ECE 6123
Advanced Signal Processing
Introduction

Peter Willett

Fall 2017

1 Filters

1.1 Standard Filters

These are the standard filter types from DSP class.

IIR (infinite-length impulse response) filter

$$y[n] = \sum_{k=0}^{M-1} b_k^* x[n-k] - \sum_{k=1}^{N-1} a_k^* y[n-k] \tag{1}$$

$$H(z) = \frac{\sum_{k=0}^{M-1} b_k^* z^{-k}}{1 + \sum_{k=1}^{N-1} a_k^* z^{-k}} \tag{2}$$

FIR (finite-length impulse response) filter.

$$y[n] = \sum_{k=0}^{M-1} b_k^* x[n-k] \tag{3}$$

$$H(z) = \sum_{k=0}^{M-1} b_k^* z^{-k} \tag{4}$$

Several things are worth mentioning.

- Please note the convention of the complex conjugate on the filter coefficients. We will (usually) work with complex arithmetic, and this turns out to be a convenient representation mostly in terms of the Hermitian transpose.
- A primary concern with IIR filters is stability. Adaptive filters change their coefficients, so one needs to be assured that no change will move the filter to an unstable configuration. In some constrained cases (such

as the adaptive notch filter[1]) this can be done, but in general it is too difficult. Hence almost all adaptive filters are FIR, and we will deal with them exclusively.

- These filters are for temporal signal processing, where causality is a concern. Noncausal signal processing ("smoothing") is of course a possibility, as is multidimensional signal processing.

An FIR filter can he written as

$$y[n] \;=\; \mathbf{w}^H \mathbf{x}_n \tag{5}$$

where

$$\mathbf{x}_n \;\equiv\; (x[n]\; x[n-1]\; x[n-2]\; \ldots\; x[n-M+1])^T \tag{6}$$

represents the input in "shift register" format as a column vector and

$$\mathbf{w} \;\equiv\; (w_0\; w_1\; w_2\; \ldots\; w_{M-1})^T \tag{7}$$

is a column vector containing the filter coefficients. It is common to use $\mathbf{w}$ for these in the optimal signal processing context as opposed to $\mathbf{b}$ as would be expected from standard DSP (4).

1.2 Adaptation

Filters adapt by small movements that we will investigate soon. That is, we have

$$y[n] \;=\; \mathbf{w}_n^H \mathbf{x}_n \tag{8}$$

where $\mathbf{w}_n$ is the filter coefficient vector at time n and

$$\mathbf{w}_n \;=\; \mathbf{w}_{n-1} \;+\; \mu(\Delta\mathbf{w})_n \tag{9}$$

The step size (presumably small) is μ and the direction $(\Delta\mathbf{w})_n$ is a vector that is a function of input, previous output and some "desired" signal $d[n]$ that $y[n]$ is being adaptive to match.

Some canonical structures are

System Identification. The adaptive filter tries to match the structure of some unknown plant. it is assumed the input to the plant is available and $d[n]$ here is the plant's output.

[1]An ANF has transfer function $H(z) = \frac{1-2bz^{-1}+z^{-2}}{1-2\alpha bz^{-1}+\alpha^2 z^{-2}}$ where the b is adapted to control the notch frequency and α is slightly less than unity.

System Inversion. The adaptive filter is placed in series with an unknown plant, and tries to match the input of that plant. The desired signal $d[n]$ here is the input delayed by some suitable number of samples to make the inversion feasible.

Prediction. The desired signal $d[n]$ here is the input signal delayed by some samples, and the goal is to represent the structure of the random process $x[n]$.

Interference Cancelation. The system tries to match whatever is "matchable" in a signal, for example in adaptive noise cancelation.

The last is rather vague, so consider the example

$$d[n] = s[n] + v_1[n] \tag{10}$$
$$x[n] = s[n] + v_2[n] \tag{11}$$

It is clear that based on $\{x[n]\}$ at least some part (i.e. $s[n]$) of $d[n]$ can be matched. The noises $v_i[n]$ remain.

