New insights into the Kalman filter beamformer: Applications to speech and robustness

Dani Cherkassky, *Student member, IEEE,* and Sharon Gannot, *Senior member, IEEE*

Abstract

Statistically optimal spatial processors (also referred to as data-dependent beamformers) are widely-used spatial focusing techniques for desired source extraction. The Kalman filter-based beamformer (KFB) [1] is a recursive Bayesian method for implementing the beamformer. This letter provides new insights into the KFB. Specifically, we adopt the KFB framework to the task of *speech* extraction. We formalize the KFB with a set of linear constraints and present its equivalence to the linearly constrained minimum power (LCMP) beamformer. We further show that the *optimal output power*, required for implementing the KFB, is merely controlling the white noise gain (WNG) of the beamformer. We also show, that in static scenarios, the adaptation rule of the KFB reduces to the simpler affine projection algorithm (APA). The analytically derived results are verified and exemplified by a simulation study.

Index Terms

Kalman filter, LCMP beamformer, microphone arrays, speech enhancement, speech extraction, adaptive beamformer, affine projection algorithm.

I. Introduction

BEAMFORMING is one of the most common techniques in microphone array processing. Typically, a beamformer is used to obtain a spatial focusing on the desired speech sources, while reducing the interfering sources and the background noise. Statistically optimal beamformers e.g., minimum variance distortionless response (MVDR) [2], [3], linearly constrained minimum variance (LCMV) [4], [5], and speech distortion weighted multichannel Wiener filter (SDW-MWF) [6], are very useful and widely-used for speech extraction in a reverberant environment [7]. Construction of the beamformers necessitates the sources' statistics and the acoustic transfer functions (ATFs) relating the sources and the microphones (or merely the respective relative transfer functions), which have to be either known or estimated from the received signals [3].

In recent years, with advances in digital processors technology, statistically optimal beamformers are considered even for dynamic and reverberant scenarios where the ATFs are time-varying [8]. Such a dynamic scenario dictates a continuous update

Manuscript received December 01, 2015; revised January 14, 2016; accepted January 18, 2016.

D. Cherkassky and S. Gannot are with the Faculty of Engineering, Bar-Ilan University, Ramat-Gan, Israel (e-mail: dani.cherkassky@gmail.com; sharon.gannot@biu.ac.il).

of the beamformer for maintaining the spatial focusing on the desired speech sources. Accordingly, a recursive solution of the beamformer criterion may facilitate the required tracking ability while minimizing the computational load.

The applicability of the constrained Kalman filter [9] as a recursive estimator of a time-varying beamformer has been addressed in the literature. A recursive solution of the MVDR criterion by a constrained Kalman filter was first proposed by Chen et al. in [1]. In [10] the Kalman filter-based beamformer (KFB) was utilized to design a robust version of the MVDR beamformer. The authors consider arbitrary but norm-bounded mismatches in the desired signal steering vector and impose a set of nonlinear constraints to maintain a distortionless response towards the desired signal even in a worst-case mismatch. Recently, the authors of [11] proposed a KFB with a set of nonlinear constraints, as a recursive scheme implementing the norm-constrained Capon beamformer. The norm-constraint was obtained by explicitly adding the norm of the beamformer to the measurement equations. In [1], [10] and [11] only stationary signals were considered. Accordingly, it is assumed that the optimal average *output power* is known or can be estimated (although in [10] it is numerically demonstrated that the KFB performs well for a wide range of values around the optimal average output power). Considering the speech extraction task, the availability of the output power cannot be assumed, as the output power, corresponding to the desired speech power, is unknown and time-varying.

In the current contribution, we adopt the KFB framework and apply it to non-stationary signals. The KFB with a set of linear constraints is formalized, and its equivalence with the LCMP beamformer is presented. We show that the output power value is only required for controlling the KFB white noise gain (WNG), known to be closely related to the beamformer robustness. Additionally, we also demonstrate that, in a static scenario, the KFB adaptation rule is reduced to affine projection algorithm (APA). In [12] it was shown that the adaptation rule of a single-channel, time-domain Kalman filter is reduced to APA in a static scenario. Here, we show that this simplification is also applicable to the multi-channel, frequency-domain case.

