BME I5100: Biomedical Signal Processing

Probabilistic Estimation
Harmonic Process

Lucas C. Parra
Biomedical Engineering Department
City College of New York

CCNY

Schedule

Week 1: Introduction
Linear, stationary, normal - the stuff biology is **not** made of.

Week 1-5: Linear systems
Impulse response
Moving Average and Auto Regressive filters
Convolution
Discrete Fourier transform and z-transform
Sampling

Week 6-7: Analog signal processing
Operational amplifier
Analog filtering

Week 8-11: Random variables and stochastic processes
Random variables
Moments and Cumulants
Multivariate distributions, Principal Components
Stochastic processes, linear prediction, AR modeling

Week 12-14: Examples of biomedical signal processing
Harmonic analysis - **estimation** circadian rhythm and speech
Linear discrimination - **detection** of evoked responses in EEG/MEG
Hidden Markov Models and Kalman Filter- **identification** and **filtering**

Probabilistic Estimation

So far we have estimated model parameters by minimizing errors between a model and observations:

- Impulse response for given input and output.
- ARMA impulse response.
- AR model coefficients assuming white noise input.

Or by measuring moments of the data:

- Mean and covariance for normal distribution
- Mean for of a Poissons process.

A systematic way of deriving algorithms to compute model parameters is to use a probability model and

- Maximum Likelihood (ML) estimation or
- Maximum *A Posteriori* (MAP) estimation

Prob. Estimation - Maximum Likelihood

Given observed data *x* and a PDF model $p(x|\Phi)$ parameterized with model parameters Φ the Maximum Likelihood estimation is

$$\hat{\phi} = \underset{\phi}{arg\,max}\; p(x|\phi)$$

The Maximum Likelihood estimates gives the

<u>model that make the observation most likely</u>.

We often find the ML estimate by solving for Φ in:

$$\frac{\partial \ln p(x|\phi)}{\partial \phi} = 0$$

Prob. Estimation - Maximum *A Posteriori*

Sometimes *prior information* on the model parameters is available in the form of a **prior probability density**

$$p_\phi(\phi)$$

This prior information can be combined with the likelihood of the data with Bayes' theorem

$$p(\phi|x) = \frac{p(x|\phi)\, p_\phi(\phi)}{p(x)}$$

This is the conditional distribution of the parameter Φ after having observed evidence *x*. It is therefor called **posterior probability density**.

Prob. Estimation - Maximum *A Posteriori*

The ML estimate can be biased towards the prior assumption using the posterior.

This gives the **maximum *a posteriori* (MAP)** estimate

$$\hat{\phi} = \arg\max_{\phi} \, p(\phi | x)$$

The MAP estimate gives the

most probable model given the observations.

Notice how this biases the ML estimate:

$$\hat{\phi} = \arg\max_{\phi} \; \ln p(\phi | x) = \arg\max_{\phi} \left[\ln p(x|\phi) + \ln p(\phi) \right]$$

An important example is the Kalman filter, where the current estimate is biased by the past data though the previous model estimate.

Prob. Estimation - Bayesian Estimator

Bayesian estimation argues that a better estimator is the **expected model** given the data rather than the most likely model:

$$\hat{\phi} = E[\phi|x] = \int d\phi \, p(\phi|x)\, \phi$$

The difficulty often lies in executing that integral analytically. The expectation is then often computed using a Monte-Carlo simulation:

- Generate samples Φ distributed according to the posterior using Gibbs Sampling.
- Use the sample average as an estimate of the expectation.

Prob. Estimation - Maximum Likelihood

Often we are given M samples $x_1, x_2, \ldots, x_M$ independently and identically distributed (i.i.d.) according to $p(x|\Phi)$. The ML estimate is then

$$\hat{\phi} = \underset{\phi}{arg\,max}\ \ln p(x_1, x_2, \ldots, x_M|\phi)$$
$$= \underset{\phi}{arg\,max} \sum_{i=1}^{M} \ln p(x_i|\phi)$$

Example: Assume i.i.d. normal samples $x_1, \ldots, x_M$ drawn from

$$p(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \equiv N(x-\mu,\sigma)$$

Prob. Estimation - Maximum Likelihood

Setting derivatives of the log-likelihood to zero

$$\frac{\partial}{\partial \mu} \sum_{i=1}^{M} \ln N(x_i - \mu, \sigma) = 0$$

$$\frac{\partial}{\partial \sigma} \sum_{i=1}^{M} \ln N(x_i - \mu, \sigma) = 0$$

gives the solutions

$$\hat{\mu} = \frac{1}{M} \sum_{i=1}^{M} x_i \qquad \hat{\sigma}^2 = \frac{1}{M} \sum_{i=1}^{M} (x_i - \hat{\mu})^2$$

The ML estimate of the mean and covariance of a Gaussian PDF model are the sample mean and sample covariance. However, notice that $E[\hat{\sigma}^2] = \sigma^2 (M-1)/M$

Prob. Estimation - ML and Least Squares

ML under Gaussian noise assumption is the same as Least squares (LS) !

