Fast Tracking of Time-Variant Systems Using Local Affine Subspaces

Till Hardenbicker and Peter Jax

July 21, 2023

Fast Tracking of Time-Variant Systems Using Local Affine Subspaces

Till Hardenbicker, Peter Jax

Institute of Communication Systems, RWTH Aachen University, Aachen, Germany
Email: {hardenbicker,jax}@iks.rwth-aachen.de

Abstract

Various audio and speech processing applications require the identification and tracking of linear acoustic systems. Previous analyses have demonstrated that in many scenarios the set of possible impulse responses forms a low dimensional manifold. Existing approaches have used this fact to improve the convergence properties of an identification algorithm, e.g., by projecting the estimated impulse response vector onto a set of lower dimensional affine subspaces that are learned from data that is known a priori. In this paper, we present a novel variant of the Kalman filter that only tracks a low dimensional system representation in a linear subspace. Experimental results show that the proposed approach is robust in adverse signal-to-noise ratios and reduces the relative system distance compared to state-of-art approaches when tracking time-variant systems.

1 Introduction

Acoustic System Identification (ASI) is a common task in digital signal processing. It arises in multiple applications such as echo cancellation [1], feedback cancellation [2, 3], the measurement of head related transfer functions [4, 5] or active noise cancellation [6]. The Kalman Filter (KF) [1, 7, 8] and the (normalized) Least Mean Squares (NLMS) algorithm [4, 9, 10] are common solutions to this task. Albeit most ASI algorithms converge well in experiments where the true system does not change, tracking of time-variant systems remains challenging. This problem is aggravated when ASI takes place with low signal-to-noise ratios. An often observed fact is that the convergence speed of ASI algorithms under noisy conditions is inversely proportional to the number of model parameters, i.e., filter coefficients [11]. This is in contrast to the fact that long impulse responses are often required to accurately model the potentially long decay of real acoustic systems.

In an acoustic environment, the number of free physical parameters is much smaller than the number of filter coefficients. For instance, using a shoebox model, the impulse response between a source and a receiver in a room is determined by their position, orientation and the room's dimension and reflection coefficients. In this case, acoustic impulse responses lie on a low dimensional manifold. This assumption is well-studied and has been exploited in system identification [12–16]. The authors of [15, 16] propose to learn such a manifold implicitly by computing Low-Dimensional Affine Subspaces (LDASs) based on a priori knowledge on impulse responses that have been measured in advance. From one perspective, these subspaces can be understood as tangent planes of the manifold. From another point of view, they are models for impulse responses in a small neighborhood [13, 15]. In [15, 16] these subspaces are combined with an adaptive filter by projecting the estimate of the impulse response vector onto the closest LDAS. It was shown that this additional projection improves ASI performance in noisy environments.

Following the established concept of ASI exploiting subspaces, we propose an algorithm for the identification of time-variant systems based on a modification of the KF. As opposed to [15, 16] our method works by performing the coefficient update within the LDAS instead of successively projecting the update onto the LDAS. Precisely, the adaptive filter does not track the actual system but its coordinates in the LDAS. Since the LDAS changes in every time step, we propose how to update the State Error Covariance (SEC) of the KF when the LDAS changes.

The remainder of this paper is structured as follows: In Section 2 we introduce the general ASI scenario together with the notation used. The proposed model of a KF in an LDAS is presented in Sections 3.1 and 3.2. In Section 4 we validate the efficacy of the devised approach for a static and a time-variant scenario.

2 System Model

Throughout this paper bold lower case letters denote column vectors and bold upper case letters denote matrices. $\widehat{a}$ denotes an estimate of a and $\boldsymbol{a}^{\mathrm{T}}$ is the transpose of $\boldsymbol{a}$. The $n \times n$ identity matrix is denoted by $\mathbf{I}_n$, whereas $\mathbf{0}_n$ and $\mathbf{1}_n$ refer to a column vector with n elements that are all zero or one, respectively. The vector diag $\mathbf{A}$ containis only the diagonal elements of matrix $\mathbf{A}$. The operators $\odot$ and $\oslash$ denote element wise multiplication and division of two vectors or matrices, respectively. The symbol $\mathcal{N}$ denotes a normal distribution.

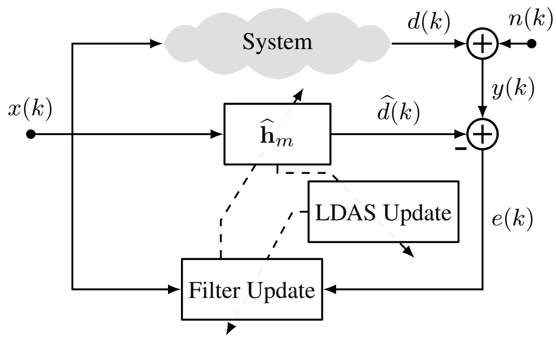

Figure 1: System theoretic block diagram for the identification of an unknown system by a filter $\widehat{\mathbf{h}}_m$.

