LECTURE NOTES IN AUDIO ANALYSIS: ON LCMV NOISE REDUCTION FILTERS FOR THE LINEAR MODEL

October 27, 2016

Mads Græsbøll Christensen
Audio Analysis Lab, AD:MT
Aalborg University

Abstract

This lecture note is concerned with how to design distortionless noise reduction filters for the linear model by imposing different constraints on the filter. The principle is basically that of LCMV and is based on the ideas first presented in [1], and the derivations herein apply to a number of different models, including the harmonic model and the harmonic chirp model when the fundamental frequency is known. As such, the note elaborates on the theory presented in our paper [2], which perhaps was a bit unclear.

Signal Model and Distortionless Filter

The model of the observed signal for addive noise is given by,

$$\mathbf{x}(n) = \mathbf{s}(n) + \mathbf{e}(n) \tag{1}$$

where $\mathbf{s}(n)$ is the speech and $\mathbf{e}(n)$ the noise. The speech signal vector is defined as

$$\mathbf{s}(n) = [\ s(n)\ s(n-1)\ \cdots\ s(n-M+1)\ ]^T, \tag{2}$$

and similarly for the observed signal, $\mathbf{x}(n)$ and the noise, $\mathbf{e}(n)$. In what follows, we will make use of the covariance matrix of the noise, which is defined as

$$\mathbf{R}_{\mathbf{e}} = \mathrm{E}\left\{\mathbf{e}(n)\mathbf{e}^H(n)\right\}. \tag{3}$$

We assume that this matrix is invertible. We now apply a noise reduction filter $\mathbf{h}$ to the observed signal to obtain an estimate, $\hat{s}(n)$, of the speech sample $s(n)$, i.e.,

$$\hat{s}(n) = \mathbf{h}^H\mathbf{x}(n) = \mathbf{h}^H\mathbf{s}(n) + \mathbf{h}^H\mathbf{e}(n). \tag{4}$$

Note that this is the first entry of the vector $\mathbf{s}(n)$. Many models can be put into the following form:

$$\mathbf{s}(n) = \mathbf{Z}(n)\mathbf{a}, \tag{5}$$

where $\mathbf{Z}(n) \in \mathbb{C}^{M\times L}$ and $\mathbf{a} \in \mathbb{C}^L$. In what follows, we require that $L \leq M$ and that $\mathbf{Z}(n)$ has full rank. Examples of models that can be put into this form include the harmonic model, and also the harmonic chirp model when the fundamental frequency is known (i.e., estimated in advance). We refer the interested reader to the papers [1] and [3] for more details regarding this. Replacing $\mathbf{s}(n)$ by this model in the above equations, we obtain

$$\begin{aligned}\hat{s}(n) &= \mathbf{h}^H\mathbf{x}(n) &(6)\\ &= \mathbf{h}^H\mathbf{Z}(n)\mathbf{a} + \mathbf{h}^H\mathbf{e}(n). &(7)\end{aligned}$$

As can be seen, the filter affects both the speech, possibly incurring speech distortion, and the noise, resulting in noise reduction. Thus, to obtain a distortionless estimate of the speech sample, we can impose the constraint

$$\mathbf{h}^H \mathbf{Z}(n) = \mathbf{b}_1^T \mathbf{Z}(n) \tag{8}$$

where $\mathbf{b}_m =$ is the mth column of the $M \times M$ identity matrix. We can now pose the noise reduction problem as the following quadratic optimization problem with linear constriants [1]:

$$\min_{\mathbf{h}} \mathbf{h}^H \mathbf{R}_{\mathbf{e}} \mathbf{h} \quad \text{s.t.} \quad \mathbf{h}^H \mathbf{Z}(n) = \mathbf{b}_1^T \mathbf{Z}(n) \tag{9}$$

It has the well-known solution

$$\mathbf{h}^\star = \mathbf{R}_{\mathbf{e}}^{-1} \mathbf{Z}(n) \left(\mathbf{Z}^H(n) \mathbf{R}_{\mathbf{e}}^{-1} \mathbf{Z}(n) \right)^{-1} \mathbf{Z}^H(n) \mathbf{b}_1. \tag{10}$$

The filter is distortionless and has optimal output SNR. Note that when $L = M$, the solution is trivial.

To obtain an estimate of an arbitrary sample, say $\hat{s}(n-m)$, we again apply a filter, i.e.,

$$\begin{aligned} \hat{s}(n-m) &= \mathbf{h}^H \mathbf{x}(n) & (11) \\ &= \mathbf{h}^H \mathbf{Z}(n) \mathbf{a} + \mathbf{h}^H \mathbf{e}(n). & (12) \end{aligned}$$

From this we see that the constraint now should be

$$\mathbf{h}^H \mathbf{Z}(n) = \mathbf{b}_m^T \mathbf{Z}(n) \tag{13}$$

and the corresponding optimal filter is given by

$$\mathbf{h}^\star = \mathbf{R}_{\mathbf{e}}^{-1} \mathbf{Z}(n) \left(\mathbf{Z}^H(n) \mathbf{R}_{\mathbf{e}}^{-1} \mathbf{Z}(n) \right)^{-1} \mathbf{Z}^H(n) \mathbf{b}_m. \tag{14}$$

We can also modify the constraints to obtain an estimate of the entire vector $\mathbf{s}(n)$ as

$$\begin{aligned} \hat{\mathbf{s}}(n) &= \mathbf{H}^H \mathbf{x}(n) & (15) \\ &= \mathbf{H}^H \mathbf{Z}(n) \mathbf{a} + \mathbf{H}^H \mathbf{e}(n). & (16) \end{aligned}$$

It then follows that for the filter to be distortionless, it must satisfy

$$\mathbf{H}^H \mathbf{Z}(n) = \mathbf{Z}(n), \tag{17}$$

which leads to the following optimization problem:

$$\min_{\mathbf{H}} \operatorname{Tr} \left\{ \mathbf{H}^H \mathbf{R}_{\mathbf{e}} \mathbf{H} \right\} \quad \text{s.t.} \quad \mathbf{H}^H \mathbf{Z}(n) = \mathbf{Z}(n). \tag{18}$$

The solution is given by

$$\mathbf{H}^\star = \mathbf{R}_{\mathbf{e}}^{-1} \mathbf{Z}(n) \left(\mathbf{Z}^H(n) \mathbf{R}_{\mathbf{e}}^{-1} \mathbf{Z}(n) \right)^{-1} \mathbf{Z}^H(n). \tag{19}$$

As we can see, we can extract a distortionless estimate of the entire speech vector, $\mathbf{s}(n)$, this way. Note that to apply the filters in practice, the noise covariance matrix, $\mathbf{R}_{\mathbf{e}}$, has to be estimated somehow.

Literature

[1] M. G. Christensen and A. Jakobsson, “Optimal filter designs for separating and enhancing periodic signals,” *IEEE Trans. Signal Process.*, vol. 58(12), pp. 5969–5983, 2010.

[2] S. M. Nørholm, J. R. Jensen, and M. G. Christensen, “Enhancement and noise statistics estimation for non-stationary voiced speech,” *IEEE Trans. Audio, Speech, Language Process.*, vol. 24(4), pp. 645–658, 2016.

[3] M. G. Christensen, J. L. Højvang, A. Jakobsson, and S. H. Jensen, “Joint fundamental frequency and order estimation using optimal filtering,” *EURASIP J. on Advances in Signal Process.*, vol. 2011(1), pp. 1–13, 2011.