Discrete-Time Signal Processing (DSP)

Chu-Song Chen

Email: song@iis.sinica.du.tw

Institute of Information Science, Academia Sinica
Institute of Networking and Multimedia, National Taiwan University

Fall 2006

What are Signals

(c.f. Kuhn 2005 and Oppenheim et al. 1999)

flow of information: generally convey information about the state or behavior of a physical system.

- measured quantity that varies with time (or position)
- electrical signal received from a transducer (microphone, thermometer, accelerometer, antenna, etc.)
- electrical signal that controls a process

continuous-time signal: Also know as analog signal.

- voltage, current, temperature, speed, speech signal, etc.

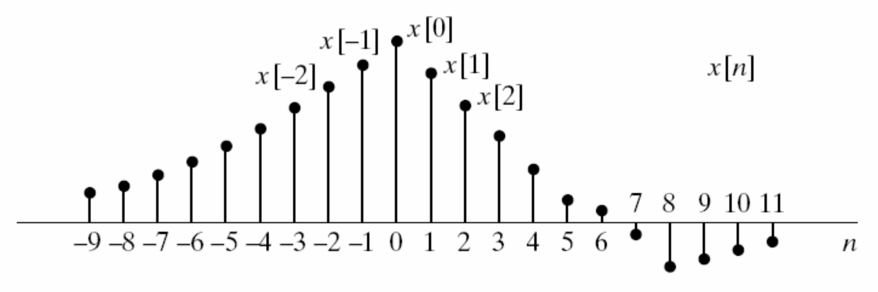
discrete-time signal: daily stock market price, daily average temperature, sampled continuous signals.

Examples of Signals

types in dimensionality:

- speech signal: represented as a function over time. -- 1D signal
- image signal: represented as a brightness function of two spatial variables. -- 2D signal
- ultra sound data or image sequence – 3D signal

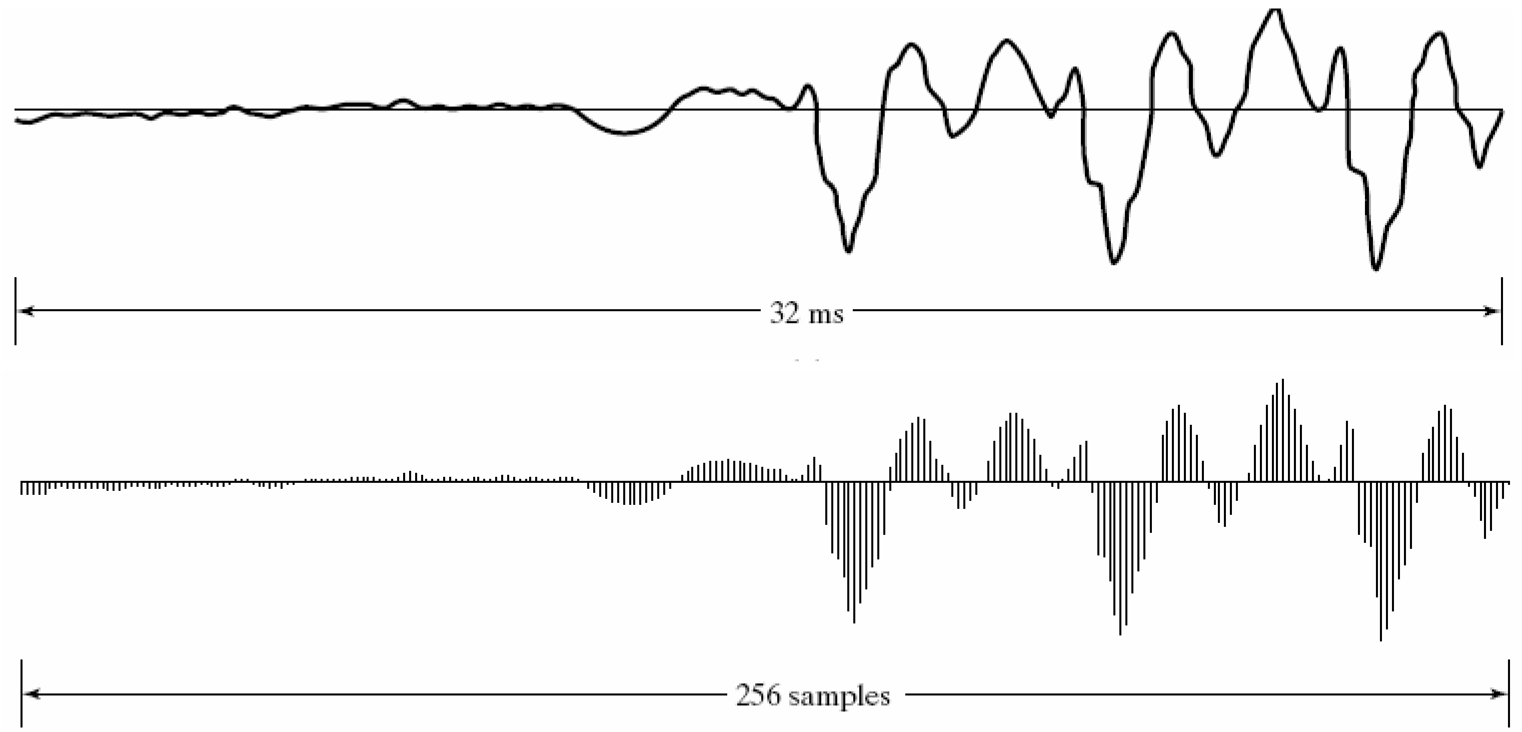
Electronics can only deal easily with time-dependent signals, therefore spatial signals, such as images, are typically first converted into a time signal with a scanning process (TV, fax, etc.)



Generation of Discrete-time Signal

In practice, discrete-time signal can often arise from periodic sampling of an analog signal.

$$x = x_a[nT], -\infty < n < \infty$$



What is Signal Processing (Kuhn 2005)

Signals may have to be transformed in order to

- amplify or filter out embedded information
- detect patterns
- prepare the signal to survive a transmission channel
- undo distortions contributed by a transmission channel
- compensate for sensor deficiencies
- find information encoded in a different domain.