Figure 1 shows a block diagram of the considered basic system identification task. An unknown system is excited by the signal $x(k)$. The system output $d(k)$ is superimposed with measurement noise $n(k)$ yielding observation $y(k)$. The goal is to model the unknown system by a filter $\widehat{\mathbf{h}}_m$ so that the error $e(k)$ between measured output $y(k)$ and prediction signal $\widehat{d}(k)$ becomes minimal. Using an estimated impulse response $\widehat{h}(k,\kappa)$, the adaptive filter predicts the estimated system output

$$\widehat{d}(k) = \sum_{\kappa=0}^{l-1} \widehat{h}(k,\kappa) x(k-\kappa) \ , \tag{1}$$

where l is the length of the adaptive filter. The error signal $e(k) = y(k) - \widehat{d}(k)$ is used to adapt the filter coefficients. The update of LDAS and filter is explained in Section 3. To express block-adaptation processes formally, we will use vector notation in the following. For better distinction to the time index, the block index is denoted by a subscript. Then, all signals except $x(k)$ are divided into non-overlapping frames of length r, such that the m-th frame of the measurement and error signal are given by

$$\begin{aligned} \mathbf{y}_m &= [y(mr-r+1) \ \dots \ y(mr-1), y(mr)]^{\mathrm{T}} \\ &= \mathbf{d}_m + \mathbf{n}_m \end{aligned} \tag{2}$$

$$\mathbf{e}_m = \mathbf{y}_m - \widehat{\mathbf{d}}_m \ . \tag{3}$$

The excitation signal is divided into overlapping frames

$$\mathbf{x}(k) = [x(k), x(k-1) \ \dots \ x(k-l+1)]^{\mathrm{T}} \tag{4}$$

that are stacked to the convolution matrix

$$\mathbf{X}_m = \left[\mathbf{x}(mr-r+1), \mathbf{x}(mr-r+2) \, \ldots \, \mathbf{x}(mr)\right]^{\mathrm{T}} . \tag{5}$$

Using

$$\widehat{\mathbf{h}}_m = \left[\widehat{h}(mr,0), \widehat{h}(mr,1) \, \ldots \, \widehat{h}(mr,l-1)\right] \tag{6}$$

(1) can be stated as

$$\widehat{\mathbf{d}}_m = \mathbf{X}_m \widehat{\mathbf{h}}_m . \tag{7}$$

3 Proposed Concept

This Section presents the contributions of this paper. Section 3.1 derives the KF equations on an arbitrary LDAS. In Section 3.2 we present how we choose the LDAS from a priori data in each time step and how the KF equations must be adapted for this choice.

3.1 Kalman Filtering on a Static Subspace

To formulate the KF in the LDAS we start by finding a state-space model [17]. If the impulse response at time step m lies in an l' dimensional affine subspace with $l' < l$, the impulse response can be expressed as

$$\mathbf{h}_m = \mathbf{V}_m \mathbf{z}_m + \overline{\mathbf{h}}_m , \tag{8}$$

where $\mathbf{z}_m \in \mathbb{R}^{l'}$ are its coordinates in the LDAS. $\mathbf{V}_m \in \mathbb{R}^{l \times l'}$ is the basis matrix whose l' column vectors span the subspace at time step m (see Section 3.2). $\overline{\mathbf{h}}_m \in \mathbb{R}^{l}$ is a support vector that controls the location of the subspace. Using (2), (7) and (8) we find the observation equation

$$\mathbf{y}_m = \mathbf{X}_m \left(\mathbf{V}_m \mathbf{z}_m + \overline{\mathbf{h}}_m\right) + \mathbf{n}_m \tag{9}$$

and identify $\mathbf{X}_m \mathbf{V}_m$ as observation matrix. For the state transition we assume a first order Markov model [1]

$$\mathbf{z}_{m+1} = \gamma \mathbf{z}_m + \boldsymbol{\delta}_m , \tag{10}$$

with scalar fading factor γ and process noise $\boldsymbol{\delta}_m$. Both measurement and process noise are assumed to be Gaussian distributed random variables, i.e., $\mathbf{n} \sim \mathcal{N}(\mathbf{0}_r, \mathbf{Q}_\mathbf{n})$ and $\boldsymbol{\delta} \sim \mathcal{N}(\mathbf{0}_{l'}, \mathbf{Q}_{\boldsymbol{\delta}})$. With this model we can formulate the Kalman equations. The superscripts $^-$ and $^+$ denote prior and posterior estimates, respectively.

