Image Processing and Analysis – Math Part

Ivar Farup

Autumn 2014

Practical Details

- Textbook: *Discrete Fourier Analysis and Wavelets* by S. A. Broughton and K. Bryan. Available in the book store and at Amazon.
- Course page in fronter.
- Lectures: check TimeEdit for details (link in Fronter)
- The lectures will be streamed using Google Hangout.

Course Structure

- Lectures
- Assignments (one per lecture) are handed in individually in fronter.
- Deadline: Announced in Fronter, at least two weeks after lecture.
- Feedback one week before deadline
- Academic honesty!

Assignments and Assessment

- Six assignments should be handed in individually in Fronter as PDF reports. The assignments will be graded together (as a portfolio), and this grade will count 40% of the final grade in the course.
- An oral exam counts the remaining 60% of the grade.
- Both the assignments and the exam must be passed to pass the course.
- The book uses MATLAB (R) for the examples. Python will be used in the lectures. You are free to choose your tool of choice for the assignments.
- Reports to be handed in should be well structured and nicely typeset. The use of LaTeX is encouraged, but not required.

Workload

- 60 ECTS is 1500–1800 hours
- 10 ECTS is 250–300 hours
- 5 ECTS is 125–150 hours
- In six sessions only, this means 20–25 hours per session
  - Prepare for lecture by reading the book chapter
  - Attend lecture
  - Do some exercises
  - Do the assignment
  - Iterate the assignment based on feedback

Image Processing and Analysis – Math Part: Vector Spaces, Signals and Images

Ivar Farup

Based upon Chapter 1 of Broughton and Bryan's *Discrete Fourier Analysis and Wavelets* and Maciej Piętka's Lecture Notes from 2010

Overview

- Common image processing problems
- Signals and images
- Vector space models for signals and images
- Decomposition into basic waveforms
- Vector spaces with inner product
- Orthogonality and orthogonal decompositions
- Signal and image digitization

Common Image Processing Problems

Image Compression

Original, 660KB

JPG, 28 KB

JPG, 14 KB

JPG, 4.7 KB

Common Image Processing Problems

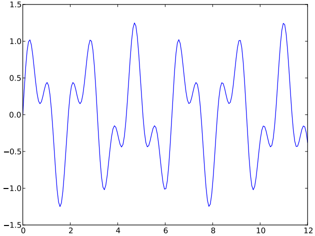
Image Denoising

Common Image Processing Problems

Edge Detection

- A real-valued function of one variable, e.g.,

$$x(t) = 0.75 \sin(3t) + 0.5 \sin(7t)$$



- Explicit formula $x(t)$ rarely known
- Real signals are complex and noisy
- Impossible to store analog signal in a digital computer

- Real-valued function $f(x, y)$ defined on a 2D region, usually rectangle
- Values of $f(x, y)$ represent intenisty/lightness of the image at (x, y)
- Typically, $f(x, y)$ is discontinuous

- A colour can be described by three numbers, e.g., R, G, and B
- These can be seen as the components of a vector:

$$\mathbf{f}(x, y) = \mathbf{e}_R f_R(x, y) + \mathbf{e}_G f_G(x, y) + \mathbf{e}_B f_B(x, y)$$

Colour

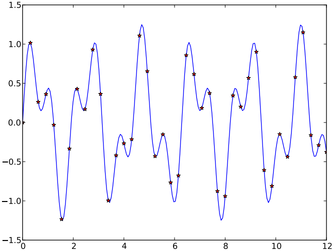
R

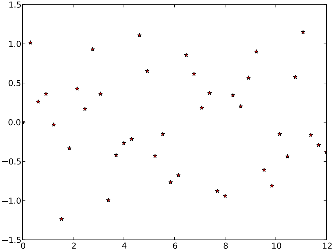
G

B

- To store a signal in a computer, it must be represented by numbers
- To get from a continuous signal to discrete numbers, sampling is used
- Sampling interval: ΔT
- Sampling rate: $1/\Delta T$
- $N + 1$ samples of the function $x(t)$ are represented by the vector