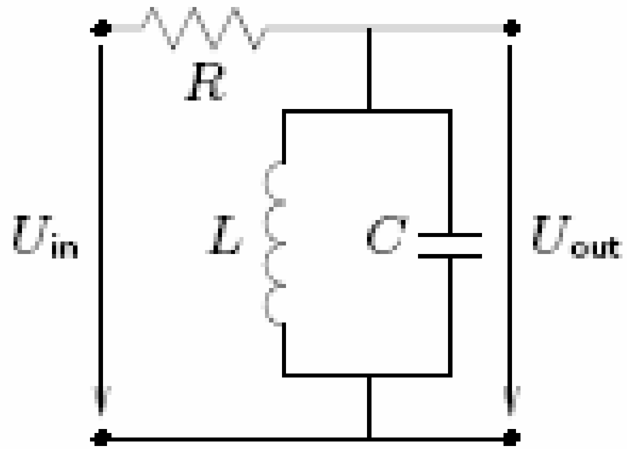
To do so, we also need:

- methods to measure, characterize, model, and simulate signals.
- mathematical tools that split common channels and transformations into easily manipulated building blocks.

Analog Electronics for Signal Processing

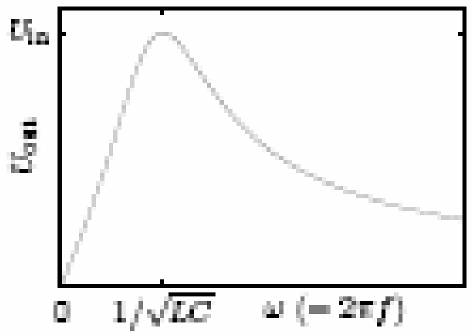
(Kuhn 2005)

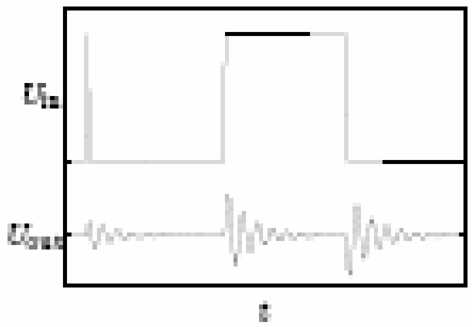
Passive networks (resistors, capacities, inductivities, crystals, nonlinear elements: diodes ...), (roughly) linear operational amplifiers



Advantages:

- passive networks are highly linear over a very large dynamic range and bandwidths.
- analog signal-processing circuits require little or no power.
- analog circuits cause little additional interference





$$\frac{U_{in} - U_{out}}{R} = \frac{1}{L}\int_{-\infty}^{t} U_{out}\,d\tau + C\frac{dU_{out}}{dt}$$

Digital Signal Processing (Kuhn 2005)

Analog/digital and digital/analog converters, CPU, DSP, ASIC, FPGA

Advantages:

- noise is easy to control after initial quantization
- highly linear (with limited dynamic range)
- complex algorithms fit into a single chip
- flexibility, parameters can be varied in software
- digital processing in insensitive to component tolerances, aging, environmental conditions, electromagnetic inference

But

- discrete time processing artifacts (aliasing, delay)
- can require significantly more power (battery, cooling)
- digital clock and switching cause interference

Typical DSP Applications (Kuhn 2005)

- **communication systems**
  modulation/demodulation, channel equalization, echo cancellation
- **consumer electronics**
  perceptual coding of audio and video on DVDs, speech synthesis, speech recognition
- **Music**
  synthetic instruments, audio effects, noise reduction
- **medical diagnostics**
  Magnetic-resonance and ultrasonic imaging, computer tomography, ECG, EEG, MEG, AED, audiology
- **Geophysics**
  seismology, oil exploration
- **astronomy**
  VLBI, speckle interferometry
- **experimental physics**
  sensor data evaluation
- **aviation**
  radar, radio navigation
- **security**
  steganography, digital watermarking, biometric identification, visual surveillance systems, signal intelligence, electronic warfare
- **engineering**
  control systems, feature extraction for pattern recognition

Syllabus
(c.f. Kuhn 2005 and Stearns 2002)

- **Signals and systems:** Discrete sequences and systems, their types and properties. Linear time-invariant systems, correlation/convolution, eigen functions of linear time-invariant systems. Review of complex arithmetics.
- **Fourier transform:** Harmonic analysis as orthogonal base functions. Forms of the Fourier transform. Convolution theorem. Dirac's delta function. Impulse trains (combs) in the time and frequency domain.
- **Discrete sequences and spectra:** Periodic sampling of continuous signals, periodic signals, aliasing, sampling and reconstruction of low-pass signals.
- **Discrete Fourier transform:** continuous versus discrete Fourier transform, symmetric, linearity, fast Fourier transform (FFT).
- **Spectral estimation:** power spectrum**.**
- **Finite and infinite impulse-response filters:** Properties of filters, implementation forms, window-based FIR design, use of analog IIR techniques (Butterworth, Chebyshev I/II, etc.)

- **Z-transform:** zeros and poles, difference equations, direct form I and II.
- **Random sequences and noise:** Random variables, stationary process, auto-correlation, cross-correlation, deterministic cross-correlation sequences, white noise.
- **Multi-rate signal processing:** decimation, interpolation, polyphase decompositions.
- **Adaptive signal processing:** mean-squared performance surface, LMS algorithm, Direct descent and the RLS algorithm.
- **Coding and Compression:** Transform coding, discrete cosine transform, multirate signal decomposition and subband coding, PCA and KL transformation.
- **Wavelet transform**: Time-frequency analysis. Discrete wavelet transform (DFT), DFT for compression.
- **Particle filtering:** hidden Markov model, state space form, Markov chain Monte Carlo (MCMC), unscented Kalman filtering, particle filtering for tracking.