For example: Assume a linear relation between an input ***x*** and an output ***y*** and allow for *independent additive zero mean Gaussian* noise ***n***:

$$\boldsymbol{y} = \boldsymbol{A}\,\boldsymbol{x} + \boldsymbol{n}$$

The likelihood of the data given model parameters is then

$$p_{x,y}(\boldsymbol{x}, \boldsymbol{y} | \boldsymbol{A}, \boldsymbol{\sigma}) = p_{\boldsymbol{n}}(\boldsymbol{y} - \boldsymbol{A}\,\boldsymbol{x} | \sigma)\; p_x(\boldsymbol{x})$$

The log likelihood, $L(\boldsymbol{A},\sigma)$, for M data points $\boldsymbol{x}_1, \boldsymbol{y}_1, \ldots, \boldsymbol{x}_M, \boldsymbol{y}_M$ given model ***A*** and σ is then

$$L(\boldsymbol{A}, \sigma) = \sum_{i=1}^{M} \ln N(\boldsymbol{y}_i - \boldsymbol{A}\,\boldsymbol{x}_i | \sigma) + const.$$

Prob. Estimation - ML and Least Squares

By inserting the definition of the Gaussian we obtain the classic LS problem:

$$\hat{\boldsymbol{A}} = \arg\max_{A} \; L(\boldsymbol{A}, \sigma) = \arg\min_{A} ||\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{X}||^2$$

where we used our usual matrix notation for the sample matrix ***X*** and ***Y***. The estimated error variance is

$$\hat{\sigma} = \arg\max_{\sigma} \; L(\boldsymbol{A}, \sigma)$$
$$= \arg\min_{\sigma} \; M \ln \sigma^d + ||\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{X}||^2 / (2\sigma^2)$$

Setting derivatives equal to zero yields the LS solution

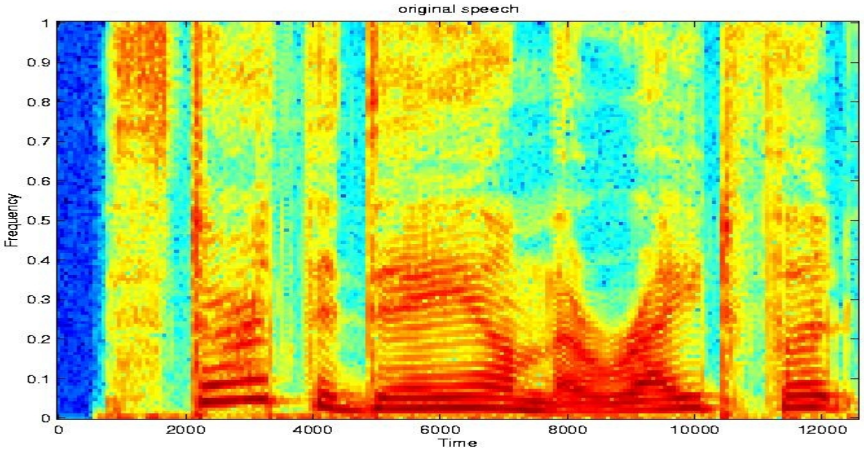
$$\hat{\boldsymbol{A}} = \boldsymbol{Y}\boldsymbol{X}^T (\boldsymbol{X}\boldsymbol{X}^T)^{-1}$$

And the obvious error estimate $\hat{\sigma}^2 = \frac{1}{Md} ||\boldsymbol{Y} - \hat{\boldsymbol{A}}\boldsymbol{X}||^2$

