Time-Domain Analysis

Part II

1. Autocorrelation Function
   • Deterministic Signal
   • Voiced speech
   • Random Signal
   • Unvoiced Speech
2. Average Magnitude Difference Function
3. Windowing issues.
4. Pitch Estimation
Autocorrelation Function

• The autocorrelation function of a deterministic signal \( x(n) \), is defined by:

\[
\phi(k) = \sum_{m=-\infty}^{\infty} x(m)x(m + k)
\]

• In practice, we would window the signal so that we can compute the above sum. Assume a rectangular window of length \( N \), then the simple form of the autocorrelation is:

\[
\phi(k) = \sum_{n=0}^{N-k-1} x(n)x(n + k)
\]

Consider the simple case of a sinusoidal signal:

\[
x(m) = A \cdot \cos \left( 2\pi \frac{f_0}{F_s} m + \theta \right)
\]

period \( P = \frac{F_s}{f_0} \)

Original \( x(m) \)

Shifted versions \( x(m-k) \)
Autocorrelation Function

\[ \phi(k) = \sum_{m=-\infty}^{\infty} x(m)x(m+k) \]

Properties of autocorrelation function

\[ \phi(k) = \phi(k+P) \] for periodic functions (period=P)

\[ \phi(k) = \phi(-k) \] symmetry

\[ \phi(k) \leq \phi(0), \forall k \] follows Cauchy’s inequality

\[ \phi(0) \] is the energy of \( x(n) \)
Example for Voiced Signal

Power Spectral Density

Autocorrelation Function

\[ R_x(\tau) \]

\[ F \rightarrow \]

\[ S_x(f) = \int_{-\infty}^{\infty} R_x(\tau) e^{-j2\pi f \tau} d\tau \]

\[ R_x(\tau) \]

\[ F^{-1} \rightarrow \]

\[ S_x(f) = \int_{-\infty}^{\infty} S_x(f) e^{j2\pi f \tau} df \]

\[ P = E[X(t)^2] = R_x(0) \]

\[ P = E[X(t)^2] = \int_{-\infty}^{\infty} S_x(f) df \]
Autocorrelation Function

- A more reasonable definition of the auto-correlation is expressed using the pair of windowed signals:

\[
R_n(k) = \sum_{m=-\infty}^{\infty} x(m)w(n - m)x(m + k)w(n - k - m)
\]

Easy to show

\[
R_n(k) = R_n(-k) = \sum_{m=-\infty}^{\infty} x(m)w(n - m)x(m - k)w(n + k - m)
\]

Another way to view the windowed autocorrelation is by considering the product of the speech signal with its delayed version, passed thru a filter whose response is:

\[
h_k(n) = w(n)w(n + k)
\]

\[
R_n(k) = \sum_{m=-\infty}^{\infty} [x(m)x(m - k)]\cdot[w(n - m)w(n + k - m)]
\]

\[
R_n(k) = \sum_{m=-\infty}^{\infty} [x(m)x(m - k)]\cdot h_k(n - m)
\]
Autocorrelation Function

e.g. N = 100
n = 300

\[ R_n(k) = \sum_{m=n-N+1}^{n-k} x(m)w(n-m)x(m+k)w(n-k-m) \]

\[ k = 0, \ldots, 50 \]

Total number of products in the sum : N – k

\[ R_n(k) = \sum_{m=201}^{300-k} x(m)w(n-m)x(m+k)w(n-m-k) \]

\[ R_n(1) = x(201)w(99)x(202)w(98) + \ldots \]
\[ x(299)w(1)x(300)w(0) \]

\[ R_n(50) = x(201)w(99)x(202)w(98) + \ldots \]
\[ x(250)w(50)x(300)w(0) \]
**Autocorrelation Function**

Recall the simple definition of the autocorrelation (not windowed)

\[ R_{unb}(k) = \frac{1}{N-k} \sum_{m=0}^{m=N-|k|-1} x(m)x(m+k) \]  
**unbiased**

\[ R_{b}(k) = \frac{1}{N} \sum_{m=0}^{m=N-|k|-1} x(m)x(m+k) \]  
**biased**

\[ R_{\text{bias}}(k) = \left(1 - \frac{k}{N}\right)R_{unb}(k) \]  
Biased for finite N, but Asymptotically unbiased for \( N \rightarrow \infty \)

