Sub-space analysis revisited

Tom Bäckström

February 5, 2007

1 Eigenanalysis

Let $\mathbf{B}$ be a $N \times N$ matrix. The *eigenvectors* $\mathbf{u}_k$ of $\mathbf{B}$ are vectors for which it holds that

$$\mathbf{B}\mathbf{u}_k = \lambda_k \mathbf{u}_k \tag{1}$$

where λ_k is the *eigenvalue* corresponding to $\mathbf{u}_k$. If $\mathbf{u}_k$ and $\mathbf{u}_h$ are both eigenvectors of $\mathbf{B}$ and $\mathbf{B}$ is symmetric ($\mathbf{B} = \mathbf{B}^T$) then

$$\lambda_h \mathbf{u}_h^T \mathbf{u}_k = \mathbf{u}_h^T \mathbf{B}^T \mathbf{u}_k = \mathbf{u}_h^T \mathbf{B} \mathbf{u}_k = \lambda_k \mathbf{u}_h^T \mathbf{u}_k. \tag{2}$$

It follows that if $\lambda_k \neq \lambda_h$, then we must have $\mathbf{u}_h^T \mathbf{u}_k = 0$. In other words, the eigenvectors are orthogonal[1]. Moreover, we can choose the scaling of the vectors $\mathbf{u}_k$ such that $||\mathbf{u}_k||^2 = \mathbf{u}_k^T \mathbf{u}_k = 1$ and then

$$\mathbf{u}_h^T \mathbf{u}_k = \begin{cases} 1 & \text{if } h = k \\ 0 & \text{if } h \neq k. \end{cases} \tag{3}$$

Define $\mathbf{U} = [\mathbf{u}_0 \dots \mathbf{u}_{N-1}]$ and

$$\mathbf{D} = \operatorname{diag}(\lambda_k) = \begin{bmatrix} \lambda_0 & 0 & \dots & 0 \\ 0 & \lambda_1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \lambda_{N-1} \end{bmatrix}. \tag{4}$$

Then[2] $\mathbf{U}^T \mathbf{U} = \mathbf{U}\mathbf{U}^T = \mathbf{I}$ and

$$\mathbf{B}\mathbf{U} = \mathbf{U}\mathbf{D}. \tag{5}$$

[1]We can readily show that eigenvectors with equal eigenvalues can be chosen in such a way that they are orthogonal, since such eigenvectors are not unique.

[2]Quite strictly speaking, we have shown only $\mathbf{U}^T \mathbf{U} = \mathbf{I}$, but also $\mathbf{U}\mathbf{U}^T = \mathbf{I}$ is true.

Multiplying from the left by $\mathbf{U}^T$ yields

$$\mathbf{U}^T\mathbf{B}\mathbf{U} = \mathbf{U}^T\mathbf{U}\mathbf{D} = \mathbf{D}. \tag{6}$$

Similarly, multiplying Eq. 5 from the right by $\mathbf{U}^T$ yields

$$\mathbf{B}\mathbf{U}\mathbf{U}^T = \boxed{\mathbf{B} = \mathbf{U}\mathbf{D}\mathbf{U}^T.} \tag{7}$$

This relation is known as the *eigendecomposition* of $\mathbf{B}$.

2 Linear model

Let us assume that we have a speech signal $\mathbf{x} = [x_0 \dots x_N]^T$ corrupted by uncorrelated additive white noise $\mathbf{v} = [v_0 \dots v_N]^T$, whereby the observed signal is $\mathbf{y} = \mathbf{x} + \mathbf{v}$. Our objective is to find an estimate $\hat{\mathbf{x}}$ of $\mathbf{x}$ using a linear model

$$\hat{\mathbf{x}} = \mathbf{H}\mathbf{y}. \tag{8}$$

An optimal model $\mathbf{H}$ would have $\mathbf{H}\mathbf{x} = \mathbf{x}$ and $\mathbf{H}\mathbf{v} = 0$, since then we would have $\hat{\mathbf{x}} = \mathbf{H}\mathbf{y} = \mathbf{H}(\mathbf{x} + \mathbf{v}) = \mathbf{x}$.