This letter is structured as follows. Section II formulates the speech extraction problem. In Section III-A we present the solution of the LCMP criterion by the KFB, demonstrate the approximation of the KFB by the APA, and present how the WNG of the KFB can be controlled. The analytical results are validated by a simulation study in Section IV. Section V concludes with a brief summary.

II. Problem formulation

Consider the problem of extracting N_d desired speech sources, contaminated by N_i interfering sources, and a stationary noise. Each of the involved signals propagates through the acoustic environment before being picked up by an arbitrary array comprising M microphones. In the short-time Fourier transform (STFT) domain, the nth source is denoted $s_n(\ell,k)$, the ATF relating the nth source and the mth microphone is denoted $h_{m,n}(\ell,k)$, and the noise at the mth microphone is denoted $v_m(\ell,k)$, where ℓ is the frame index, and k is the frequency index. The received signals in the STFT domain can be formulated in a vector:

$$\mathbf{z}(\ell,k)=\mathbf{H}(\ell,k)\mathbf{s}(\ell,k)+\mathbf{v}(\ell,k), \tag{1}$$

where $\mathbf{s}(\ell,k)=[s_1(\ell,k),\ldots,s_N(\ell,k)]^T$, $\mathbf{H}(\ell,k)\in\mathbb{C}^{M\times N}$ such that $[\mathbf{H}(\ell,k)]_{m,n}=h_{m,n}(\ell,k)$, $N=N_\text{d}+N_\text{i}$ is the total number of sources of interest, and $\mathbf{v}(\ell,k)=[v_1(\ell,k),\ldots,v_M(\ell,k)]^T$ is an additive and stationary noise, uncorrelated with any of the other sources.

The extraction of the desired signals can be accomplished by applying a beamformer $\mathbf{w}(\ell,k)$ to the microphone signals, resulting in $y(\ell,k)=\mathbf{w}^H(\ell,k)\mathbf{z}(\ell,k)$. Assuming $M\geq N$, $\mathbf{w}(\ell,k)$ can be chosen to satisfy the LCMP criterion[1] [13]:

$$\mathbf{w}(\ell,k)=\underset{\mathbf{w}}{\operatorname{argmin}}\left\{\mathbf{w}^H(\ell,k)\mathbf{\Phi}_{\mathrm{zz}}(\ell,k)\mathbf{w}(\ell,k)\right\} \quad \text{subject to } \mathbf{H}^H(\ell,k)\mathbf{w}(\ell,k)=\mathbf{g}(\ell,k), \tag{2}$$

where $\mathbf{\Phi}_{\mathrm{zz}}(\ell,k)$ is the power spectral density (PSD) matrix of $\mathbf{z}(\ell,k)$, and $\mathbf{g}(\ell,k)\in\mathbb{C}^{N\times 1}$ is the constraint vector. The well-known solution to (2) is given by [13], [5]:

$$\mathbf{w}_{\mathrm{LCMP}}(\ell,k)=\mathbf{\Phi}_{\mathrm{zz}}^{-1}(\ell,k)\mathbf{H}(\ell,k)\times \left(\mathbf{H}^H(\ell,k)\mathbf{\Phi}_{\mathrm{zz}}^{-1}(\ell,k)\mathbf{H}(\ell,k)\right)^{-1}\mathbf{g}(\ell,k). \tag{3}$$

Based on (1) and the constraint set, the desired signals component at the beamformer output is given by $d(\ell,k)=\mathbf{g}^H(\ell,k)\mathbf{s}(\ell,k)$. In a dynamic scenario, where the ATFs matrix is time-varying, (3) should be computed for each frame index ℓ. Consequently, the solution in (3) imposes high computational load in a dynamic scenarios. In the following section we will show that (2) can be recursively solved by applying the constrained Kalman filter [9].

III. KALMAN FILTER BEAMFORMER

In this section we derive a linearly constrained KFB. Since the beamformer is applied bin-wise, the frequency index is omitted for the sake of conciseness.

