Relative Transfer Function Estimation Based on Extended Kalman Filter

Thomas Dietzen, Marc Moonen, Toon van Waterschoot

KU Leuven, Dept. of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, Leuven, Belgium.

***Abstract*—Relative transfer function estimation is a required in a number of signal processing tasks such as the design of a beamformer. In this technical report, we present a state space model of the relative transfer functions employing a non-linear observation equation based on the concept of a blocking matrix. The relative transfer functions are estimated using the the extended Kalman filter. In order to avoid convergence to local minima, we propose to adjust the state equation parameters based on the assumption that the mean power over all RTFs lies in the proximity of one.**

I. Signal Model

In the short-time Fourier transform (STFT) domain, with l and k indexing the frame and the frequency bin, respectively, let $y_m(l,k)$ with $m = 1,\dots,M$ denote the mth microphone signal. In the following, we treat all frequency bins independently and hence omit the frequency index. We define the stacked microphone signal vector[1]

$$\mathbf{y}(l) = \begin{pmatrix} y_1(l) & \cdots & y_M(l) \end{pmatrix}^T = \begin{pmatrix} y_1(l) & \mathbf{y}_{2:M}^T(l) \end{pmatrix}^T, \quad (1)$$

composed of the speech component $\mathbf{x}(l)$ and the noise component $\mathbf{v}(l)$, defined analogously to (1),

$$\mathbf{y}(l) = \mathbf{x}(l) + \mathbf{v}(l). \quad (2)$$

The speech components of the individual microphone signals are related by the possibly time-varying relative transfer functions (RTFs) $\mathbf{h}(l)$, defined relative to the transfer function of the first microphone, i.e.

$$\mathbf{x}(l) = \mathbf{h}(l)x_1(l), \quad (3)$$

$$\mathbf{h}(l) = \begin{pmatrix} 1 & h_2(l) & \cdots & h_M(l) \end{pmatrix}^T = \begin{pmatrix} 1 & \mathbf{h}_{2:M}^T(l) \end{pmatrix}^T. \quad (4)$$

We assume $\mathbf{x}(l)$ and $\mathbf{v}(l)$ to be uncorrelated,

$$\mathbf{\Psi}_y(l) = \mathbf{\Psi}_x(l) + \mathbf{\Psi}_v(l), \quad (5)$$

with the microphone signal correlation matrix $\mathbf{\Psi}_y(l) = E\left\{\mathbf{y}(l)\mathbf{y}^H(l)\right\}$ and the speech and noise component correlation matrices $\mathbf{\Psi}_x(l)$ and $\mathbf{\Psi}_v(l)$ defined analogously. We further assume that $\mathbf{v}(l)$ can be modeled as a white signal, i.e.

$$\mathbf{\Psi}_{v|i}(l) = E\left\{\mathbf{v}(l-i)\mathbf{v}^H(l)\right\} = \mathbf{0} \quad \forall \quad i \neq 0. \quad (6)$$

Due to random phase changes from one frame to the next, this assumption can be considered to be justified if the frame length and the frame shift are sufficiently large.

[1]Notation: in the following, vectors are denoted by lower-case boldface letters, matrices by upper-case boldface letters, with zero and identity matrices denoted by $\mathbf{0}$ and $\mathbf{I}$, respectively. The superscripts $(\cdot)^T$, $(\cdot)^*$, and $(\cdot)^H$ denote the transpose, the conjugate, and the conjugate transpose operation, respectively. The expected value operation is denoted by $E\{\cdot\}$.

II. Extended Kalman Filter

We intend to estimate the RTFs $\mathbf{h}_{2:M}(l)$ from the microphone signals $\mathbf{y}(l)$ and the presumably known noise correlation matrix $\mathbf{\Psi}_v(l)$ using Kalman filter theory. We define the process and observation equation in Section II-A, whereby the observation equation will be non-linear. The extended Kalman filter update equations, based on the linearization of the observation equation, are presented in Section II-B.