**Autocorrelation voiced speech**  
**Female**

\[ F'_0 = \frac{1}{T_0} \]

\( T_0 \) and time index

\( \text{normalized frequency} \)
Autocorrelation voiced speech

Math equation:

\[ 3F_0 = \frac{1}{3T} \]

Image description:

- Sixteen.wav filtered at 600Hz

Graphs showing autocorrelation and spectrum analysis of voiced speech.
The case of a random process

When $X(t)$ is a stationary random process, then:

The mean of $X(t)$ is

$$\mu_X(t) = E[X(t)]$$

$$= \int_{-\infty}^{\infty} x f_{X(t)}(x) \, dx$$

$$= \mu_X$$

$f_{X(t)}(x)$: the probability density function (pdf)

The autocorrelation function of $X(t)$ is

$$R_X(t_1, t_2) = E[X(t_1)X(t_2)]$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f_{X(t_1), X(t_2)}(x_1, x_2) \, dx_1 \, dx_2$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f_{X(0), X(t_1-t_2)}(x_1, x_2) \, dx_1 \, dx_2$$

$$= R_X(t_2 - t_1)$$
For convenience of notation, we redefine:

\[ R_x(\tau) = E[X(t)X(t-\tau)], \quad \text{for all } t \]

1. The mean-square value
   \[ R_x(0) = E[X^2(t)], \quad \tau = 0 \]

2. The autocorrelation is symmetrical:
   \[ R_x(\tau) = R(-\tau) \]

3. The largest value is at lag 0
   \[ |R_x(\tau)| \leq R_x(0) \]

The case of a random process

The \( R_x(\tau) \) provides the interdependence information of two random variables obtained from \( X(t) \) at times \( \tau \) seconds apart.

![Graph showing the autocorrelation function with slowly and rapidly fluctuating random processes.](image)
Random process: sinusoid with random phase

\[ X(t) = \cos(2\pi f_c t + \Theta) \quad \Theta \sim U(-\pi, \pi) \]

\[ f_\Theta(\theta) = \begin{cases} 
\frac{1}{2\pi}, & -\pi \leq \theta \leq \pi \\
0, & \text{elsewhere} 
\end{cases} \]

Random process: sinusoid with random phase

\[ X(t) = A\cos(2\pi f_c t + \Theta) \]

\[ R_X(\tau) = E[X(t + \tau)X(t)] = \frac{A^2}{2} \cos(2\pi f_c t) \]
Random process: white noise

N=1000;
X=randn(1,N);

m=50;
Rx=Rx_est(X,m);

plot(X)
plot([-m:m],Rx)

Power Spectral Density

for j=1:100,
    Rx=Rx_est(X,m);
    Sx=fftshift(abs(fft(Rx)));
end;
Sx_av=Sx_av+Sx;
Sx_av=Sx_av/100;
Autocorrelation – unvoiced speech

Average Magnitude Difference Function

- Instead of taking average over magnitude, we do it over magnitude difference

\[ \Delta M_p(p) = \sum_{m=n-N+1}^{n} |x(m) - x(m - p)| \]

- Operator \( T(x) \) is the concatenation of a linear filter \( H(z) = 1 - z^{-p} \) and taking absolute value \(|x|\)
- It will be used in a computationally efficient solution to pitch estimation
**Summary**

- **Short-time energy**
  \[ E_n = \sum_{m=n-N+1}^{n} x^2(m) \]

- **Short-time amplitude**
  \[ M_n = \sum_{m=n-N+1}^{n} |x(m)| \]

- **Short-time average zero-cross (ZC) rate**
  \[ Z_n = \sum_{m=n-N+1}^{n} |\text{sgn}[x(m)] - \text{sgn}[x(m-1)]| \]

- **Short-time amplitude difference**
  \[ \Delta M_n(p) = \sum_{m=n-N+1}^{n} |x(m) - x(m-p)| \]

- **Short-time auto-correlation**
  \[ R_n(k) = \sum_{m=-\infty}^{\infty} x(m)w(n-m)x(m+k)w(n-k-m) \]
Window Effect

Why do we window the signal:

- Obvious advantage: localizing a part of speech for analysis.