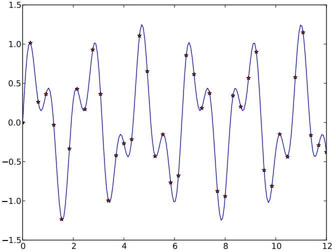
$$\mathbf{x} = (x_0, x_1, \ldots, x_N)$$

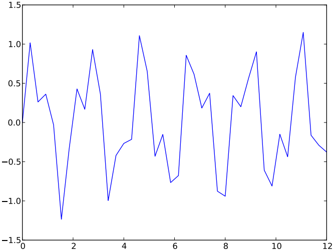




- Causes information loss
- Interpolation needed to restore signal

Signals and Images
Reconstruction





- Simple linear interpolation is not always enough
- Smarter solution sought for. . .

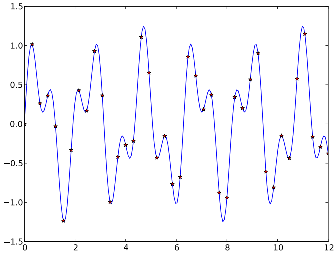
- $x(t)$ and $f(x, y)$ cannot be measured with infinite presicion
- Limitations of representation to, e.g., 8 bits: 0–255
- Intermediate values are rounded to nearest integer
- Introduces quantisation error

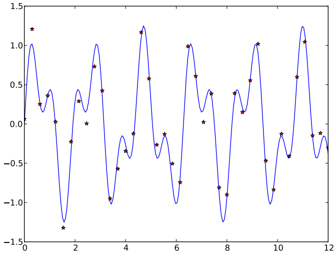
- All measurements (mic, CCD) are subject to noise
- Let $\mathbf{x}$ be the original noise-free sampled signal
- Noisy samples $y_n = x_n + \epsilon_n$
- Different types of noise:
  - Impulse noise
  - Gaussian noise:

$$p(\epsilon) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{(\epsilon - \mu)^2}{2\sigma^2}$$

(μ = mean, often zero, and σ^2 = variance)

Gaussian noise with $\sigma = 0.1$





Signals and Images

Noise

20% Impulse noise

Gaussian noise, $\sigma = 0.2$

- Noise
- Sampling
- Quantisation

- A vector space over $\mathbb{R}$ (or $\mathbb{C}$) is a set V with two operations defined: vector addition and scalar multiplication
- For all vectors $\mathbf{u}, \mathbf{v} \in V$, also $\mathbf{u} + \mathbf{v} \in V$.
- For all $\mathbf{u} \in V$ and $a \in \mathbb{R}$ (or $\mathbb{C}$), $a\mathbf{u} \in V$

$$
\begin{aligned}
\mathbf{u} + \mathbf{v} &= \mathbf{v} + \mathbf{u} \quad \text{(Commutative)} \\
(\mathbf{u} + \mathbf{v}) + \mathbf{w} &= \mathbf{u} + (\mathbf{v} + \mathbf{w}) \quad \text{(Assosiative)} \\
\mathbf{u} + \mathbf{0} &= \mathbf{u} \quad \text{(Zero vector)} \\
1\mathbf{u} &= \mathbf{u} \quad \text{(Multiplicative identity)} \\
\mathbf{u} + (-\mathbf{u}) &= 0 \quad \text{(Additive inverse)} \\
(ab)\mathbf{u} &= a(b\mathbf{u}) \quad \text{(Distributive, multiplication)} \\
a(\mathbf{u} + \mathbf{v}) &= a\mathbf{u} + a\mathbf{v} \quad \text{(Distributive, addition)}
\end{aligned}
$$