Prediction

$$\widehat{\mathbf{d}}_m = \mathbf{X}_m \left(\mathbf{V}_m \widehat{\mathbf{z}}_m^- + \overline{\mathbf{h}}_m\right) \tag{11a}$$

$$\mathbf{e}_m = \mathbf{y}_m - \widehat{\mathbf{d}}_m \tag{11b}$$

Measurement Update

$$\mathbf{K}_m = \mathbf{P}_m^- \mathbf{V}_m^{\mathrm{T}} \mathbf{X}_m^{\mathrm{T}} \left(\mathbf{X}_m \mathbf{V}_m \mathbf{P}_m^- \mathbf{V}_m^{\mathrm{T}} \mathbf{X}_m^{\mathrm{T}} + \mathbf{Q}_\mathbf{n}\right)^{-1} \tag{12a}$$

$$\Delta \mathbf{z}_m = \mathbf{K}_m \mathbf{e}_m \tag{12b}$$

$$\widehat{\mathbf{z}}_m^+ = \widehat{\mathbf{z}}_m^- + \Delta \mathbf{z}_m \tag{12c}$$

$$\mathbf{P}_m^+ = \left(\mathbf{I}_{l'} - \mathbf{K}_m \mathbf{X}_m \mathbf{V}_m\right) \mathbf{P}_m^- \tag{12d}$$

Time Update

$$\widehat{\mathbf{z}}_{m+1}^- = \gamma \widehat{\mathbf{z}}_m^+ \tag{13a}$$

$$\mathbf{P}_{m+1}^- = \gamma^2 \mathbf{P}_m^+ + \mathbf{Q}_{\boldsymbol{\delta}} . \tag{13b}$$

An impulse response estimate in full space is obtained by

$$\widehat{\mathbf{h}}_m^+ = \mathbf{V}_m \widehat{\mathbf{z}}_m^+ + \overline{\mathbf{h}}_m . \tag{14}$$

These equations correspond to a block time domain KF [8], but the estimation problem is rotated in an LDAS spanned by the columns of the basis matrix $\mathbf{V}_m$. For the special case $\mathbf{V}_m = \mathbf{I}_l$ and $\overline{\mathbf{h}}_m = \mathbf{0}_l$ the proposed approach and the block time domain KF are identical. Then, the state $\mathbf{z}_m$ equals the impulse response. Comparing KF with NLMS, the Kalman Gain $\mathbf{K}_m$ is often considered an optimal step size that reflects the current uncertainty of the adaptive filter [18]. This uncertainty is given by $\mathbf{P}$ and is estimated recursively and depends on an estimate of $\mathbf{Q}_{\boldsymbol{\delta}}$ [17]. When the estimate of $\mathbf{Q}_{\boldsymbol{\delta}}$ is imperfect and the measurement is corrupted by noise, the ASI algorithm may adapt into a wrong direction which leads to instabilities. In (12), however, the SEC has only $l' \times l'$ entries so that the filter can only adapt within the subspace spanned by $\mathbf{V}_m$. Hence, the risk of adaptation in wrong directions is mitigated. Since the algorithm only has to track $l' < l$ variables we expect that the proposed model can track changes faster.

In the next Section we will present how to determine the basis matrix for the proposed approach.

3.2 Time-Variant Subspace Model

In the manifold framework, where a nonlinear function maps $\mathbf{z}_m$ onto $\mathbf{h}_m$, $\mathbf{V}_m$ and $\overline{\mathbf{h}}_m$ can be understood as the linearization of this function. However, in what follows, we only consider LDASs that are formed by the principal components of reference impulse responses that are collected in advance. The LDAS is updated at each time step m. Then, state and SEC of the KF have to account for this change of basis. We group this update of basis vectors $\mathbf{V}$, state $\mathbf{z}$ and SEC $\mathbf{P}$ into a subspace update. From informal experiments we conclude that the subspace update best takes place between measurement update and time update. We denote state and SEC after the subspace update by $\mathbf{z}_m^=$ and $\mathbf{P}_m^=$, respectively.