Lectures: 12 times.

References:

- S. D. Stearns, Digital Signal Processing with Examples in MATLAB, CRC Press, 2003. (main textbook, but not dominant)
- B. A. Shenoi, Introduction to Signal Processing and Filter Design, Wiley, 2006.
- S. Salivahanan, A. Vallavaraj, and C. Gnanapriya, Digital Signal Procesing, McGraw-Hill, 2002.
- A. V. Oppenheim and R. W. Schafer, Discrete Time Signal Processing, 2nd ed., Prentice Hall, 1999.
- J. H. McClellan, R. W. Schafer, and M. A. Yoder, Signal Processing First, Prentice Hall, 2004. (suitable for beginners)
- S. K. Mitra, Digital Signal Processing, A Computer-Based Approach, McGraw-Hill, 2002.
- Markus Kuhn, Digital Signal Processing slides in Cambridge, http://www.c1.cam.ac.uk/Teaching/2005/DSP
- Some relevant papers …

Main journals and conferences in this field

Journal

- IEEE Transactions on Signal Processing
- Signal Processing
- EUROSIP Journal on Applied Signal Processing
- ...

Conference

- IEEE ICASSP (International Conference on Acoustics, Speech, and Signal Processing)

Evaluations in this course

Homework - about three times.
Tests: twice
Term project

Review of complex exponential

(c.f. Kuhn 2005 and Oppenheim et al. 1999)

geometric series is used repeatedly to simplify expressions in DSP.

$$\sum_{n=0}^{n-1} x^n = 1 + x + x^2 + \ldots + x^{N-1} = \frac{1-x^n}{1-x}$$

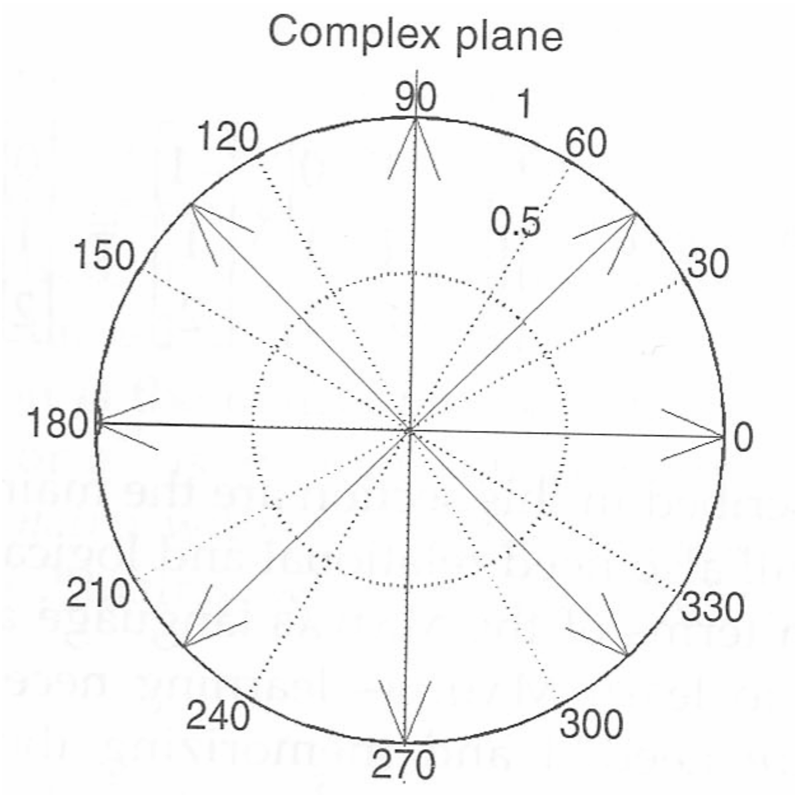
- if the magnitude of x is less than one, then

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}, \qquad |x| < 1$$

In DSP, the geometric series is often a complex exponential variable of the form e^{jk}, where j = $\sqrt{-1}$

For example

$$\sum_{n=0}^{N-1} e^{j\frac{2\pi n}{N}} = \frac{1-e^{j2\pi}}{1-e^{j\frac{2\pi n}{N}}} = 0 \qquad (1)$$



Trigonometric Identities

$$\sin\theta = \frac{1}{2j}(e^{j\theta} - e^{-j\theta})$$

$$\cos\theta = \frac{1}{2}(e^{j\theta} + e^{-j\theta})$$

$$e^{j\theta} = \cos\theta + j\sin\theta$$

$$\sin(\theta+\alpha) = \sin\theta\cos\alpha + \cos\theta\sin\alpha$$

$$\cos(\theta+\alpha) = \cos\theta\cos\alpha - \sin\theta\sin\alpha$$

$$\sin\theta\sin\alpha = \frac{1}{2}(\cos(\theta-\alpha) - \cos(\theta+\alpha))$$

$$\cos\theta\cos\alpha = \frac{1}{2}(\cos(\theta+\alpha) + \cos(\theta-\alpha))$$

$$\sin\theta\cos\alpha = \frac{1}{2}(\sin(\theta+\alpha) + \sin(\theta-\alpha))$$

$$\sin^2\theta + \cos^2\theta = 1$$

$$\cos^2\theta - \sin^2\theta = \cos 2\theta$$

$$\sin^2\theta = \frac{1}{2}(1 - \cos 2\theta)$$

$$\cos^2\theta = \frac{1}{2}(1 + \cos 2\theta)$$

$$\tan\theta = \frac{\sin\theta}{\cos\theta}$$

Trigonometric functions, especially sine and cosine functions, appear in different combinations in all kinds of harmonic analysis: Fourier series, Fourier transforms, etc.

Advantages of complex exponential

The identities that give sine and cosine functions in terms of exponentials are important – because they allow us to find sums of sines and cosines using the geometric series.