- Finite sampled audio signal: $\mathbb{R}^N$
  - Vector elements are the samples
  - Addition component by component: mixing
  - Multiplication by scalar: amplification
  - Zero vector: silence
  - Additive inverse: negative of all components
- Sampled image: $\mathbb{R}^{M \times N}$
  - Matrix elements are the samples
  - Addition: blending
  - Multiplication by scalar: contrast
  - Zero vector: black image
  - Additive inverse: negative image

- Bounded, infinite time: $L^{\infty}(\mathbb{N})$
  - $\mathbf{x} = (x_0, x_1, \ldots)$, $x_i \in \mathbb{R}$
  - $|x_k| \leq M$ for all $k \geq 0$.
- Finite energy, infinite time: $L^2(\mathbb{N})$
  - $\mathbf{x} = (x_0, x_1, \ldots)$, $x_i \in \mathbb{R}$
  - Finite energy:

$$\sum_{k=0}^{\infty} |x_k|^2 < \infty$$

- A problem is defined in some setting where it is difficult to solve
- Transform the problem to a new domain
- Solve the transformed problem
- Transform the solution back to the original domain

$$\begin{aligned}(1,1) &= 1 \times (1,0) + 1 \times (0,1) \\ &= \sqrt{2} \times \frac{(1,1)}{\sqrt{2}} + 0 \times \frac{(-1,1)}{\sqrt{2}}\end{aligned}$$

$$e^{i\omega t} = \cos(\omega t) + i\sin(\omega t)$$

- Expresses a wave with angular frequency $\omega = 2\pi f$, where f is the frequency
- Periodic in t with period $\lambda = 2\pi/|\omega|$

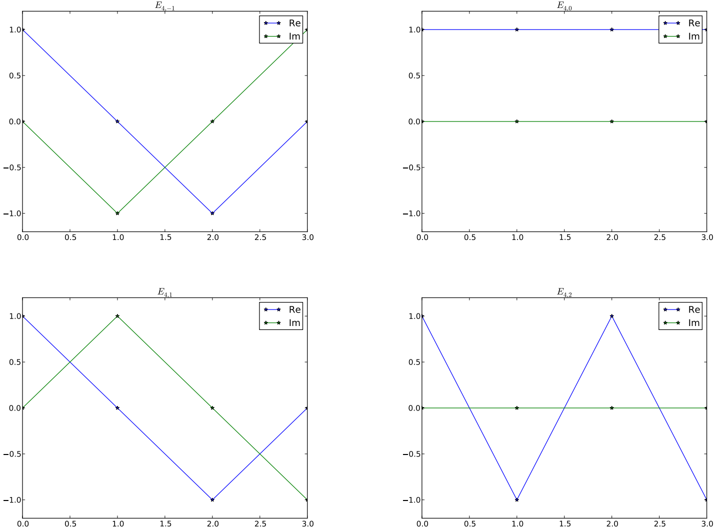
- Signal $\mathbf{x} = (x_0, x_1, \ldots, x_{N-1}) \in \mathbb{R}^N$
- Samples taken at $t_n = nT/N$, $n = 0, 1, \ldots, N-1$
- Sampling interval $\Delta T = T/N$
- Appropriate basic waveforms: $\exp(2\pi ikt/T)$, sampled at $t = nT/N$, for any $k \in \mathbb{Z}$
- Basic waveform vectors (fixed N, any $k \in \mathbb{Z}$):

$$\mathbf{E}_{N,k} = \begin{pmatrix} \exp(2\pi ik0/N) \\ \exp(2\pi ik1/N) \\ \cdots \\ \exp(2\pi ik(N-1)/N) \end{pmatrix}$$