3.2.1 Update of nearest neighbors

We define a set of training data

$$\mathcal{H} = \left\{\mathbf{h}_b^{\mathrm{tr}} \,|\, b \in 1, 2, \ldots B\right\} , \tag{15}$$

where each $\mathbf{h}_b^{\mathrm{tr}}$ is one of B impulse response vectors that have been recorded in advance. Similarly to [16] we use K nearest neighbors $\mathcal{U} \in \mathcal{H}$ where the cardinality $|\mathcal{U}| = K$ and

$$\left\|\mathbf{h}_{\in}^{\mathrm{tr}} - \widehat{\mathbf{h}}_m^+\right\| \leq \left\|\mathbf{h}_{\notin}^{\mathrm{tr}} - \widehat{\mathbf{h}}_m^+\right\| \quad \forall\, \mathbf{h}_{\in}^{\mathrm{tr}} \in \mathcal{U}, \mathbf{h}_{\notin}^{\mathrm{tr}} \notin \mathcal{U} . \tag{16}$$

The new support vector is easily found as

$$\overline{\mathbf{h}}_{m+1} = \frac{1}{K} \sum_{\mathbf{h} \in \mathcal{U}} \mathbf{h} . \tag{17}$$

In order to perform a Principal Component Analysis (PCA) we estimate covariance of the nearest neighbors

$$\mathbf{Q}_\mathcal{U} = \frac{1}{K-1} \sum_{\mathbf{h} \in \mathcal{U}} \left(\mathbf{h} - \overline{\mathbf{h}}_{m+1}\right) \left(\mathbf{h} - \overline{\mathbf{h}}_{m+1}\right)^{\mathrm{T}} \tag{18}$$

so that the approximation by Eigenvalue Decomposition (EVD)

$$\mathbf{Q}_\mathcal{U} \approx \mathbf{V}_{m+1} \mathbf{Q}_\mathbf{z} \mathbf{V}_{m+1}^{\mathrm{T}} \tag{19}$$

yields the matrix $\mathbf{V}_{m+1}$ whose l' columns span the subspace for the next timestep. The $l' \times l'$ diagonal matrix $\mathbf{Q}_\mathbf{z}$ contains the variance that the principal direction vectors within $\mathbf{V}_m$ express. For now, this procedure exhibits a high computational complexity, caused by the neighborhood search and the EVD in each time step. A reduction of complexity is subject of future research.

3.2.2 State update

After the principal components and the vector basis have changed, $\widehat{\mathbf{z}}_m^+$ needs to be projected onto the new LDAS. Since $\mathbf{V}_m$ is orthonormal due to the EVD, we can define the change-of-basis matrix

$$\mathbf{M} = \mathbf{V}_{m+1}^{\mathrm{T}} \mathbf{V}_m . \tag{20}$$

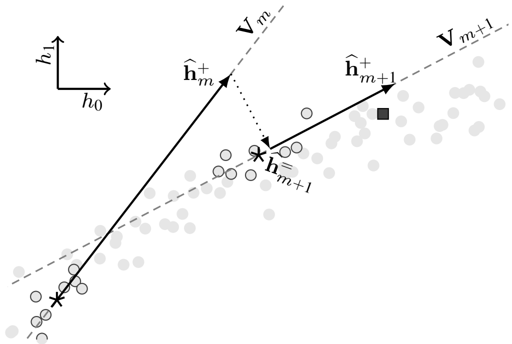

Figure 2: Schematical visualization of measurement and subspace update for $l = 2$ and $l' = 1$. Training data $\mathcal{H}$ is shown by grey dots, ○ are nearest neighbors $\mathcal{U}$ and ⋆ are the support vectors $\overline{\mathbf{h}}$. → shows the measurement update and ⋯► shows the subspace update. Dashed lines visualize the LDAS. ■ is the true system.

Then, the projected state is

$$\widehat{\mathbf{z}}_{m+1}^{=} = \mathbf{M}\widehat{\mathbf{z}}_m^{+} + \mathbf{V}_{m+1}^{\mathrm{T}}\left(\overline{\mathbf{h}}_m - \overline{\mathbf{h}}_{m+1}\right) . \tag{21}$$

While $\mathbf{M}$ accounts for the change of basis, the second addend in (21) causes $\widehat{\mathbf{z}}_{m+1}^{=}$ to be close to the origin, such that most of the energy of $\widehat{\mathbf{h}}$ is expressed by $\overline{\mathbf{h}}$. Figure 2 gives an intuitive interpretation of subspace update and state update for the two-dimensional case.