Eg. from (1), we have

$$\sum_{n=0}^{N-1} \sin\left(\frac{2\pi n}{N}\right) = 0 \qquad \sum_{n=0}^{N-1} \cos\left(\frac{2\pi n}{N}\right) = 0$$

ie. a sum of equally spaced samples of any sine or cosine function is zero, provided the sum is over a cycle (or a number of cycles), of the function.

Why are sine waves useful?

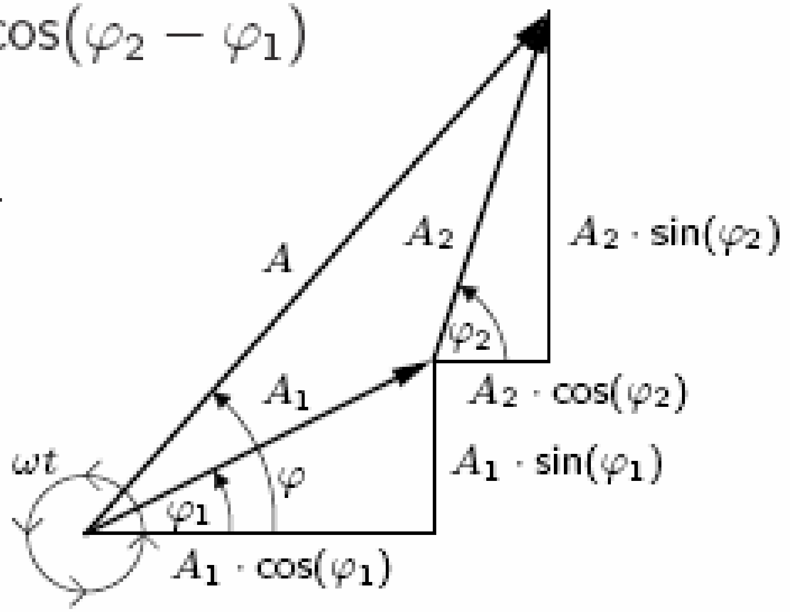
1) Adding together sine waves of equal frequency, but arbitrary amplitude and phase, results in another sine wave of the same frequency:

$$A_1 \cdot \sin(\omega t + \varphi_1) + A_2 \cdot \sin(\omega t + \varphi_2) \; = \; A \cdot \sin(\omega t + \varphi)$$

with

$$A \; = \; \sqrt{A_1^2 + A_2^2 + 2A_1A_2\cos(\varphi_2 - \varphi_1)}$$

$$\tan\varphi \; = \; \frac{A_1\sin\varphi_1 + A_2\sin\varphi_2}{A_1\cos\varphi_1 + A_2\cos\varphi_2}$$



Sine waves of any phase can be from from sin and cos alone:

$$A \cdot \sin(\omega t + \varphi) =$$
$$a \cdot \sin(\omega t) + b \cdot \cos(\omega t)$$

with $a = A \cdot \cos(\varphi)$, $b = A \cdot \sin(\varphi)$ and $A = \sqrt{a^2 + b^2}$, $\tan\varphi = \frac{b}{a}$.

2) Sine waves are orthogonal to each other:

$$\int_{-\infty}^{\infty} \sin(\omega_1 t + \varphi_1) \cdot \sin(\omega_2 t + \varphi_2)\,\mathrm{d}t = 0$$

$$\Longleftrightarrow \quad \omega_1 \neq \omega_2 \quad \vee \quad \varphi_1 - \varphi_2 = (2k+1)\pi \quad (k \in \mathbb{Z})$$

They can be used to form an orthogonal function basis for a transform.

Why are complex numbers so useful?

1) They give us all n solutions ("roots") of equations involving polynomials up to degree n (the $\sqrt{-1} = \mathrm{j}$ story).

2) They form the "great unifying theory" that combines sine functions and exponential functions:

$$\begin{aligned}\cos(\omega t) &= \frac{1}{2}\left(\mathrm{e}^{\mathrm{j}\omega t} + \mathrm{e}^{-\mathrm{j}\omega t}\right)\\ \sin(\omega t) &= \frac{1}{2\mathrm{j}}\left(\mathrm{e}^{\mathrm{j}\omega t} - \mathrm{e}^{-\mathrm{j}\omega t}\right)\end{aligned}$$

or

$$\cos(\omega t + \varphi) = \frac{1}{2}\left(\mathrm{e}^{\mathrm{j}\omega t+\varphi} + \mathrm{e}^{-\mathrm{j}\omega t-\varphi}\right)$$

or

$$\begin{aligned}\cos(\omega n + \varphi) &= \Re\{\mathrm{e}^{\mathrm{j}\omega n+\varphi}\} = \Re\{(\mathrm{e}^{\mathrm{j}\omega})^n \cdot \mathrm{e}^{\mathrm{j}\varphi}\}\\ \sin(\omega n + \varphi) &= \Im\{\mathrm{e}^{\mathrm{j}\omega n+\varphi}\} = \Im\{(\mathrm{e}^{\mathrm{j}\omega})^n \cdot \mathrm{e}^{\mathrm{j}\varphi}\}\end{aligned}$$

We can now represent sine waves as projections of a rotating complex vector. This allows us to represent sine-wave sequences as exponential sequences with basis $e^{j\omega}$.

A phase shift in such a sequence corresponds to a rotation of a complex vector.

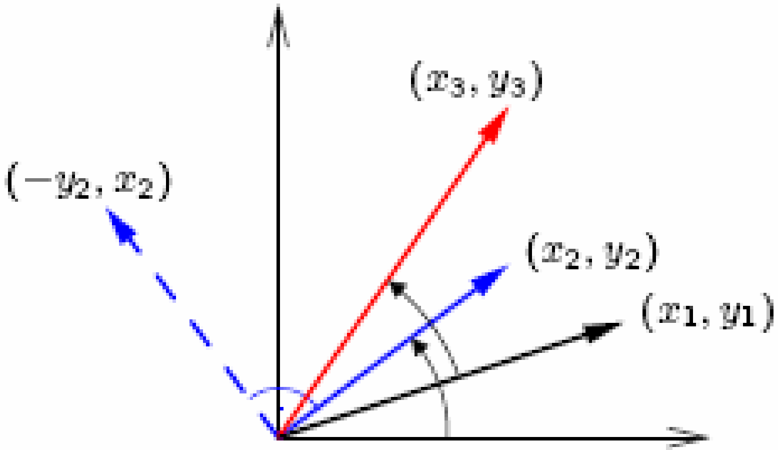
3) Complex multiplication allows us to modify the amplitude and phase of a complex rotating vector using a single operation and value.