- Corresponds to analog waveform with $\omega = 2\pi k/T$

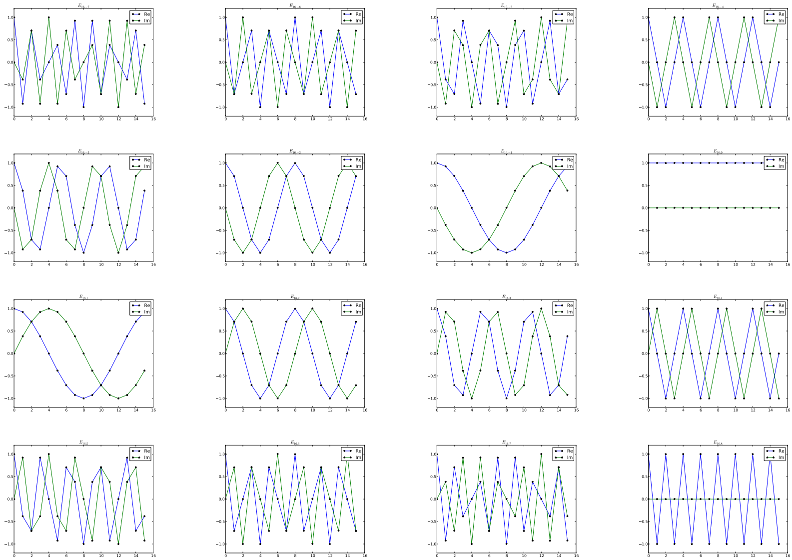
Basic Waveforms

Visualisations



Basic Waveforms

Visualisations



The components of $\mathbf{E}_{N,k}$ are complex N-th roots of unity:

$$\begin{aligned}
\mathbf{E}_{N,k}(m)^N &= \exp(2\pi ikm/N)^N \\
&= \exp(2\pi ikm) \\
&= \exp(2\pi i)^{km} \\
&= 1
\end{aligned}$$

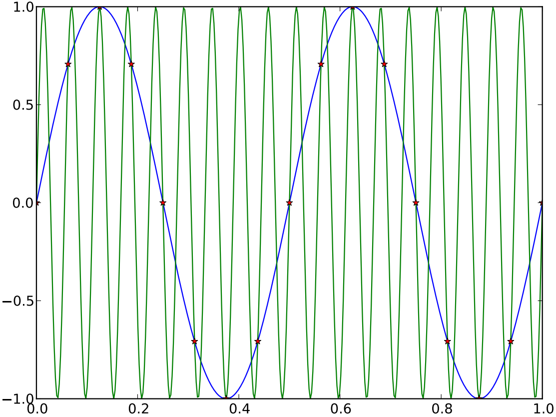
- Aliasing:

$$\begin{aligned}\mathbf{E}_{N,k+nN}(m) &= \exp(2\pi i(k+nN)m/N) \\ &= \exp(2\pi ikm/N)\exp(2\pi ikn) \\ &= \exp(2\pi ikm/N) \\ &= \mathbf{E}_{N,k}(m)\end{aligned}$$

- We need only use $-N/2 < k \leq N/2$
- Any signal outside this range is aliased

Basic Waveforms

Aliasing



- $k_{\max} = N/2$ is the maximum allowed frequency index
- Frequencies greater than $k_{\max}/T = 1/(2\Delta T)$ cannot be reconstructed from the samples
- $1/(2\Delta T)$ is the Nyquist frequency (half the sample rate)
- Shannon: Perfect reconstruction is possible when the highest frequency present in the analog signal was smaller than the Nyquist frequency
- Sampling rate of CD audio: 44100 Hz
- Human hearing: 12–20000 Hz
- To avoid aliasing, the signal must be filtered *before* sampling

- Generalisation of the dot product
- Adds geometric notions to vector space formalism:
  - Length
  - Distance
  - Angle
  - Orthogonality

