Linnæus University
School of Computer Science, Physics and Mathematics

Master Thesis

Image Compression by Using Haar Wavelet Transform and Singular Value Decomposition

Zunera Idrees
1982-08-15
Eliza Hashemiaghjekandi
1979-08-18
Subject: Mathematics
Level: Advance
Course code: 4MA11E

Abstract

The rise in digital technology has also rose the use of digital images. The digital images require much storage space. The compression techniques are used to compress the data so that it takes up less storage space. In this regard wavelets play important role. In this thesis, we studied the Haar wavelet system, which is a complete orthonormal system in $L^2(\mathbb{R})$. This system consists of the functions φ the father wavelet, and ψ the mother wavelet. The Haar wavelet transformation is an example of multiresolution analysis. Our purpose is to use the Haar wavelet basis to compress an image data. The method of averaging and differencing is used to construct the Haar wavelet basis. We have shown that averaging and differencing method is an application of Haar wavelet transform. After discussing the compression by using Haar wavelet transform we used another method to compress that is based on singular value decomposition. We used mathematical software MATLAB to compress the image data by using Haar wavelet transformation, and singular value decomposition.

Key-words: Multi resolution analysis; Haar wavelet transformation(MRA); Average and difference method; Singular value decomposition(SVD).

Acknowledgment

This thesis is written as the part of Magister in Mathematics and Modelling at Linnaeus University, Sweden. Our deepest gratitude goes to our supervisor Irina Askeritova for her support and guidance. She supported us throughout the writing of this thesis. We would also like to thank the respectable faculty members at the department of mathematics. At the end, we thank our family and friends for their support and encouragement.

Contents

1 Introduction **4**

2 Basics Concepts of Linear Algebra **5**

3 Multiresolution Analysis **8**
3.1 Definition of Multiresolution Analysis 9

4 Haar Wavelet Transform **15**
4.1 Wavelet Transformation of a Signal 18
4.2 Haar Normalized Basis 21
4.3 Two-Dimensional Haar Wavelet Transform 23

5 Averaging and Differencing Technique **24**
5.1 Transformation of The Image Matrix 24
5.2 Averaging and Differencing Technique with a Row vector 24
5.3 Transformation of Image Matrix using Linear Algebra 26
5.4 Approximation and Detail Coefficients in One and Two Dimensional Transformation 29
5.5 Image Results using Transformation Matrices 32

6 Application of Haar Wavelet **38**
6.1 Compression 38

7 Compressed Results **40**
7.1 Gray Compressed Pictures 40
7.2 Colored Compressed Pictures 40

8 Normalized Transformation **45**
8.1 Normalized Wavelet Transform 45
8.2 Compression 46
8.3 Rebuilding of Normalized Compressed Image 47

9 Singular Value Decomposition **49**
9.1 Diagonalization 49
9.2 Diagonalization of Symmetric Matrices 50
9.3 Singular Value Decomposition 51
9.4 Application of Singular Value Decomposition 53
9.5 Compressed Results 55

10 Conclusion **57**

References **58**

1 Introduction

In this thesis, we study the Haar wavelets and Singular Value Decomposition, and apply them to compress the image data. The Haar wavelet is a certain sequence of functions, and it is known as the first wavelet ever. This sequence was first introduced by Alfréd Haar. The Haar wavelets are frequently used in signal processing. The digital images require huge amount of storage space because they have redundant data. To save storage space, different compression techniques are used to get rid of redundant data. When the large amount of data is downloaded from the internet, it requires much storage space and time. It also needs a fast internet connection. This workload can be reduced if the data is compressed. For instance, a grey scale image of size 256×256 has 65536 elements to store. By using the compression techniques, we can compress the pixels but the image quality is affected a little. Now the computers are capable to store a large amount of data. The low Internet connection can take a considerable amount of time to download the large amount of data. The wavelet transform increases the speed of this procedure. When some one clicks on an image to download, the computer recalls the wave transformed matrix from the computer memory. It proceeds with all approximation coefficients and then detailed coefficients. At the end it sends the progressively smaller coefficients. As soon as the computer receives this information it starts to reconstruct the image in progressively greater detail until the original image is reconstructed. Data compression has a long history. Federal Bureau of Investigation (FBI) is using the compression technique to manage the massive collection of finger print cards. There are more than 200 million finger print cards and the amount of cards is increasing at the rate of $30,000 - 50,000$ new cards. For instance, the finger print stores about $70,000$ pixels and that require about 0.6 MB storage space. The finger prints of both hands require 6 MB storage space. FBI's current archive requires more than 200 terabytes(a terabyte is 10^2 bytes). This requires much storage space, and the current computers are not capable to store this vast amount of data. For the requirement of time and storage space these files must be compressed to send them over a modem. Wavelets are used to compress the data. The discrete wavelets are used to compress the finger prints data.

The structure of the thesis is: Chapter 2 consists of basic definitions that are helpful to understand the theory of Haar wavelet transform and singular value decomposition methods. In chapter 3, we discuss the Multiresolution analysis, which is the complete orthonormal system in $L^2(\mathbb{R})$ consisting of the set of dilates and translates of a single function ψ. Chapter 4 represents the normalized and non-normalized Haar wavelet system, which is the first example of orthonormal wavelet basis. The Haar wavelet basis are constructed on the interval $[0,1]$ to compress the one and two dimensional signals. Chapter 5, chapter 6, chapter 7 and chapter 8 consist of the application of normalized and non-normalized Haar wavelet transform. In chapter 9, the method of singular value decomposition is used to compress the image data.

2 Basics Concepts of Linear Algebra

In this chapter, we recall some basic definitions and theorems from the linear algebra that will help us to understand the concepts of Haar wavelet transform and Singular value decomposition. This material in this chapter is taken from the books [5] and [4].

Vector Space

Let us remind first the basic definition of a vector space.

Definition 2.1. Vector Space

A vector space over the field $\mathbb{F}$ is a non-empty set V closed with respect to binary operations called addition and multiplication satisfying the following axioms:

1. Closure
   For any $u,\ v\ \in\ V$ the vector $u+v \in V$.

2. Commutativity
   For any $u,\ v,\ w\ \in\ V$,
$$u+v=v+u.$$

3. Associativity
   For any $u,\ v\ \in V$,
$$(u+v)+w=u+(v+w).$$

4. Existence of Additive Identity
   There is a zero vector $0 \in V$ such that $u+0=u$.

5. Existence of Additive Inverse
   For every vector $u \in V$, there is a vector $-u \in V$ such that $u+(-u)=0$.

6. Closure for Scalar Multiplication
   For any $u\ \in\ V$ and $\alpha\ \in\ \mathbb{F}$ the vector $\alpha u \in V$.

7. Associativity of scalar multiplication
   For any $u\ \in\ V$ and $\alpha,\ \beta\ \in\ \mathbb{F}$,
$$\alpha(\beta u)=(\alpha\beta)u.$$

8. Multiplicative Identity
   For any $u\ \in\ V$, there exist an element $1\ \in\ V$ such that $1u=u$.

1. First Distributive Property
   For any $u,\ v\ \in\ V$ and $\alpha\ \in\ \mathbb{F}$,
$$\alpha(u+v)=\alpha u+\alpha v.$$

2. Second Distributive property
   For any $u,\ v\ \in\ V$ and $\alpha\ \in\ \mathbb{F}$,
$$(u+v)\alpha=\alpha u+\alpha v.$$

The subspace of vector space is defined as:

Definition 2.2. Subspace

Let V be a vector space and W is a subset of V. If the subset W itself is a vector space, then we say that W is a subspace of the vector space V.

Definition 2.3. Sum and Direct Sum

Let us suppose that V be a vector space over the field $\mathbb{F}$. If U and W are subspaces of V, then

$$U+W=\{u+w: u\in U, w\in W\}.$$

The vector space V is said to be the direct sum of subspaces U and W if for every element $v\in V$ there exists unique elements $u\in U$ and $w\in W$ such that $v=u+w$. The direct sum is denoted by $V=U\oplus W$. The direct sum can be generalized for 'n' subspaces. If $U_1,U_2,\cdots,U_n$ are subspaces of V, then the direct sum for $U_1,U_2,\cdots,U_n$ subspaces is denoted as $V=\bigoplus_{i=1}^{n}$.

Theorem 2.1. *Let U and W be subspaces of a vector space V. Then V is a direct sum of U and W if $U+W=V$ and $U\cap W=\{0\}$.*

Definition 2.4. Spanning Set

Let V be a vector space over a field $\mathbb{F}$, then the set of all linear combinations of vectors $w_1,w_2,\cdots,w_n$, denoted by span$\{w_1,\ w_2,\ \cdots,\ w_n\}$ is

$$\text{span}\{w_1,\ w_2,\ \cdots,\ w_n\}=\{w: w=\sum_{j=1}^{n}\beta_j w_j,\ \beta_j\ \in\ \mathbb{F}\text{ for all } j=1,\ 2,\ \cdots,\ n.\}$$

Definition 2.5. Basis

Let W be a subspace of a vector space V. A basis for W is set of a linearly independent vectors $\{e_1,e_2,\cdots,e_n\}\in W$ such that W=span$\{e_1,e_2,\cdots,e_n\}$.

Lemma 2.2. *Let V be a vector space over the field $\mathbb{F}$ and let W be a non-empty subset of V.*

If W is finite, that is $W=\{w_1,w_2,\cdots,w_n\}$ for some $n\in\mathbb{N}$ and $w_j\neq w_k$ for $j\neq k$. Then W is a basis for V if and only if for each $v\in V$, there exists unique $\alpha_1,\alpha_2,\cdots,\alpha_n\in\mathbb{F}$ such that

$$v=\sum_{j=1}^{n}\alpha_j w_j.$$

In this thesis, we need a definition of a linear operator.

Definition 2.6. Linear Operator

Let U and V be vector spaces. Then an operator $T: U\to V$ is linear if

i.
$$T(u_1+u_2)=T(u_1)+T(u_2)\text{for all} u_1,u_2\in U.$$

ii.
$$T(\alpha u)=\alpha T(u)\text{ for all}\alpha\in\mathbb{F}\text{and} u\in U.$$

Inner Product Space

In this section, we give a brief introduction to an inner product space.

Definition 2.7. Inner Product Space

Let V be a vector space over the field $\mathbb{F}$. The inner product $< \cdot, \cdot >: V \times V \to \mathbb{R}$ is a function that maps $V \times V \to \mathbb{R}$ with the following properties:

1. For any $u \in V, < u, u >= 0 \Leftrightarrow < u, u > \geq 0$, the positive definiteness.
2. For any $u,\ v,\ w \in V$
$$< u+v, w >=< u, w > + < v, w >,$$
3. For any $\alpha\ \in \mathbb{F}$ and $u,\ v \in V$
$$< \alpha u, v >= \alpha < u, v >,$$
4. For any $u,\ v \in V$
$$< u, v >=< v, u >,$$

A vector space V with an inner product is called an inner product space.

Example 2.1. Let $f,\ g \in V$. Let V be a space of real valued functions on $[0,1]$. In our thesis, we will consider the inner product of $f,\ g$ in the space $L^2(\mathbb{R})$ as

$$< f, g >= \int_0^1 f(t)\overline{g(t)}dt,$$

$< f, g >$ satisfies all the properties (1 - 4) of inner product.

Definition 2.8. Orthogonality

Let $u,\ v$ be two vectors in an inner product space V. The vector u is orthogonal to vector v, denoted by $u \perp v$, if $< u, v >= 0$.

Definition 2.9. Orthonormal Sets

A set of vectors $\{v_1, \cdots v_i\} \in \mathbb{R}^n$ is orthonormal if

$$< v_i,\ v_j >= \begin{cases} 0, & \text{if } i \neq j \\ 1, & \text{if } i = j \end{cases}$$

3 Multiresolution Analysis

In this chapter, we describe the construction of a complete orthonormal wavelet system in $L^2(\mathbb{R})$, which consists of a set of translates and dilates of a single function ψ. Our aim is to use the basic construction of multiresolution analysis to construct a wavelet system . First, we discuss some definitions that we need to introduce the wavelets. The material in this chapter is taken from [3], [4], [6] and [9].

Definition 3.1. Translation
If $f : \mathbb{R} \to \mathbb{R}$ and $y \in \mathbb{R}$, then the translation $T_y : \mathbb{R} \to \mathbb{R}$, is an operator defined by

$$T_y f(x) = f(x-y).$$

Definition 3.2. t-Dilation
If $f : \mathbb{R} \to \mathbb{R}$ and $t > 0$, then the t-dilation $f_t : \mathbb{R} \to \mathbb{R}$, is defined by

$$f_t(x) = \frac{1}{t} f(\frac{x}{t}).$$

It is clear that by changing of variables, we have,

$$\begin{aligned}\int_{\mathbb{R}} f_t(x)dx &= \int_{\mathbb{R}} f(\frac{x}{t})\frac{dx}{t} \\ &= \int_{\mathbb{R}} f(x)dx.\end{aligned}$$

Definition 3.3. Closure of a Set
Let (X, τ) be a topological space, and let S be a subset of X. Then the closure of S is the intersection of all closed sets containing S and it is denoted by $\bar{S}$.

Definition 3.4. Support of a Function
The support of a function $f : \mathbb{R} \to \mathbb{C}$ is the closure of a set of all points at which f is not zero. We denote it by supp f that is,

$$\text{supp} f = \overline{\{x \in \mathbb{R} : f(x) \neq 0\}}. \tag{3.1}$$

We say that f has a compact support if there exists $r < \infty$, such that $\text{supp} f \subseteq [-r, r]$.

The theory of multiresolution analysis explained a systematic method to generate the orthogonal wavelets. The first basic concept of multiresolution analysis was introduced by Mallat. In this thesis, our purpose is to introduce the multiresolution analysis for the construction of Haar wavelet. To explain the theory of multiresolution analysis we need some notations.

Let us consider two functions φ and $\psi \in L^2(\mathbb{R})$, and for any $j, k \in \mathbb{Z}$, by translation and dilation, we define $\varphi_{j,k},\ \psi_{j,k} \in L^2(\mathbb{R})$ as

$$\varphi_{j,k}(x) = 2^{\frac{j}{2}} \varphi(2^j x - k),$$

and

$$\psi_{j,k}(x) = 2^{\frac{j}{2}} \psi(2^j x - k).$$

The L^2-norms of the functions $\varphi_{j,k}$ and $\psi_{j,k}$ are equal for all $j, k \in \mathbb{Z}$. Indeed, by substituting $y = (2^j x - k)$, we have

$$\begin{aligned}||\psi_{j,k}||^2 &= \int_{\mathbb{R}} |2^{j/2}\psi(2^j x - k)|^2 dx \\ &= \int_{\mathbb{R}} 2^j |\psi(2^j x - k)|^2 dx \\ &= \int_{R} |\psi(y)|^2 dy = ||\psi||^2.\end{aligned}$$

Similarly, $||\varphi_{j,k}||^2 = ||\varphi||^2$.

First, we discuss the graph of $\psi_{j,k}(x)$ by noticing the dilation involved in the definition. We can obtain the graph of $\psi(2^j x)$ by constructing the graph of $\psi(x)$. Suppose $\psi(x)$ has a compact support, supp $\psi(x) = [-r, r]$. It means that $\psi(x) = 0$ for all x such that $|x| > r$. Then $\psi(2^j x)$ has compact support,

$$\text{supp } \psi(2^j x) = [-r/2^j, r/2^j].$$

We can sketch the graph of $\psi(2^j x - k) = \psi(2^j(x - 2^{-j}k))$ from translating the graph of $\psi(2^j x)$ by $2^{-j}k$. So we can say that $\psi(2^j x - k)$ has compact support supp $\psi(2^j x - k) = [2^{-j}k - 2^{-j}r, 2^{-j}k + 2^{-j}r]$, that is, $\psi(2^j x - k) = 0$ for all x such that $|2^j x - k| > r$. And also by multiplying the factor $2^{\frac{j}{2}}$ to $\psi(2^j x - k)$, the graph of $\psi_{j,k}$ is obtained by stretching in the Y direction.