Rotation of a 2D vector in (x, y)-form is notationally slightly messy, but fortunately $j^2 = -1$ does exactly what is required here:

$$\begin{pmatrix} x_3 \\ y_3 \end{pmatrix} = \begin{pmatrix} x_2 & -y_2 \\ y_2 & x_2 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} x_1x_2 - y_1y_2 \\ x_1y_2 + x_2y_1 \end{pmatrix}$$

$$z_1 = x_1 + jy_1, \quad z_2 = x_2 + jy_2$$

$$z_1 \cdot z_2 = x_1x_2 - y_1y_2 + j(x_1y_2 + x_2y_1)$$



Complex phasors

Amplitude and phase are two distinct characteristics of a sine function that are inconvenient to keep separate notationally.

Complex functions (and discrete sequences) of the form

$$A \cdot \mathrm{e}^{\mathrm{j}\omega t+\varphi} = A \cdot [\cos(\omega t+\varphi) + \mathrm{j} \cdot \sin(\omega t+\varphi)]$$

(where $\mathrm{j}^2 = -1$) are able to represent both amplitude and phase in one single algebraic object.

Thanks to complex multiplication, we can also incorporate in one single factor both a multiplicative change of amplitude and an additive change of phase of such a function.

Least Squares and Orthogonality

(c.f. Stearns, 2003, Chap. 2)

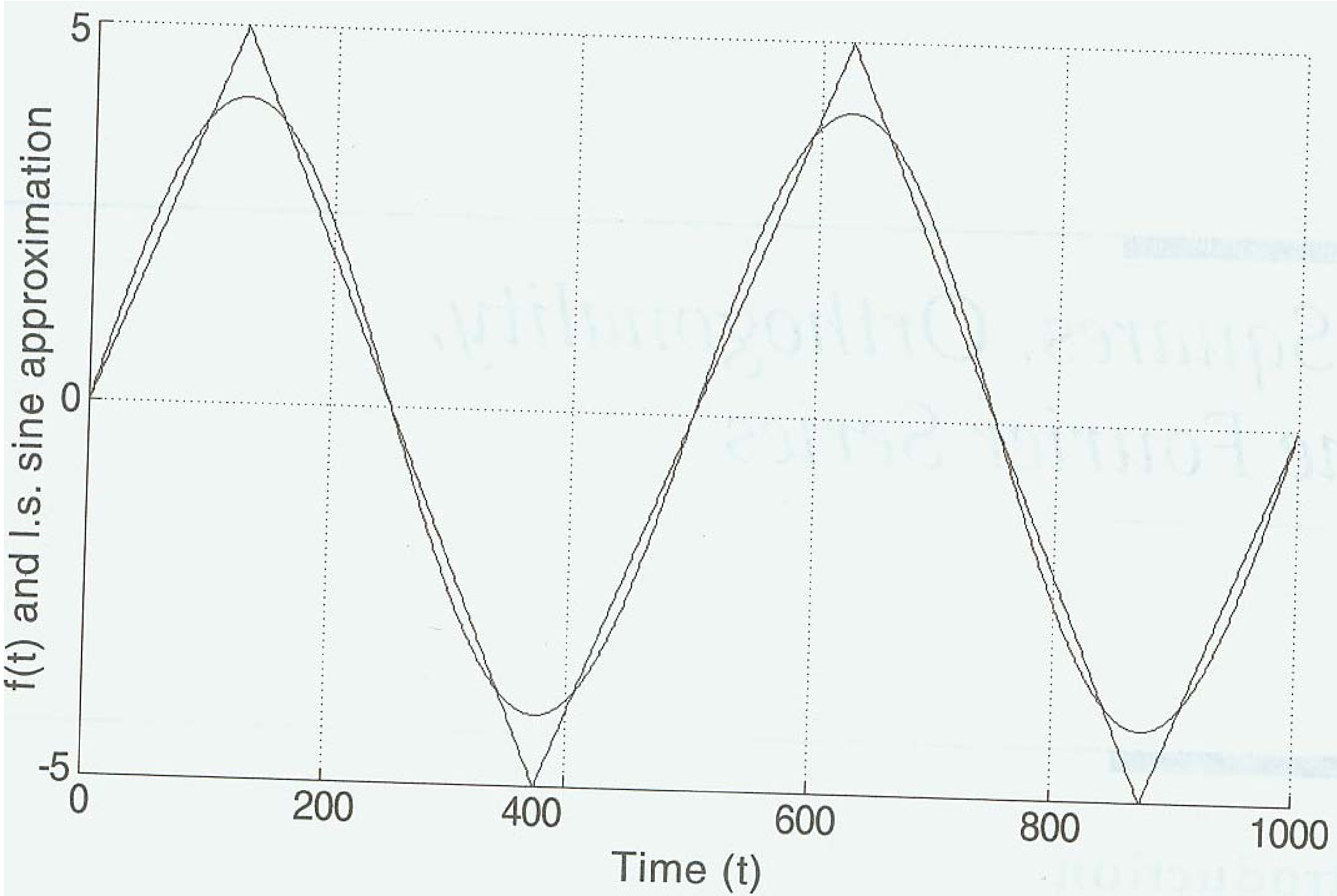
least squares:

Suppose we have two continuous functions, f(t) and g(c,t), where c is a parameter (or a set of parameters). If c is selected to minimize the total squared error (TSE)