- Let V be a vector space over $\mathbb{C}$ (or $\mathbb{R}$)
- The inner product is a function from $V \times V$ to $\mathbb{C}$ (or $\mathbb{R}$)
- The inner product of the vectors $\mathbf{v}$ and $\mathbf{w}$ is denoted $(\mathbf{v}, \mathbf{w})$
- The inner product satisfies
  - $(\mathbf{v}, \mathbf{w}) = \overline{(\mathbf{w}, \mathbf{v})}$ (conjugate symmetry)
  - $(a\mathbf{u} + b\mathbf{v}, \mathbf{w}) = a(\mathbf{u}, \mathbf{v}) + b(\mathbf{v}, \mathbf{w})$ (linearity in first argument)
  - $(\mathbf{v}, \mathbf{v}) \geq 0$ for all $\mathbf{v} \in V$, and $(\mathbf{v}, \mathbf{v}) = 0 \iff \mathbf{v} = \mathbf{0}$
- A vector space with an inner product is called an *inner product space*

- In case of a vector space over $\mathbb{R}$, $(\mathbf{v}, \mathbf{w}) = (\mathbf{w}, \mathbf{v})$
- $(\mathbf{v}, \mathbf{v})$ is always real-valued (due to conjucate symmetry)
- $(\mathbf{w}, a\mathbf{u} + b\mathbf{v}) = \bar{a}(\mathbf{w}, \mathbf{u}) + \bar{b}(\mathbf{w}, \mathbf{v})$ (conjugate linearity in the second argument)

- A *norm* quantifies the size/length of vectors
- A norm on a vector space V is a function $||\mathbf{v}||$ from V to $\mathbb{R}$ satisfying
  - $||\mathbf{v}|| \geq 0$ for all $\mathbf{v} \in V$
  - $||\mathbf{v}|| = 0 \iff \mathbf{v} = 0$
  - $||a\mathbf{v}|| = |a| \, ||\mathbf{v}||$
- Every inner product space has a 'natural' norm $||\mathbf{v}|| = \sqrt{(\mathbf{v}, \mathbf{v})}$
- $||\mathbf{v}||^2$ is often denoted the 'energy' of $\mathbf{v}$

- Vectors $\mathbf{x} = (x_1, \ldots, x_N)$ and $\mathbf{y} = (y_1, \ldots, y_N) \in \mathbb{C}^N$
- Dot product on $\mathbb{C}^N$:

$$(\mathbf{x}, \mathbf{y}) = x_1\overline{y_1} + \cdots + x_N\overline{y_N} = \sum_{i=1}^{N} x_i\overline{y_i}$$

- Corresponding norm (Euclidean norm):

$$||\mathbf{x}|| = \sqrt{|x_1|^2 + \cdots + |x_N|^2}$$

- Corresponding distance between two vectors $\mathbf{x}$ and $\mathbf{y}$:

$$||\mathbf{x} - \mathbf{y}|| = \sqrt{|x_1 - y_1|^2 + \cdots + |x_N - y_N|^2}$$

- A and B matrices with entries a_{jk} and b_{jk}
- Inner product defined as

$$(A, B) = \sum_{j=1}^{m} \sum_{k=1}^{n} a_{jk} \overline{b_{jk}}$$

- Corresponding norm (Frobenius norm):

$$||A|| = \sqrt{\sum_{j=1}^{m} \sum_{k=1}^{n} |a_{jk}|^2}$$

- In essence identical to the Euclidean norm on $\mathbb{C}^{mn}$

- The *p-norm*

$$||\mathbf{x}||_p = (|x_1|^p + \cdots + |x_n|^p)^{1/p}$$

- The *1-norm* or *city-block distance*:

$$||\mathbf{x}||_1 = |x_1| + \cdots + |x_n|$$

- The *maximum norm*:

$$||\mathbf{x}||_\infty = \max_{1 \le j \le n} |x_j|$$

- The *maximum norm* for matrices:

$$||A||_\infty = \max_{1 \le j \le m} \sum_{k=1}^{n} |a_{jk}|$$

- Let V be an inner product space
- Two vectors $\mathbf{v}, \mathbf{w} \in V$ are orthogonal if $(\mathbf{v}, \mathbf{w}) = 0$