A. The LCMP-KFB

The unknown evolution of the time-varying beamformer $\mathbf{w}(\ell)$ can be modeled by a state-vector obeying a first-order Markov process [14]. Accordingly, the underlying process of interest is assumed to satisfy the following recursive *model equation*:

$$\mathbf{w}(\ell)=\gamma\mathbf{w}(\ell-1)+\mathbf{v}_{\mathrm{w}}(\ell), \tag{4}$$

where $\mathbf{v}_{\mathrm{w}}(\ell)$ is the model driving-noise vector, and γ is the forgetting factor. We model $\mathbf{v}_{\mathrm{w}}(\ell)$ by a zero-mean Gaussian random process uncorrelated with $\mathbf{w}(\ell-1)$. The covariance matrix of $\mathbf{v}_{\mathrm{w}}(\ell)$ is assumed to be $\mathbf{Q}=\sigma_{\mathrm{w}}^2\mathbf{I}_M$, with $\mathbf{I}_M$ an $M\times M$ identity matrix. The values of γ and σ_{w}^2 are controlling the dynamic behavior of the beamformer $\mathbf{w}(\ell)$. Typically, the forgetting factor γ is set to a value very close to '1.0' [15]. The value of σ_{w}^2 determines the tradeoff between tracking capabilities and the misalignment at convergence of the Kalman filter [12], [16].

The *measurement equation* of the beamformer $\mathbf{w}(\ell)$ can be straightforwardly defined as:

$$\begin{bmatrix} d(\ell) \\ \mathbf{g}(\ell)\end{bmatrix}=\begin{bmatrix}\mathbf{w}^H(\ell)\mathbf{z}(\ell) \\ \mathbf{H}^H(\ell)\mathbf{w}(\ell)\end{bmatrix}+\begin{bmatrix} d(\ell)-y(\ell) \\ \mathbf{v}_{\mathrm{c}}(\ell)\end{bmatrix}, \tag{5}$$

[1]In the literature, LCMP and LCMV terms are often interchanged. Here we adopt the distinction as defined by Van Trees [13].

Algorithm 1: Kalman filter based beamformer

Initialization: $\hat{\mathbf{w}}_{0|0} = \mathbf{0}$, $\mathbf{P}_{0|0} = \frac{\sigma_{\mathrm{w}}^2}{1-\gamma^2}\mathbf{I}_M$
Input: $\hat{\mathbf{w}}_{\ell-1|\ell-1}$, $\mathbf{P}_{\ell-1|\ell-1}$, $\mathbf{z}(\ell)$
begin Propagation step:
  $\hat{\mathbf{w}}_{\ell|\ell-1} = \gamma \hat{\mathbf{w}}_{\ell-1|\ell-1}$,
  $\mathbf{P}_{\ell|\ell-1} = \gamma^2 \mathbf{P}_{\ell-1|\ell-1} + \mathbf{Q}$.
end
begin Update step:
  $\mathbf{K}_\ell = \mathbf{P}_{\ell|\ell-1}\mathbf{F}_\ell^H \left(\mathbf{F}_\ell \mathbf{P}_{\ell|\ell-1}\mathbf{F}_\ell^H + \mathbf{R}_\ell\right)^{-1}$,
  $\hat{\mathbf{w}}_{\ell|\ell} = \hat{\mathbf{w}}_{\ell|\ell-1} + \mathbf{K}_\ell \left(\mathbf{x}_\ell - \mathbf{F}_\ell \hat{\mathbf{w}}_{\ell|\ell-1}\right)$,
  $\mathbf{P}_{\ell|\ell} = \left(\mathbf{I}_\mathbf{M} - \mathbf{K}_\ell \mathbf{F}_\ell\right)\mathbf{P}_{\ell|\ell-1}$.
end
Output: $\hat{\mathbf{w}}_{\ell|\ell}$, $\mathbf{P}_{\ell|\ell}$, $y(\ell) = (\hat{\mathbf{w}}_{\ell|\ell})^H \mathbf{z}(\ell)$

where $\mathbf{v}_{\mathrm{c}}(\ell)$ is the constraints' errors vector and $d(\ell) - y(\ell)$ models the inevitable difference between the desired and the actual outputs of the beamformer. Despite the fact that (5) is a valid definition for the measurement equation, it is impractical, as $d(\ell)$ is unavailable. However, we can reformulate the measurement equation as follows:

