EECS 559

Advanced Signal Processing

http://www.eecs.umich.edu/~cscott

Course Overview:

Spectral Estimation

and

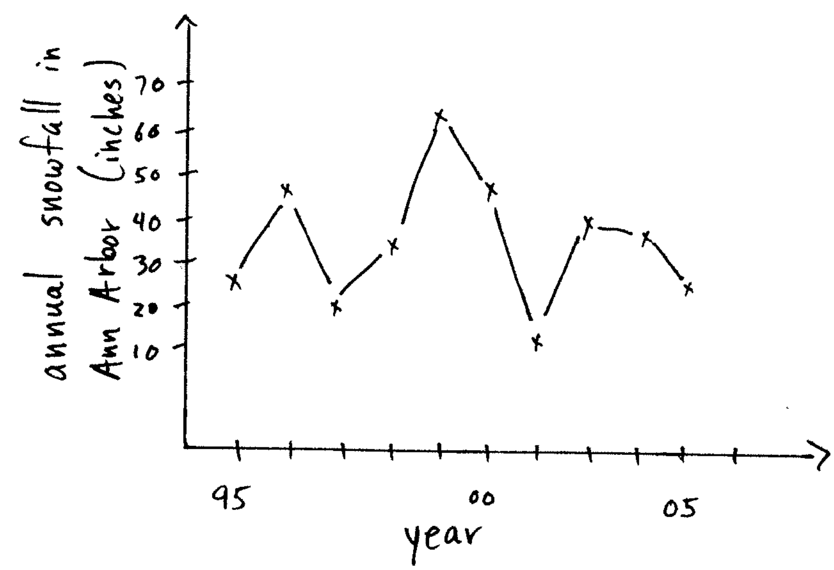
Pattern Recognition

Overview of Spectral Estimation

Spectral Estimation is the problem of estimating the power spectral density of a random process.

What's a random process?

A random process is a non-deterministic signal or time series $x[n]$



Other examples:

What's a power spectral density (PSD), aka power spectrum aka spectrum, and why estimate it?

Assuming a RP is wide-sense stationary and ergodic, the PSD is

$$P_{xx}(f) = \sum_{k=-\infty}^{\infty} r_{xx}[k]\, e^{-2\pi i f k}, \quad -\tfrac{1}{2} \le f \le \tfrac{1}{2}$$

where

$$r_{xx}[k] = E\{x[n]\, x[n+k]\}$$

is the <u>autocorrelation function</u> of $x[n]$.