A. State Equations

We define the RTF vector $\mathbf{h}_{2:M}(l)$ as the hidden state to be estimated and model the state evolution in form of a first-order difference equation, called process equation, i.e.

$$\mathbf{h}_{2:M}(l) = \alpha(l)\mathbf{h}_{2:M}(l-1) + \mathbf{h}_{2:M|\Delta}(l), \quad (7)$$

with $\alpha(l) \leq 1$ the state transition factor and $\mathbf{h}_{2:M|\Delta}(l)$ the unknown white process noise with presumably known correlation matrix $\mathbf{\Psi}_{h_{2:M|\Delta}}(l) = \psi_{h_{2:M|\Delta}}(l)\mathbf{I}$. In practice, rather than representing a physically motivated model, both $\alpha(l)$ and $\psi_{h_{2:M|\Delta}}(l)$ are commonly considered tuning parameters. In our algorithm, in order to control the convergence behavior of the extended Kalman filter, we update $\alpha(l)$ and $\psi_{h_{2:M|\Delta}}(l)$ based on the power of the recent state estimate. The update strategy will be discussed in Section II-B.

We define $\mathbf{y}_{2:M}(l)$ to be the observed signal and $\mathbf{v}(l)$ to be the observation noise with presumably known correlation matrix $\mathbf{\Psi}_v(l)$. Similarly to the process noise, the observation noise is required to be white, cf. (6). A relation between the observed signal $\mathbf{y}_{2:M}(l)$, the hidden state $\mathbf{h}_{2:M}(l)$, and the observation noise $\mathbf{v}(l)$ may be established using the concept of a blocking matrix, denoted by $\mathbf{B}(l)$ in the following. Per definition, the ideal blocking matrix is orthogonal to $\mathbf{h}(l)$, i.e. $\mathbf{B}^H(l)\mathbf{h}(l) = \mathbf{0}$, such that $\mathbf{B}(l)$ perfectly blocks the speech component $\mathbf{x}(l) = \mathbf{h}(l)x_1(l)$ of $\mathbf{y}(l)$ in (2), i.e. we yield

$$\mathbf{B}^H(l)\mathbf{y}(l) = \mathbf{B}^H(l)\mathbf{v}(l). \quad (8)$$

A blocking matrix design satisfying (8) is given by

$$\mathbf{B}(l) = \begin{pmatrix} -\mathbf{h}_{2:M}^H(l) \\ \mathbf{I} \end{pmatrix}. \quad (9)$$

The observation equation is then derived from (8) by inserting (9) on the left-hand side and rearranging,

$$\mathbf{y}_{2:M}(l) = y_1(l)\mathbf{h}_{2:M}(l) + \mathbf{B}^H(l)\mathbf{v}(l). \quad (10)$$

Note that due to the product $\mathbf{B}^H(l)\mathbf{v}(l)$, where $\mathbf{B}(l)$ depends on the state $\mathbf{h}_{2:M}(l)$ to be estimated, the observed signal $\mathbf{y}_{2:M}(l)$ in (10) is non-linearly related to $\mathbf{h}_{2:M}(l)$ and the observation noise $\mathbf{v}(l)$. Following the extended Kalman filter paradigm, we linearize $\mathbf{B}^H(l)\mathbf{v}(l)$ around the current Kalman filter state estimate $\hat{\mathbf{h}}_{2:M}(l)$ and $\mathbf{v}(l) = \mathbf{0}$. Effectively, this corresponds to substituting $\mathbf{B}(l)$ by $\hat{\mathbf{B}}(l)$, where $\hat{\mathbf{B}}(l)$ is the estimate of $\mathbf{B}(l)$ based on $\hat{\mathbf{h}}_{2:M}(l)$, such that (10) is linearly approximated by

$$\mathbf{y}_{2:M}(l) \approx y_1(l)\mathbf{h}_{2:M}(l) + \hat{\mathbf{B}}^H(l)\mathbf{v}(l). \tag{11}$$

B. Update Equations

The update equations of the extended Kalman filter are derived by applying standard Kalman filter theory to the linearized system given by (7) and (11), yielding

$$\hat{\mathbf{h}}_{2:M}(l) = \alpha(l)\hat{\mathbf{h}}^+_{2:M}(l-1), \tag{12}$$
$$\mathbf{\Psi}_{h_{2:M}}(l) = \alpha^2(l)\mathbf{\Psi}^+_{h_{2:M}}(l-1) + \psi_{h_{2:M|\Delta}}(l)\mathbf{I}, \tag{13}$$
$$\begin{aligned}\mathbf{u}(l) &= \mathbf{y}_{2:M}(l) - y_1(l)\hat{\mathbf{h}}_{2:M}(l)\\ &= \hat{\mathbf{B}}^H(l)\mathbf{y}(l),\end{aligned} \tag{14}$$
$$\mathbf{\Psi}_u(l) = \|y_1(l)\|^2\mathbf{\Psi}_{h_{2:M}}(l) + \hat{\mathbf{B}}^H(l)\mathbf{\Psi}_v(l)\hat{\mathbf{B}}(l), \tag{15}$$
$$\mathbf{K}(l) = y_1^*(l)\mathbf{\Psi}_{h_{2:M}}(l)\mathbf{\Psi}_u^{-1}(l), \tag{16}$$
$$\hat{\mathbf{h}}^+_{2:M}(l) = \hat{\mathbf{h}}_{2:M}(l) + \mathbf{K}(l)\mathbf{u}(l), \tag{17}$$
$$\mathbf{\Psi}^+_{h_{2:M}}(l) = \mathbf{\Psi}_{h_{2:M}}(l) - y_1(l)\mathbf{K}(l)\mathbf{\Psi}_{h_{2:M}}(l). \tag{18}$$