$$\begin{bmatrix} 0 \\ \mathbf{g}(\ell) \end{bmatrix} = \begin{bmatrix} \mathbf{w}^H(\ell)\mathbf{z}(\ell) \\ \mathbf{H}^H(\ell)\mathbf{w}(\ell) \end{bmatrix} + \begin{bmatrix} v_{\mathrm{r}}(\ell) \\ \mathbf{v}_{\mathrm{c}}(\ell) \end{bmatrix}, \tag{6}$$

where $v_{\mathrm{r}}(\ell) = d(\ell) - y(\ell) - d(\ell) = -y(\ell)$. The formulation in (6) is more practical, as only the statistical properties of $v_{\mathrm{r}}(\ell)$ have to be known rather than $d(\ell)$ itself. The measurement equation (6) can be recast in a matrix form:

$$\mathbf{x}(\ell) = \mathbf{F}(\ell)\mathbf{w}(\ell) + \mathbf{v}_{\mathrm{m}}(\ell), \tag{7}$$

where

$$\mathbf{x}(\ell) \triangleq \begin{bmatrix} 0 \\ \mathbf{g}(\ell) \end{bmatrix}, \ \mathbf{F}(\ell) \triangleq \begin{bmatrix} \mathbf{z}^H(\ell) \\ \mathbf{H}^H(\ell) \end{bmatrix}, \ \mathbf{v}_{\mathrm{m}}(\ell) \triangleq \begin{bmatrix} v_{\mathrm{r}}^*(\ell) \\ \mathbf{v}_{\mathrm{c}}(\ell) \end{bmatrix}.$$

The measurement noise $\mathbf{v}_{\mathrm{m}}(\ell)$ is assumed to be zero-mean with a diagonal, time-varying covariance matrix $\mathbf{R}(\ell) \in \mathbb{R}^{(N+1)\times(N+1)}$:

$$\mathbf{R}(\ell) = \begin{bmatrix} \sigma_{\mathrm{r}}^2(\ell) & 0 & 0 & 0 \\ 0 & \sigma_{\mathrm{c},1}^2(\ell) & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \sigma_{\mathrm{c,N}}^2(\ell) \end{bmatrix}. \tag{8}$$

The *state-space* model (4), (7) can be recursively solved for $\mathbf{w}(\ell)$ in the linear minimum mean square error (MMSE) sense, by an application of the Kalman filter as described in Algorithm 1. However, in order to implement Algorithm 1 the entries of $\mathbf{R}$ have to be set. The values of $\left\{\sigma_{\mathrm{c},1}^2(\ell), \cdots, \sigma_{\mathrm{c,N}}^2(\ell)\right\}$ define the allowed deviation from the constraints. For understanding the role of $\sigma_r^2(\ell)$ one can easily verify using (1), (2) and (6), that in case the constraints are satisfied (i.e. $\mathbf{v}_{\mathrm{c}}(\ell) = 0$) the following equation holds: $v_{\mathrm{r}}(\ell) = -(d(\ell) + \mathbf{w}^H(\ell)\mathbf{v}(\ell))$. Accordingly, $\sigma_{\mathrm{r}}^2(\ell)$ should be set to the power of the desired signal (which is of course unknown, and time-varying) plus the power of the residual noise. Nevertheless, we claim that $\sigma_{\mathrm{r}}^2(\ell)$ can be set to a constant value, as will be shown in the sequel. Moreover, we will show that the entries of $\mathbf{R}$ impose a WNG constraint on the beamformer weights.

B. The WNG of the LCMP-KFB

To prove our claim with respect to the role of $\sigma_{\rm r}^2$ we start with the observation that a solution of the state-space model (4), (7) using the Kalman filter, assuming no constraints errors, namely $\mathbf{v}_{\rm c}(\ell)=\mathbf{0}$, is identical to the following constrained minimization problem [10]:

$$\mathbf{w}(\ell)=\underset{\mathbf{w}}{\operatorname{argmin}}\left\{\mathrm{E}\left[|0-\mathbf{w}^H(\ell)\mathbf{z}(\ell)|^2\right]\right\}\quad \text{subject to } \mathbf{H}^H(\ell)\mathbf{w}(\ell)=\mathbf{g}(\ell). \tag{9}$$

By definition, $\mathrm{E}\left[|0-\mathbf{w}^H(\ell)\mathbf{z}(\ell)|^2\right]=\mathbf{w}^H(\ell)\mathbf{\Phi}_{\rm zz}(\ell)\mathbf{w}(\ell)$, and thus the proposed KFB is a recursive solution of the LCMP criterion (2). This is already a first clue that the actual value of $\sigma_{\rm r}^2$ may not have an essential role in KFB implementation, as it is not required for the LCMP implementation (3).