In words, the PSD is the

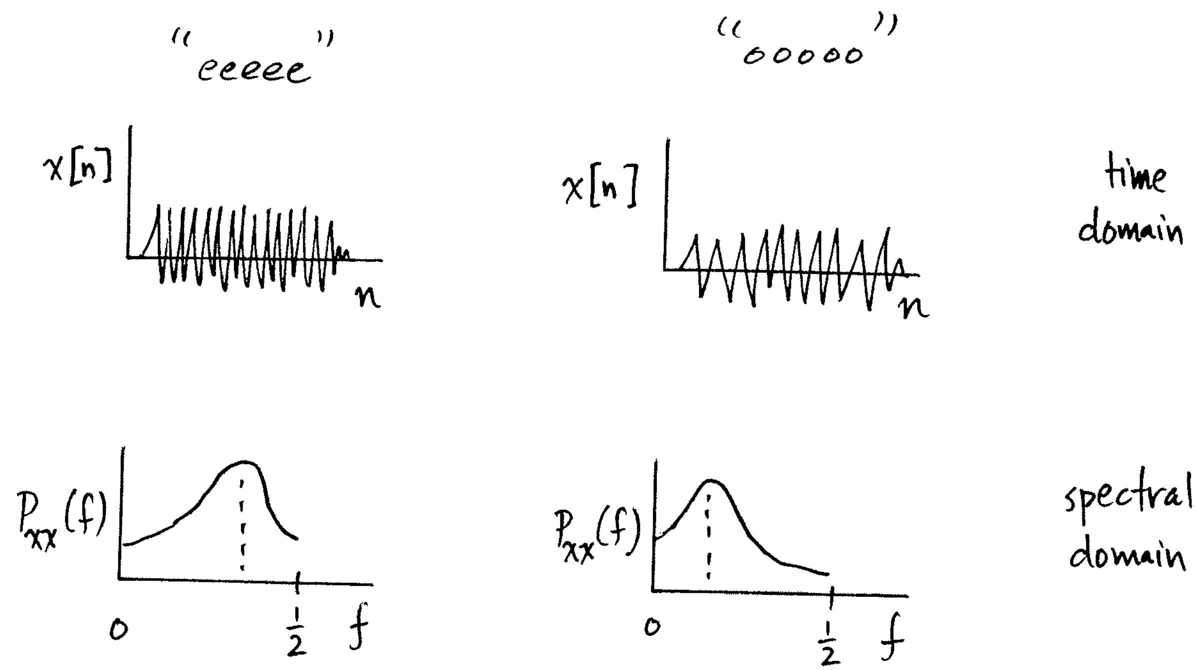
________ ____ ________ ___________

of the ACF.

Just as the DTFT represents the frequency content of a deterministic signal, so does the PSD capture the frequency content of a random process.

Therefore, if we know the PSD of a RP, we know the <u>periodicities</u> of the RP, which are often the key to understanding and analysis.

Example: Speech



Differences clearer in spectral domain.

Where does estimation come in?

The PSD depends on $r_{xx}[k] = E\{x[n]x[n+k]\}$, $k = \ldots, -2, -1, 0, 1, 2, \ldots$. Typically we only observe a finite-duration realization $\{x[0], x[1], \ldots, x[N-1]\}$.

Why should I study spectral estimation?

Because it's a useful tool for understanding signals, and can be applied to many, many, many different applications. We will touch on some in lecture, and through homework and projects.

Some broad uses of SE include:

- data exploration and visualization
- hypothesis generation for signal detection
- time-series forecasting
- linear predictive coding
- feature extraction for pattern recognition

Keep in mind that S.E. is rarely the end goal.

Methods for Spectral Estimation

That's what this course is about!

Broadly speaking, there are two flavors of spectral estimators:

- **parametric**: Assumes a specific probabilistic mechanism governing the RP that is defined by a finite set of parameters. Given those parameters, the PSD is known. This reduces the problem to parameter estimation.

- **nonparametric** (classical): estimate ACF $r_{xx}[k]$ and take DFT. Apply window functions to reduce bias.

Major theme of course: There are many methods for spectral estimation and it is not clear which is the "best." It's not even clear what "best" means.