If the function φ, ψ are centered at the point 0 and constructed on a scale which is comparable to 1. This means that the mass of the function is mostly located around the origin of length about n, where the n is a small positive integer. So, the functions $\varphi_{j,k}$ and $\psi_{j,k}$ are centered near the point $2^{-j}k$ on a scale comparable to 2^{-j}.

3.1 Definition of Multiresolution Analysis

In this section, we will define multiresolution analysis and its properties. To design a multiresolution analysis we need a set of nested spaces, by selecting the functions $\varphi_{j,k}(x)$, for any $j, k \in \mathbb{Z}$, we determine the nested spaces V_j for any $j \in \mathbb{Z}$.

Definition 3.5. Multiresolution Analysis

A multiresolution analysis(MRA) with father wavelet φ consists of a sequence of closed subspaces $\{V_j\}_{j\in\mathbb{Z}}$ of $L^2(\mathbb{R})$ satisfying the following properties:

a. **Monotonicity**
The sequence is increasing. $V_j \subseteq V_{j+1}$ for all $j \in \mathbb{Z}$.

b. **Existence of the Scaling Function**
There exists a function $\varphi \in V_0$, such that the set $\{\varphi(\cdot - k) : k \in \mathbb{Z}\}$ is an orthonormal basis for V_0, that is,

$$V_0 = \{\sum_{k\in\mathbb{Z}} c(k)\varphi_{0,k} : c = (c(k))_{k\in\mathbb{Z}} \in l^2(\mathbb{Z})\}. \tag{3.2}$$

c. **Dilation Property**
$f(x) \in V_0$ if and only if $f(2^j x) \in V_j$, for each $j \in \mathbb{Z}$.

d. **Trivial Intersection Property**
$\bigcap_{j\in\mathbb{Z}} V_j = \{0\}$.

e. **Density**
$\overline{\bigcup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R})$.

Definition 3.6. Wavelets

A complete orthonormal set in $L^2(R)$ of the form $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$, is called a wavelet system, and the functions $\psi_{j,k}$ are called wavelets. The function $\psi \in L^2(\mathbb{R})$ is called the mother wavelet, and the function $\varphi \in L^2(\mathbb{R})$ is called the scaling function or father wavelet.

Definition 3.7. Wavelet Identity

If $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$, is a wavelet system, then any function $f \in L^2(\mathbb{R})$ (see Theorem 4.10 [4]) can be represented as a series

$$f = \sum_{j,k\in\mathbb{Z}} \langle f, \psi_{j,k}\rangle \psi_{j,k}, \tag{3.3}$$

which is called the wavelet identity. And the map taking f to the sequence of coefficients $\{\langle f, \psi_{j,k}\rangle\}_{j,k\in\mathbb{Z}}$ is called the wavelet transform.

So the wavelet identity breaks down f into its components at different scales 2^{-j}, centered at different locations $2^{-j}k$, for $j,k \in \mathbb{Z}$.

To understand the definition of multiresolution analysis, we consider an example.

Example 3.1. Let $I_{j,k}$ be a dyadic interval

$$I_{j,k} = [2^{-j}k, 2^{-j}(k+1)), \text{for any} j,k \in \mathbb{Z} \tag{3.4}$$

Let us consider

$$V_j = \{f \in L^2(\mathbb{R}) : f \text{ is constant on } I_{j,k}, \text{for all} j,k \in \mathbb{Z}\}. \tag{3.5}$$

We can show that $\{V_j\}_{j\in\mathbb{Z}} \subseteq L^2(\mathbb{R})$ satisfies the MRA properties.

a. First, we show that $\{V_j\}_{j\in\mathbb{Z}}$ has monotonicity. Let us consider $f \in V_j$, which means f is a constant on dyadic intervals of length 2^{-j}, so f is a constant on dyadic intervals of length 2^{-j-1}, hence $f \in V_{j+1}$. In other words, if $V_0 = \{f : f = \sum_{k\in\mathbb{Z}} c_k \varphi_{0,k}\}$ be the vector space of a constant functions over the interval $[k, k+1)$, then $V_1 = \{f : f = \sum_{k\in\mathbb{Z}} c_k \varphi_{1,k}\}$ will be the vector space of functions that are equal to constants over the intervals $[k, \frac{k+1}{2^j})$ and $[\frac{k+1}{2^j}, k+1)$. Generally, the space V_j and $j \in \mathbb{Z}$, includes all constant on the interval $[0,1)$ with 2^j subintervals.

   So every vector in V_j is contained in V_{j+1}, thus the spaces V_j are nested

   $$\cdots - V_1 \subset V_0 \subset V_1 \subset V_2 \subset \cdots \subset V_j \subset V_{j+1},$$

   and $\{V_j\}_{j\in\mathbb{Z}}$ is an increasing sequence.

b. To show the existence of the scaling function, consider

   $$\varphi(x) = \begin{cases} 1, & 0 \le x \le 1 \\ 0, & \text{otherwise.} \end{cases} \tag{3.6}$$

   Note that, $\{\varphi_{0,k}\}_{k\,\in\,\mathbb{Z}}$, where $\varphi_{0,k} = \varphi(\cdot - k)$ is an orthonormal system. Indeed

   $$\begin{aligned} \langle \varphi_{0,k}, \varphi_{0,k'}\rangle &= \int_{-\infty}^{+\infty} \varphi(x-k)\varphi(x-k')dx \\ &= \begin{cases} 1, & \text{if } k = k' \\ 0, & \text{if } k \neq k', \end{cases} \end{aligned}$$

   because the supports of different $\varphi_{0,k}$ do not overlap and

   $$||\varphi_{0,k}||^2 =< \varphi_{0,k}, \varphi_{0,k} >= ||\varphi||^2 = 1$$

Moreover, the orthonormality of the set $\{\varphi_{0,k}\}_{k\in\mathbb{Z}}$ implies the orthonormality of the set $\{\varphi_{j,k}\}_{k\in\mathbb{Z}}$, for each $j\in\mathbb{Z}$. It follows from the definition and by change of variables,

$$
\begin{aligned}
<\varphi_{j,k},\varphi_{j,k'}> &= <\varphi_{0,k},\varphi_{0,k'}> \\
&= \int_{-\infty}^{+\infty}\varphi(x-k)\varphi(x-k')dx \\
&= \begin{cases} 1, & \text{if } k=k' \\ 0, & \text{if } k\neq k'. \end{cases}
\end{aligned}
$$

In general, the set $\{\varphi_{j,k}\}_{k\in\mathbb{Z}}$ is a complete orthonormal system for the subspace $V_j,\ j\in\mathbb{Z}$, despite the union of these systems is not a complete orthonormal system for $L^2(\mathbb{R})$, because $\{\varphi_{j,k}\}_{j,k\in\mathbb{Z}}$ are not orthogonal at different levels for any $j\in\mathbb{Z}$.

c. The dilation property can be shown from the definitions of $V_j,\ j\in\mathbb{Z}$. By definition, $f\in V_j$ is equal to some constants on the interval $[2^{-j}k,2^{-j}(k+1)]$. Recall that if $f\in V_0$ then

$$f(x)=\{c_k,\ x\in(k,k+1)\}$$

Therefore

$$f(2^jx)=\{c_k,\ x\in(\frac{k}{2^j},\frac{k+1}{2^j})\},$$

hence, $f\in V_0$ if and only if $f(2^jx)\in V_j$.

d. To show that $\bigcap_{j\in\mathbb{Z}}V_j=\{0\}$, we consider $f\in\bigcap V_j$. We need to prove that $f\in\bigcap V_j$ is equal to zero. If $f\in\bigcap V_j$, then f is constant on the intervals $[0,2^{-j})$ and $[-2^{-j},0)$ for any $j\in\mathbb{Z}$. Let $j\to-\infty$, so f is constant on $[0,+\infty)$ and $[-\infty,0)$. On the other hand we know $f\in L^2(\mathbb{R})$, so

$$\int_{-\infty}^{+\infty}|f(t)|^2dt=c^2\int_{-\infty}^{+\infty}dt<+\infty.$$

Thus f must be 0 on $\mathbb{R}$.

e. To show the density property, we should prove that for any $f\in L^2(\mathbb{R})$, there exists a sequence $\{f_n\}_{n\in\mathbb{N}}$ such that each $f_n\in\bigcup_{j\in\mathbb{Z}}V_j$, and $||f_n-f||\to0$ as $n\to+\infty$.

The function $\varphi\in L^1(\mathbb{R})$, then the Fourier transformation,

$$\widehat{\varphi}(\omega)=\int_{\mathbb{R}}\varphi(t)e^{-\imath\,\omega\,t}dt$$

is bounded and continuous at 0, where

$$\widehat{\varphi}(0)=\int_{\mathbb{R}}\varphi(t)e^0dt=1.$$

By Plancherel's formula,

$$||\widehat{f}||^2=\frac{1}{(2\pi)}||f||^2.$$

$f\in L^2(\mathbb{R})$ implies that $\widehat{f}\in L^2(\mathbb{R})$. Consider

$$
\begin{aligned}
I &= \int_{-\infty}^{+\infty}|\widehat{f}(t)|^2dt \\
&= \lim_{R\to+\infty}\int_{-R}^{+R}|\widehat{f}(t)|^2dt\leq\infty,
\end{aligned}
$$

so for any $\varepsilon > 0$, there exist R such that

$$I - \int_{-R}^{+R} |\widehat{f}(t)|^2 dt < \varepsilon^2.$$

It means

$$\int_{-\infty}^{-R} |\widehat{f}(t)|^2 dt + \int_{+R}^{+\infty} |\widehat{f}(t)|^2 dt = \int_{|t|>R} |\widehat{f}(t)|^2 dt < \varepsilon^2.$$

Define $\widehat{h}$ by

$$\widehat{h}(t) = \begin{cases} \widehat{f}(t), & \text{if } |t| \leq R \\ 0, & \text{if } |t| > R. \end{cases} \tag{3.7}$$

We can write

$$||f - h|| = \frac{1}{\sqrt{2\pi}} ||\widehat{f} - \widehat{h}|| < \varepsilon. \tag{3.8}$$

Suppose $P_j : L^2(\mathbb{R}) \to V_j$ be the orthogonal projection of h on V_j. Then

$$P_j(h) = \sum_{k \in \mathbb{Z}} < h, \varphi_{j,k} > \varphi_{j,k},$$

since $h - P_j(h)$ is orthogonal to every element of V_j, we have

$$||h||^2 = ||h - P_j(h) + P_j(h)||^2 = ||h - P_j(h)||^2 + ||P_j(h)||^2.$$

We should show that $||P_j(h)|| \to ||h||$ as $j \to +\infty$. By Plancherel's formula

$$| < h, \varphi_{j,k} > |^2 = \frac{1}{(2\pi)^2} | < \widehat{h}, \widehat{\varphi}_{j,k} > |^2.$$

And also

$$\widehat{\varphi}_{j,k}(\omega) = \int_R 2^{\frac{j}{2}} \varphi(2^j t - k) e^{-i\omega t} dt, \tag{3.9}$$

by substituting $2^j t - k = T$ in (3.9), we get

$$\begin{aligned} \widehat{\varphi}_{j,k}(\omega) &= \int_R 2^{\frac{-j}{2}} \varphi(T) e^{-i\omega(\frac{T+k}{2^j})} dT, \\ &= 2^{\frac{-j}{2}} e^{-i\omega\frac{k}{2^j}} \int_R \varphi(T) e^{-i\omega(\frac{T}{2^j})} dT, \\ &= 2^{\frac{-j}{2}} e^{-i\omega\frac{k}{2^j}} \widehat{\varphi}\left(\frac{\omega}{2^j}\right). \end{aligned}$$

Hence

$$\begin{aligned} | < h, \varphi_{j,k} > |^2 &= \frac{1}{(2\pi)^2} 2^{-j} \left| \int_R \widehat{h}(t) \overline{\widehat{\varphi}(\frac{t}{2^j})} e^{\frac{ikt}{2^j}} dt \right|^2 \\ &= \frac{1}{(2\pi)^2} 2^j \left| \sum_{l \in \mathbb{Z}} \int_{(2l-1)\pi}^{(2l+1)\pi} \widehat{h}(2^j x) \overline{\widehat{\varphi}(x)} e^{ikx} dx \right|^2 \\ &= 2^j \left| \sum_{l \in \mathbb{Z}} \frac{1}{2\pi} \int_{-\pi}^{\pi} \widehat{h}(2^j(\theta + 2\pi l)) \overline{\widehat{\varphi}(\theta + 2\pi l)} e^{ik\theta} d\theta \right|^2, \end{aligned}$$

let us consider

$$F(\theta) = \sum_{l \in \mathbb{Z}} \widehat{h}(2^j(\theta + 2\pi l)) \overline{\widehat{\varphi}(\theta + 2\pi l)}, \tag{3.10}$$

on $[-\pi, \pi)$. We can write

$$|< h, \varphi_{j,k} >|^2 = \left| \frac{2^j}{2\pi} \int_{-\pi}^{\pi} F(\theta) e^{ik} \right|^2 \tag{3.11}$$

As we know

$$P_j(h) = \sum_{k \in \mathbb{Z}} < h, \varphi_{j,k} > \varphi_{j,k},$$

therefore, from (3.11), we get

$$||P_j(h)||^2 = \sum_{k \in \mathbb{Z}} |< h, \varphi_{j,k} >|^2 = \frac{2^j}{2\pi} \int_{-\pi}^{\pi} |F(\theta)|^2 d\theta. \tag{3.12}$$

We choose J such that $2^J > \frac{R}{\pi}$. If $j > J$, then at every point of θ in equation (3.10), the sum reduces to the single nonzero term. To show this we let, $\widehat{h}(2^j(\theta + 2\pi m)) \neq 0$ and $\widehat{h}(2^j(\theta + 2\pi n)) \neq 0$. By equation (3.7), $\widehat{h}(t) = 0$ for $|t| > R$, implies that $|\theta + 2\pi m| \leq \frac{R}{2^j}$ and $|\theta + 2\pi n| \leq \frac{R}{2^j}$. By triangle inequality, for $j > J$

$$|2\pi(m-n)| \leq |2\pi m + \theta| + |2\pi n + \theta| \leq \frac{R}{2^j} + \frac{R}{2^j} < 2\pi$$

which is possible if $m = n$. Therefore

$$|F(\theta)|^2 = \sum \left| \widehat{h}(2^j(\theta + 2\pi l)) \overline{\widehat{\varphi}(\theta + 2\pi l)} \right|^2, \tag{3.13}$$

by substituting $|F(\theta)|^2$ into the expression for (3.12), we have

$$\begin{aligned}
||P_j(h)||^2 &= \frac{2^j}{2\pi} \int_{-\pi}^{\pi} \pi \left| \sum_{l \in \mathbb{Z}} \widehat{h}(2^j(\theta + 2\pi l)) \widehat{\varphi}(\theta + 2\pi l) \right|^2 d\theta \\
&= \frac{2^j}{2\pi} \int_{\mathbb{R}} |\widehat{h}(2^j y)|^2 \, |\widehat{\varphi}(y)|^2 \, dy \\
&= \frac{1}{2\pi} \int_{\mathbb{R}} \left| \widehat{h}(t) \right|^2 \left| \widehat{\varphi}(\frac{t}{2^j}) \right|^2 dt.
\end{aligned}$$

Since $\widehat{\varphi}$ is bounded, so $|\widehat{\varphi}(t)| \leq c$, for all t. Therefore

$$||\widehat{h}(t)||^2 ||\widehat{\varphi}(\frac{t}{2^j})||^2 \leq c ||\widehat{h}(t)||^2.$$

Since $h \in L^2(\mathbb{R})$

$$\int_{\mathbb{R}} c |\widehat{h}(t)|^2 dt < +\infty.$$

We also consider that $\widehat{\varphi}$ is continuous at zero and $\widehat{\varphi}(0) = 1$. So for each t, $||\widehat{h}(t)||^2 ||\widehat{\varphi}(\frac{t}{2^j})||^2$ converges to $||\widehat{h}(t)||^2$ as $j \to +\infty$. Hence

$$||P_j(h)||^2 \leq \frac{1}{2\pi} \int_{\mathbb{R}} |\widehat{h}(t)|^2 dt = \frac{1}{2\pi} ||\widehat{h}||^2 = ||h||^2.$$

Also, by lemma (4.41 iii [4],) $P_j(f) = f$. Since we assumed that $f \in V_j$, therefore

$$\begin{aligned} ||h - P_j(h)|| &= ||h - f + f - P_j(h)|| \\ &= ||h - f + P_j(f) - P_j(h)|| \\ &= ||h - f|| + ||P_j(f - h)|| < \varepsilon + \varepsilon \end{aligned} \tag{3.14}$$

and

$$||f - P_j(h)|| \leq ||h - P_j(h)|| + ||f - h||.$$

By equations (3.8) and (3.14) we obtain

$$||f - P_j(h)|| < 2\varepsilon.$$

Since $P_j(h) \in V_j$, so $\cup_{j \in \mathbb{Z}} V_j$ is dense in $L^2(\mathbb{R})$.