In the sequel, we consider a static scenario where $\mathbf{w}(\ell)=\mathbf{w}$. A somewhat similar scenario was considered in [12], with an application to a single channel echo-canceler. It was assumed in [12] that the covariance matrix of the filter coefficients estimation errors of a single-channel, time-domain Kalman filter tends to become diagonal when the filter has started to converge. In our case, a multi channel, frequency-domain Kalman filter is applied. Assuming that the KFB has started to converge, we expect the spatial covariance of the estimation errors of the beamformer weights $\mathbf{P}_{\ell|\ell-1}$ to be close to zero with variance approximately equal to a positive, small value ϵ_ℓ:

$$\mathbf{P}_{\ell|\ell-1}\approx\epsilon_\ell\cdot\mathbf{I}_M. \tag{10}$$

By substituting the above approximation into the beamformer update equation in Algorithm 1, the following update rule results in:

$$\begin{aligned}\hat{\mathbf{w}}_{\ell|\ell}&=\hat{\mathbf{w}}_{\ell|\ell-1}+\mathbf{K}_\ell\mathbf{e}_\ell=\\&=\hat{\mathbf{w}}_{\ell|\ell-1}+\mathbf{F}_\ell^H\left(\mathbf{F}_\ell\mathbf{F}_\ell^H+\frac{1}{\epsilon_\ell}\mathbf{R}_\ell\right)^{-1}\mathbf{e}_\ell,\end{aligned} \tag{11}$$

with $\mathbf{e}_\ell=\mathbf{x}_\ell-\mathbf{F}_\ell\hat{\mathbf{w}}_{\ell|\ell-1}$ the *a priori* error vector.

Observing (11), and noticing that $\mathbf{R}_\ell$ is a diagonal matrix, we conclude that in a static scenario the KFB adaptation rule simplifies to a regularized APA [14]. In our case, the transition matrix $\mathbf{F}_\ell$ serves as the *input signal matrix* in a conventional APA recursion and $\frac{1}{\epsilon_\ell}\mathbf{R}_\ell$ is a time-varying regularization term. The above analysis suggests that the entries of $\mathbf{R}$ are solely used for regularization in the static scenario. Several works addressed the APA regularization issue. In [17] it was shown that the optimal regularization term is proportional to the signal to noise ratio (SNR) of the input signal. In [18] an adaptive regularization by a time-varying, diagonal matrix was proposed. However, an optimal regularization is beyond the scope of the current contribution.

At this stage, it may be already intuitively clear that the entries of $\mathbf{R}$ impose a constraint on the norm of the beamformer. For explicitly demonstrating this, let us assume for the sake of simplicity that $\mathbf{R}=\sigma^2\mathbf{I}_{N+1}$ and rewrite the update rule in (11) after convergence as:

$$\hat{\mathbf{w}}_{\ell|\ell} = \mathbf{G}_\ell \hat{\mathbf{w}}_{\ell|\ell-1} + \mathbf{u}_\ell, \tag{12}$$

where

$$\mathbf{G}_\ell = \mathbf{I}_M - \mathbf{F}_\ell^H \left(\mathbf{F}_\ell \mathbf{F}_\ell^H + \frac{\sigma^2}{\epsilon_\ell} \mathbf{I}_{N+1} \right)^{-1} \mathbf{F}_\ell, \tag{13}$$

$$\mathbf{u}_\ell = \mathbf{F}_\ell^H \left(\mathbf{F}_\ell \mathbf{F}_\ell^H + \frac{\sigma^2}{\epsilon_\ell} \mathbf{I}_{N+1} \right)^{-1} \mathbf{x}_\ell. \tag{14}$$