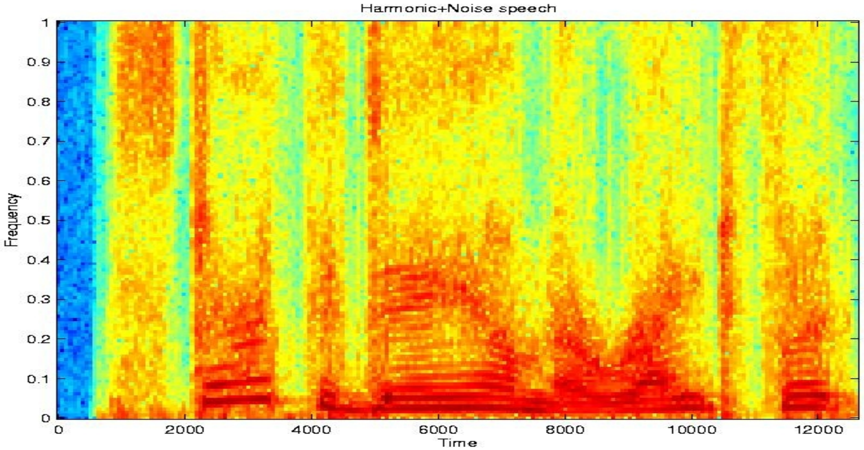
Harmonic Analysis

We will model speech by computing ML estimate in small frames assuming that speech has a harmonic component and a AR regular process.

Original



Synthesized



Harmonic

Noise

Harmonic Analysis

Harmonic model is sum of sinusoids with frequency being multiples of basic pitch frequency ω

$$h(t)=\sum_{k=1}^{K} b_k \sin(k\omega t)+c_k \cos(k\omega t)$$

The coefficients b_k and c_k encode the amplitude and phase of each harmonic component:

$$b\sin(\alpha)+c\cos(\alpha)=a\sin(\alpha+\phi)$$

$$h(t)=\boldsymbol{s}^T(\omega t)\boldsymbol{b}$$

In matrix notation we can write

$$\begin{bmatrix} h[1] \\ h[2] \\ \vdots \\ h[T] \end{bmatrix} = \begin{bmatrix} \sin(\omega) & \sin(\omega 2) & \dots & \sin(\omega T) \\ \vdots & \vdots & & \vdots \\ \sin(\omega K) & \sin(K\omega 2) & \dots & \sin(K\omega T) \\ \cos(\omega) & \cos(\omega 2) & \dots & \cos(\omega T) \\ \vdots & \vdots & & \vdots \\ \cos(\omega K) & \cos(K\omega 2) & \dots & \cos(K\omega T) \end{bmatrix}^T \begin{bmatrix} b_1 \\ \vdots \\ b_K \\ c_1 \\ \vdots \\ c_K \end{bmatrix}$$

$$\boldsymbol{h}=\boldsymbol{S}^T\boldsymbol{b}$$

Harmonic Analysis

We model a given signal $x[t]$ as a deterministic harmonic process plus independent additive zero mean white Gaussian noise $n[t]$

$$x[t] = h[t] + n[t]$$

The log likelihood of $x[t]$ given model parameters $\boldsymbol{b}$ is then

$$L(\boldsymbol{b}, \omega, \sigma) = \sum_t \ln N(h(t) - x[t] \,|\sigma)$$

$$= \sum_t \ln N(\boldsymbol{s}^T(\omega t)\boldsymbol{b} - x[t] \,|\sigma)$$

The ML solution for $\boldsymbol{b}$ for a **given pitch ω** is then (a LS solutions)

$$\hat{\boldsymbol{b}} = \arg\max_{\boldsymbol{b}} \; L(\boldsymbol{b}, \omega, \sigma)$$

$$= \arg\min_{\boldsymbol{b}} \; ||\boldsymbol{x} - \boldsymbol{S}^T \boldsymbol{b}||^2 = \boldsymbol{S}^{-T}\boldsymbol{x}$$

Harmonic Analysis

The variance is again the estimation error

$$\hat{\sigma}^2 = ||\boldsymbol{x} - \boldsymbol{S}^T \hat{\boldsymbol{b}}||^2 / T$$

However, finding the solutions for the best pitch ω is a non-convex optimization problem!

$$\hat{\boldsymbol{\omega}} = \underset{\omega}{arg\,max}\ \ L(\hat{\boldsymbol{b}}(\omega), \omega, \sigma)$$