4 Haar Wavelet Transform

In this chapter, our purpose is to construct the Haar Wavelet basis and use these basis to compress an image. The Haar system is an orthonormal system on the interval $[0,1]$. The Haar wavelet is a first known wavelet which has an orthonormal basis. It is a sequence of functions supported on the small subintervals of length $[0,1]$. The Haar basis functions are step functions with jump discontinuities. The Haar wavelet transform is used to compress one and two dimensional signals. The material in this chapter is taken from [3], [4], [6] and [9]. First, we discuss some related definitions to the wavelets.

Definition 4.1. Haar Wavelet
The Haar wavelet is constructed from the MRA, which is generated by the scaling function $\varphi = \chi_{[0,1)}(x)$ for $j,k \in \mathbb{Z}$

$$\varphi = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & otherwise, \end{cases} \tag{4.1}$$

we can also define a family of shifted and translated scaling function $\{\varphi_{j,k}(x)\}_{j,k\in\mathbb{Z}}$ by

$$\varphi_{j,k}(x) = 2^{j/2}\varphi(2^j x - k),$$

and it is shown in 4.1

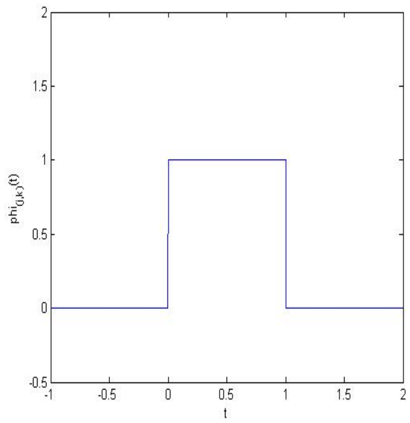


Figure 4.1: Scaling Function or Father Wavelet

it is clear that

$$\varphi(2^j x - k) = \begin{cases} 1 & k2^{-j} \leq x < (k+1)2^{-j} \\ 0 & otherwise. \end{cases}$$

This collection can be introduced as the system of Haar scaling functions.

Definition 4.2. Haar Function
Let $\psi(x) = \chi_{[0,1/2)}(x) - \chi_{[1/2,1)}(x)$ be the Haar function,

$$\psi(x) = \begin{cases} 1 & 0 \leq x \leq 1/2 \\ -1 & 1/2 \leq x \leq 1/2 \\ 0 & otherwise. \end{cases} \tag{4.2}$$

For each $j,k \in \mathbb{Z}$ by translation and dilation we can define

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k).$$

The collection $\{\psi_{j,k}(x)\}_{j,k\in\mathbb{Z}}$ is introduced as the Haar system on $\mathbb{R}$. The Haar scaling function can be shown in the 4.2

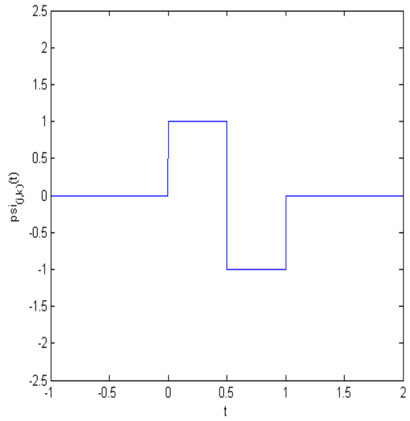


Figure 4.2: Haar Scaling Function

Properties of Haar Wavelet

a. Functions $\varphi_{j,k}(x)$ and $\psi_{j,k}(x)$ are supported on the dyadic interval $I_{j,k} = [k2^{-j}, (k+1)2^{-j})$ since for each $j,k \in \mathbb{Z}$

$$\varphi_{j,k}(x) = 2^{j/2}\chi_{I_{j,k}}(x),$$

and

$$\psi_{j,k}(t) = 2^{j/2}(\chi_{I_{j+1,2k}}(x) - \chi_{I_{j+1,2k+1}}(x)).$$

It means that they are not vanish on $I_{j,k}$.

b. The Haar system is an orthonormal system on $\mathbb{R}$.

Proof. Let $j \in \mathbb{Z}$ be fixed, for any $k,k' \in \mathbb{Z}$

$$I_{j,k} \cap I_{j,k'} = \begin{cases} \emptyset & k \neq k' \\ I_{j,k} & k = k'. \end{cases}$$

If $k \neq k'$, then $\psi_{j,k}(t)$ and $\psi_{j,k'}(t)$ are supported on disjoint intervals. So

$$< \psi_{j,k}(x), \psi_{j,k'}(x) >= \int_{\mathbb{R}} \psi_{j,k}(x)\psi_{j,k'}(x) = 0.$$

If $k = k'$, then

$$< \psi_{j,k}(x), \psi_{j,k'}(x) >= \int_{I_{j,k}} \psi_{j,k}(x)\psi_{j,k}(x) = \int_{I_{j,k}} |\psi_{j,k}(x)|^2 = 1,$$

so the Haar system is an orthonormal system.

Now to show the orthonormality between scales, suppose $k,k',j,j' \in \mathbb{Z}$, with $j \neq j'$. So there are following possibilities:

1. $I_{j,k} \cap I_{j,k'} = \phi$. It is obvious that the product $\psi_{j,k}(x)\psi_{j',k'}(x)$ for all x is zero. So

$$< \psi_{j,k}(x), \psi_{j',k'}(x) >= \int_{\mathbb{R}} \psi_{j,k}(x)\psi_{j',k'}(x) = 0.$$

2. Consider $j' > j$, and the intervals $I_{j,k}$ and $I_{j',k'}$ are not disjoint, then $I_{j,k} \supseteq I_{j',k'}$. So $I_{j',k'}$ contains the first or second half of $I_{j,k}$. Hence,

$$< \psi_{j,k}(x), \psi_{j',k'}(x) >= \int_{I_{j,k}} \psi_{j,k}(x)\psi_{j',k'}(x)dx = \int_{I_{j,k}} \psi_{j,k}(x)dx = 0.$$

□

c. The system $\{\psi_{k,j} | j,k \in \mathbb{Z}\}$ is complete in $L^2(\mathbb{R})$.

Proof. To show the completeness of this system, we need to show that span$\{\psi_{k,j} : j,k \in \mathbb{Z}\}$ is dense in $L^2(\mathbb{R})$. Let us consider the vector space V_j i.e

$$V_j = span\{\chi_{I_{j,k}} : k \in \mathbb{Z}\}.$$

It is obvious that simple functions are dense in $L^2(\mathbb{R})$. Hence, the set of all finite linear combinations of characteristic functions of intervals are also dense in $L^2(\mathbb{R})$. And also we know that for every interval I, the function χ_I can be approximated by functions in $\bigcup_{j\in\mathbb{Z}} V_j$. So

$$\overline{\text{span}\{V_j : j \in \mathbb{Z}\}} = L^2(\mathbb{R})$$

As we mentioned, the sequence $\{V_j\}_{j\in\mathbb{Z}}$ is increasing.i.e for every $j \in \mathbb{Z}$, $V_j \subseteq V_{j+1}$. This property holds for the Haar scaling function.
Moreover, $\bigcap_{j\in\mathbb{Z}} V_j = 0$. To show this, let us consider a function $f \in \bigcap_{j\in\mathbb{Z}} V_j$. It is clear that f is a constant on an arbitrarily intervals, and $f \in L^2(\mathbb{R})$. Therefor $f = 0$.
In the next step of proof, we should define the vector space W_j

$$W_j = span\{\psi_{j,k} : k \in \mathbb{Z}\}.$$

We have shown above that the system $\{\psi_{j,k} : k \in \mathbb{Z}\}$ is an orthonormal system in $L^2(R)$. So $W_j \perp W_j'$ for $j \neq j'$. Then it is enough to show that

$$\oplus_{j\in\mathbb{Z}} W_j = L^2(\mathbb{R}).$$

By definition,

$$\begin{aligned}
\varphi_{0,0}(x) &= \chi_{[0,1)}(x) \\
&= \frac{1}{2}\chi_{[0,2)}(x) + \frac{1}{2}(\chi_{[0,1)}(x) - \chi_{[1,2)}(x)) \\
&= \frac{1}{\sqrt{2}}\varphi_{-1,0}(x) + \frac{1}{\sqrt{2}}\psi_{-1,0}(x) \in V_{-1} + W_{-1}.
\end{aligned}$$

By translation we can see

$$\varphi_{0,k}(x)\frac{1}{\sqrt{2}}\varphi_{-1,k}(x) + \frac{1}{\sqrt{2}}\psi_{-1,k}(x) \in V_{-1} + W_{-1}.$$

So $V_0 \subseteq V_{-1} + W_{-1}$. And also by the above relation $V_{-1} \subseteq V_0$, and $W_{-1} \subseteq V_0$, which means $V_{-1} + W_{-1} \subseteq V_0$.
Hence
$$V_0 = V_{-1} + W_{-1}.$$
As we know $V_{-1} \perp W_{-1}$. Therefore the above sum is a direct sum.
$$V_0 = V_{-1} \oplus W_{-1}.$$
So by repeating these arguments at the level N, we obtain the orthogonal decomposition
$$V_N = V_{N-1} \oplus W_{N-1},$$
which is the very important property of the Haar basis. By induction
$$V_N = V_{N-1} \oplus W_{N-1} = (V_{N-2} \oplus W_{N-2}) \oplus W_{N-1} = V_M \oplus (\oplus_{N-1}^{j=M} W_j),$$
for any $M \leq N-1$. So by above relation , if $M \to -\infty$, then $V_M \to 0$. Hence
$$V_N = \oplus_{N-1}^{j=-\infty} W_j.$$
Let $N \to +\infty$, we have
$$L^2(\mathbb{R}) = \oplus_{j \in \mathbb{Z}} W_j.$$
Hence, the property is done. □

4.1 Wavelet Transformation of a Signal

Let us consider a signal f. For simplicity we will consider $f \in \mathbb{R}^8$ (We can expand the procedure to any finite dimensions,) so
$$f = (c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8).$$
By the definition (3.5) we can represent f as
$$f(x) = \sum_{j,k \in \mathbb{Z}} < f, \psi_{j,k} > \psi_{j,k},$$
where $< f, \psi_{j,k} >$ are called the Haar wavelet coefficients.

So according to the definition (4.1) and (4.2)
$$f = f_1 \varphi_{0,0} + f_2 \psi_{0,0} + f_3 \psi_{1,0} + f_4 \psi_{1,1} + f_5 \psi_{2,0} + f_6 \psi_{2,1} + f_7 \psi_{2,2} + f_8 \psi_{2,3},$$
where
$$\begin{aligned} f_1 &=< f, \varphi_{0,0} >= \int_0^1 f(x) \varphi_{0,0}(x) dx \\ &= \int_0^{\frac{1}{8}} c_1 \varphi_{0,0}(x) dx + \int_{\frac{1}{8}}^{\frac{2}{8}} c_2 \varphi_{0,0}(x) dx + \int_{\frac{2}{8}}^{\frac{3}{8}} c_3 \varphi_{0,0}(x) dx \\ &+ \int_{\frac{3}{8}}^{\frac{4}{8}} c_4 \varphi_{0,0}(x) dx + \int_{\frac{4}{8}}^{\frac{5}{8}} c_5 \varphi_{0,0}(x) dx + \int_{\frac{5}{8}}^{\frac{6}{8}} c_6 \varphi_{0,0}(x) dx + \int_{\frac{6}{8}}^{\frac{7}{8}} c_7 \varphi_{0,0}(x) dx + \int_{\frac{7}{8}}^{1} c_8 \varphi_{0,0}(x) dx \\ &= \frac{c_1 + c_2 + c_3 + c_4 + c_5 + c_6 + c_7 + c_8}{8}. \end{aligned}$$

$$
\begin{aligned}
f_2 &= < f, \psi_{0,0} >= \int_0^1 f(x)\psi_{0,0}(x)dx \\
&= \int_0^{\frac{1}{8}} c_1\psi_{0,0}(x)dx + \int_{\frac{1}{8}}^{\frac{2}{8}} c_2\psi_{0,0}(x)dx + \int_{\frac{2}{8}}^{\frac{3}{8}} c_3\psi_{0,0}(x)dx \\
&+ \int_{\frac{3}{8}}^{\frac{4}{8}} c_4\psi_{0,0}(x)dx + \int_{\frac{4}{8}}^{\frac{5}{8}} c_5\psi_{0,0}(x)dx + \int_{\frac{5}{8}}^{\frac{6}{8}} c_6\psi_{0,0}(x)dx + \int_{\frac{6}{8}}^{\frac{7}{8}} c_7\psi_{0,0}(x)dx + \int_{\frac{7}{8}}^{1} c_8\psi_{0,0}(x)dx \\
&= \frac{c_1 + c_2 + c_3 + c_4 - c_5 - c_6 - c_7 - c_8}{8}.
\end{aligned}
$$

$$
\begin{aligned}
f_3 &= < f, \psi_{1,0} >= \int_0^1 f(x)\psi_{1,0}(x)dx \\
&= \int_0^{\frac{1}{8}} c_1\psi_{1,0}(x)dx + \int_{\frac{1}{8}}^{\frac{2}{8}} c_2\psi_{1,0}(x)dx + \int_{\frac{2}{8}}^{\frac{3}{8}} c_3\psi_{1,0}(x)dx \\
&+ \int_{\frac{3}{8}}^{\frac{4}{8}} c_4\psi_{1,0}(x)dx + \int_{\frac{4}{8}}^{\frac{5}{8}} c_5\psi_{1,0}(x)dx + \int_{\frac{5}{8}}^{\frac{6}{8}} c_6\psi_{1,0}(x)dx + \int_{\frac{6}{8}}^{\frac{7}{8}} c_7\psi_{1,0}(x)dx + \int_{\frac{7}{8}}^{1} c_8\psi_{1,0}(x)dx \\
&= \frac{c_1 + c_2 - c_3 - c_4}{8}.
\end{aligned}
$$

$$
\begin{aligned}
f_4 &= < f, \psi_{1,1} >= \int_0^1 f(x)\psi_{1,1}(x)dx \\
&= \int_0^{\frac{1}{8}} c_1\psi_{1,1}(x)dx + \int_{\frac{1}{8}}^{\frac{2}{8}} c_2\psi_{1,1}(x)dx + \int_{\frac{2}{8}}^{\frac{3}{8}} c_3\psi_{1,1}(x)dx \\
&+ \int_{\frac{3}{8}}^{\frac{4}{8}} c_4\psi_{1,1}(x)dx + \int_{\frac{4}{8}}^{\frac{5}{8}} c_5\psi_{1,1}(x)dx + \int_{\frac{5}{8}}^{\frac{6}{8}} c_6\psi_{1,1}(x)dx + \int_{\frac{6}{8}}^{\frac{7}{8}} c_7\psi_{1,1}(x)dx + \int_{\frac{7}{8}}^{1} c_8\psi_{1,1}(x)dx \\
&= \frac{c_5 + c_6 - c_7 - c_8}{8}.
\end{aligned}
$$