2 Correlation

2.1 Definitions and Properties

This will be very important. We'll assume wide-sense stationarity (wss) for analysis and design and that unless otherwise stated means of zero. We define

$$r[m] \equiv \mathcal{E}\{x[n]x^*[n-m]\} \tag{12}$$

where the convention is important, and we might refer to this as $r_{xx}[m]$ if there is confusion. It is easy to see that

$$r[-m] = r^*[m] \tag{13}$$

As for cross-correlation we define

$$r_{xy}[m] \equiv \mathcal{E}\{x[n]y^*[n-m]\} \tag{14}$$

for two random signals $x[n]$ & $y[n]$. In matrix form we have

$$\mathbf{R} \equiv \mathcal{E}\{\mathbf{x}_n\mathbf{x}_{n-m}^H\} \tag{15}$$

and it is important to stress that $\mathbf{R}$ can be so defined whether $\mathbf{x}_n$ represents a "vectorized" scalar time process as in (6) or whether it is a vector time process[2]. Cross-correlation matrices are be defined similarly; we will define cross-correlation vectors shortly. Note that the $(i,j)^{th}$ element of $\mathbf{R}$ is

$$\mathbf{R}(i,j) \;=\; \mathcal{E}\{\mathbf{x}_n(i)\mathbf{x}_n^*(j)\} \tag{16}$$

which is probably obvious, but in the case of a vectorized *wss* we have $\mathbf{R}(i,j) \;=\; \mathcal{E}\{x[n+1-i]x^*[n+1-j]\}$.

Here are some properties of the correlation matrix. When the proof is obvious it is suppressed.

- It is Hermitian: $\mathbf{R}^H = \mathbf{R}$.

- If $x[n]$ represents a "vectorized" scalar time process as in (6), then the correlation matrix has a special form

$$\mathbf{R} \;=\; \begin{pmatrix} r[0] & r[1] & r[2] & r[3] & \dots & r[M-1] \\ r[-1] & r[0] & r[1] & r[2] & \dots & r[M-2] \\ r[-2] & r[-1] & r[0] & r[1] & \dots & r[M-3] \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ r[-(M-1)] & r[-(M-2)] & r[-(M-3)] & r[-(M-4)] & \dots & r[0] \end{pmatrix} \tag{17}$$

  which is called "Toeplitz." A Toeplitz matrix has constant elements along all super- and sub-diagonals.

- It is non-negative definite.

$$\mathbf{w}^H\mathbf{R}\mathbf{w} = \mathcal{E}\{\mathbf{w}^H\mathbf{x}_n\mathbf{x}_n^H\mathbf{w}\} = \mathcal{E}\{|y[n]|^2\} \geq 0 \tag{18}$$

- Define the "backwards" vector

$$\mathbf{x}_n^B \;\equiv\; (x[n-M+1]\; x[n-M+2]\; x[n-M+3]\; \dots\; x[n])^T \tag{19}$$

  Then

$$\mathbf{R}_B \;\equiv\; \mathcal{E}\{\mathbf{x}_n^B(\mathbf{x}_{n-m}^B)^*\} \;=\; \mathbf{R}^* \;=\; \mathbf{R}^T \tag{20}$$

Please note that the text is for some reason fond of referring to the random process under study as $u[n]$, which I think is a little confusing in light of more typical the unit step nomenclature.

---

[2] An example of a vector time process is that the i^{th} element of $\mathbf{x}_n$ is the measurement from the i^{th} microphone in an array at time n.

2.2 Autoregressive Models

Although we will soon see them again in the context of optimization, we have enough ammunition now to understand them in terms of models. An autoregressive (AR) model is a special case of (2) with unity numerator; that is,

$$y[n] = x[n] - \sum_{k=1}^{N-1} a_k^* y[n-k] \tag{21}$$

$$y[n] = x[n] - \mathbf{a}^H \mathbf{y}_{n-1} \tag{22}$$

$$H(z) = \frac{\sigma_x^2}{1 + \sum_{k=1}^{N-1} a_k^* z^{-k}} \tag{23}$$

where the input $x[n]$ is assumed to be white (and usually but not necessarily Gaussian) with power σ_x^2. Define

$$\mathbf{r} \equiv \mathcal{E}\{\mathbf{y}_{n-1} y[n]^*\} \tag{24}$$

$$= (r[-1]\, r[-2]\ \ldots\ r[-M])^T \tag{25}$$

$$= (r[1]\, r[2]\ \ldots\ r[M])^H \tag{26}$$

Then from (23) we can write

$$\mathbf{r} = \mathcal{E}\{\mathbf{y}_{n-1} y[n]^*\} \tag{27}$$

$$= \mathcal{E}\{\mathbf{y}_{n-1}(x[n] - \mathbf{a}^H \mathbf{y}_{n-1}])^*\} \tag{28}$$

$$= -\mathbf{R}\mathbf{a} \tag{29}$$

in which the only subtlety is that $\mathbf{y}_{n-1}$ and $x[n]$ be independent – the latter is a "future" input to the AR filter. Repeating the last, we have

$$\mathbf{R}\mathbf{a} = -\mathbf{r} \tag{30}$$

in which (30) represents the celebrated "Yule-Walker" equations. Note that since all quantities can be estimated from the data $\{y[n]\}$ (30) provides a way to estimate an AR process from its realization[3].