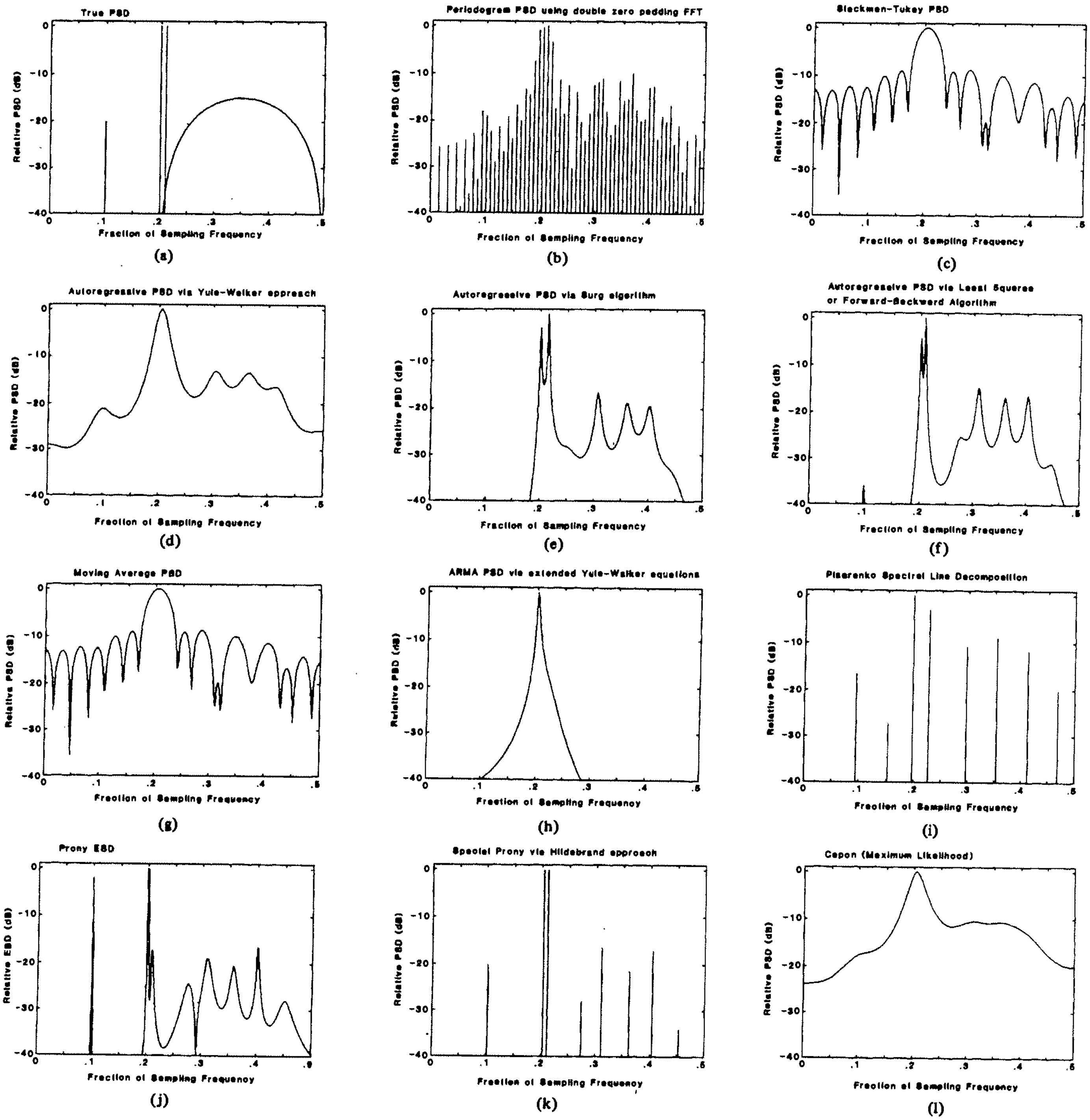


Fig. 16. Illustration of various spectra for the same 64-point sample sequence.

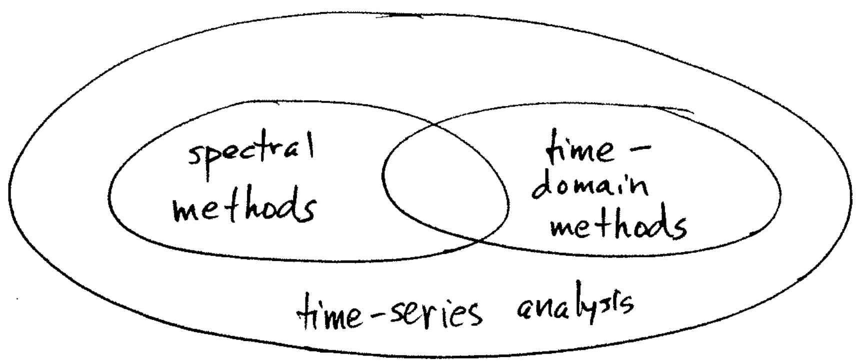
sinusoid components have very sharp responses at the sinusoid frequencies, with a broad response at the higher end of the spectrum. Table IV shows the actual parameter estimates obtained with the Prony method. The actual amplitudes for the sinusoids of frequencies 0.1, 0.2, and 0.21 were 0.1, 1., and 1., respectively. The most accurate estimates of the three sinusoid powers and frequencies is provided by the spectral line decomposition variant of the Prony method, pictured in Fig. 16(k). This is no surprise since this technique is the least squares approach that assumes a sinusoidal model. It is a discrete spectrum so that the broad-band process is not well modeled, although several lines are present to indicate spectral power in this region. Table V lists the actual parameter estimates obtained with this procedure.