$$
\begin{aligned}
f_5 &= < f, \psi_{2,0} > \\
&= \int_0^1 f(x)\psi_{2,0}(x)dx \\
&= \int_0^{\frac{1}{8}} c_1\psi_{2,0}(x)dx + \int_{\frac{1}{8}}^{\frac{2}{8}} c_2\psi_{2,0}(x)dx + \int_{\frac{2}{8}}^{\frac{3}{8}} c_3\psi_{2,0}(x)dx \\
&+ \int_{\frac{3}{8}}^{\frac{4}{8}} c_4\psi_{2,0}(x)dx + \int_{\frac{4}{8}}^{\frac{5}{8}} c_5\psi_{2,0}(x)dx + \int_{\frac{5}{8}}^{\frac{6}{8}} c_6\psi_{2,0}(x)dx + \int_{\frac{6}{8}}^{\frac{7}{8}} c_7\psi_{2,0}(x)dx + \int_{\frac{7}{8}}^{1} c_8\psi_{2,0}(x)dx \\
&= \frac{c_1 - c_2}{8}.
\end{aligned}
$$

$$
\begin{aligned}
f_6 &= < f, \psi_{2,1} >= \int_0^1 f(x)\psi_{2,1}(x)dx \\
&= \int_0^{\frac{1}{8}} c_1\psi_{2,1}(x)dx + \int_{\frac{1}{8}}^{\frac{2}{8}} c_2\psi_{2,1}(x)dx + \int_{\frac{2}{8}}^{\frac{3}{8}} c_3\psi_{2,1}(x)dx \\
&+ \int_{\frac{3}{8}}^{\frac{4}{8}} c_4\psi_{2,1}(x)dx + \int_{\frac{4}{8}}^{\frac{5}{8}} c_5\psi_{2,1}(x)dx + \int_{\frac{5}{8}}^{\frac{6}{8}} c_6\psi_{2,1}(x)dx + \int_{\frac{6}{8}}^{\frac{7}{8}} c_7\psi_{2,1}(x)dx + \int_{\frac{7}{8}}^{1} c_8\psi_{2,1}(x)dx \\
&= \frac{c_3 - c_4}{8}.
\end{aligned}
$$

$$
\begin{aligned}
f_7 &= < f, \psi_{2,2} > = \int_0^1 f(x)\psi_{2,2}(x)dx \\
&= \int_0^{\frac{1}{8}} c_1\psi_{2,2}(x)dx + \int_{\frac{1}{8}}^{\frac{2}{8}} c_2\psi_{2,2}(x)dx + \int_{\frac{2}{8}}^{\frac{3}{8}} c_3\psi_{2,2}(x)dx \\
&+ \int_{\frac{3}{8}}^{\frac{4}{8}} c_4\psi_{2,2}(x)dx + \int_{\frac{4}{8}}^{\frac{5}{8}} c_5\psi_{2,2}(x)dx + \int_{\frac{5}{8}}^{\frac{6}{8}} c_6\psi_{2,2}(x)dx + \int_{\frac{6}{8}}^{\frac{7}{8}} c_7\psi_{2,2}(x)dx + \int_{\frac{7}{8}}^{1} c_8\psi_{2,2}(x)dx \\
&= \frac{c_5 - c_6}{8}.
\end{aligned}
$$

$$
\begin{aligned}
f_8 &= < f, \psi_{2,3} > \\
&= \int_0^1 f(x)\psi_{2,3}(x)dx \\
&= \int_0^{\frac{1}{8}} c_1\psi_{2,3}(x)dx + \int_{\frac{1}{8}}^{\frac{2}{8}} c_2\psi_{2,3}(x)dx + \int_{\frac{2}{8}}^{\frac{3}{8}} c_3\psi_{2,3}(x)dx \\
&+ \int_{\frac{3}{8}}^{\frac{4}{8}} c_4\psi_{2,3}(x)dx + \int_{\frac{4}{8}}^{\frac{5}{8}} c_5\psi_{2,3}(x)dx + \int_{\frac{5}{8}}^{\frac{6}{8}} c_6\psi_{2,3}(x)dx + \int_{\frac{6}{8}}^{\frac{7}{8}} c_7\psi_{2,3}(x)dx + \int_{\frac{7}{8}}^{1} c_8\psi_{2,3}(x)dx \\
&= \frac{c_7 - c_8}{8}.
\end{aligned}
$$

Hence,

$$
\begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \\ f_5 \\ f_6 \\ f_7 \\ f_8 \end{bmatrix} = \frac{1}{8}
\begin{bmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\
1 & 1 & -1 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & -1 & -1 \\
1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & -1
\end{bmatrix}
\begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \\ c_6 \\ c_7 \\ c_8 \end{bmatrix} = A
\begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \\ c_6 \\ c_7 \\ c_8 \end{bmatrix},
$$

where

$$
A_H = \frac{1}{8}
\begin{bmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\
1 & 1 & -1 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & -1 & -1 \\
1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & -1
\end{bmatrix},
$$

the matrix A_H is called the Haar transformation matrix. Note that the first row corresponds to the basis vector $\varphi_{0,0}$, the second row corresponds to the basis vector $\psi_{0,0}$, the third row corresponds to the basis vector $\psi_{1,0}$, and so on.
The matrix A_H is the multiplication of the following three matrices A_{H_1}, A_{H_2} and A_{H_3}. The Haar wavelet coefficients of these matrices can be obtained in the similar way by using the Haar scaling functions in appropriate dyadic interval.

$$A_{H_1} = \frac{1}{2}\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix},$$

$$A_{H_2} = \frac{1}{2}\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$$A_{H_3} = \frac{1}{2}\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

4.2 Haar Normalized Basis

As we mentioned, the constant factor $2^{\frac{j}{2}}$ is chosen to normalize a Haar basis function. Thus by multiplying each non normalized coefficient with $2^{\frac{j}{2}}$, we obtain the normalized coefficients.

Therefore the normalized Haar transform matrix is obtained by the following calculates

$$f = f_1\varphi_{0,0} + f_2\psi_{0,0} + f_3\psi_{1,0} + f_4\psi_{1,1} + f_5\psi_{2,0} + f_6\psi_{2,1} + f_7\psi_{2,2} + f_8\psi_{2,3},$$

where

$$\begin{aligned} f_1 = <f, \varphi_{0,0}> = \int_0^1 f(x)\varphi_{0,0}(x)dx = \int_0^{\frac{1}{8}} c_1\varphi_{0,0}(x)dx + \int_{\frac{1}{8}}^{\frac{2}{8}} c_2\varphi_{0,0}(x)dx + \int_{\frac{2}{8}}^{\frac{3}{8}} c_3\varphi_{0,0}(x)dx \\ + \int_{\frac{3}{8}}^{\frac{4}{8}} c_4\varphi_{0,0}(x)dx + \int_{\frac{4}{8}}^{\frac{5}{8}} c_5\varphi_{0,0}(x)dx + \int_{\frac{5}{8}}^{\frac{6}{8}} c_6\varphi_{0,0}(x)dx + \int_{\frac{6}{8}}^{\frac{7}{8}} c_7\varphi_{0,0}(x)dx + \int_{\frac{7}{8}}^{1} c_8\varphi_{0,0}(x)dx \\ = \frac{c_1 + c_2 + c_3 + c_4 + c_5 + c_6 + c_7 + c_8}{8}. \end{aligned}$$

$$\begin{aligned} f_2 = <f, \psi_{0,0}> = \int_0^1 f(x)\psi_{0,0}(x)dx = \int_0^{\frac{1}{8}} c_1\psi_{0,0}(x)dx + \int_{\frac{1}{8}}^{\frac{2}{8}} c_2\psi_{0,0}(x)dx + \int_{\frac{2}{8}}^{\frac{3}{8}} c_3\psi_{0,0}(x)dx \\ + \int_{\frac{3}{8}}^{\frac{4}{8}} c_4\psi_{0,0}(x)dx + \int_{\frac{4}{8}}^{\frac{5}{8}} c_5\psi_{0,0}(x)dx + \int_{\frac{5}{8}}^{\frac{6}{8}} c_6\psi_{0,0}(x)dx + \int_{\frac{6}{8}}^{\frac{7}{8}} c_7\psi_{0,0}(x)dx + \int_{\frac{7}{8}}^{1} c_8\psi_{0,0}(x)dx \\ = \frac{c_1 + c_2 + c_3 + c_4 - c_5 - c_6 - c_7 - c_8}{8}. \end{aligned}$$

$$
\begin{aligned}
f_3 =< f, \psi_{1,0} >= \int_0^1 f(x)\psi_{1,0}(x)dx = \int_0^{\frac{1}{8}} c_1\psi_{1,0}(x)dx + \int_{\frac{1}{8}}^{\frac{2}{8}} c_2\psi_{1,0}(x)dx + \int_{\frac{2}{8}}^{\frac{3}{8}} c_3\psi_{1,0}(x)dx \\
+ \int_{\frac{3}{8}}^{\frac{4}{8}} c_4\psi_{1,0}(x)dx + \int_{\frac{4}{8}}^{\frac{5}{8}} c_5\psi_{1,0}(x)dx + \int_{\frac{5}{8}}^{\frac{6}{8}} c_6\psi_{1,0}(x)dx + \int_{\frac{6}{8}}^{\frac{7}{8}} c_7\psi_{1,0}(x)dx + \int_{\frac{7}{8}}^{1} c_8\psi_{1,0}(x)dx \\
= \frac{\sqrt{2}c_1 + \sqrt{2}c_2 - \sqrt{2}c_3 - \sqrt{2}c_4}{8}.
\end{aligned}
$$

$$
\begin{aligned}
f_4 =< f, \psi_{1,1} >= \int_0^1 f(x)\psi_{1,1}(x)dx = \int_0^{\frac{1}{8}} c_1\psi_{1,1}(x)dx + \int_{\frac{1}{8}}^{\frac{2}{8}} c_2\psi_{1,1}(x)dx + \int_{\frac{2}{8}}^{\frac{3}{8}} c_3\psi_{1,1}(x)dx \\
+ \int_{\frac{3}{8}}^{\frac{4}{8}} c_4\psi_{1,1}(x)dx + \int_{\frac{4}{8}}^{\frac{5}{8}} c_5\psi_{1,1}(x)dx + \int_{\frac{5}{8}}^{\frac{6}{8}} c_6\psi_{1,1}(x)dx + \int_{\frac{6}{8}}^{\frac{7}{8}} c_7\psi_{1,1}(x)dx + \int_{\frac{7}{8}}^{1} c_8\psi_{1,1}(x)dx \\
= \frac{\sqrt{2}c_5 + \sqrt{2}c_6 - \sqrt{2}c_7 - \sqrt{2}c_8}{8}.
\end{aligned}
$$

$$
\begin{aligned}
f_5 =< f, \psi_{2,0} >= \int_0^1 f(x)\psi_{2,0}(x)dx = \int_0^{\frac{1}{8}} c_1\psi_{2,0}(x)dx + \int_{\frac{1}{8}}^{\frac{2}{8}} c_2\psi_{2,0}(x)dx + \int_{\frac{2}{8}}^{\frac{3}{8}} c_3\psi_{2,0}(x)dx \\
+ \int_{\frac{3}{8}}^{\frac{4}{8}} c_4\psi_{2,0}(x)dx + \int_{\frac{4}{8}}^{\frac{5}{8}} c_5\psi_{2,0}(x)dx + \int_{\frac{5}{8}}^{\frac{6}{8}} c_6\psi_{2,0}(x)dx + \int_{\frac{6}{8}}^{\frac{7}{8}} c_7\psi_{2,0}(x)dx + \int_{\frac{7}{8}}^{1} c_8\psi_{2,0}(x)dx \\
= \frac{2c_1 - 2c_2}{8}.
\end{aligned}
$$

$$
\begin{aligned}
f_6 =< f, \psi_{2,1} >= \int_0^1 f(x)\psi_{2,1}(x)dx = \int_0^{\frac{1}{8}} c_1\psi_{2,1}(x)dx + \int_{\frac{1}{8}}^{\frac{2}{8}} c_2\psi_{2,1}(x)dx + \int_{\frac{2}{8}}^{\frac{3}{8}} c_3\psi_{2,1}(x)dx \\
+ \int_{\frac{3}{8}}^{\frac{4}{8}} c_4\psi_{2,1}(x)dx + \int_{\frac{4}{8}}^{\frac{5}{8}} c_5\psi_{2,1}(x)dx + \int_{\frac{5}{8}}^{\frac{6}{8}} c_6\psi_{2,1}(x)dx + \int_{\frac{6}{8}}^{\frac{7}{8}} c_7\psi_{2,1}(x)dx + \int_{\frac{7}{8}}^{1} c_8\psi_{2,1}(x)dx \\
= \frac{2c_3 - 2c_4}{8}.
\end{aligned}
$$

$$
\begin{aligned}
f_7 =< f, \psi_{2,2} >= \int_0^1 f(x)\psi_{2,2}(x)dx = \int_0^{\frac{1}{8}} c_1\psi_{2,2}(x)dx + \int_{\frac{1}{8}}^{\frac{2}{8}} c_2\psi_{2,2}(x)dx + \int_{\frac{2}{8}}^{\frac{3}{8}} c_3\psi_{2,2}(x)dx \\
+ \int_{\frac{3}{8}}^{\frac{4}{8}} c_4\psi_{2,2}(x)dx + \int_{\frac{4}{8}}^{\frac{5}{8}} c_5\psi_{2,2}(x)dx + \int_{\frac{5}{8}}^{\frac{6}{8}} c_6\psi_{2,2}(x)dx + \int_{\frac{6}{8}}^{\frac{7}{8}} c_7\psi_{2,2}(x)dx + \int_{\frac{7}{8}}^{1} c_8\psi_{2,2}(x)dx \\
= \frac{2c_5 - 2c_6}{8}.
\end{aligned}
$$

$$
\begin{aligned}
f_8 =< f, \psi_{2,3} >= \int_0^1 f(x)\psi_{2,3}(x)dx = \int_0^{\frac{1}{8}} c_1\psi_{2,3}(x)dx + \int_{\frac{1}{8}}^{\frac{2}{8}} c_2\psi_{2,3}(x)dx + \int_{\frac{2}{8}}^{\frac{3}{8}} c_3\psi_{2,3}(x)dx \\
+ \int_{\frac{3}{8}}^{\frac{4}{8}} c_4\psi_{2,3}(x)dx + \int_{\frac{4}{8}}^{\frac{5}{8}} c_5\psi_{2,3}(x)dx + \int_{\frac{5}{8}}^{\frac{6}{8}} c_6\psi_{2,3}(x)dx + \int_{\frac{6}{8}}^{\frac{7}{8}} c_7\psi_{2,3}(x)dx + \int_{\frac{7}{8}}^{1} c_8\psi_{2,3}(x)dx \\
= \frac{2c_7 - 2c_8}{8}.
\end{aligned}
$$

So

$$A_H = \frac{1}{8}\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ \sqrt{2} & \sqrt{2} & -\sqrt{2} & -\sqrt{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sqrt{2} & \sqrt{2} & -\sqrt{2} & -\sqrt{2} \\ 2 & -2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & -2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & -2 \end{bmatrix}.$$

4.3 Two-Dimensional Haar Wavelet Transform

The two-dimensional Haar transform can be obtain easily from the one-dimensional. Just we should consider matrix rather than a vector. There are two types of two-dimensional wavelet transform.

The first type is standard decomposition of an image, in which we apply the one dimensional Haar transform to each row of pixel values. This operation obtains an average value with detail coefficients. Then we apply the one dimensional Haar transform to each column. So when we consider one-dimensional Haar basis, we easily get the two-dimensional basis.

The second type is nonstandard decomposition. In this way, first we apply horizontal pairwise averaging and differencing to each row of the image, then we apply vertical pairwise averaging and differencing to each column of the obtained image.