The standard dot product in $\mathbb{R}^3$ and the Euclidean norm

- Let the directions of $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3$ form the angle θ

$$\mathbf{x} \cdot \mathbf{y} = \sum_{j=1}^{3} x_j y_j = ||\mathbf{x}|| \, ||\mathbf{y}|| \cos \theta$$

- Two non-zero vectors $\mathbf{x}$ and $\mathbf{y}$ are orthogonal if and only if they are perpendicular, $\theta = \pi/2$

- A subset $S \subset V$ of vectors in V is orthogonal if $(\mathbf{v}, \mathbf{w}) = 0$ for every pair of distinct vectors $\mathbf{v}, \mathbf{w} \in S$
- An orthogonal set S is *orthonormal* if $||\mathbf{v}|| = 1$ for each $\mathbf{v} \in S$

- The standard basis in $\mathbb{R}^N$: $S = \{\mathbf{e}_1, \ldots, \mathbf{e}_N\}$
- $\mathbf{e}_k = (0, \ldots, 0, 1, 0, \ldots, 0)$ with 1 in the k-th place
- The set is orthogonal under the Euclidean inner product,

$$(\mathbf{e}_j, \mathbf{e}_k) = 0 \text{ when } j \neq k$$

- The set is orthonormal since

$$(\mathbf{e}_j, \mathbf{e}_j) = 1$$

- The set S of basic discrete waveforms in $\mathbb{C}^N$: $S = \{\mathbf{E}_{N,k}\}$ with fixed N and k in the range $-N/2 < k \leq N/2$
- S is orthogonal:

$$(\mathbf{E}_{N,k}, \mathbf{E}_{N,l}) = 0 \text{ when } k \neq l$$

$$
\begin{aligned}
(\mathbf{E}_{N,k}, \mathbf{E}_{N,l}) &= \sum_{r=0}^{N-1} e^{2\pi ikr/N}\overline{e^{2\pi ilr/N}} \\
&= \sum_{r=0}^{N-1} e^{2\pi ikr/N} e^{-2\pi ilr/N} \\
&= \sum_{r=0}^{N-1} e^{2\pi i(k-l)r/N} \\
&= \sum_{r=0}^{N-1} \left(e^{2\pi i(k-l)/N}\right)^r
\end{aligned}
$$

The sum of the N terms of a geometric series with the common ratio $z = e^{2\pi i(k-l)/N} \neq 1$

$$
\begin{aligned}
(\mathbf{E}_{N,k}, \mathbf{E}_{N_l}) &= \frac{1 - z^N}{1 - z} \\
&= \frac{1 - e^{2\pi i(k-l)}}{1 - e^{2\pi i(k-l)/N}} \\
&= 0
\end{aligned}
$$

when $k \neq l$

- Let V be a vector space over $\mathbb{C}$ or $\mathbb{R}$
- Consider a finite subset $S = (\mathbf{v}_1, \ldots, \mathbf{v}_N) \subset V$
- If for any such subset, the only solution to

$$\alpha_1 \mathbf{v}_1 + \cdots + \alpha_N \mathbf{v}_N = 0$$

  is $\alpha_k = 0$ for all $1 \leq k \leq N$, then S is said to be *linearly independent*
- No vector can be expressed as a combination of the remaining vectors
- Example: any two non-zero, non-parallel vectors in the plane are linearly independent
- Any orthogonal set of non-zero vectors is linearly independent

- Let S be a set of vectors in a vector space V over $\mathbb{C}$ of $\mathbb{R}$
- If every vector $\mathbf{v} \in V$ can be expressed as a finite linear combination of elements of S:

$$\mathbf{v} = \alpha_1 \mathbf{v}_1 + \cdots + \alpha_N \mathbf{v}_N$$

  for suitable $\alpha_k \in \mathbb{C}$ (or $\mathbb{R}$) and $\mathbf{v}_k \in S$, then S is said to *span* the vector space V
- If a spanning set S is linearly independent, it is a *basis* for V
- Any vector in V can be built from elements of the basis in a unique way