Eq. (12)–(13) are referred to as the time update of the state estimate $\hat{\mathbf{h}}_{2:M}(l)$ and the state error correlation matrix $\mathbf{\Psi}_{h_{2:M}}(l)$, which is determined by $\alpha(l)$ and $\psi_{h_{2:M|\Delta}}(l)$. In (14)–(16), the blocking matrix output signal $\mathbf{u}(l)$, its correlation matrix $\mathbf{\Psi}_u(l)$, and the Kalman gain $\mathbf{K}(l)$ are computed. Eq. (17)–(18) are referred to as the measurement update of the state estimate $\hat{\mathbf{h}}^+_{2:M}(l)$ and the state error correlation matrix $\mathbf{\Psi}^+_{h_{2:M}}(l)$, which is determined by $\mathbf{u}(l)$ and $\mathbf{K}(l)$. The extended Kalman filter is initialized with $\hat{\mathbf{h}}^+_{2:M}(0)$ and $\mathbf{\Psi}^+_{h_{2:M}}(0)$.

Over time, the state estimate should converge to the hidden state, i.e. $\hat{\mathbf{h}}_{2:M}(l) \to \mathbf{h}_{2:M}(l)$ and $\mathbf{\Psi}_{h_{2:M}}(l) \to \mathbf{0}$, such that $\mathbf{u}(l) \to \mathbf{B}^H(l)\mathbf{v}(l)$ and $\mathbf{\Psi}_u(l) \to \mathbf{B}^H(l)\mathbf{\Psi}_v(l)\mathbf{B}(l)$. However, due to the linear approximation (11) of the originally non-linear observation equation (10), the algorithm needs to be carefully controlled in order to avoid local minima. This can be done via the process equation parameters $\alpha(l)$ and $\psi_{h_{2:M|\Delta}}(l)$, cf. (12)–(13). We adjust both parameters based on the assumption that the mean power over all RTFs, given by $\|\mathbf{h}_{2:M}(l)\|^2/(M-1)$, lies somewhere in the proximity of one. In order to promote estimates of that kind, we first compute the power of the recent state estimate $\hat{\mathbf{h}}^+_{2:M}(l-1)$ in dB,

$$\eta(l) = 10\log_{10}\left(\frac{\|\hat{\mathbf{h}}^+_{2:M}(l-1)\|^2}{M-1}\right)\text{dB}, \tag{19}$$

and then, using a Gaussian function, update $\alpha(l)$ and $\psi_{h_{2:M|\Delta}}(l)$ in the following manner,

$$\phi_{h_{2:M|\Delta}}(l) = 1 - \exp\left(-\frac{\eta^2(l)}{2\sigma^2}\right), \tag{20}$$
$$\psi_{h_{2:M|\Delta}}(l) = \beta\psi_{h_{2:M|\Delta}}(l-1) + (1-\beta)\phi_{h_{2:M|\Delta}}(l), \tag{21}$$
$$\alpha(l) = \sqrt{1 - \psi_{h_{2:M|\Delta}}(l)}, \tag{22}$$

where σ in (20) controls the width of the Gaussian function, $\beta \leq 1$ in (21) is an exponential smoothing parameter and the initial value $\psi_{h_{2:M|\Delta}}(0) \leq 1$. Note that both $\psi_{h_{2:M|\Delta}}(l)$ and $\alpha(l)$ reside in the interval $[0, 1]$. If $\eta(l)$ remains close to 0 dB, then $\psi_{h_{2:M|\Delta}}(l)$ remains close to zero and $\alpha(l)$ close to one, expressing trust in the recent state estimate. In contrast, if $\eta(l)$ diverges from 0 dB, then $\psi_{h_{2:M|\Delta}}(l)$ increases while $\alpha(l)$ decreases, expressing distrust in the recent state estimate.