5 Averaging and Differencing Technique

In the previous chapter, the Haar wavelet basis are used to get Haar transformation matrix A_H. The matrix A_H is the multiplication of A_{H_1}, A_{H_2} and A_{H_3}. For instance, for two dimensional Haar transformation, the matrix A_{H_1} will first average and difference the columns of an image matrix and then it will average and diffidence the rows of column transformed matrix. In this chapter, another method, called Averaging and differencing will be used to transform the image matrix. This method is an application of Haar wavelet. We will discuss this method in detail to make it understand, and to know how the Haar wavelet works to compress an image data. The material in this chapter is taken mainly from [8] and [1].

5.1 Transformation of The Image Matrix

We can view the elements or pixels of an image in matrix form by using the MATLAB and other mathematical programs. Generally, we have two types of images to store in the computer, the gray scale image and the colored image. In the computer, each image is stored in the matrix form. For instance, if we have an image of the dimension 512×512, then this image will be stored as a matrix of the dimension 512×512. Each of the gray and the colored image has different number of elements to store in the system. For example if we have a gray scale image of 256×256 pixels, then this image has 65,536 elements to store. On the other hand a 640×480 pixel color image has nearly a million elements to store. To store the large images in the system we need large storage space. We can overcome this problem with the help of the Haar wavelet. By using the Haar wavelet, we can compress an image that take less storage space. We suppose that our image has $2^j \times 2^j$ pixels for $j = 1, 2, \cdots, n$. Now, if we choose $j = 9$ then the image will consist of $2^9 \times 2^9 = 512 \times 512$ elements. This shows that the image matrix will be of the order 512×512. As we mentioned before, first the Haar wavelet transforms the columns of an image matrix and then it transforms the rows of the column-transformed matrix. This transformed matrix is called row -column transformed matrix and it is used for the image compression. In this sections, we will use averaging and differencing method to transform the image matrix.
The transformation of an image matrix is done in several steps, if we have an image of 2^j resolution then we will have j steps to transform the matrix. For instance, if we choose $j = 3$, then we will have three steps to get the transformed matrix.

5.2 Averaging and Differencing Technique with a Row vector

Let us consider an 8×8 image matrix . The image matrix $I = [c_{ij}]$ for $i = 1, 2, 3, \cdots, 8$ and $j = 1, 2, \cdots, 8$ is given as

$$I = \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} & c_{15} & c_{16} & c_{17} & c_{18} \\ c_{21} & c_{22} & c_{23} & c_{24} & c_{25} & c_{26} & c_{27} & c_{28} \\ c_{31} & c_{32} & c_{33} & c_{34} & c_{35} & c_{36} & c_{37} & c_{38} \\ c_{41} & c_{42} & c_{43} & c_{44} & c_{45} & c_{46} & c_{47} & c_{48} \\ c_{51} & c_{52} & c_{53} & c_{54} & c_{55} & c_{56} & c_{57} & c_{58} \\ c_{61} & c_{62} & c_{63} & c_{64} & c_{65} & c_{66} & c_{67} & c_{68} \\ c_{71} & c_{72} & c_{73} & c_{74} & c_{75} & c_{76} & c_{77} & c_{78} \\ c_{81} & c_{82} & c_{83} & c_{84} & c_{85} & c_{86} & c_{87} & c_{88} \end{bmatrix}.$$

For convenience, we will describe the averaging and differencing technique by considering a row vector from the matrix I. Now, we choose the first row vector from the matrix

I, which is given as

$$f = \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} & c_{15} & c_{16} & c_{17} & c_{18}. \end{bmatrix}$$

In this row vector we have $8 = 2^3$ elements, which shows that $j = 3$ and we have three steps to get the transformed matrix. These three steps are given below.

Step 1. We transform f into f_1.

$$f_1 = \begin{matrix} [\frac{c_{11}+c_{12}}{2} & \frac{c_{13}+c_{14}}{2} & \frac{c_{15}+c_{16}}{2} \; \frac{c_{17}+c_{18}}{2} \; c_{11}-\frac{c_{11}+c_{12}}{2} \; c_{13}-\frac{c_{13}+c_{14}}{2} \\ c_{15}-\frac{c_{15}+c_{16}}{2} \; c_{17}-\frac{c_{17}+c_{18}}{2}]. \end{matrix}$$

First, we will take the pair of elements from f and take the average of each pair. As a result we will get four elements and these elements will be the corresponding first four elements of f_1, as it is mentioned above in f_1. These four numbers of f_1 are called approximation coefficients. After that we subtract these approximation coefficient from the first element of each pair of f. By doing this we will get four more elements and these elements will be the last four elements of f_1. These last four elements of f_1 are called detailed or spars coefficients. Finally, the f_1 is given as below,

$$f_1 = [\tfrac{c_{11}+c_{12}}{2} \; \tfrac{c_{13}+c_{14}}{2} \; \tfrac{c_{15}+c_{16}}{2} \; \tfrac{c_{17}+c_{18}}{2} \; \tfrac{c_{11}-c_{12}}{2} \; \tfrac{c_{13}-c_{14}}{2} \; \tfrac{c_{15}-c_{16}}{2} \; \tfrac{c_{17}-c_{18}}{2}].$$

The approximation coefficients store all the important information about the image and the detailed coefficients stores detailed information.

Step 2. In this step we will apply the averaging and differencing operation to f_1 as we performed to f to get f_1. But, here we will only apply this operation on the approximation coefficients of f_1 while the detailed coefficients will remain the same. So, by doing this we can transform f_1 into f_2 which is given as,

$$f_2 = \begin{matrix} [\frac{\frac{c_{11}+c_{12}}{2}+\frac{c_{13}+c_{14}}{2}}{2} & \frac{\frac{c_{15}+c_{16}}{2}+\frac{c_{17}+c_{18}}{2}}{2} & \frac{c_{11}+c_{12}}{2}-\frac{\frac{c_{11}+c_{12}}{2}+\frac{c_{13}+c_{14}}{2}}{2} & \frac{c_{15}+c_{16}}{2}-\frac{\frac{c_{15}+c_{16}}{2}+\frac{c_{17}+c_{18}}{2}}{2} \\ \frac{c_{11}-c_{12}}{2} & \frac{c_{13}-c_{14}}{2} & \frac{c_{15}-c_{16}}{2} & \frac{c_{17}-c_{18}}{2}]. \end{matrix}$$

Step 3. In the final step, we will again keep all the detailed coefficients, and we will apply the same steps to f_2 to get the transformed vector f_3. Again we will consider the approximation coefficients of f_2 and apply the operation of averaging and differencing on them. If we choose the first two approximation coefficients of f_2 and take the average, this will give us the first entry of f_3. In the next step, subtract this averaged quantity from the first element of the corresponding pair. Similarly, by doing the same with rest of the approximation coefficients, we get transformed vector f_3. The transformed vector f_3 is given as below,

$$f_3 = \begin{matrix} [\frac{\frac{c_{11}+\cdots+c_{14}}{4}+\frac{c_{15}+\cdots+c_{18}}{4}}{2} & \frac{\frac{c_{11}+\cdots+c_{14}}{4}-\frac{c_{15}+\cdots+c_{18}}{4}}{2} & \frac{c_{11}+c_{12}-c_{13}-c_{14}}{4} & \frac{c_{15}+c_{16}-c_{17}-c_{18}}{4} \\ \frac{c_{11}-c_{12}}{2} & \frac{c_{13}-c_{14}}{2} & \frac{c_{15}-c_{16}}{2} & \frac{c_{17}-c_{18}}{2}]. \end{matrix}$$

after simplification, we can write f_3 as

$$f_3 = \begin{matrix} [\frac{c_{11}+\cdots+c_{18}}{8} & \frac{c_{11}+\cdots+c_{14}-c_{15}-\cdots-c_{18}}{8} & \frac{c_{11}+c_{12}-c_{13}-c_{14}}{4} & \frac{c_{15}+c_{16}-c_{17}-c_{18}}{4} \\ \frac{c_{11}-c_{12}}{2} & \frac{c_{13}-c_{14}}{2} & \frac{c_{15}-c_{16}}{2} & \frac{c_{17}-c_{18}}{2}]. \end{matrix}$$

The method of averaging and differencing is very effective. For large image matrices the averaging and differencing technique by this procedure becomes quite tedious and complicated to compute. We can perform averaging and differencing to the large image matrix quite easily and quickly by using the matrix multiplication. In the next section, we will discuss the transformation of the image matrix by using the multiplication of some matrices.

5.3 Transformation of Image Matrix using Linear Algebra

We have described the transformation of a row vector f intof_1, f_1intof_2, and$f_2$intof_3 in three steps. When we perform the averaging and the differencing technique to the image matrix, and if the image matrix is large enough, the averaging and differencing will become hard and complicated for numerical computation. With the help of linear algebra, we can develop a way to get a transformed image matrix only in one step by using single equation. For this we need transformation matrices which can be developed from f_1, f_2, f_3. Let's say A_1, A_2, A_3 are the transformation matrices. In this section, we will see how these matrices are transformed from f_1, f_2 and f_3. The transformation of vector f into f_1 can be written as

$$f_1 = (A_1 f^T)^T. \tag{5.1}$$

The transformed vector f_1 can be written in the form of the matrix A_1. First we write f_1 in the matrix form, which is given as

$$f_1 = \left(\begin{bmatrix} 1/2 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 1/2 & -1/2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & -1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & -1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & -1/2 \end{bmatrix} \begin{bmatrix} c_{11} \\ c_{12} \\ c_{13} \\ c_{14} \\ c_{15} \\ c_{16} \\ c_{17} \\ c_{18} \end{bmatrix} \right)^T.$$

The transformation matrix A_1 can be written as

$$A_1 = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ \\ 1/2 & -1/2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & -1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & -1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & -1/2 \end{bmatrix} = \begin{bmatrix} A_{\frac{S}{2}\times S} \\ \\ D_{\frac{S}{2}\times S} \end{bmatrix}.$$

By using the relation (5.1), the transformation matrix A_1 will average and difference the vector f only once, the resulting vector is f_1. The transformation matrix A_1 is of size 8×8 but we can extended it up to the size $2^j \times 2^j$where$j = 1, 2, \cdots, n$. to get a transformation matrix of any size. The matrix A_1 will divide the image matrix into two blocks $A_{\frac{S}{2}\times S}$ and $D_{\frac{S}{2}\times S}$, where $S = 2^j, j = 1, 2, \cdots, n$. The upper block $A_{\frac{S}{2}\times S}$ is called the block of approximation coefficients and the lower block $D_{\frac{S}{2}\times S}$ is called the block of detailed coefficients. Now, to transform vector f_1 into f_2 we have

$$f_2 = (A_2 f_1^T)^T. \tag{5.2}$$

The vector f_2 can be written in the matrix form as follows.

$$f_2 = \left(\begin{bmatrix} 1/2 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 1/2 & -1/2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & -1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{c_{11}+c_{12}}{2} \\ \frac{c_{13}+c_{14}}{2} \\ \frac{c_{15}+c_{16}}{2} \\ \frac{c_{17}+c_{18}}{2} \\ \frac{c_{11}-c_{12}}{2} \\ \frac{c_{13}-c_{14}}{2} \\ \frac{c_{15}-c_{16}}{2} \\ \frac{c_{17}-c_{18}}{2} \end{bmatrix} \right)^T .$$

Hence the transformation matrix A_2 can be written as

$$A_2 = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ \\ 1/2 & -1/2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & -1/2 & 0 & 0 & 0 & 0 \\ \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} A_{\frac{S}{4}\times\frac{S}{4}} \\ \\ D_{\frac{S}{4}\times\frac{S}{4}} \\ \\ I_{\frac{S}{2}\times\frac{S}{2}} \end{bmatrix} .$$

The transformation matrix A_2 will Average and difference the row vector f_1 twice. The resulting vector f_2 will contain three blocks. As we know, this vector $f_1$contains two blocks after first transformation. The matrix $A_{\frac{S}{2}\times\frac{S}{2}}$ further divide the block of approximation coefficients into two sub blocks $A_{\frac{S}{4}\times\frac{S}{4}}$ and $D_{\frac{S}{4}\times\frac{S}{4}}$. The upper block contains approximation coefficients, the middle block contains detailed coefficients and the lower block is arranged in a way that it will keep the same detailed coefficients of f_1 and we denote this block as $I_{\frac{S}{2}\times\frac{S}{2}}$. Now, to transform f_2 into f_3 we have

$$f_3 = (A_3 f_2^T)^T. \tag{5.3}$$

We can get the transformation matrix A_3 from third step of averaging and differencing. First we write f_3 in the matrix form as we did in the case of f_1 and f_2. Similarly the matrix A_3 can also be written as blocks of approximation and detailed coefficients as we described above for A_2. By doing so we can get A_3, which is given as follows

$$A_3 = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 \\ \\ 1/2 & -1/2 & 0 & 0 & 0 & 0 & 0 & 0 \\ \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} A_{\frac{S}{8}\times\frac{S}{8}} \\ \\ D_{\frac{S}{8}\times\frac{S}{8}} \\ I_{\frac{S}{2}+\frac{S}{4}\times\frac{S}{4}+\frac{S}{2}} \end{bmatrix} .$$

We got f_3 after three steps by averaging and differencing method. On the other hand we can get f_3 in one step by doing multiplication of transformation matrices A_1, A_2, A_3 i.e

$f_3 = W \cdot f$ if

$$W = A_3 \cdot A_2 \cdot A_1.$$

Finally, the matrix W becomes

$$W = \begin{bmatrix} 1/8 & 1/8 & 1/8 & 1/8 & 1/8 & 1/8 & 1/8 & 1/8 \\ \\ 1/8 & 1/8 & 1/8 & 1/8 & -1/8 & -1/8 & -1/8 & -1/8 \\ 1/4 & 1/4 & -1/4 & -1/4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/4 & 1/4 & -1/4 & -1/4 \\ 1/2 & -1/2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & -1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & -1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & -1/2 \end{bmatrix} = \begin{bmatrix} A \\ \\ D \end{bmatrix}.$$

The matrix W will be used to transform an image matrix. The transformed matrix will contain two blocks A and D. The matrix A contains approximation coefficients and the matrix D contains detailed coefficients. The matrix W is obtained by using the averaging and differencing method, and the matrix A_H, in section(4), is obtained by using the Haar wavelet basis. We can observe that the upper block which contains the approximation coefficients is same. Where as the block of detailed coefficients is different. The difference between the detail coefficients does not effect the image result. we can also observe that the transformation matrix A_{H_1}, and the A_1 are similar but the other matrices have different detail coefficients. So the averaging and differencing method is used as the application of Haar wavelet.

5.3.1 **Reconstruction of the Image Matrix**

The individual matrices A_1,A_2,A_3 are invertible because each column of these matrices that comprise W is orthogonal to every other column. Thus

$$W^{-1} = A_1^{-1} \cdot A_2^{-1} \cdot A_3^{-1}.$$

By multiplying f_3 with W^{-1}, i.e $f = f_3 \cdot W^{-1}$, we can get the original vector or image matrix back.

As we have stated above, Haar wavelet does this transformation to each column and then repeat the procedure to each row of the image matrix. This means that we have two transformations, **one dimensional and two dimensional** transformation. We will discuss the Haar transformed matrices by using the following example.

Example 5.1. To describe the one and two dimensional transformation, we consider an 8×8 image matrix. Let I is an original image matrix given below

$$I = \begin{bmatrix} 90 & 84 & 76 & 67 & 59 & 55 & 60 & 65 \\ 92 & 83 & 78 & 68 & 61 & 54 & 59 & 68 \\ 92 & 80 & 73 & 68 & 62 & 56 & 62 & 71 \\ 91 & 81 & 73 & 66 & 62 & 59 & 66 & 68 \\ 89 & 80 & 72 & 68 & 63 & 65 & 68 & 71 \\ 87 & 77 & 71 & 68 & 64 & 70 & 68 & 76 \\ 84 & 79 & 72 & 70 & 66 & 71 & 73 & 77 \\ 85 & 81 & 77 & 76 & 73 & 76 & 73 & 75 \end{bmatrix}.$$

First, we will describe the one dimensional transformation by using the transformation matrix W and then the two dimensional transformation.