3.2.3 Covariance Update

When the span of the LDAS changes, this also has to be accounted for in the SEC matrix. This includes the case that $\mathbf{V}_{m+1}$ spans directions that have not been spanned by $\mathbf{V}_m$. Precisely, $\mathbf{V}_{m+1}$ can be decomposed into the span of the current basis

$$\mathbf{V}_{\parallel,m+1} = \mathbf{V}_m\mathbf{M}^{\mathrm{T}} \tag{22}$$

and the orthogonal complement

$$\mathbf{V}_{\perp,m+1} = \mathbf{V}_{m+1} - \mathbf{V}_{\parallel,m+1} . \tag{23}$$

The same can be done for the SEC matrix:

$$\mathbf{P}_{m+1}^{=} = \underbrace{\mathbf{M}\mathbf{P}_m^{+}\mathbf{M}^{\mathrm{T}}}_{\mathbf{P}_{\parallel,m+1}} + \underbrace{\mathbf{V}_{\perp,m+1}^{\mathrm{T}}\mathbf{Q}_{\mathcal{U}}\mathbf{V}_{\perp,m+1}}_{\mathbf{P}_{\perp,m+1}} . \tag{24}$$

Here, $\mathbf{P}_{\parallel,m+1}$ corresponds to the SEC that is known from the current time step m. For the novel directions however, no information is available so that $\mathbf{P}_{\perp,m+1}$ has to be initialized in these directions. For this initialization, we use the neighborhood covariance $\mathbf{Q}_{\mathcal{U}}$.

3.2.4 Initialization

Prior to the first iteration, $\mathbf{z}$, $\mathbf{P}$, $\mathbf{V}$ and $\overline{\mathbf{h}}$ need to be initialized. To do so, we perform the PCA in (17) – (19) with the difference that instead of the neighborhood $\mathcal{U}$, the entire training set $\mathcal{H}$ is used. E.g., $\mathbf{Q}_{\mathcal{H}}$ is the training data covariance. The resulting $\overline{\mathbf{h}}$ and $\mathbf{V}$ span a subspace that optimally expresses the entire training set in a least square sense. The EVD in (19) also yields the explained variances $\mathbf{Q}_{\mathbf{z}}$ that are used to initialize the SEC $\mathbf{P}_0^-$. When the Kalman Gain $\mathbf{K}_m$ in (12a) is interpreted as a step size, this choice of $\mathbf{P}_0^-$ causes a higher initial step size in directions where the training data has higher variance [18].

4 Experiments

In this Section, the algorithm described above is evaluated and compared to similar approaches.[1] We investigate a generic scenario, where a static and a time-variant acoustic path between a source and a receiver in a reverberant environment have to be identified. Before, we explain the generation of training data and give details about the simulation setup.

[1] Simulations were performed with computing resources granted by RWTH Aachen University under project rwth1260.

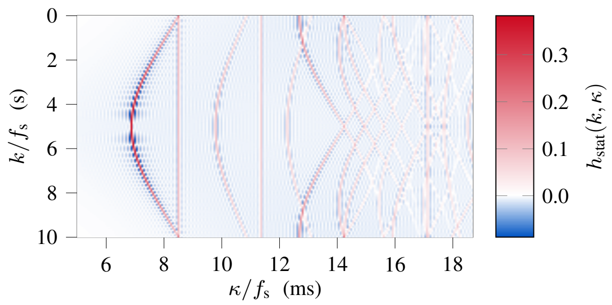

Figure 3: First reflections of the impulse response over time.

4.1 Generation of Training Data

As stated in the algorithmic description, a set of training data is needed which contains candidate impulse responses whose affine combinations can reflect the true impulse response. In [16] it was argued that in real applications, these candidates can be gathered during runtime when the current estimate is deemed reliable in terms of a low SEC. We used the shoe box model from [19] to create a training set and impulse responses for validation. The simulated room has the dimensions [3 m, 4 m, 2.5 m] and the absorption coefficients were chosen to obtain a reverberation time T_{60} of about 0.125 s. For all impulse responses, the receiver was placed at $\mathbf{r}_{\mathrm{rec}} = [2\,\mathrm{m}, 2\,\mathrm{m}, 1\,\mathrm{m}]$. The set of possible source positions $\mathbf{r}_{\mathrm{src}}$ in the training set were located on a sphere around the receiver, with a fixed distance of 0.5 m. The directional vector $\mathbf{r}_{\mathrm{src}} - \mathbf{r}_{\mathrm{rec}}$ was drawn from a spherical uniform distribution with radius 0.5. The size of the training set was chosen to $B = 10000$.

Since the proposed algorithm is inherently initialized with the mean value of all training data and all training data have the same direct path, a significant fraction of the acoustic path's energy would be identified before the first iteration. In order to eliminate this effect we simulate a source with a cardioid characteristic and orient the source so that it points away from the receiver. Thus, there is no direct path between source and receiver. The receiver has omnidirectional characteristic.