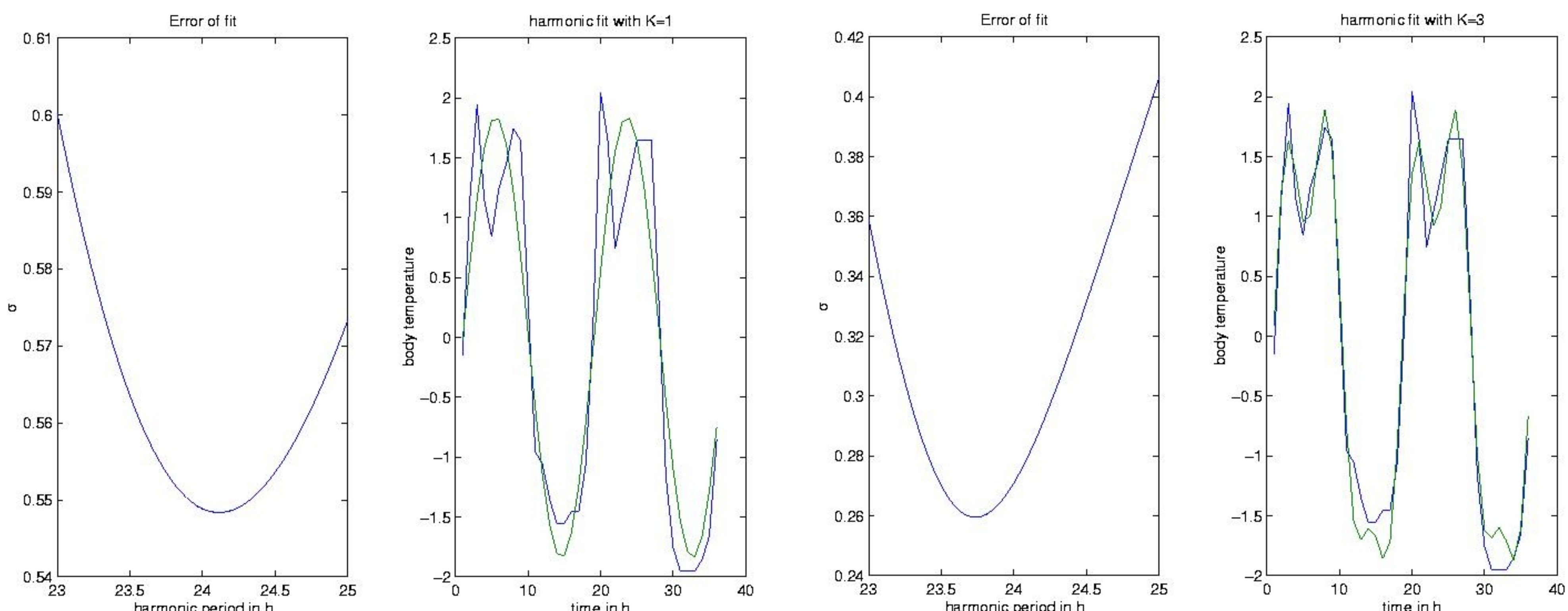
$$= \underset{\omega}{arg\,min}\ \ ||\boldsymbol{x} - \boldsymbol{S}^T(\omega)\hat{\boldsymbol{b}}(\omega)||^2 = ?$$

Exhaustive search is really the only solution.

Harmonic Analysis

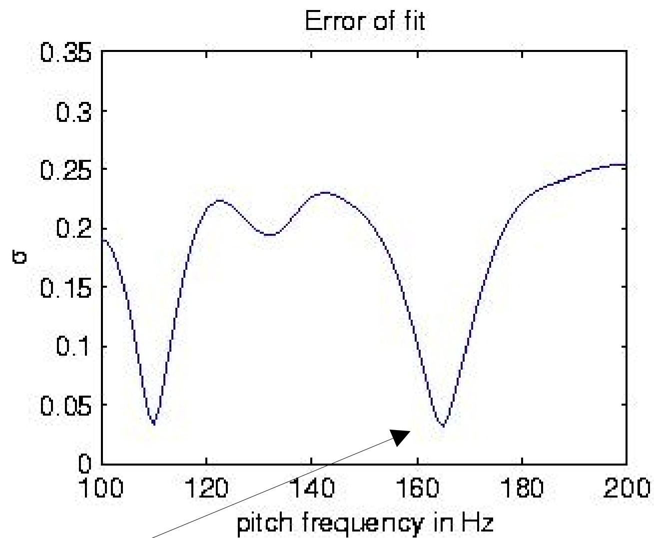
Example: Model Circadian rhythm with harmonic sine and cosine.