1. **One Dimensional Transformation**

   In one dimensional transformation, first we transform the columns which is called **column transformed** matrix. Since $W = A_3 \cdot A_2 \cdot A_1$, by multiplying W and I we get

$$Q = W \cdot I,$$

$$Q = \begin{bmatrix} 88.7500 & 80.6250 & 74.0000 & 68.8750 & 63.7500 & 63.2500 & 66.1250 & 71.3750 \\ 2.5000 & 1.3750 & 1.0000 & -1.6250 & -2.7500 & -7.2500 & -4.3750 & -3.3750 \\ -0.2500 & 1.5000 & 2.0000 & 0.2500 & -1.0000 & -1.5000 & -2.2500 & -1.5000 \\ 1.7500 & -0.7500 & -1.5000 & -2.5000 & -3.0000 & -3.0000 & -2.5000 & -1.2500 \\ -1.0000 & 0.5000 & -1.0000 & -0.5000 & -1.0000 & 0.5000 & 0.5000 & -1.5000 \\ 0.5000 & -0.5000 & 0 & 1.0000 & 0 & -1.5000 & -2.0000 & 1.5000 \\ 1.0000 & 1.5000 & 0.5000 & 0 & -0.5000 & -2.5000 & 0 & -2.5000 \\ -0.5000 & -1.0000 & -2.5000 & -3.0000 & -3.5000 & -2.5000 & 0 & 1.0000 \end{bmatrix} = \begin{bmatrix} A \\ D \end{bmatrix},$$

   Hence Q is the column transformed matrix. It contains two blocks A and D. The first block of Q contains approximation coefficients, and second block contains detailed coefficients.

2. **Two Dimensional Transformation**

   After getting Q, we perform the averaging and differencing to the rows of Q. For this, we first take the transpose of the matrix W and then multiply it by the matrix Q. This yields

$$\begin{aligned} T &= W \cdot I \cdot W^T \\ &= Q \cdot W^T, \end{aligned} \tag{5.4}$$

$$T = \begin{bmatrix} 72.0938 & 5.9688 & 6.6250 & -2.6250 & 4.0625 & 2.5625 & 0.2500 & -2.6250 \\ -1.8125 & 2.6250 & 1.1250 & -0.5625 & 0.5625 & 1.3125 & 2.2500 & -0.5000 \\ -0.3438 & 1.2188 & -0.2500 & 0.3125 & -0.8750 & 0.8750 & 0.2500 & -0.3750 \\ -1.5938 & 0.8438 & 1.2500 & -0.5625 & 1.2500 & 0.5000 & 0 & -0.6250 \\ -0.4375 & -0.0625 & 0.2500 & 0.1250 & -0.7500 & -0.2500 & -0.7500 & 1.0000 \\ -0.1250 & 0.3750 & -0.2500 & -0.2500 & 0.5000 & -0.5000 & 0.7500 & -1.7500 \\ -0.3125 & 1.0625 & 0.5000 & -0.1250 & -0.2500 & 0.2500 & 1.0000 & 1.2500 \\ -1.5000 & -0.2500 & 10000 & -1.7500 & 0.2500 & 0.2500 & -0.5000 & -0.5000 \end{bmatrix}.$$

   The transformed matrix T is also called **row-and-column transformed** matrix. So T is the matrix which will be used for Haar compression and it has only one approximation coefficient. We can calculate the column-row transformed matrix of any size by the procedure described above.

   We can get back the original image matrix from the row-and-column transformed matrix.

   By taking the inverse of (5.4), we can get the matrix I which is our original image matrix.

$$I = W^{-1} \cdot T \cdot (W^t)^{-1}. \tag{5.5}$$

5.4 Approximation and Detail Coefficients in One and Two Dimensional Transformation

In this section, we will discuss approximation and detail coefficients. First, we will discuss the approximation coefficients in one dimensional transformation and later on in two dimensional transformation.

5.4.1 One Dimensional Transformation

We consider again the image matrix I of size $2^j \times 2^j$, for $j = 3$ from the previous section. By applying the averaging and differencing method on this image matrix we will see how the approximation and detailed coefficients look like. Since, we have discussed before that for a matrix of the size $2^j \times 2^j$, we have three steps to get the transformed matrix T. So we will do average and difference step by step.

1. **Averaging and Differencing Once**

   First, we do the average and difference once, it will divide the $2^j \times 2^j$ matrix into two blocks A and D each of dimension $2^j/2 \times 2^j$. For this, we use the equation

$$\begin{aligned} Q_1 &= A_1 \cdot I \\ &= W \cdot I, \end{aligned}$$

   consider $W = A_1$ be the transformation matrix. The matrix Q_1 becomes

$$Q_1 = \begin{bmatrix} 91.0000 & 83.5000 & 77.0000 & 67.5000 & 60.0000 & 54.5000 & 59.5000 & 66.5000 \\ 91.5000 & 80.5000 & 73.0000 & 67.0000 & 62.0000 & 57.5000 & 64.0000 & 69.5000 \\ 88.0000 & 78.5000 & 71.5000 & 68.0000 & 63.5000 & 67.5000 & 68.0000 & 73.5000 \\ 84.5000 & 80.0000 & 74.5000 & 73.0000 & 69.5000 & 73.5000 & 73.0000 & 76.0000 \\ -1.0000 & 0.5000 & -1.0000 & -0.5000 & -1.0000 & 0.5000 & 0.5000 & -1.5000 \\ 0.5000 & -0.5000 & 0 & 1.0000 & 0 & -1.5000 & -2.0000 & 1.5000 \\ 1.0000 & 1.5000 & 0.5000 & 0 & -0.5000 & -2.5000 & 0 & -2.5000 \\ -0.5000 & -1.0000 & -2.5000 & -3.0000 & -3.5000 & -2.5000 & 0 & 1.0000 \end{bmatrix} = \begin{bmatrix} A \\ D \end{bmatrix}.$$

   The upper block A consists of the approximation coefficients which contains all the important information of the image. The lower block D contains detailed coefficients.

2. **Averaging and Differencing Twice**

   To get the twice transformed matrix with average and difference method, we use

$$\begin{aligned} Q_2 &= A_2 \cdot A_1 \cdot I \\ &= W \cdot I. \end{aligned}$$

   Where $W = A_2 \cdot A_1$ The average and difference twice will again divide the block A into to two sub blocks A_s and D_s each of dimension $2^j/4 \times 2^j$ where $j = 3$. The lower block of detailed coefficient D will remain the same. Hence, the matrix Q_2 becomes,

$$Q_2 = \begin{bmatrix} 91.2500 & 82.0000 & 75.0000 & 67.2500 & 61.0000 & 56.0000 & 61.7500 & 68.0000 \\ 86.2500 & 79.2500 & 73.0000 & 70.5000 & 66.5000 & 70.5000 & 70.5000 & 74.7500 \\ -0.2500 & 1.5000 & 2.0000 & 0.2500 & -1.0000 & -1.5000 & -2.2500 & -1.5000 \\ 1.7500 & -0.7500 & -1.5000 & -2.5000 & -3.0000 & -3.0000 & -2.5000 & -1.2500 \\ -1.0000 & 0.5000 & -1.0000 & -0.5000 & -1.0000 & 0.5000 & 0.5000 & -1.5000 \\ 0.5000 & -0.5000 & 0 & 1.0000 & 0 & -1.5000 & -2.0000 & 1.5000 \\ 1.0000 & 1.5000 & 0.5000 & 0 & -0.5000 & -2.5000 & 0 & -2.5000 \\ -0.5000 & -1.0000 & -2.5000 & -3.0000 & -3.5000 & -2.5000 & 0 & 1.0000 \end{bmatrix} = \begin{bmatrix} A_s \\ D_s \\ D \end{bmatrix}$$

   The upper block A_s contains the approximation coefficients, and the other two blocks D_s and D contains detailed coefficients.

3. **Averaging and Differencing Thrice**

If we do average and difference thrice that is $Q_3 = A_3 \cdot A_2 \cdot A_1 \cdot I = W \cdot I$ on the image matrix I, where $W = A_3 \cdot A_2 \cdot A_1$. We get the following matrix

$$Q_3 = \begin{bmatrix} 88.7500 & 80.6250 & 74.0000 & 68.8750 & 63.7500 & 63.2500 & 66.1250 & 71.3750 \\ 2.5000 & 1.3750 & 1.0000 & -1.6250 & -2.7500 & -7.2500 & -4.3750 & -3.3750 \\ -0.2500 & 1.5000 & 2.0000 & 0.2500 & -1.0000 & -1.5000 & -2.2500 & -1.5000 \\ 1.7500 & -0.7500 & -1.5000 & -2.5000 & -3.0000 & -3.0000 & -2.5000 & -1.2500 \\ -1.0000 & 0.5000 & -1.0000 & -0.5000 & -1.0000 & 0.5000 & 0.5000 & -1.5000 \\ 0.5000 & -0.5000 & 0 & 1.0000 & 0 & -1.5000 & -2.0000 & 1.5000 \\ 1.0000 & 1.5000 & 0.5000 & 0 & -0.5000 & -2.5000 & 0 & -2.5000 \\ -0.5000 & -1.0000 & -2.5000 & -3.0000 & -3.5000 & -2.5000 & 0 & 1.0000 \end{bmatrix} = \begin{bmatrix} A_{s_1} \\ D_{s_1} \\ D \end{bmatrix}.$$

Averaging and differencing thrice divides the block A_s into two sub blocks A_{s_1} and D_{s_1} of size $2^j/2 \times 2^j$. Hence, we have only one block of size $2^j/8 \times 2^j$ which contain the approximation coefficients while the rest of the blocks contains the detail coefficients.

5.4.2 Two Dimensional Transformation

For two dimensional transformation or row-column transformation, we will discuss Q_1, Q_2 and Q_3 step by step.

1. **Averaging and Differencing Once**

   First, we do the average and difference once, it will divide our $2^j \times 2^j$ for $j = 3$, matrix into four blocks of the dimension $2^j/2 \times 2^j/2$. For this we use the equation

$$\begin{aligned} T_1 &= A_1 \cdot I \cdot A_1^T \\ &= Q \cdot W^T. \end{aligned}$$

   Where $Q = A_1 \cdot I$ and $W = A_1$. The matrix T_1 becomes

$$T_1 = \begin{bmatrix} 87.2500 & 72.2500 & 57.2500 & 63.0000 & 3.7500 & 4.7500 & 2.7500 & -3.5000 \\ 86.0000 & 70.0000 & 59.7500 & 66.7500 & 5.5000 & 3.0000 & 2.2500 & -2.7500 \\ 83.2500 & 69.7500 & 65.5000 & 70.7500 & 4.7500 & 1.7500 & -2.0000 & -2.7500 \\ 82.2500 & 73.7500 & 71.5000 & 74.5000 & 2.2500 & 0.7500 & -2.0000 & -1.5000 \\ -0.2500 & -0.7500 & -0.2500 & -0.5000 & -0.7500 & -0.2500 & -0.7500 & 1.0000 \\ 0 & 0.5000 & -0.7500 & -0.250 & 0.5000 & -0.5000 & 0.7500 & -1.7500 \\ 1.2500 & 0.2500 & -1.5000 & -1.2500 & -0.2500 & 0.2500 & 1.0000 & 1.2500 \\ -0.7500 & -2.7500 & -3.0000 & 0.5000 & 0.2500 & 0.2500 & -0.5000 & -0.5000 \end{bmatrix} = \begin{bmatrix} A & D_1 \\ D_2 & D_3 \end{bmatrix}$$

   The block A consists on the approximation coefficients, and the other three blocks contains the detailed coefficients.

2. **Averaging and Differencing Twice**

   To get the transformed matrix twice with average and difference method, we use

$$\begin{aligned} T_2 &= A_2 \cdot A_1 \cdot I \cdot A_2^T \cdot A_1^T \\ &= Q_2 \cdot W^T. \end{aligned}$$

   where $Q_2 = A_2 \cdot A_1 \cdot I$ and $W = A_2 \cdot A_1$ The average and difference twice further divides the block A in the upper left half corner of the matrix T_1 into four sub blocks of size $2^j/4 \times 2^j/4$. Hence the matrix T_2 becomes,

$$T_2 = \begin{bmatrix} 78.8750 & 61.6875 & 7.7500 & -3.1875 & 4.6250 & 3.8750 & 2.5000 & -3.1250 \\ 77.2500 & 70.5625 & 5.5000 & -2.0625 & 3.5000 & 1.2500 & -2.0000 & -2.1250 \\ 0.8750 & -1.5625 & -0.2500 & 0.3125 & -0.8750 & 0.8750 & 0.2500 & -0.3750 \\ -0.7500 & -2.4375 & 1.2500 & -0.5625 & 1.2500 & 0.5000 & 0 & -0.6250 \\ -0.5000 & -0.3750 & 0.2500 & 0.1250 & -0.7500 & -0.2500 & -0.7500 & 1.0000 \\ 0.2500 & -0.5000 & -0.2500 & -0.2500 & 0.5000 & -0.5000 & 0.7500 & -1.7500 \\ 0.7500 & -1.3750 & 0.5000 & -0.1250 & -0.2500 & 0.2500 & 1.0000 & 1.2500 \\ -1.7500 & -1.2500 & 1.0000 & -1.7500 & 0.2500 & 0.2500 & -0.5000 & -0.5000 \end{bmatrix} = \begin{bmatrix} A_s & D_1 \\ D_2 & D_3 \end{bmatrix}$$

From the above T_2 matrix we observe that the block A_s of size $2^j/2 \times 2^j/2$ and it contains approximation coefficients. And the other blocks contain detailed coefficients.

3. **Averaging and Differencing Thrice**

   If we do average and difference thrice i.e $T_3 = Q_3 \cdot W$, where $Q_3 = A_3 \cdot A_2 \cdot A_1 \cdot I$ and $W = A_3 \cdot A_2 \cdot A_1$. We get the following matrix

$$T_3 = \begin{bmatrix} 72.0938 & 5.9688 & 6.6250 & -2.6250 & 4.0625 & 2.5625 & 0.2500 & -2.6250 \\ -1.8125 & 2.6250 & 1.1250 & -0.5625 & 0.5625 & 1.3125 & 2.2500 & -0.5000 \\ -0.3438 & 1.2188 & -0.2500 & 0.3125 & -0.8750 & 0.8750 & 0.2500 & -0.3750 \\ -1.5938 & 0.8438 & 1.2500 & -0.5625 & 1.2500 & 0.5000 & 0 & -0.6250 \\ -0.4375 & -0.0625 & 0.2500 & 0.1250 & -0.7500 & -0.2500 & -0.7500 & 1.0000 \\ -0.1250 & 0.3750 & -0.2500 & -0.2500 & 0.5000 & -0.5000 & 0.7500 & -1.7500 \\ -0.3125 & 1.0625 & 0.5000 & -0.1250 & -0.2500 & 0.2500 & 1.0000 & 1.2500 \\ -1.5000 & -0.2500 & 10000 & -1.7500 & 0.2500 & 0.2500 & -0.5000 & -0.5000 \end{bmatrix}.$$

   Averaging and differencing thrice divides the block of size $2^j/4 \times 2^j/4$ that contains approximation coefficients into $2^j/8 \times 2^j/8$ sub matrices. While the remaining matrices contain the detail coefficients.

5.5 Image Results using Transformation Matrices

In the section 5.4, we considered 8×8 matrix and applied average and difference method to transform the image data in one and two dimensional respectively. We considered a small matrix because it is difficult to show the results on the large size of matrices like a matrix of size512×512. Therefore, for simplicity, we used 8×8 image matrix. The procedure described above for the transformation of image matrix will same for all size of the image matrices. For instance, we will consider two different images of size $2^j \times 2^j$ resolution, where $j = 9$. To transform the image matrix, first we will extend the transformation matrices up to the size of the original image matrix. Later, we will apply the transformed matrix to the image to show that how the approximation and detailed coefficients stores the image information. The whole procedure will be carried out with the help of MATLAB. For this we will consider two different pictures. The figure 5.1a and 5.1b are the original images of the panda and the Swedish parliament.