3 Eigenstuff

3.1 Basic Material

For a general $M \times M$ (square) matrix $\mathbf{A}$ the equation

$$\mathbf{A}\mathbf{q} = \lambda\mathbf{q} \tag{31}$$

[3]The power σ_x^2 needs to be computed separately. We will address this later.

has M solutions in λ, although these may be complex. This is easy to see as (31) implies that the determinant of $(\mathbf{A}-\lambda\mathbf{I})$, which is an M^{th}-order polynomial in λ and hence has M roots, is zero. It is also easy to re-write all solutions of (31) as

$$\begin{aligned}\mathbf{AQ} &= \mathbf{Q\Lambda} &(32)\\ \mathbf{A} &= \mathbf{Q\Lambda Q}^{-1} &(33)\end{aligned}$$

in which

$$\mathbf{Q} \equiv \begin{pmatrix} \uparrow & \uparrow & \dots & \uparrow \\ \mathbf{q}_1 & \mathbf{q}_2 & \ddots & \mathbf{q}_M \\ \downarrow & \downarrow & \dots & \downarrow \end{pmatrix} \qquad \mathbf{\Lambda} \equiv \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_M \end{pmatrix} \tag{34}$$

are the eigenvectors arranged into a column and a diagonal matrix of eigenvalues. We have not shown that $\mathbf{Q}^{-1}$ in general exists for (33), but that is not in scope for this course. By convention eigenvectors are scaled to have unit length.

3.2 Hermitian Matrices

Our matrices will usually be correlation matrices, and these are Hermitian. We have the following:

Eigenvalues are non-negative and real.

$$0 \le \mathbf{q}_i^H \mathbf{R} \mathbf{q}_i = \mathbf{q}_i^H (\lambda_i \mathbf{q}_i) = \lambda_i |\mathbf{q}_i|^2 \tag{35}$$

Eigenvectors are orthonormal.

$$\begin{aligned}\mathbf{q}_i^H \mathbf{R} \mathbf{q}_j &= \mathbf{q}_i^H \mathbf{R} \mathbf{q}_j &(36)\\ \mathbf{q}_i^H (\lambda_j \mathbf{q}_j) &= (\mathbf{q}_i \lambda_i)^H \mathbf{q}_j &(37)\\ \lambda_i \mathbf{q}_i^H \mathbf{q}_j &= \lambda_j \mathbf{q}_i^H \mathbf{q}_j &(38)\end{aligned}$$

For this to be true either $\lambda_i = \lambda_j$ or $\mathbf{q}_i^H \mathbf{q}_j = 0$. For distinct eigenvalues the latter must be true. For $N \le M$ repeated eigenvalues there is a subspace of dimension N (orthogonal to all the eigenvectors with different eigenvalues) that is an eigen-space: any vector within it has the eigen-property (31). By convention we take an orthonormal basis of that eigen-space as the eigenvectors; it doesn't matter much which such basis.

Diagonalization. The analog to (33) is

$$\mathbf{R} \;=\; \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^H \tag{39}$$

since $\mathbf{Q}^{-1} = \mathbf{Q}^H$ – see previous property of orthonormality. Actually

$$\mathbf{R} \;=\; \sum_{i=1}^{M} \lambda_i \mathbf{q}_i^H \mathbf{q}_i \tag{40}$$

is a rather nice way of expressing the same thing.

Matrix trace is sum of eigenvalues and determinant is product. This comes from (39), but actually applies to any matrix $\mathbf{A}$.