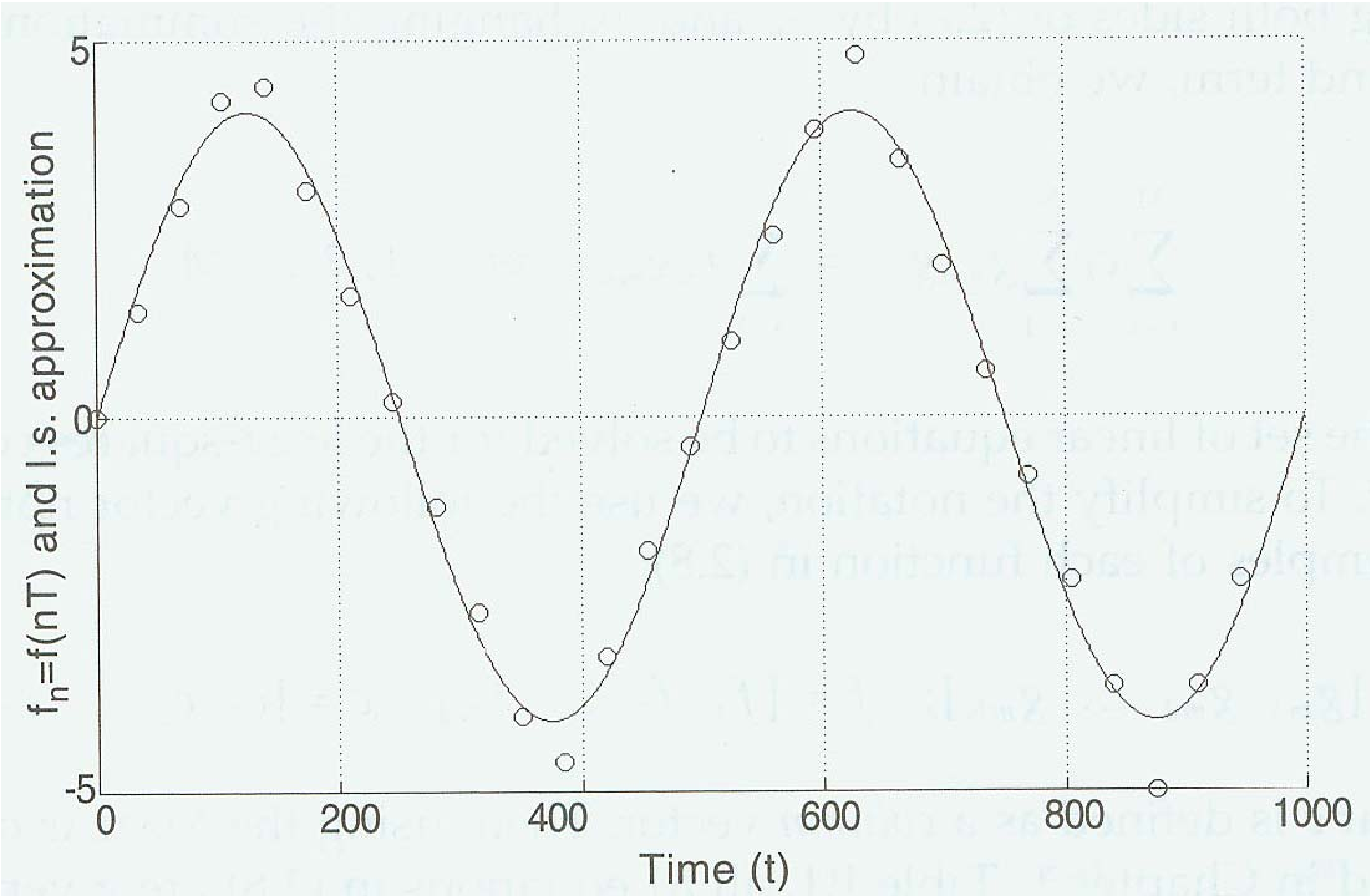
$$TSE = \int_{t_1}^{t_2} \left(f(t) - \hat{f}(c,t) \right)^2 dt$$

assume $\hat{f} = g$

An example of continuous least-squares approximation



In DSP, least squares approximations are made more often to discrete (sampled) data, rather than to continuous data



If the approximating function is again g(c,t), the total squared error in the discrete case is now given as

$$TSE = \sum_{n=1}^{N} \left(f_n - \hat{f}(c, nT) \right)^2$$

where f_n is the nth element of f, and T is the time step (interval between samples).

Assume that there are M basis functions (or bases), $g_1, \ldots, g_M$, to represent g.

$$\hat{f}(c, nT) = c_1 g_{1n} + c_2 g_{2n} + c_3 g_{3n} + \cdots + c_M g_{Mn}$$

$$TSE = \sum_{n=1}^{N} \left(f_n - \sum_{m=1}^{M} c_m g_{mn} \right)^2$$

$$\nabla(\mathrm{TSE}) = \left[\frac{\partial \mathrm{TSE}}{\partial c_1} \;\; \frac{\partial \mathrm{TSE}}{\partial c_2} \;\; \cdots \;\; \frac{\partial \mathrm{TSE}}{\partial c_M}\right] = [0 \;\; 0 \;\; \ldots \;\; 0]$$

$$\frac{\partial \mathrm{TSE}}{\partial c_m} = -2\sum_{n=1}^{N} g_{mn}\left[f_n - \sum_{k=1}^{M} c_k g_{kn}\right]$$

$$= -2\left[\sum_{n=1}^{N} g_{mn} f_n - \sum_{n=1}^{N}\sum_{k=1}^{M} c_k g_{kn}\right] = 0; \quad m = 1, 2, \ldots, M$$

Let us denote that

$$G_m = [g_{m1} \;\; g_{m2} \;\; \ldots \;\; g_{mN}]; \quad f = [f_1 \;\; f_2 \;\; \ldots \;\; f_N];$$

$$c = [c_1 \;\; c_2 \;\; \ldots \;\; c_M]'$$ (where $'$ means matrix transpose)

$$\begin{bmatrix} G_1 * G_1' & G_1 * G_2' & \cdots & G_1 * G_M' \\ G_2 * G_1' & G_2 * G_2' & \cdots & G_2 * G_M' \\ \vdots & & & \vdots \\ G_M * G_1' & G_M * G_2' & \cdots & G_M * G_M' \end{bmatrix} * \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_M \end{bmatrix} = \begin{bmatrix} f * G_1' \\ f * G_2' \\ \vdots \\ f * G_M' \end{bmatrix}$$

$$G = [G_1' \ G_2' \ \ldots \ G_M']$$

$$G' * G * c = (f * G)'$$

$$\boxed{c = (G' * G) \backslash (f * G)'}$$

This is represented in MATLAB form, where x = A\b means that x is the solution of the linear equation system Ax = b.
In this case, $c = (G^T G)^{-1} G^T b$, when $G^T G$ is nonsingular.