It can be easily verified that $\mathbf{u}_\ell$, as defined in (14), is a solution of the following regularized minimization problem:

$$\mathbf{u}_\ell = \underset{\mathbf{u}_\ell}{\operatorname{argmin}} \left\{ \|\mathbf{x}_\ell - \mathbf{F}_\ell \mathbf{u}_\ell\|_2^2 + \frac{\sigma^2}{\epsilon_\ell} \|\mathbf{u}_\ell\|_2^2 \right\}. \tag{15}$$

Observing (15), it is clear that σ^2 imposes a norm constraint on $\mathbf{u}_\ell$. Considering (12), $\hat{\mathbf{w}}_{\ell|\ell}$ is a summation of an update term $\mathbf{u}_\ell$ and $\mathbf{G}_\ell \hat{\mathbf{w}}_{\ell|\ell-1}$. In case $\sigma^2 = 0$, $\mathbf{G}_\ell$ is a projection matrix to the null subspace of the column-space of $\mathbf{F}_\ell$, while for $\sigma^2 \to \infty$, $\mathbf{G}_\ell = \mathbf{I}_M$. In both cases $\mathbf{G}_\ell$ cannot enlarge the norm of $\hat{\mathbf{w}}_{\ell|\ell-1}$. For $0 < \sigma^2 < \infty$ we postulate that $\|\mathbf{G}_\ell \hat{\mathbf{w}}_{\ell|\ell-1}\|_2 \leq \|\hat{\mathbf{w}}_{\ell|\ell-1}\|_2$. Accordingly, since σ^2 imposes a norm constraint on the update term $\mathbf{u}_\ell$, since $\hat{\mathbf{w}}_{1|1} = \mathbf{u}_1$ and considering the recursion $\hat{\mathbf{w}}_{\ell|\ell-1} = \gamma \hat{\mathbf{w}}_{\ell-1|\ell-1}$ with $\gamma < 1.0$, we conclude that σ^2 also imposes a norm constraint on $\hat{\mathbf{w}}_{\ell|\ell}$ and hence controls the sensitivity of the KFB, proportional to $\|\hat{\mathbf{w}}_{\ell|\ell}\|_2^2$ (and equals to the reciprocal of the WNG).

In conclusion, we reformulated the LCMP beamforming problem into state-space equations, which are optimally solved (in the linear MMSE sense) by the Kalman filter. Several design parameters were used for deriving the KFB. While γ, and σ_{w}^2 should be set in accordance with the dynamics of the scenario, the components of $\mathbf{R}$ control the sensitivity of the KFB to deviation from nominal design values.

IV. Experimental Study

In this section we verify the KFB formulation and exemplify its properties, derived in the previous section. However, it should be stressed that the following can by no means serve as a comprehensive comparison between the closed-form LCMP beamformer and KFB.

A. Setup

In the following examples we are using a linear array with $M = 6$ microphones, the sampling frequency of the system is set to 16 KHz, and the STFT analysis window length is set to 1024 samples, with 50% overlap between successive frames. We simulate a static scenario with desired speech source impinging on the array from the broadside, an interfering speech source impinging on the array from an angle of arrival (AOA) equal to $\theta = 60^\circ$, and a stationary (fan) noise impinging on the array with AOA equal to $\theta = 120^\circ$. Microphone signals are further corrupted by a low level, additive and spatially-white sensor noise. The signal to interference ratio (SIR) is set to 0 dB, the SNR is set to 8 dB, and the level of the sensor noise is set to -48 dB relative to the desired speech level. The constraint vector $\mathbf{g}$ is set to impose a distortionless response towards the

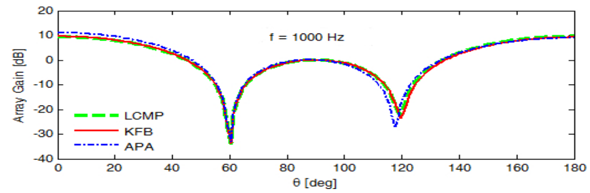


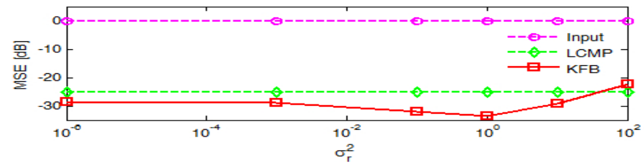
Fig. 1: Closed-form LCMP, KFB, and APA beampatterns.

