Harmonic analysis

Basic idea: Interpret a time or space series as a summation of contributions from harmonic functions - each with a unique temporal or spatial scale.

Recall the normal equations for the least squares best fit of 'y' to predictors 'x'

$$y = a_0 + a_1 x_1 + a_2 x_2 \cdots a_N x_N$$

where $\bar{y} = a_0 + a_1 \bar{x} + a_2 \bar{x_2} \ldots a_N \bar{x_N}$

and:

$$\overline{x_1' y'} = a_1 \overline{x_1'^2} + a_2 \overline{x_1' x_2'} + a_3 \overline{x_1' x_3'} \cdots$$

$$\overline{x_2' y'} = a_1 \overline{x_1' x_2'} + a_2 \overline{x_2'} + a_3 \overline{x_2' x_3'} \cdots$$

In the case of harmonic analysis, the functions to be fit (i.e. the x's), or basis functions, are of the form:

$$\cos\left(2\pi k \frac{t}{T}\right) \quad \text{and} \quad \sin\left(2\pi k \frac{t}{T}\right)$$

$$k = 1, 2, 3, \ldots$$

Hence y can be written as:

$$y = a_0 + \sum_{k=1}^{N/2} A_k \cos\left(2\pi k \frac{t}{T}\right) + \sum_{k=1}^{N/2} B_k \sin\left(2\pi k \frac{t}{T}\right)$$

$k = 1, 2, 3 \ldots$

T = total time length of the record
N = # of grid points or time steps
A_k & B_k = regression coefficients for each predictor.

Notes

- Each predictor is a harmonic function with frequency k/T, and hence fits into the interval $[0, T]$ 'k' times

- If $k=1$, the frequency is $1/T$ and hence predictor completes 2π radians in $[0, T]$

* - If $k = N/2$, the frequency is the highest that can be resolved, the predictor completes π radians in one timestep → Nyquist frequency

- for a given k, fitting k cosine and k sine waves with variable amplitudes A & B into $[0, T]$: Equivalent to fitting k cosine waves (w/ no sine waves) with variable amp & phase

The amplitude of each predictor is found by solving the normal equations (like in regression)

Discrete Fourier transform

In the special case of evenly spaced data points (e.g. a constant time interval)

Consider domain $[0, T]$ where 0 and T coincide with data points $i=1$ and $i=N+1$ where N is an even number.

1) The functions assume the form

$$\cos\left(2\pi k \frac{i\Delta t}{T}\right) \quad \text{and} \quad \sin\left(2\pi k \frac{i\Delta t}{T}\right)$$

where Δt is the spacing between grid points.

2) The average of each sine / cosine function on the interval $[0, T]$ is zero. Hence $a_0 = \bar{y}$

3) The harmonic functions are mutually orthogonal on $[0, T]$ because the correlation between sine and cosine functions is zero. Therefore, the off-diagonal elements in covariance matrix $= 0$.

Hence, the normal equations reduce to a set of equations of the form:

$$a_K = \frac{\overline{x_K' y'}}{\overline{x_K'^2}}$$

4) the functions (the K's) each have variance $\overline{x_K'^2} = \frac{1}{2}$ except for $A_{N/2}$ and $B_{N/2}$, of which the former takes on a value of $\pm$ unity at each data point, and ~~the~~ latter a value of zero at each grid point.

Hence, the variance of the predictor with amplitude $A_{N/2}$, $B_{N/2}$ is one and zero, respectively.

5) From #3, the solutions for each predictor's amplitude:

$$a_K = \frac{\overline{x_K' y'}}{\overline{x_K'^2}} \quad \text{reduce to.}$$

$$A_K = \frac{2}{N} \sum_{i=1}^{N} y_i \cos\left(2\pi K \frac{i\Delta t}{T}\right)$$

$$B_K = \frac{2}{N} \sum_{i=1}^{N} y_i \sin\left(2\pi K \frac{i\Delta t}{T}\right)$$

$$K = 1, N/2 - 1$$

$$A_{N/2} = \frac{1}{N} \sum_{i=1}^{N} y_i \cos\left(\pi N \frac{\Delta t}{T}\right)$$

$$B_{N/2} = 0.$$

The function y can now be written as a sum of sines and cosines:

$$y(t) = \bar{y} + \sum_{K=1}^{N/2-1} A_K \cos\left(2\pi K \frac{t}{T}\right)$$

$$+ \sum_{K=1}^{N/2-1} B_K \sin\left(2\pi K \frac{t}{T}\right)$$

$$+ A_{N/2} \cos\left(\pi N \frac{t}{T}\right)$$

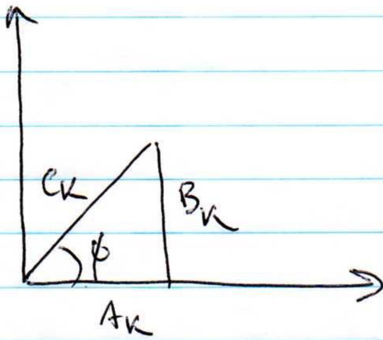
Can also write this in amplitude / phase form:

$$y(t) = \bar{y} + \sum_{K=1}^{N/2-1} C_K \cos\left(2\pi K \frac{(t-t_K)}{T}\right) + A_{N/2} \cos\left(\pi N \frac{t}{T}\right)$$

where $C_K^2 = A_K^2 + B_K^2$

$$t_K = \frac{T}{2\pi K} \tan^{-1}\left(\frac{B_K}{A_K}\right)$$

In geometric terms:



$$\phi = \tan^{-1}\left(\frac{B_K}{A_K}\right)$$

Fraction of variance explained for each K:

$$r^2(y, x_K) = \frac{\left(\overline{x'_K y'}\right)^2}{\overline{x_K'^2}\ \overline{y_K'^2}}$$

Using the expressions for A_K and B_K, and the fact that $\overline{x_K'^2} = \frac{1}{2}$

$r^2(y, x_K)$ is $\dfrac{A_K^2}{2\overline{y_K'^2}}$ for the cosine functions

$r^2(y, x_K)$ is $\dfrac{B_K^2}{2\overline{y_K'^2}}$ for the sine functions:

... $\dfrac{A_{N/2}^2}{\overline{y'^2}}$ for $N/2$

To get the actual (not fraction) variance explained:

$$\frac{A_k^2 + B_k^2}{2\,\cancel{\overline{y'^2}}} \cdot \cancel{\overline{y'^2}} = \frac{C_k^2}{2}$$

$$\boxed{\frac{A_k^2 + B_k^2}{2} = \frac{C_k^2}{2}}$$

Explained variance by a particular k (spectral power)

for $k = 1 \Rightarrow N/2 - 1$

$A^2_{N/2}$ for $k = N/2$ (Nyquist freq.)

Notes on expressing power spectrum coefficients.

Can also express $y(t)$ with complex notation:

$$y(t) = \bar{y} + \sum_{k=1}^{N/2-1} H_k e^{i(2\pi k \frac{t}{T})}$$

Recall $e^{i\theta} = \cos\theta + i\sin\theta$

Real part of $H_k = A_k$
Imaginary part $H_k = B_k$

The term with the "k" is more compactly expressed in terms of angular frequency (ω)

$$\boxed{\omega = \frac{2\pi k}{T} \quad \text{or} \quad \frac{2\pi k}{N}}$$

* Be aware of these notational conventions in the Hartmann and Wilks references — or you may find yourself confused!