Matrix derivation: Least squares can be derived via another way by using matrix derivations:

TSE = $$\sum_{n=1}^{N}\left(f_n - \sum_{m=1}^{M} c_m g_{mn}\right)^2$$

Let b = $f^T = [f_1, \ldots, f_n]^T$,

then TSE = $||b - Gc||^2 = (b - Gc)^T(b - Gc)$.

$$\frac{\partial TSE}{\partial c} = 2\left(\frac{\partial(b - Gc)}{\partial c}\right)^T (b - Gc) = -2G^T(b - Gc) = 0$$

$$\Rightarrow G^T b = G^T G c$$

When G^TG is nonsingular, $c = \left(G^T G\right)^{-1} G^T b$

Orthogonal bases (or orthogonal basis functions):

In many cases, we hope the bases to be 'orthogonal' to each other. (if two row vectors a and b are orthogonal, then the inner product ab' = 0)

Advantage: suppose the n functions are mutually orthogonal with respect to the N samples,

$$G_m * G'_k = 0; \quad m \neq k$$

then each equation in solving the least squares becomes

$$G_m * G'_m * c_m = f * G'_m; \quad m = 1, 2, \ldots, M$$

the solution of c becomes very simple:

$$c_m = \sum_{n=1}^{N} f_n g_{mn} \Big/ \sum_{n=1}^{N} g_{mn}^2; \quad m = 1, 2, \ldots, M$$

An intuitive explanation: orthographic projection

The solution of c is the "orthographic projection" of the input vector f onto the subspace formed by the orthogonal bases.

We can change the number of bases, M, and the solution still remains as the same form.

- Choosing the number of bases to represent a signal establish the fundamental concept of signal compression.

Discrete Fourier Series

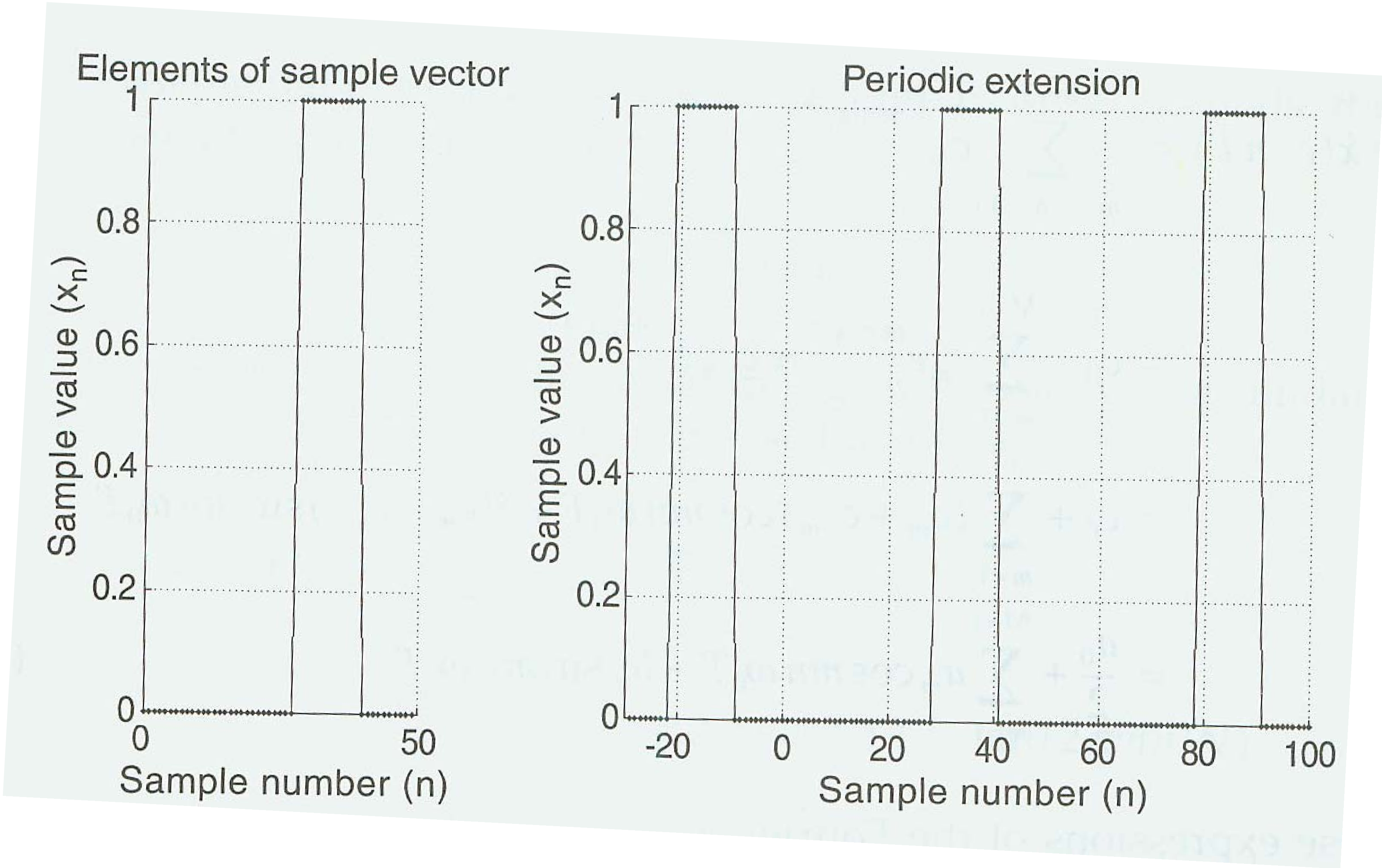
(c.f. Stearns, 2003, Chap. 2)

Harmonic analysis:

A discrete Foruier series consisits of combinations of sampled sine and cosine functions. It forms the basis of a branch of mathematics called harmonic analysis, which is applicable to the study of all kinds of natural phenomena, including the motion of stars and planets and atoms, acoustic waves, radio waves, etc.