- A vector space can have many different bases; often infinitely many
- If a basis for V is finite: $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_N\}$, then V is called *finite-dimensional*
  - The dimension of a finite-dimensional vector space equals the number of vectors in the basis
  - Any set of N linearly independent vectors in V is a basis
- $\mathbb{R}^N$ and $\mathbb{C}^N$ are N-dimensional vector spaces over $\mathbb{R}$ and $\mathbb{C}$, respectively

An orthogonal basis for a vector space V is particularly important and useful

- The standard basis for $\mathbb{R}^N$ and $\mathbb{C}^N$:

$$\mathbf{e}_k = (0, \ldots, 0, 1, 0, \ldots, 0) \text{ with 1 in the } k\text{-th place}$$

- The set of discrete basic waveforms $\{\mathbf{E}_{N,k}\}$ with fixed N and k in the range $-N/2 < k \leq N/2$:

$$\mathbf{E}_{N,k}(m) = e^{2\pi ikm/N}, \text{ with } 0 \leq m \leq N-1$$

- Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_N\}$ be an orthogonal basis for V
- Any $\mathbf{v} \in V$ can be expressed as

$$\mathbf{v} = \sum_{k=1}^{N} \alpha_k \mathbf{v}_k$$

  with $\alpha_k = (\mathbf{v}, \mathbf{v}_k)/||\mathbf{v}_k||^2$
- Proof: Take the inner product of both sids with any $\mathbf{v}_m \in S$

Decomposition formula in the $\{\mathbf{E}_{N,k}\}$ basis for $\mathbb{C}^N$

- The norm of the basis vector $\mathbf{E}_{N,k}$:

$$||\mathbf{E}_{N,k}||^2 = (\mathbf{E}_{N,k}, \mathbf{E}_{N,k}) = \sum_{r=0}^{N-1} e^{2\pi ikr/N}\overline{e^{2\pi ikr/N}}$$

$$= \sum_{r=0}^{N-1} e^{2\pi ikr/N} e^{-2\pi ikr/N} = \sum_{r=0}^{N-1} 1 = N$$

- Thus,

$$\mathbf{x} = \frac{1}{N}\sum_{k=0}^{N-1}(\mathbf{x}, \mathbf{E}_{N,k})\mathbf{E}_{N,k}$$

for any $\mathbf{x} \in \mathbb{C}^N$

- Signal and image digitization is affected by quantisation errors
- Finite presision is used to store numbers in a computer
- Shorter data types are preferred over longer types
  - Less storage space required
- Integer types preferred over floating point
  - More efficient processing in hardware

Quantisation Map: A function from the samples to their quantised representation. Typically, $q : \mathbb{R} \to \{0, 1, \ldots, r-1\}$, where r is the number of *quantisation intervals*

Dequantisation Map: A function from the quantised number representation of the signal back to the space of the signal. Typically, $\tilde{q} : \{0, 1, \ldots, r-1\} \to \mathbb{R}$

Code Book: The set $\{z_k\}_{k=0,1,\ldots,r-1}$ where $z_k = \tilde{q}(k)$ is called the code book

Reconstruction: The approximate reconstruction of $\mathbf{x}$ is $\tilde{\mathbf{x}}$ with components

$$\tilde{x}_j = \tilde{q}(q(x_j)) = z_{q(x_j)}$$

$$mD = 100\frac{||\mathbf{x} - \tilde{\mathbf{x}}||^2}{||\mathbf{x}||^2}$$

- The percentage of the total signal energy
- By adjusting the quantisation map and the dequantisation map, one can minimise distortion for any particular signal or image

Signal and Image Quantisation

Example

6 bits

5 bits

4 bits

3 bits

Do the following exercises in the text book: 1.5, 1.6, 1.7, 1.12, 1.13, 1.19 (use a computer), 1.20, 1.23, 1.30