desired speech, and a null towards the interfering speech. The response to the stationary noise is unconstrained. An error-free estimation of the ATFs is considered for both beamformers. Accordingly, the inputs of $\mathbf{v}_\mathrm{c}$ are set to a very low value of 10^{-9}, i.e. the constraints are perfectly satisfied by the KFB. The parameters that control the dynamic behavior of the KFB are set to $\gamma = 0.9, \sigma_\mathrm{w}^2 = 10$, however, these values are of minor importance in a static scenario.

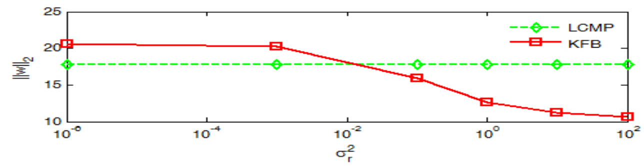
B. Results

The first simulative study is exemplifying the equivalence between the KFB and the closed-form LCMP beamformer, as well as validating the approximation of the KFB adaptation rule by the APA. For this test, a free-field environment is considered. The LCMP beamformer is calculated using (3), the KFB is calculated using Algorithm 1 (with $\sigma_\mathrm{r}^2 = 1$), and the APA is calculated using (11). The resulting spatial responses $B(\theta, k) = \mathbf{w}^H(k)\mathbf{H}(\theta, k)$, are depicted in Fig. 1, for a frequency of 1000 Hz. As can be readily observed, the beampatterns of the LCMP and the KFB are practically identical. While the APA approximation is quite accurate, the differences may be attributed to the approximation in (10). It should be noted that the presented beampatterns are only obtained after the KFB and the APA had converged.

The second simulative study is analyzing the effect of σ_r^2 on the KFB performance. A reverberant enclosure with $T_{60} = 0.3$ sec is considered and the ATFs are simulated using [19]. We compare the performance of the KFB for various values of σ_r^2 to the performance of the closed-form LCMP beamformer. The RMS power of the input signal is kept constant at 0.45. The mean square error (MSE) between $y(\ell)$ and $d(\ell)$, is used as a figure-of-merit to facilitate the comparison. The results are presented in Fig. 2 and Fig. 3. As depicted in Fig. 2a the KFB performs well for a wide range of σ_r^2, while Fig. 2b demonstrates how the average l_2 norm of the KFB $\|\mathbf{w}\|_2 = \frac{1}{K}\sum_{k=0}^{K-1} \|\mathbf{w}(k)\|_2$ is controlled by σ_r^2. It may be noted that in this example, for some values of σ_r^2 the KFB outperforms the closed-form LCMP while maintaining a lower l_2 norm, and hence a higher WNG. Specifically, this is valid for $\sigma_\mathrm{r}^2 = 1$. We therefore compare the response of the KFB (with $\sigma_\mathrm{r}^2 = 1$) to the sources of interest as a function of frequency, to the corresponding response of the LCMP in Fig. 3. From Fig. 3a we deduce that, both beamformers maintain a distortionless response to the desired speech signal across the entire frequency band, and that the attenuation level of the interfering speech obtained by the KFB is slightly worse compared with the LCMP, but still practically very good. For the fan noise attenuation, depicted in Fig. 3b, both beamformers demonstrate poor attenuation of the high frequency components. This may be attributed to the low noise level at the higher frequency range. However, the KFB seems to better attenuate the unconstrained stationary fan noise, especially at the low-medium frequency range, which in turn leads to the better MSE performance.

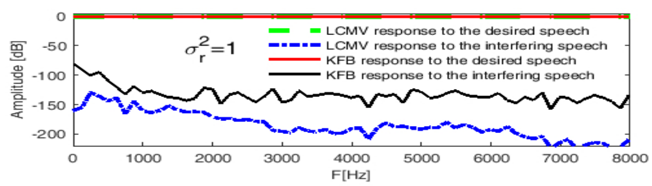


(a) Desired speech extraction performance.