Let $x = [x_1, \ldots, x_{N-1}]$.

If we say the fundamental period of x is N samples, we image that the samples of x repeat, over and over again, in the time domain.



Sample vector and periodic extension; N=50

The fundamental period is N samples, or NT seconds, where T is the time step in seconds.

The fundamental frequency is the reciprocal of the fundamental period, f_0 = 1/NT Hertz (HZ). 'Hertz' means "cycles per second."

Another unit of frequency besides f is

$$\omega = 2\pi f \text{ rad/s (radians per second)}$$

Fourier Series (a least-square approximation using sine and cosine bases)

$$\hat{x}_n \equiv \hat{x}(c, nT) = \sum_{m=-(M-1)}^{M-1} c_m e^{jmn\omega_0 T}$$

$$= c_0 + \sum_{m=1}^{M-1} c_m e^{jmn\omega_0 T} + c_{-m} e^{-jmn\omega_0 T}$$

$$= c_0 + \sum_{m=1}^{M-1} (c_m + c_{-m}) \cos mn\,\omega_0 T + j(c_m - c_{-m}) \sin mn\,\omega_0 T$$

$$= \frac{a_0}{2} + \sum_{m=1}^{M-1} a_m \cos mn\,\omega_0 T + b_m \sin mn\,\omega_0 T$$

Equivalence of Fourier Series Coefficients

TABLE 2.1

Equivalence of Fourier Series Coefficients

$[a, b]$ in terms of c	$a_0 = 2c_0;\quad a_m = c_m + c_{-m};\quad m > 0$ $b_0 = 0;\quad b_m = j(c_m - c_{-m});\quad m > 0$
c in terms of $[a, b]$	$c_0 = a_0/2;\quad c_m = (a_m + jb_{-m})/2;\quad m < 0$ $c_m = c'_{-m} = (a_m - jb_{-m})/2;\quad m > 0$

If the fundamental period $2\pi/w_0$ covers N samples or NT seconds, then the fundamental frequency must be

$$\omega_0 = \frac{2\pi}{NT} \quad \text{rad/s}$$

With this substitution to indicate sampling over exactly one fundamental period:

$$\begin{aligned}
\hat{x}_n \equiv \hat{x}(c, nT) &= \sum_{m=-(M-1)}^{M-1} c_m e^{j2\pi mn/N} \\
&= \frac{a_0}{2} + \sum_{m=0}^{M-1} a_m \cos(2\pi mn/N) + b_m \sin(2\pi mn/N)
\end{aligned}$$

In this form, the harmonic functions are orthogonal with respect to the N samples of x:

$$\sum_{n=0}^{N-1} \cos(2\pi mn/N)\sin(2\pi kn/N) = 0; \quad m, k \geq 0$$

$$\sum_{n=0}^{N-1} \cos(2\pi mn/N)\cos(2\pi kn/N) = 0; \quad m, k \geq 0 \text{ and } m \neq k$$

$$\sum_{n=0}^{N-1} \sin(2\pi mn/N)\sin(2\pi kn/N) = 0; \quad m, k \geq 0 \text{ and } m \neq k$$

These results can be proved by using the trigonometric identities and the geometric series application.

- We can use least squares principle to determine the best coefficients a_m and b_m.

By applying the orthographic projection, the least-squares Fourier coefficients are

$$a_m = \frac{2}{N}\sum_{n=0}^{N-1} x_n \cos(2\pi mn/N); \quad 0 \le m \le M-1$$

$$b_m = \frac{2}{N}\sum_{n=0}^{N-1} x_n \sin(2\pi mn/N); \quad 1 \le m \le M-1$$

When we use the complex exponential as bases, the coefficients c_m can be determined by a_m and b_m as:

$$c_m = \frac{1}{N}\sum_{n=0}^{N-1} x_n e^{-j2\pi mn/N}; \quad m \ge 0$$

$$c_{-m} = c'_m$$

' means the complex conjugate.

or equivalently

$$c_m = \frac{\omega_0}{2\pi}\sum_{n=0}^{N-1} x_n e^{-jm\omega_0 nT} T; \quad 0 \le m < \infty; \quad c_{-m} = c'_m$$

The results also suggest a continuous form of the Fourier series. We can image decreasing the time step, T, toward zero, and at the same time increasing N in a way such that the period, NT, remains constant. Thje samples (x_n) or x(t) are thus packed more densely, so that, in the limit, we have the Fourier series for a continuous periodic function:

$$\boxed{x(t) = \sum_{m=-\infty}^{\infty} c_m e^{jm\omega_0 t}; \quad c_m = \frac{\omega_0}{2\pi}\int_0^{2\pi/\omega_0} x(t) e^{-jm\omega_0 t}\, dt}$$

Sometimes, for the sake of symmetry, c_m is given by an integral around t=0:

$$c_m = \frac{\omega_0}{2\pi}\int_{-\pi/\omega_0}^{\pi/\omega_0} x(t)e^{-jm\omega_0 t}\,dt$$

The continuous forms of the Fourier series are, nevertheless, applicable to a wide range of natural periodic phenomena.

We have introduced two forms of the discrete Fourier series, and show how to calculate the coefficients when the samples are taken over one fundamental period of the data.