Let us assume that the speech signal $\mathbf{x}$ has such a structure that it can be written as

$$\mathbf{x} = \mathbf{A}\mathbf{c} \tag{9}$$

where $\mathbf{A}$ is an $N \times M$ matrix with $N > M$, that is, $\mathbf{A}$ is a rank-deficient matrix and the above equation has more equations the free parameters in $\mathbf{c} = [c_0 \dots c_{M-1}]^T$. The description of $\mathbf{x}$ using $\mathbf{c}$ is thus a compression of the information in to a more compact form.

We can then try to find such a vector $\mathbf{c}$ that $\mathbf{A}\mathbf{c}$ explains as much of $\mathbf{y}$ as possible, that is,

$$\hat{\mathbf{c}} = \arg\min_{\mathbf{c}} ||\mathbf{y} - \mathbf{A}\mathbf{c}||^2. \tag{10}$$

The solution is readily found as the pseudo-inverse

$$\hat{\mathbf{c}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y} \tag{11}$$

whereby the estimated signal is

$$\hat{\mathbf{x}} = \mathbf{A}\hat{\mathbf{c}} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}. \tag{12}$$

Our linear model is then $\mathbf{H} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$, but unfortunately, we do not know how to choose $\mathbf{A}$.

In order to find a $\mathbf{A}$, let us first study the autocorrelation of $\mathbf{x}$

$$\mathbf{R}_x = E[\mathbf{x}\mathbf{x}^T] = E[\mathbf{A}\mathbf{c}\mathbf{c}^T\mathbf{A}^T] = \mathbf{A}E[\mathbf{c}\mathbf{c}^T]\mathbf{A}^T = \mathbf{A}\mathbf{R}_c\mathbf{A}^T. \tag{13}$$

The eigendecomposition of $\mathbf{R}_c$ is $\mathbf{R}_c = \mathbf{V}\mathbf{D}_1\mathbf{V}^T$ and thus

$$\mathbf{R}_x = \mathbf{A}\mathbf{R}_c\mathbf{A}^T = \mathbf{A}\mathbf{V}\mathbf{D}_1\mathbf{V}^T\mathbf{A}^T = \mathbf{U}_1\mathbf{D}_1\mathbf{U}_1^T \tag{14}$$

where $\mathbf{U}_1 = \mathbf{A}\mathbf{V}$. Then

$$\mathbf{U}_1\mathbf{V}^T = \mathbf{A}\mathbf{V}^T\mathbf{V} = \mathbf{A}. \tag{15}$$

The eigendecomposition of $\mathbf{R}_x$ is

$$\mathbf{R}_x = \mathbf{U}\mathbf{D}\mathbf{U}^T. \tag{16}$$

Notice that while $\mathbf{D}_1$ is $M \times M$, matrix $\mathbf{D}$ is $N \times N$. We therefore can make the choice that $\mathbf{U}$ can be partitioned as $\mathbf{U} = [\mathbf{U}_1\, \mathbf{U}_2]$ and we have

$$\begin{aligned} \mathbf{R}_x = \mathbf{U}\mathbf{D}\mathbf{U}^T &= [\mathbf{U}_1\, \mathbf{U}_2] \begin{bmatrix} \mathbf{D}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2 \end{bmatrix} \begin{bmatrix} \mathbf{U}_1^T \\ \mathbf{U}_2^T \end{bmatrix} \\ &= \mathbf{U}_1\mathbf{D}_1\mathbf{U}_1^T + \mathbf{U}_2\mathbf{D}_2\mathbf{U}_2^T. \end{aligned} \tag{17}$$