4.2 Experimental Setup

As test signals $x(k)$ and $n(k)$ we generated $2 \cdot 50$ noise signals of length 10 s with a sampling frequency $f_s = 16\,\mathrm{kHz}$. All samples were drawn from a uniform distribution.We considered a scenario where the source moves on an arc in the horizontal plane. The corresponding impulse responses were simulated using the settings described in the previous Section. For the static impulse response h_{stat} at time k the relative source position was at

$$\mathbf{r}_{\mathrm{src}}(k) - \mathbf{r}_{\mathrm{rec}} = 0.5\,[\cos\phi(k)\,,\ \sin\phi(k)\,,\ 0] \tag{25}$$

and the azimuthal angle ϕ is increased uniformly from 0 to π. For simplicity, we updated the position every 4 ms ($\Delta\phi = 0.072°$). The first reflections of h_{stat} for all source positions are depicted in Figure 3. For moving sources the signal at the receiver consists of reflections originating from past positions of the sound source so that the simulated signal at the receiver is obtained by [20]

$$d(k) = \sum_{\kappa=0}^{l-1} h_{\mathrm{stat}}(k-\kappa,\kappa)\,x(k-\kappa) . \tag{26}$$

To validate the convergence behaviour alone, we also consider a time-invariant scenario where the source is fixed at $\phi = 0$. The resulting signals $d(k)$ from both scenarios were mixed with $n(k)$ to obtain Echo-To-Noise Ratios (ENRs) of $\pm 10\,\mathrm{dB}$. The ENR is defined as

$$\frac{\mathrm{ENR}}{\mathrm{dB}} = 10\log_{10}\frac{\sum_k d^2(k)}{\sum_k n^2(k)} . \tag{27}$$

The resulting signals are fed to the proposed algorithm and the reference algorithms presented in the next Section. After each filter update we compute the relative system distance

$$\frac{D_m}{\text{dB}} = 10\log_{10}\frac{\left\|\widehat{\mathbf{h}}_m^+ - \mathbf{h}_{m,\text{eff}}\right\|_2^2}{\left\|\mathbf{h}_{m,\text{eff}}\right\|_2^2}\,. \tag{28}$$

From (26) we can see that the effective impulse response vector to be identified reads

$$\mathbf{h}_{m,\text{eff}} = [h(mr,0)\,,\dots h(mr-l+1,l-1)]\,. \tag{29}$$

The number of filter coefficients l for the ground truth and the adaptive filters is set to 2000. Based on preliminary experiments, the subspace dimension is $l' = 200$, the number of neighbors is $K = 2l'$ and the frame shift is $r = 64$.

4.3 Reference Algorithms

As a reference, we consider the (block) time domain KF [8]. It is given by the equations (11) - (13) when $\mathbf{V}$ is set fix to be $\mathbf{I}_l$ and $\overline{\mathbf{h}}$ is $\mathbf{0}_l$. In preliminary experiments we observed that the subspace approach converges significantly faster when the SEC matrix is initialized as described in Section 3.2.4. Hence, for a fair comparison, we initialized the time domain KF with $\mathbf{P}_0 = \mathbf{Q}_{\mathcal{H}}$.

To validate the effect of multiple local subspaces, we also consider a KF on one constant LDAS. To do so, we initialize the proposed KF as described in 3.2.4 but omit the subspace update such that $\mathbf{V}_m$ and $\overline{\mathbf{h}}_m$ are constant.

In [15, 16] it is proposed to project the posterior state estimate $\mathbf{h}_m^+$ of any adaptive filter onto the current subspace. In [16], the authors also use K nearest neighbors to span the local subspace. Hence, as a reference we consider a time domain KF whose state is projected onto the LDAS in each iteration:

$$\widehat{\mathbf{h}}_{m,\text{proj}} = \mathbf{V}_m\mathbf{V}_m^{\mathrm{T}}\left(\widehat{\mathbf{h}}_m^+ - \overline{\mathbf{h}}_m\right) + \overline{\mathbf{h}}_m\,. \tag{30}$$

The algorithm is modified slightly for the sake of a fair comparison. In [16] the authors propose to span the subspace without use of an EVD. In our implementation, however, we find the subspace by (17) – (19). Moreover, in [16] a soft projection is proposed:

$$\widehat{\mathbf{h}}_{m,\text{comb}} = \mathbf{w}\odot\widehat{\mathbf{h}}_{m,\text{proj}} + (\mathbf{1}_l - \mathbf{w})\odot\widehat{\mathbf{h}}_m^+\,. \tag{31}$$

Transferring this concept into our implementation, we choose the projection weights

$$\mathbf{w} = \operatorname{diag}\mathbf{P}_m^+ \oslash \left(\operatorname{diag}\mathbf{P}_m^+ + \operatorname{diag}\mathbf{Q}_{\widehat{\mathbf{h}}}\right) \tag{32}$$

where $\mathbf{Q}_{\widehat{\mathbf{h}}}$ is the recursively estimated covariance matrix of the filter state $\widehat{\mathbf{h}}_m^+$. Lastly, our implementation is entirely in time domain. To investigate the proposed covariance update (24), we also simulate our algorithm without this update, e.g., $\mathbf{P}_{m+1}^= = \mathbf{P}_m^+$.