The maximum likelihood spectrum, shown in Fig. 16(l) has a smooth spectrum. It cannot resolve the two closely spaced sinusoidal components. The smooth nature of the MLSE spectrum, being the equivalent of an average of all the AR spectra from order 1 to 16, is typical of this method.

If more accurate frequency estimation of noisy sinusoids

Some Context

Spectral estimation is a topic within the broader topic of time-series analysis.



Signal processors tend to focus more on frequency-domain methods for a number of reasons; but primarily because of an interest in applications involving signals characterized by their frequency content (e.g., speech).

By contrast, statisticians tend to place more emphasis on time-domain methods for similar reasons (e.g., economic forecasting).

In reality, the two perspectives share much in common and should be seen in a complementary way.

Overview of Pattern Recognition

A <u>pattern</u> is a vector $\underline{x} \in \mathbb{R}^d$.

Typically the elements of $\underline{x}$ are measurements obtained from some real-world phenomenon.

<u>Example</u> $\underline{x} = (x_1, x_2, x_3, x_4)$.

x_1 = # characters in an email message

x_2 = # of occurences of "loan" in an email

x_3 = # " " " "viagra" " " "

x_4 = # of emails previously received from sender

Each email received corresponds to such an $\underline{x}$.

The elements of $\underline{x}$ are called <u>features</u>.

Suppose you want to design a SPAM filter. What would you do? In principle, the best decision rule could be a complex function of the features.

<u>Pattern recognition</u> will refer to the probem of designing a <u>classifier</u> from <u>labeled training data</u>.

Back to SPAM filtering: Given example emails

$\underline{x}_1, x_2, \cdots, \underline{x}_M$ SPAM

$\underline{x}_{M+1}, \underline{x}_{M+2}, \cdots, \underline{x}_{M+N}$ NOT SPAM.

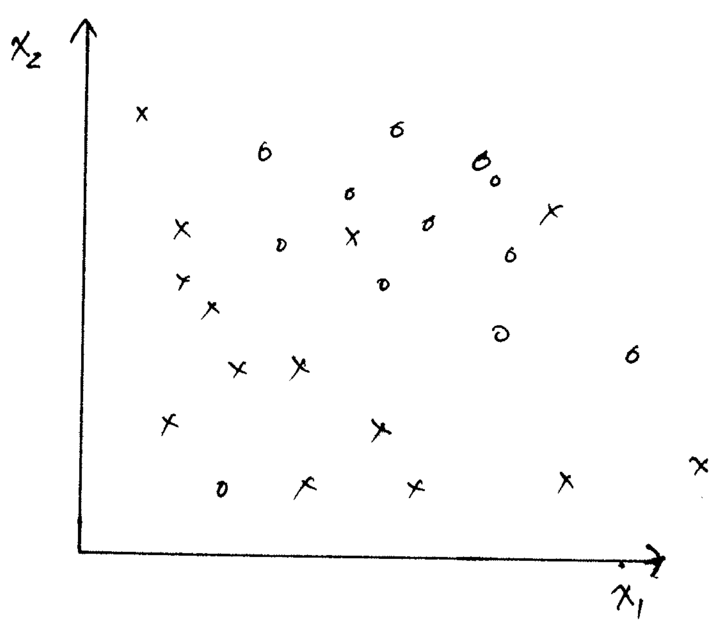
how should you build a classifier?

<u>Example</u> $\underline{x} = (x_1, x_2)$

$x_1 =$ hours spent studying for exam

$x_2 =$ " " working homework

Will a student pass a class?



o = passed class

x = failed class

Training data from last year's class