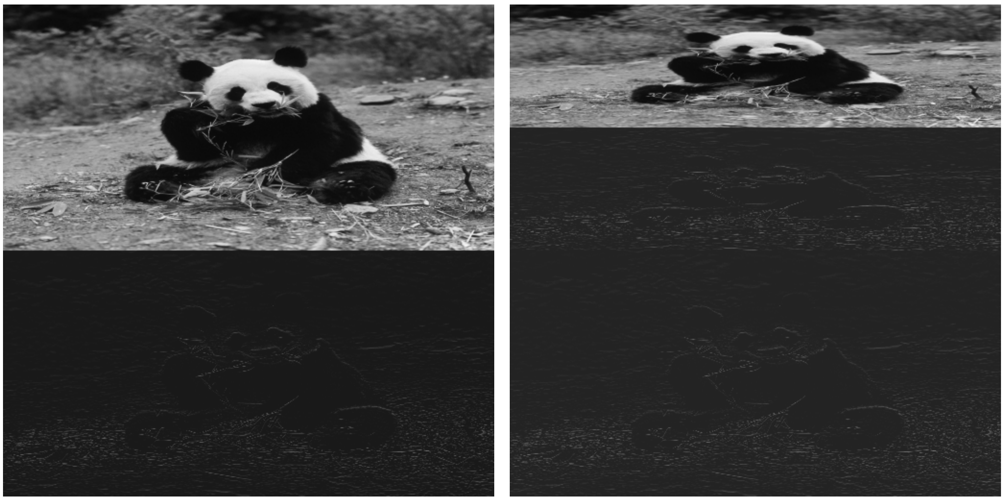
(a) The Panda　　　(b) Swedish Praliment

Figure 5.1: Original Images

When an image matrix is transformed some of the detailed coefficients keep the horizontal information and the some of them keep the vertical information. Therefore, to show the horizontal and vertical information, we consider the picture of Swedish parliament, which contains the horizontal and vertical lines. These pictures will be shown in the next section. In section 7, we will discuss the compressed results of gray images as well as the colored images of both pictures.

5.5.1 **One Dimensional Transformation**

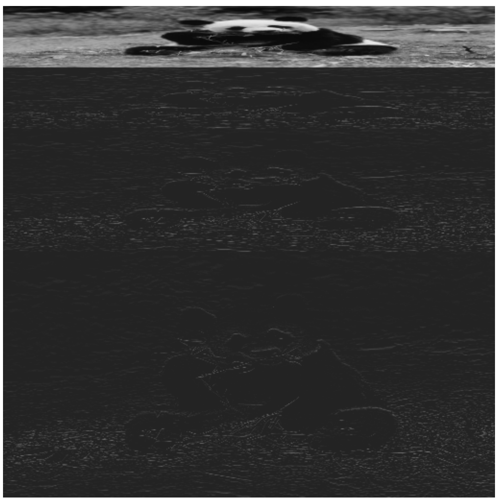
By averaging and differencing once, twice and thrice, we can convert the transformation matrices Q_1, Q_2 and Q_3 into images to observe the approximation and detailed coefficients.



(a) AD Once (b) AD Twice

Figure 5.2: Averaging and Differencing Once and Twice

As we have discussed before that when we do average and difference in one dimensional transformation, it divides our image matrix into two sub matrices.



(a) AD Thrice

Figure 5.3: Averaging and Differencing Thrice

In our case the image matrix is of the dimension 512×512, so average and difference method divides this matrix into two sub-blocks of size 256×512 where the upper block contains approximation coefficients, and the lower block contains the detailed coefficients. In the upper part where we have approximation coefficients the image is clearly visible and in the lower part where we have detailed coefficients we have some detail about the image. Similarly, the average and difference third time divides the 256×512 matrix into two sub blocks of size 128×512. The upper block of size 128×512 contains approximation coefficients and the other two matrices of size 128×512 and 256×512 contain detailed coefficients.

Similarly, we can do the same thing step by step. When we do the average and difference 9th time, our image matrix will be divided into 9 sub blocks. In the upper block we have the sub block of size 1×512 and it will contain the approximation coefficients whereas the remaining blocks will contain the detail coefficients.

The results after the averaging and differencing once, twice and thrice can be seen in the figures above, we used the abbreviation AD for averaging and differencing.

5.5.2 Two Dimensional Transformation

In this section, we consider the pictures 5.1a and 5.1b. This time we will do averaging and differencing in two dimensional transformation.

We have discussed the transformed matrices T_1, T_2 and T_3 in two dimensional transformation in the section 5.4.2. Here we will apply this technique on the original image of panda and Swedish parliament. When we do the average and difference once in two dimensional transformation, the image matrix will be divided into four sub blocks of size 256×256. The block in the upper left corner contains approximation coefficients, and the other three blocks contain detailed coefficients.

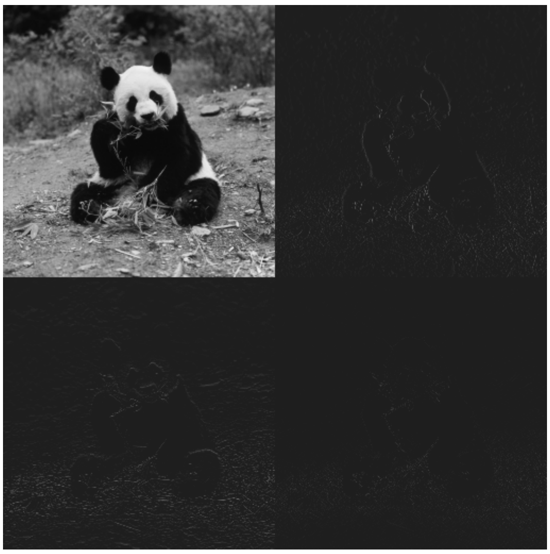
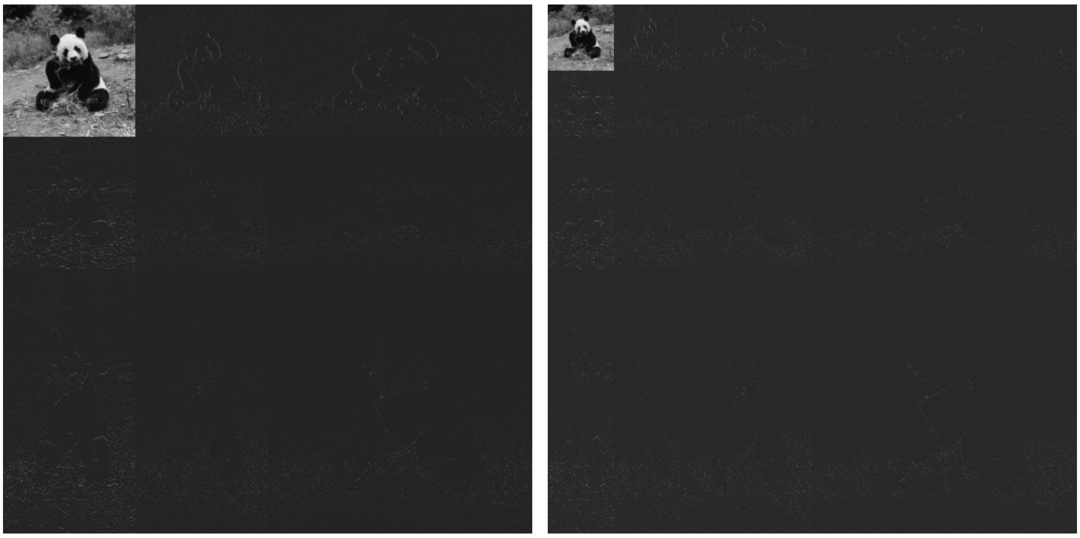


Figure 5.4: AD Once

We can observe that the image is visible in the part where we have approximation coefficients, and the information about the image is less where we have detailed coefficients. To show the detailed part clearly we added a coefficient in the detailed blocks of the transformed matrix.

Now, if we do the same thing again, then the block of size 256×256 in the upper left half will be divided further into four sub blocks of size 128×128. And the matrix of size 128×128 in the upper left half will contain approximation coefficients where as the



(a) AD Twice

(b) AD Thrice

Figure 5.5: Averaging and Differencing Twice and Thrice

remaining blocks will contain detailed coefficients. The results after the averaging and differencing once, twice and thrice can be seen in the figures 5.6, 5.7a and 5.7b.

In the picture 5.6 we can see the four blocks represent the approximation and detailed coefficients. The upper left corner of the image represent the approximation coefficients and the upper right corner contains the detailed coefficients and we can also observe that these detailed coefficients store the vertical information about the image and the lower left corner contains the horizontal information about the image.

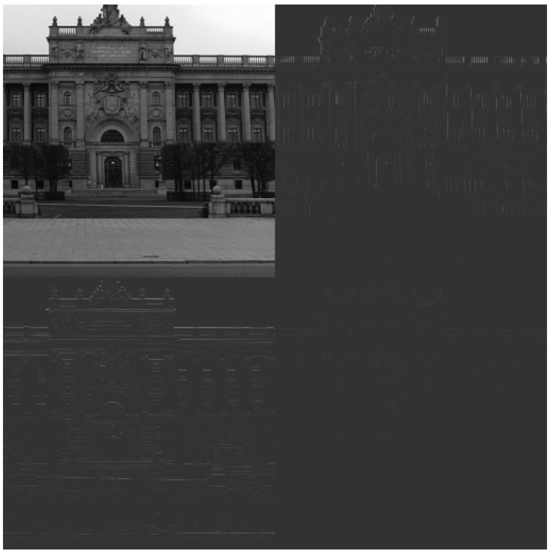
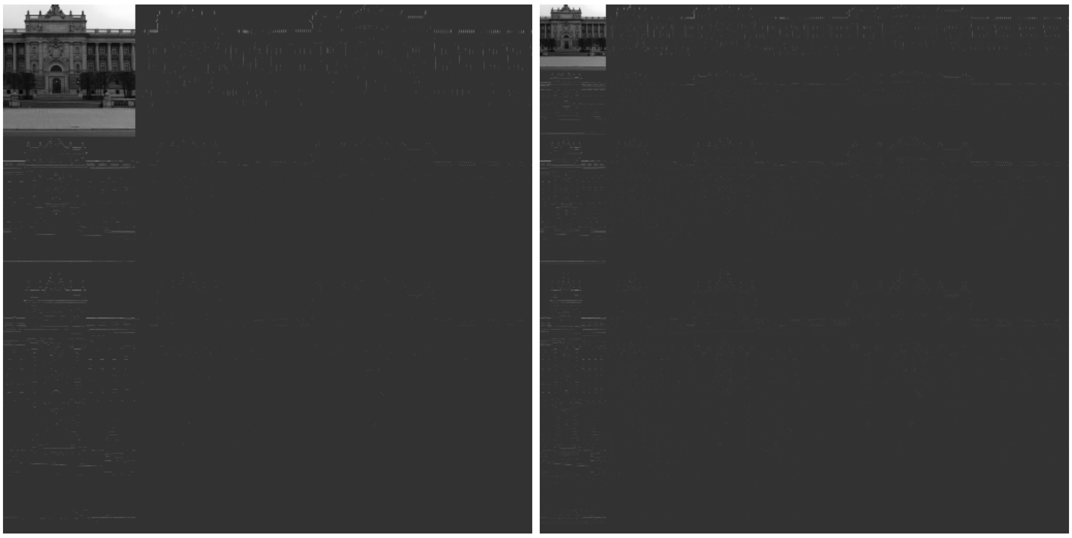


Figure 5.6: AD Once



(a) AD Twice

(b) AD Thrice

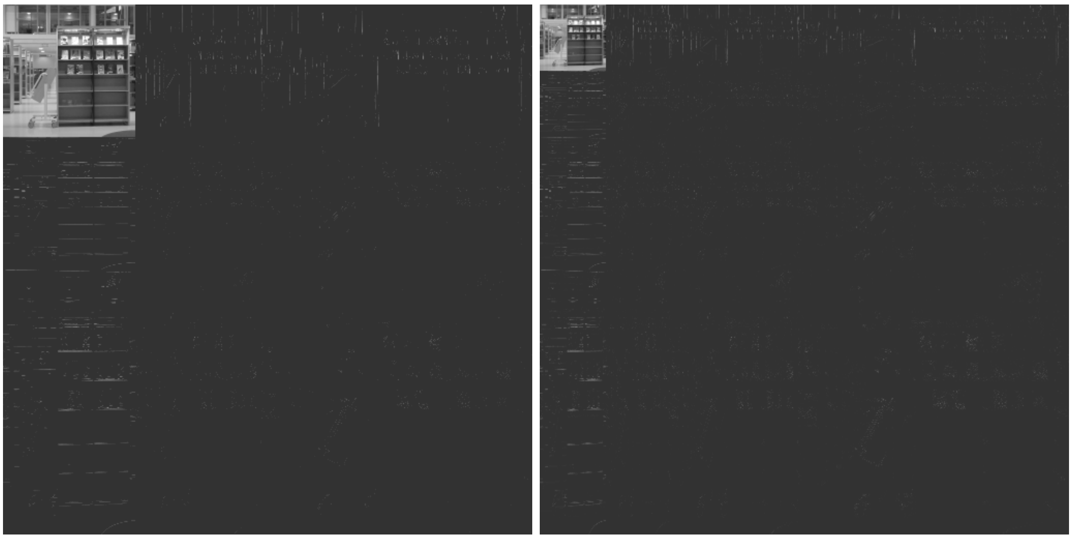
Figure 5.7: Averaging and Differencing Twice and Thrice

We can observe the horizontal and vertical information in the pictures 5.7a and 5.7b as well. To show the image results clearly we brighten the detailed part in each picture. The horizontal and detailed coefficients can also observe in pictures 5.8b, 5.9a, and 5.9b.

(a) Original Image

(b) AD Once

Figure 5.8: Original Image and Averaging and Differencing Once



(a) AD Twice (b) AD Thrice

Figure 5.9: Averaging and Differencing Twice and Thrice

6 Application of Haar Wavelet

Haar wavelets have broad applications, in this chapter we will describe some of the application of Haar wavelet. Haar wavelet is used to compress an image, to download a picture form Internet. The material in this chapter is taken from [7] and [1].

6.1 Compression

Digital technology is increasing day by day. Therefore, the use of digital images is also increasing rapidly. Digital images require a large amount of storage space. When we download a picture from Internet it takes time. To store and transmit data in an efficient manner we need to reduce the redundancy of the image data. The wavelets are used to compress an image data so that it take less storage space. The images contain redundant data. The image compression reduces the pixels of an image and we can store other images too. It also reduces the time which is required to download a picture and a web page from the Internet. To get rid of redundant data we compress the transformed matrix T and change this matrix into sparse matrix by thresholding because spars matrices are those matrices which take up very less memory in the system. Before going into the details of wavelet compression, we define sparse matrix.

Definition 6.1. Sparse Matrix
If a matrix $S = \{a_{ij}\}$ has high number of zero entries, then this matrix is called sparse matrix.

To compress the image we need a sparse matrix, the sparse matrices have advantage because they take less storage space. For example, let I is the original image matrix and T is the Haar transformed matrix. The matrix T can be made sparse by choosing a small value.

For the wavelet compression, we need to choose a non-negative small value ε, where ε is called threshold value. In the next step the elements of the transformed matrix T will be replaced by zero if the absolute value of the elements in T are less than or equal to the ε. Mathematically we can write , if $T = \{a_{ij}\}$ then

$$B_{ij} = \begin{cases} a_{ij} & \text{if } |a_{ij}| > \varepsilon \\ 0 & \text{if } |a_{ij}| \leq \varepsilon. \end{cases}$$

We have two types of compression depend upon choosing the epsilon value. They are

1. **Lossless Compression**

   If we choose $\varepsilon = 0$, this means that we can not modify any of the elements and we will not lose original information. This compression is called lossless compression.