4.4 Implementation Details

To eliminate the influence of an imperfect measurement noise estimation, the (generally time-variant) measurement noise covariance in (12a) is set to the constant value $\mathbf{Q}_\mathbf{n} = \sigma_n^2\mathbf{I}_r$. The noise power σ_n^2 is found as $\mathrm{E}\left\{n^2(k)\right\}$ and is known and static in our setup. For the online estimation of the process noise covariance $\mathbf{Q}_{\boldsymbol{\delta}}$ we follow [21] in using the recursive average of the weight update.

$$\mathbf{Q}_{\boldsymbol{\delta},m+1} = \alpha\mathbf{Q}_{\boldsymbol{\delta},m} + (1-\alpha)\operatorname{diag}\left(\Delta\mathbf{z}_m^2\right)\,. \tag{33}$$

Here, $\Delta\mathbf{z}_m^2$ is the element wise squared state update and α is a scalar smoothing factor. In line with [21], we also use a fading factor $\gamma = 1$. For the recursive estimation of $\mathbf{Q}_{\boldsymbol{\delta}}$ and $\mathbf{Q}_{\widehat{\mathbf{h}}}$ we set $\alpha = 0.973$, which corresponds to a time constant of 150 ms.

5 Results

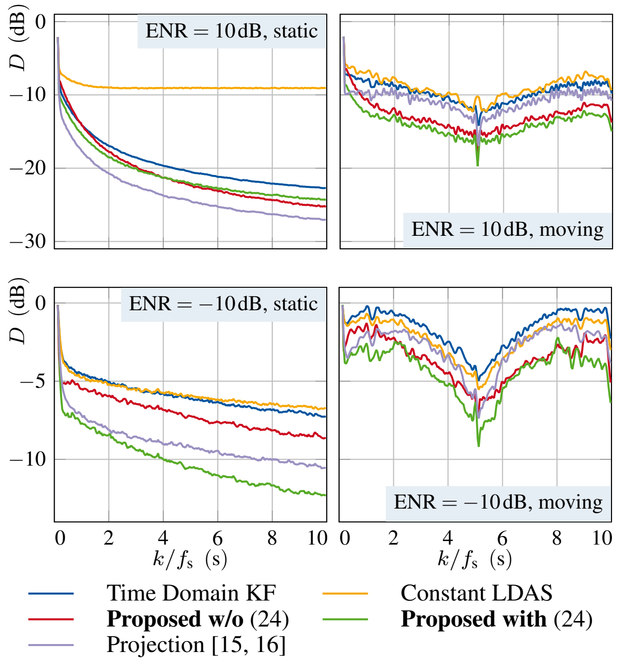

Figure 4: Relative system distance for the considered algorithms.

Figure 4 shows the relative system distance over time. In the static scenario with ENR = 10 dB, the KF on a single subspace cannot achieve a low relative system distance because the global subspace cannot model the impulse response well. In the remaining three scenarios, the KF on one subspace exhibits a performance comparable to the time domain KF. From this we conclude that the rotation into a single LDAS alone has no significant benefit over the time domain KF. The models that exploit local LDAS, however, are superior in all experiments.

In the static scenario with ENR = 10 dB the projection algorithm [16] is superior to the proposed approaches. A reason might be that due to the proposed soft projection the algorithm is not limited to the local subspace. The proposed variants however achieve lower system distances on the time-variant scenarios and when the ENR is low. As expected, they are more robust against imperfect estimation of the process noise covariance, since only the covariance in relevant directions is considered. Imperfectly estimated covariances in directions that exceed the subspace are discarded and hence do not influence the measurement update. Comparing the proposed algorithm with and without the covariance update (24), it can be seen that the covariance update improves system identification.

In the time-variant scenarios all algorithms achieve their lowest relative system distance at around 5 s. Figure 3 shows that at this time the location of the first reflection in the impulse response does not change, which favors convergence.