- Less obvious disadvantage: how will the windowing affect the analysis results?
  - Compare the localization property of two extreme cases: N=1 vs. N=\infty
    - N=1: best in time but worst in frequency
    - N=\infty: best in frequency but worst in time

- Fundamentally speaking, windowing is the pursuit of a better tradeoff between time and frequency localization

Window Parameters

- Window length
  - Not too short: too few samples are not enough to resolve the uncertainty
  - Not too long: too many samples would introduce more uncertainty

- Window shape
  - Rectangular window is conceptually the simplest, but suffers from spectral leakage problem

- It is always a tradeoff: there is no universally optimal window; which window is more effective depends on specific applications
Rectangular Window

Various Windows

- Hamming/Hanning window
  - Also known as raised Cosine window
- Blackman window
  - Based on two cosine terms
- Bartlett window
  - Also known as triangular window
- Kaiser-Bessel window
  - Based on Bessel functions
Hanning / Hamming Windows

N-point Hann window

\[ W(n) = 0.5 + 0.5 \cos\left(\frac{2\pi n}{N}\right) \]

N-point Hamming window

\[ W(n) = 0.54 + 0.46 \cos\left(\frac{2\pi n}{N}\right) \]

N-point raised cosine window

\[ W(n) = a + (1-a) \cos\left(\frac{2\pi n}{N}\right) \]

Hamming

Window Kaiser

- **Time domain**
- **Frequency domain**

Leakage Factor: 0.03 %
Relative sidelobe attenuation: 42.5 dB
Mainlobe width (3dB): 0.035dB
**Blackman Window**

\[
W(n) = 0.42 + 0.5 \cos\left(\frac{\pi n}{N}\right) + 0.08 \cos\left(2\pi \frac{n}{N}\right)
\]

**Barlett Window**

\[
W(n) = 1 - \frac{|n - N/2|}{N}
\]
Kaiser-Bessel Window

\[ w_k(n) = \begin{cases} 
\frac{I_0 \left( B \left( 1 - \left[ \frac{2n}{(N-1)} \right]^2 \right)^{1/2} \right)}{I_0(0)} & -\frac{(N-1)}{2} \leq n \leq \frac{(N-1)}{2} \\
0 & \text{otherwise}
\end{cases} \]

B determines the tradeoff between the main lobe and sidelobes and is often specified as a half integer multiple of \( \pi \).

\[ I_0(x) = 1 + \sum_{L=0}^{\infty} \frac{\left( \frac{x}{2} \right)^{2L}}{(L!)^2} \]

zero order modified Bessel function of the first kind
• How to estimate the period for sinusoids?
  – Fourier Transform (will be covered by the next lectures)
• In the time domain, we might consider
  – Autocorrelation function – recall its periodic property
  – Average Magnitude Difference Function (AMDF) – essentially follow the same motivation but computationally more efficient
Pitch Period Estimation

We assume the signal is periodic (i.e. repeats itself), and let $L$ be the estimate of the pitch period.

We then 'predict' the current period, from the previous one.

$$\tilde{x}(n) = c \cdot x(n - L)$$

And the prediction error, for the estimate $L$ is then
Pitch Estimation using prediction

\[ e(n) = x(n) - \tilde{x}(n) = x(n) - c \cdot x(n - L) \]

Energy Of error
\[ E[e^2(n)] = E\left[ (x(n) - c \cdot x(n - L))^2 \right] \]
\[ E[e^2(n)] = E[x^2(n)] + c^2 E[x^2(n - L)] - 2c \cdot E[x(n)x(n - L)] \]

Minimize Energy Of error
\[ \frac{\partial}{\partial c} E[e^2(n)] = 2c E[x^2(n - L)] - 2E[x(n)x(n - L)] \]
\[ \frac{\partial}{\partial c} E[e^2(n)] = 0 \quad \Rightarrow \quad c = \frac{E[x(n)x(n - L)]}{E[x^2(n - L)]} = \frac{R_x(0, L)}{R_x(L, L)} \]

Pitch Estimation using prediction

- For a given range of estimates for the pitch \( L = L_1 \) to \( L_2 \)
- Compute the autocorrelation at lags 0, \( L \)
- Compute the optimum gain \( c \), and the prediction error:

\[ c = \frac{R_x(0, L)}{R_x(L, L)} \]

- Choose the value of \( L \) (pitch) that has the minimum error.

\[ MinErr = R_x(0, 0) - \frac{R_x^2(0, L)}{R_x(L, L)} \]