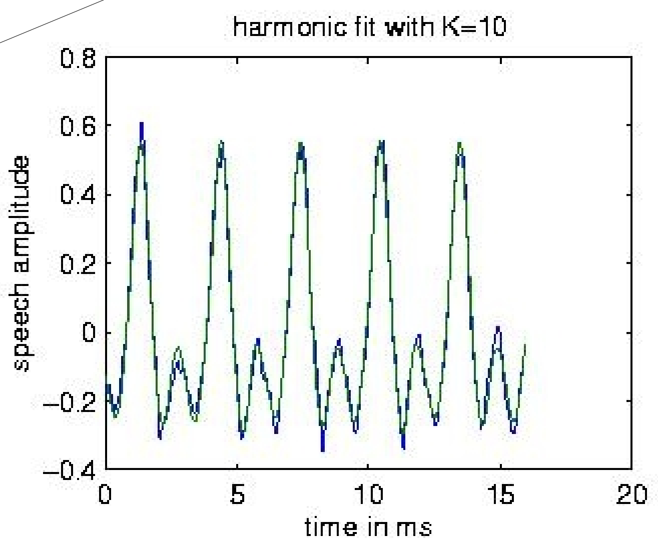
```
for i=1:length(f)                              % for all suspect freqs
  phase = [0:T-1]'*[1:K]*2*pi*f(i)/fs; % nicely scaled phase
  S = [sin(phase) cos(phase)]';            % harmonic basis S'
  b = S'\x;                                % LS solution for b
  s(i) = std(x-S'*b);                      % Error of fit
end
[m,i]=min(s); 1./f(i)                         % and the winner is f(i)
```

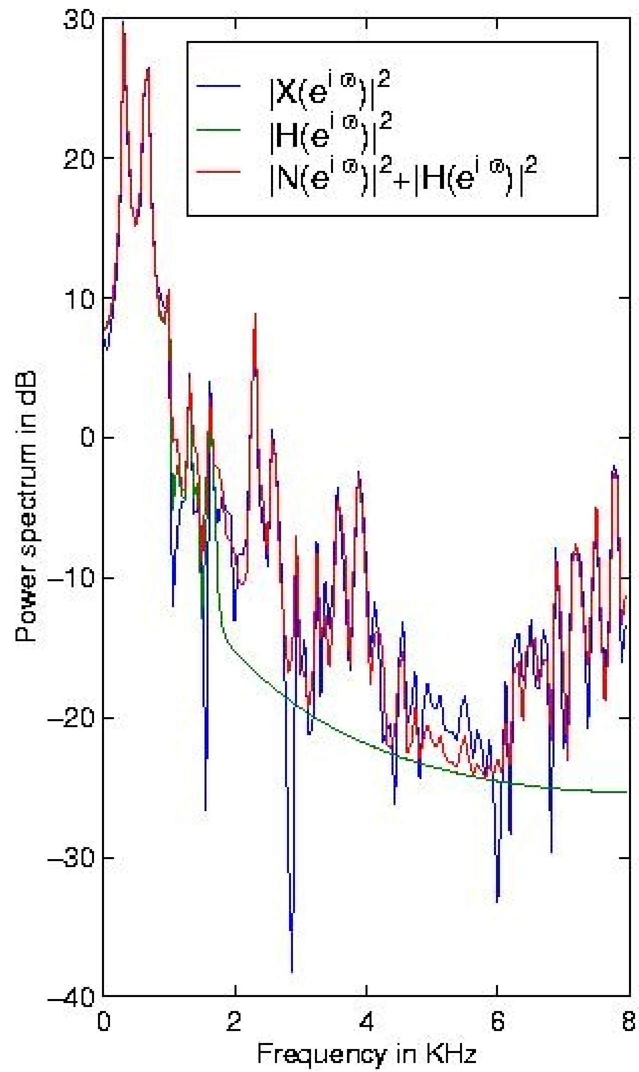


Harmonic Analysis

Example: Model of 16 ms of speech as harmonic process.







Note that the power spectrum is additive for orthogonal processes.

Best fit at ω=165Hz

Harmonic Analysis

Assignment 10:

- Select a 16 ms segment of voiced speech (clear harmonic structure).
- Model the that segment as a deterministic harmonic process.
- What is you estimate for pitch frequency ω?
- How many harmonics (K) seem appropriate for your signal?

- Measure how orthogonal the harmonic process $h[t]$ is from the remaining noise process $n[t]$.
- Model the remaining noise $n[t]$ as an AR process of order P.
- Show the combined model spectra (harmonic + noise) and overlay with to the raw Fourier spectrum.
- Try a few different K and P. What do you think is a good choice?