6 Summary and Outlook

In this paper, we have introduced an ASI algorithm that operates on a low dimensional affine subspace. This subspace is updated in every time step, using the principal directions of neighboring reference impulse responses. We presented update rules for the state and the state error covariance to account for this change of basis. Simulative results under stationary and time-variant conditions underline the benefits of the proposed algorithm. It also shows robustness against low echo-to-noise ratios. The results also suggest an improvement of the tracking performance in adverse echo-to-noise conditions.

References

[1] 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.

[2] H. Schepker, L. T. T. Tran, S. Nordholm, and S. Doclo, "Improving adaptive feedback cancellation in hearing aids using an affine combination of filters," in *International Conference on Acoustics, Speech, and Signal Processing*, pp. 231–235, 2016.

[3] S. Kühl, C. Anemueller, C. Antweiler, F. Heese, P. Vicinus, and P. Jax, "Feedback cancellation for IP-based teleconferencing systems," in *ITG Conference on Speech Communication*, pp. 1–5, VDE, 2021.

[4] G. Enzner, "Analysis and optimal control of LMS-type adaptive filtering for continuous-azimuth acquisition of head related impulse responses," in *International Conference on Acoustics, Speech, and Signal Processing*, pp. 393–396, IEEE, 2008.

[5] S. Nagel, T. Kabzinski, S. Kühl, C. Antweiler, and P. Jax, "Acoustic head-tracking for acquisition of head-related transfer functions with unconstrained subject movement," in *International Conference on Audio for Virtual and Augmented Reality*, Audio Engineering Society, 2018.

[6] J. Fabry, F. König, S. Liebich, and P. Jax, "Online secondary path estimation with masked auxiliary noise for active noise control," in *International Workshop on Acoustic Echo and Noise Control*, pp. 331–335, IEEE, 2018.

[7] G. Enzner, "Bayesian inference model for applications of time-varying acoustic system identification," in *European Signal Processing Conference*, pp. 2126–2130, IEEE, 2010.

[8] T. Kabzinski and P. Jax, "A unified perspective on time-domain and frequency-domain Kalman filters for acoustic system identification," in *European Signal Processing Conference*, pp. 90–94, IEEE, 2022.

[9] S. Haykin, *Adaptive filter theory*. Pearson Education Limited, fifth ed., 2014.

[10] C. Antweiler, J. Grunwald, and H. Quack, "Approximation of optimal step size control for acoustic echo cancellation," in *International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, pp. 295–298, IEEE, 1997.

[11] P. Vary and R. Martin, *Digital speech transmission: Enhancement, coding and error concealment*. John Wiley & Sons, 2006.

[12] B. Laufer-Goldshtein, R. Talmon, and S. Gannot, "A study on manifolds of acoustic responses," in *Latent Variable Analysis and Signal Separation*, pp. 203–210, Springer, 2015.

[13] M. Fozunbal, T. Kalker, and R. W. Schafer, "Multi-channel echo control by model learning," in *International Workshop on Acoustic Echo and Noise Control*, IEEE, 2008.

[14] R. Talmon, I. Cohen, S. Gannot, and R. R. Coifman, "Diffusion maps for signal processing: A deeper look at manifold-learning techniques based on kernels and graphs," *Signal Processing Magazine*, vol. 30, no. 4, pp. 75–86, 2013.

[15] T. Haubner, A. Brendel, and W. Kellermann, "Online supervised acoustic system identification exploiting prelearned local affine subspace models," in *International Workshop on Machine Learning for Signal Processing*, pp. 1–6, IEEE, 2020.

[16] T. Haubner, A. Brendel, and W. Kellermann, "Online acoustic system identification exploiting Kalman filtering and an adaptive impulse response subspace model," *Journal of Signal Processing Systems*, vol. 94, no. 2, pp. 147–160, 2022.

[17] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, 1960.

[18] D. P. Mandic, S. Kanna, and A. G. Constantinides, "On the intrinsic relationship between the least mean square and Kalman filters [lecture notes]," *Signal Processing Magazine*, vol. 32, no. 6, pp. 117–122, 2015.

[19] R. Scheibler, E. Bezzam, and I. Dokmanic, "Pyroomacoustics: A Python package for audio room simulation and array processing algorithms," in *International Conference on Acoustics, Speech, and Signal Processing*, IEEE, 2018.

[20] C. Urbanietz and G. Enzner, "Binaural rendering of dynamic head and sound source orientation using high-resolution HRTF and retarded time," in *International Conference on Acoustics, Speech, and Signal Processing*, pp. 566–570, 2018.

[21] C. Paleologu, J. Benesty, S. Ciochină, and S. L. Grant, "A Kalman filter with individual control factors for echo cancellation," in *International Conference on Acoustics, Speech, and Signal Processing*, pp. 5974–5978, 2014.