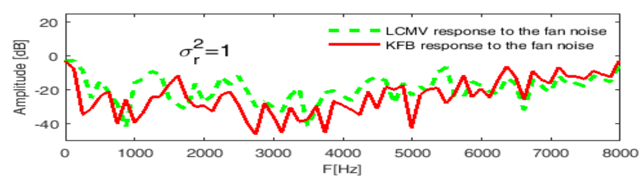


(b) Frequency-averaged WNG for LCMP and KFB.

Fig. 2: Closed-form LCMP vs. KFB



(a) Responses to constrained sources.



(b) Response to the unconstrained noise source.

Fig. 3: Closed-form LCMP vs. KFB frequency responses.

V. Summary and Conclusions

In this contribution we adopted the KFB framework for a speech extraction task, by formalizing it with a set of linear constraints and presenting its equivalence to the LCMP beamformer. We also showed that in a static scenario the proposed KFB update rule is reduced to the APA, and demonstrated analytically and by simulation how the variance of the noise in the measurement equation controls the WNG of the KFB. The dynamic properties of the KFB were not considered and will be an issue for a future study.

References

[1] Y. H. Chen and C. T. Chiang, "Adaptive beamforming using the constrained Kalman filter," *IEEE Transactions on Antennas and Propagation,*, vol. 41, no. 11, pp. 1576–1580, 1993.

[2] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.

[3] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Transactions on Signal Processing*, vol. 49, no. 8, pp. 1614–1626, 2001.

[4] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 5, no. 2, pp. 4–24, 1988.

[5] S. Markovich-Golan, S. Gannot, and I. Cohen, "Multichannel eigenspace beamforming in a reverberant noisy environment with multiple interfering speech signals," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 6, pp. 1071–1086, 2009.

[6] S. Doclo and M. Moonen, "Gsvd-based optimal filtering for single and multimicrophone speech enhancement," *IEEE Transactions on Signal Processing,*, vol. 50, no. 9, pp. 2230–2244, 2002.

[7] S. Gannot and I. Cohen, "Adaptive beamforming and postfitering," in *Springer Handbook of Speech Processing*. New York: Springer, 2007.

[8] S. Markovich-Golan, S. Gannot, and I. Cohen, "Subspace tracking of multiple sources and its application to speakers extraction," in *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), 2010*. IEEE, 2010, pp. 201–204.

[9] D. Simon, "Kalman filtering with state constraints: a survey of linear and nonlinear algorithms," *IET Control Theory & Applications*, vol. 4, no. 8, pp. 1303–1318, 2010.

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

[11] J. S. Hu and M. T. Lee, "Norm constrained Capon beamforming using multi-rank signal models with Kalman filter implementation," *IEEE Transactions on Antennas and Propagation,*, vol. 62, no. 9, pp. 4574–4583, 2014.

[12] C. Paleologu, J. Benesty, and S. Ciochina, "Study of the general Kalman filter for echo cancellation.," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 8, pp. 1539–1549, 2013.

[13] H. L. Van Trees, *Detection, estimation, and modulation theory, optimum array processing*, Wiley, New York, 2004.

[14] S. Haykin, *Adaptive filter theory*, Pearson Education India, 2007.

[15] H. A. P. Blom and Y. Bar-Shalom, "The interacting multiple model algorithm for systems with Markovian switching coefficients," *IEEE Transactions on Automatic Control*, vol. 33, no. 8, pp. 780–783, 1988.

[16] G. Enzner and P. Vary, "Frequency-domain adaptive Kalman filter for acoustic echo control in hands-free telephones," *Signal Processing*, vol. 86, no. 6, pp. 1140–1156, 2006.

[17] C. Paleologu, J. Benesty, and S. Ciochina, "Regularization of the affine projection algorithm," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 58, no. 6, pp. 366–370, 2011.

[18] Y. S. Choi, H. C. Shin, and W. J. Song, "Adaptive regularization matrix for affine projection algorithm," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 54, no. 12, pp. 1087–1091, 2007.

[19] E.A.P. Habets, "Room impulse response generator," *Tech. Rep, Technische Universiteit Eindhoven*, 2006.