Since Eqs. 14 and 17 must both hold, we must have $\mathbf{D}_2 = 0$. In other words, the speech signal does not have non-zero eigenvalues in the space spanned by $\mathbf{U}_2$ and it corresponds to the noise subspace. Note also that

$$\begin{aligned} \underset{N\times N}{\mathbf{I}} &= \mathbf{U}\mathbf{U}^T = \mathbf{U}_1\mathbf{U}_1^T + \mathbf{U}_2\mathbf{U}_2^T \\ \underset{M\times M}{\mathbf{I}} &= \mathbf{U}_1^T\mathbf{U}_1 \\ \underset{(N-M)\times(N-M)}{\mathbf{I}} &= \mathbf{U}_2^T\mathbf{U}_2 \end{aligned} \tag{18}$$

Inserting Eq. 15 into Eq. 12 yields

$$\begin{aligned} \hat{\mathbf{x}} &= \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y} = \mathbf{U}_1\mathbf{V}^T(\mathbf{V}\mathbf{U}_1^T\mathbf{U}_1\mathbf{V}^T)^{-1}\mathbf{V}\mathbf{U}_1^T\mathbf{y} \\ &= \mathbf{U}_1\mathbf{V}^T\mathbf{V}(\mathbf{U}_1^T\mathbf{U}_1)^{-1}\mathbf{V}^T\mathbf{V}\mathbf{U}_1^T\mathbf{y} = \mathbf{U}_1(\mathbf{U}_1^T\mathbf{U}_1)^{-1}\mathbf{U}_1^T(\mathbf{y}) = \mathbf{U}_1\mathbf{U}_1^T\mathbf{y} \end{aligned} \tag{19}$$

since $\mathbf{V}\mathbf{V}^T = \mathbf{V}^T\mathbf{V} = \mathbf{I}$ and $\mathbf{U}_1^T\mathbf{U}_1 = \mathbf{I}$. Therefore, our linear filter is $\mathbf{H} = \mathbf{U}_1^T\mathbf{U}_1$. However, we obtained matrix $\mathbf{U}_1$ from $\mathbf{R}_x$ which we do not have. Fortunately, we can see that the eigenmatrix of $\mathbf{R}_y$ and $\mathbf{R}_x$ is equal by the following derivation.

The covariance of $\mathbf{y}$ is

$$\begin{aligned} \mathbf{R}_y &= E[\mathbf{y}\mathbf{y}^T] = E[(\mathbf{x}+\mathbf{v})(\mathbf{x}+\mathbf{v})^T] \\ &= E[\mathbf{x}\mathbf{x}^T] + E[\mathbf{v}\mathbf{v}^T] + E[\mathbf{x}\mathbf{v}^T + \mathbf{v}\mathbf{x}^T] = \mathbf{R}_x + \mathbf{I}\sigma_v^2 \end{aligned} \tag{20}$$

since $\mathbf{x}$ and $\mathbf{v}$ are uncorrelated and $\mathbf{v}$ is white $\mathbf{R}_v = \sigma_v^2\mathbf{I}$. Since $\mathbf{U}$ is the eigenmatrix of $\mathbf{R}_x$ we have

$$\mathbf{R}_y\mathbf{U} = \mathbf{R}_x\mathbf{U} + \sigma_v^2\mathbf{U} = (\mathbf{D} + \sigma_v^2)\mathbf{U} = \hat{\mathbf{D}}\mathbf{U} \tag{21}$$

where $\hat{\mathbf{D}} = \mathrm{diag}(\lambda_k + \sigma_v^2)$. It follows that the eigenmatrix of $\mathbf{R}_y$ is $\mathbf{U}$ and the eigenvalues are $\hat{\lambda}_k = \lambda_k + \sigma_v^2$.

In other words, all eigenvalues that are (approximately) equal to (or smaller than) σ_v^2 belong to the noise subspace. It follows that we can find $\mathbf{U}_1$ by finding those eigenvectors of $\mathbf{R}_y$ that correspond to eigenvalues λ_k larger than σ_v^2, that is $\lambda_k > \sigma_v^2$.