multimedia communications. He is the inventor of many important technologies. In particular, he was the lead researcher at Bell Labs who conceived and designed the world-first real-time hands-free full-duplex stereophonic teleconferencing system. Also, he conceived and designed the world-first PC-based multi-party hands-free full-duplex stereo conferencing system over IP networks.

He was the co-chair of the 1999 International Workshop on Acoustic Echo and Noise Control and the general co-chair of the 2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. He is the recipient, with Morgan and Sondhi, of the IEEE Signal Processing Society 2001 Best Paper Award. He is the recipient, with Chen, Huang, and Doclo, of the IEEE Signal Processing Society 2008 Best Paper Award. He is also the co-author of a paper for which Huang received the IEEE Signal Processing Society 2002 Young Author Best Paper Award. In 2010, he received the "Gheorghe Cartianu Award" from the Romanian Academy. In 2011, he received the Best Paper Award from the IEEE WASPAA for a paper that he co-authored with Chen.

Jingdong Chen received the Ph.D. degree in pattern recognition and intelligence control from the Chinese Academy of Sciences in 1998.

From 1998 to 1999, he was with ATR Interpreting Telecommunications Research Laboratories, Kyoto, Japan, where he conducted research on speech synthesis, speech analysis, as well as objective measurements for evaluating speech synthesis. He then joined the Griffith University, Brisbane, Australia, where he engaged in research on robust speech recognition and signal processing. From 2000 to 2001, he worked at ATR Spoken Language Translation Research Laboratories on robust speech recognition and speech enhancement. From 2001 to 2009, he was a Member of Technical Staff at Bell Laboratories, Murray Hill, New Jersey, working on acoustic signal processing for telecommunications. He subsequently joined WeVoice Inc. in New Jersey, serving as the Chief Scientist. He is currently a professor at the Northwestern Polytechnical University in Xi'an, China. His research interests include acoustic signal processing, adaptive signal processing, speech enhancement, adaptive noise/echo control, microphone array signal processing, signal separation, and speech communication. Dr. Chen is currently an Associate Editor of the IEEE Transactions on Audio, Speech, and Language Processing, an associate member of the IEEE Signal Processing Society (SPS) Technical Committee (TC) on Audio and Acoustic Signal Processing (AASP), and a member of the editorial advisory board of the Open Signal Processing Journal. He was the Technical Program Co-Chair of the 2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA) and the Technical Program Chair of IEEE TENCON 2013, and helped organize many other conferences. He co-authored the books *Study and Design of Differential Microphone Arrays* (Springer-Verlag, 2013), *Speech Enhancement in the STFT Domain* (Springer-Verlag, 2011), *Optimal Time-Domain Noise Reduction Filters: A Theoretical Study* (Springer-Verlag, 2011), *Speech Enhancement in the Karhunen-Loève Expansion Domain* (Morgan&Claypool, 2011), *Noise Reduction in Speech Processing* (Springer-Verlag, 2009), *Microphone Array Signal Processing* (Springer-Verlag, 2008), and *Acoustic MIMO Signal Processing* (Springer-Verlag, 2006). He is also a co-editor/co-author of the book *Speech Enhancement* (Berlin, Germany: Springer-Verlag, 2005) and a section co-editor of the reference *Springer Handbook of Speech Processing* (Springer-Verlag, Berlin, 2007).

Dr. Chen received the 2008 Best Paper Award from the IEEE Signal Processing Society (with Benesty, Huang, and Doclo), the best paper award from the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA) in 2011 (with Benesty), the Bell Labs Role Model Teamwork Award twice, respectively, in 2009 and 2007, the NASA Tech Brief Award twice, respectively, in 2010 and 2009, the Japan Trust International Research Grant from the Japan Key Technology Center in 1998, the Young Author Best Paper Award from the 5th National Conference on Man-Machine Speech Communications in 1998, and the CAS (Chinese Academy of Sciences) President's Award in 1998.