3.3 Relation to Power Spectrum

It's perhaps not obvious, but there is only one non-trivial situation where the eigenstuff and DFT have a strong relationship. This is when the correlation is "circulant" for a Toeplitz matrix, meaning

$$r[m] \;=\; r[M+m] \tag{41}$$

In the case that the process is real, this means $r[m] = r[M-m]$ as well: the top row of the Toeplitz matrix is symmetric around its midpoint. Consider

$$\mathbf{q}_p \;=\; \left(1\; e^{j2p\pi/M}\; e^{j4p\pi/M}\; e^{j6p\pi/M}\; \ldots\; e^{j(M-1)p2\pi/M}\right)^H \tag{42}$$

Then the $(m+1)^{st}$ element of the product $\mathbf{R}\mathbf{q}_p$ is

$$(\mathbf{R}\mathbf{q}_p)(m+1) \;=\; \sum_{k=0}^{M-1} r(k-m)e^{-jkp2\pi/M} \tag{43}$$

$$= \sum_{k=0}^{m-1} r(k-m)e^{-jkp2\pi/M} + \sum_{k=m}^{M-1} r(k-m)e^{-jkp2\pi/M} \tag{44}$$

$$= \sum_{k=0}^{m-1} r(M+k-m)e^{-jkp2\pi/M} + \sum_{k=m}^{M-1} r(k-m)e^{-jkp2\pi/M} \tag{45}$$

$$
\begin{aligned}
&= \sum_{k=M-m}^{M-1} r(k)e^{-j(k+m-M)p2\pi/M} \\
&\qquad\qquad + \sum_{k=0}^{M-m-1} r(k)e^{-j(k+m)p2\pi/M} &(46)\\
&= e^{-jmp2\pi/M} \sum_{k=M-m}^{M-1} r(k)e^{-jkp2\pi/M} \\
&\qquad\qquad + \sum_{k=0}^{M-m-1} r(k)e^{-jkp2\pi/M} &(47)\\
&= e^{-jmp2\pi/M} \left(\sum_{k=0}^{M-1} r(k)e^{-jkp2\pi/M} \right) &(48)\\
&= S(p)e^{-jmp2\pi/M} &(49)
\end{aligned}
$$

which implies that the $\mathbf{q}_p$, which is the p^{th} DFT vector, is an eigenvector with eigenvalue the p^{th} element of the power spectrum. One could go backwards from (39) and show that the circulant condition must be true if the DFT relationship holds.

But the DFT and frequency analysis have a fairly strong relationship to Toeplitz covariance matrices, as we shall see. One example is this:

$$
\begin{aligned}
\lambda_i &= \mathbf{q}_i^H \mathbf{R} \mathbf{q}_i &(50)\\
&= \sum_{k=1}^{M} \sum_{l=1}^{M} \mathbf{q}_i(k)^* \mathbf{q}_i(l) r[k-l] &(51)\\
&= \sum_{k=1}^{M} \sum_{l=1}^{M} \mathbf{q}_i(k)^* \mathbf{q}_i(l) \frac{1}{2\pi} \int_{-\pi}^{\pi} S(\omega) e^{j\omega(k-l)} d\omega &(52)\\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} S(\omega) |Q_i(\omega)|^2 d\omega &(53)
\end{aligned}
$$

where

$$
Q_i(\omega) \equiv \sum_{k=0}^{M-1} \mathbf{q}_i(k+1) e^{-j\omega k} \qquad (54)
$$

is the DFT of the eigenvector. Now since

$$
\frac{1}{2\pi} \int_{-\pi}^{\pi} |Q_i(\omega)|^2 d\omega = \sum_{k=1}^{M} |\mathbf{q}_i(k)|^2 \qquad (55)
$$

by Parseval and since this is unity, (53) tells us that

$$
\min_{\omega} \{S(\omega)\} \le \lambda_i \le \max_{\omega} \{S(\omega)\} \qquad (56)
$$

which is nicer looking than it is useful, unfortunately.

At this point it is probably worth looking at a particular case, that of a sinusoid in noise. Suppose

$$x[n] \ = \ ae^{j\omega n} \ + \ \nu[n] \tag{57}$$

where a and $\{\nu[n]\}$ are complex Gaussian, respectively a random variable with variance σ_a^2 and a white noise process with power σ_ν^2. Then with

$$\gamma_\omega \ \equiv \ (1 \, e^{j\omega} \, e^{j2\omega} \ \ldots \ e^{j(M-1)\omega})^T \tag{58}$$

we have

$$\mathbf{R} \ = \ \sigma_a^2 \gamma(\omega)\gamma(\omega)^H \ + \ \sigma_\nu^2 \mathbf{I} \tag{59}$$

The eigenstuff is dominated by one eigenvalue equal to $M\sigma_a^2 + \sigma_\nu^2$ with eigenvector proportional to $\gamma(\omega)$. The other eigenvectors are orthogonal to $\gamma(\omega)$ (of course!) and have common eigenvalue σ_ν^2.