2. **Lossy Compression**

   If we choose $\varepsilon > 0$, this means that we modify some of the elements and in this way we lose some of the information. This compression is called lossy compression.

By making the comparison between lossless and lossy compression, we observe the following things. In lossless compression we do not loose original information and we get original image back. In lossy compression we loose some of our original information and it reduces the image quality. We can not get the original image back, we can only make the approximation of the original image. This method removes the details of image which are not noticeable by human eyes. Our task is to describe the image compression technique by using Haar wavelets.

The matrix T that we considered in the section 5.3.1 gets the following form after choosing some threshold value ε.

$$S = \begin{bmatrix} 72.0938 & 5.9688 & 6.6250 & -2.6250 & 4.0625 & 2.5625 & 0 & -2.6250 \\ -1.8125 & 2.6250 & 1.1250 & 0 & 0 & 1.3125 & 2.2500 & 0 \\ 0 & 1.2188 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1.5938 & 0 & 1.2500 & 0 & 1.2500 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1.7500 \\ 0 & 1.0625 & 0 & 0 & 0 & 0 & 0 & 1.2500 \\ -1.5000 & 0 & 0 & -1.7500 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

We can observe that the matrix S is sparse enough and it will take less storage space. In the compressed matrix S the negative, positive and zero coefficients will correspond to the gray, white and black pixels. The zeros will represent the black and the high positive values will represent the white pixels.

Now we summaries our discussion about wavelet compression in the following steps.

1. Conversion of image into matrix(I).
2. To calculate the row column transformed matrix (T).
3. Thresholding.
4. Reconstruction of image matrix.

To get reconstructed matrix R from the sparse matrix S, we use equation (5.5) again,

$$I = W^{-1} \cdot T \cdot (W^T)^{-1}.$$

Here the matrix W is orthogonal and we know that the inverse of an orthogonal matrix is equal to its own transpose, the above equation hence becomes

$$R = W^T \cdot S \cdot W. \tag{6.1}$$

For instance, the reconstructed matrix R is the following

$$R = \begin{bmatrix} 91.9063 & 83.7813 & 76.2188 & 68.4688 & 60.0938 & 55.5938 & 60.4688 & 65.7188 \\ 91.9063 & 83.7813 & 76.2188 & 68.4688 & 60.0938 & 55.5938 & 60.4688 & 65.7188 \\ 89.4688 & 81.3438 & 73.7813 & 66.0313 & 62.5313 & 58.0313 & 61.1563 & 69.9063 \\ 89.4688 & 81.3438 & 73.7813 & 66.0313 & 62.5313 & 58.0313 & 64.6563 & 66.4063 \\ 88.7813 & 78.1563 & 71.2188 & 68.7188 & 63.0313 & 67.5313 & 69.1563 & 71.9063 \\ 86.6563 & 6.0313 & 69.0938 & 66.5938 & 65.1563 & 69.6563 & 68.7813 & 76.5313 \\ 84.4063 & 78.7813 & 74.3438 & 71.8438 & 64.0313 & 68.5313 & 72.4063 & 77.6563 \\ 87.4063 & 1.7813 & 77.3438 & 74.8438 & 70.5313 & 75.0313 & 71.9063 & 77.1563 \end{bmatrix}.$$

We can observe that the reconstructed matrix is little bit different from the original image matrix I. So we lost some information about the image. The difference between the original image matrix and reconstructed matrix will depend upon choosing the epsilon value. If the epsilon is large then the difference will also be large. If the epsilon is zero we do not loss any information about the image and we will get back the original image matrix.

Compression ratio is used to measure the level of compression. Compression ratio is defined as

Definition 6.2. Compression Ratio

The ratio of the non-zero entries in the transformed matrix T to the number of non zero entries in the compressed matrix S is called compression ratio.

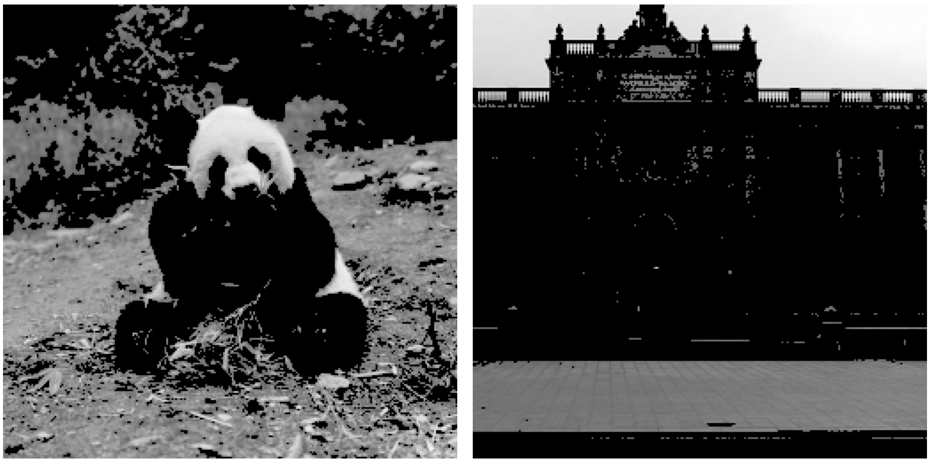
According to [1] a compression ration of $10:1$ or greater means the compressed matrix S is sparse enough to have a significant savings in terms of storage and transmission time.

7 Compressed Results

In the previous section, we presented image results by using the transformation matrices. Now we will choose different threshold values(ε) to compress the image. We will discuss the results step by step.

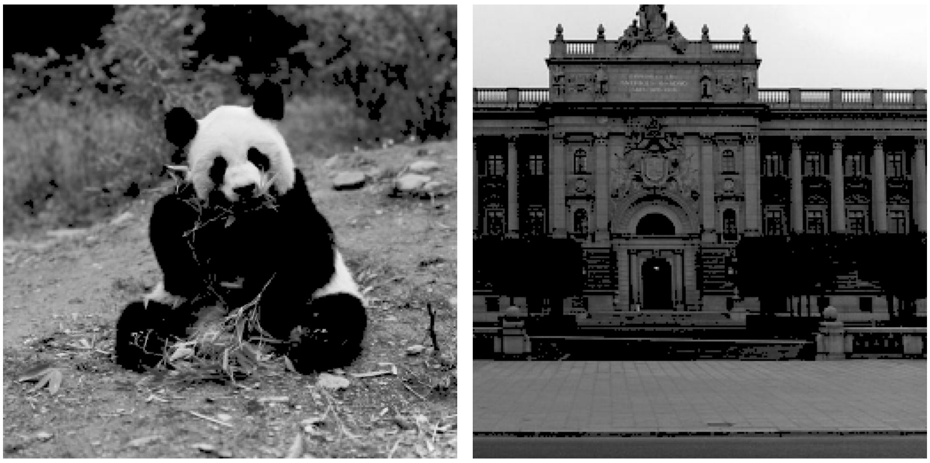
7.1 Gray Compressed Pictures

In the pictures below we choose different threshold values and compress the images. For the same epsilon the compression ratio for different pictures is different. We used the abbreviations CR for compression ratio in the pictures below:



(a) ε=100, CR=249094 : 33047= 8 : 1 (b) ε=100, CR=226419 : 20567= 11: 1

Figure 7.1: Averaging and Differencing Once



(a) ε=50, CR = 249094 : 50905 = 5 : 1 (b) ε=50, CR=226419 : 45314= 5: 1

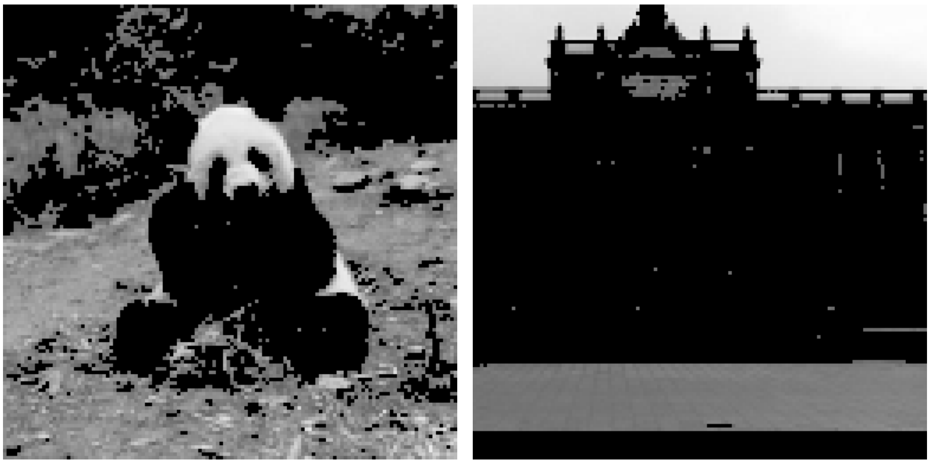
Figure 7.2: Averaging and Differencing Once

7.2 Colored Compressed Pictures

In this section, we compress colored pictures. The colored picture is divided into three different matrices and we use the same procedure for each matrix as we did for gray

(a) ε=10,CR = 249094 : 81502 = 3 : 1 (b) ε=10,CR = 226419 : 65281 = 3 : 1

Figure 7.3: Averaging and Differencing Once



(a) ε = 100, CR= 250060 : 8392 = 30 : 1 (b) ε=100, CR=222438 : 4970= 45: 1

Figure 7.4: Averaging and Differencing Twice

images. Then we combine these three compressed matrices to get the colored image back. The figures 7.8a and 7.8b are the original colored images.

(a) $\varepsilon = 50$, CR= 250060 : 12895 = 19 : 1

(b) ε=50, CR=222438 : 11521= 19: 1

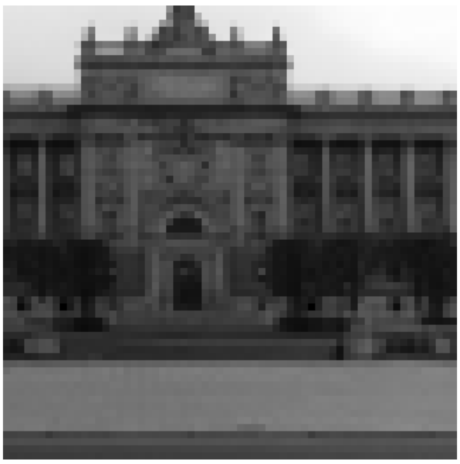
Figure 7.5: Averaging and Differencing Twice

(a) ε= 10, CR= 250060 : 32403 = 8 : 1

(b) ε=10, CR=222438 : 16360= 14: 1

Figure 7.6: Averaging and Differencing Twice

(a) ε= 10, CR= 250649 : 4096 = 61 : 1



(b) ε=10, CR=222438 : 4094= 54: 1

Figure 7.7: Averaging and Differencing Thrice

(a) Original Image of Panda

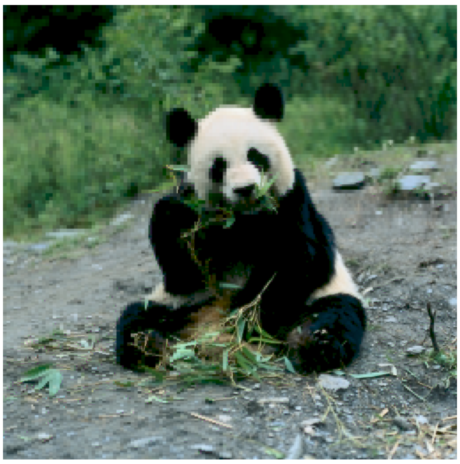
(b) Original Image of Swedish Parliament

Figure 7.8: Original Images

(a) CR = 798900:157815= 5 :1

(b) CR=686243:182457= 4:1

Figure 7.9: Average and Difference Once



(a) CR = 758534:89145= 9 :1

(b) CR=693496:101811=7:1

Figure 7.10: Average and Difference Twice

(a) CR = 762574:74628 = 10 :1 (b) CR=695379:84795=8:1

Figure 7.11: Average and Difference Thrice

8 Normalized Transformation

In this chapter, we are going to discuss another way of transformation called normalized transformation. In the chapter 5 and chapter 6 we have discussed averaging and differencing technique and transformations in details, here we will discuss normalized transformation briefly. The material in this chapter is taken from [4] and [1].

8.1 Normalized Wavelet Transform

In the chapter 5, we formed the non normalized transformation matrices A_i and then these matrices were used to form the transformation matrix W. After that we discussed one dimensional and two dimensional transformation with the help of examples.

Now, first we make the transformation matrices A_1, A_2 and A_3 normalized, and then we proceed to the single transformation matrix W. These normalized matrices will help us to make the transformation matrices powerful. These transformation matrices are then used to form the transformation matrix W. The transformation matrices will be orthogonal. Orthogonality means that the length of each column of the matrix is one and each column is orthogonal to the other columns of the matrix. Hence, the transformation matrix W will also be orthogonal as it is comprised by the orthogonal matrices. The transformation done in such a way is called normalized wavelet transform.

Now we discuss some benefits which the orthogonal transformation has.

1. The orthogonal matrix increases the speed of the calculations that helps us to reconstruct the image matrix, since the inverse of an orthogonal matrix is equal to its transpose.

2. If we compress an image with the help of normalized transformation, then this will be more close to the original image as compare to the non normalized transformation.

Now, we rewrite our transformation matrices A_1, A_2 and A_3 in the normalized form. By dividing each row of the matrix by its length, we can transform a matrix into the normalized form. Hence the normalized transformation matrices are:

$$A_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix},$$

$$A_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$$A_3 = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Now we form the orthogonal transformation matrix W using the equation

$$W = A_3 \cdot A_2 \cdot A_3,$$

as a result the matrix W becomes

$$W = \begin{bmatrix} \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} \\ \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{-1}{\sqrt{8}} & \frac{-1}{\sqrt{8}} & \frac{-1}{\sqrt{8}} & \frac{-1}{\sqrt{8}} \\ \frac{1}{2} & \frac{1}{2} & \frac{-1}{2} & \frac{-1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & \frac{-1}{2} & \frac{-1}{2} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix}.$$

Now we are in the position to form the one dimensional transformed matrix Q and two dimensional transformed matrix T. To form the matrix Q, we use the equation

$$Q = W \cdot I,$$

where I is the image matrix this can be seen in the section 5.4.1. Similarly, we can use the equation

$$T = W \cdot I \cdot W^T,$$

it has already been discussed in the section 5.4.2. After getting the transformed matrices Q and T we can do the one dimensional and two dimensional compression. We have already done the compression in the chapter 6 in details. Here we will discuss the compression briefly. The rest of the process is same and the difference is that here we have orthogonal matrices.

8.2 Compression

For the compression we again consider the picture of panda. We will apply the same operations as in the chapter 6 to compress the picture by using the normalized transformation matrices.

In the section 5.2 we have discussed the compression by averaging and differencing method in details. We will show some results of normalized compressed pictures in one dimensional and two dimensional transformations. Here we will just apply the operations on the image matrix,

8.3 Rebuilding of Normalized Compressed Image

The pictures 8.1a is compressed with compression ratio 5 : 1. The picture 7.2a has the same compression ratio, which is non normalized compressed image. We can observe that the normalized compressed result is more closer to the original image. The image 8.2a can be compare to the image 7.6a

(a) AD once

Figure 8.1: Normalized Rebuilt Image, CR=5:1

(a) AD Twice

Figure 8.2: Normalized Rebuilt Image, CR= 8:1

9 Singular Value Decomposition

In this chapter an important and special factorization, the singular value decomposition will be discussed. Before going into the details of the singular value decomposition we discuss some basic decompositions. The material in this chapter is taken from [6] and [2].

9.1 Diagonalization

A process of converting a square matrix into a diagonal matrix is called diagonalization. With the help of diagonalization a matrix A which contains information about eigenvalues and eigenvectors can be factorized in a useful factorization of the type

$$A = PDP^{-1}.$$

Definition 9.1. If a square matrix A is similar to a diagonal matrix, that is, if

$$A = PDP^{-1}$$

where P is invertible matrix and D is diagonal matrix then A is called diagonalizable.

The following theorem is about this suitable factorization and tells us how to construct this factorization.

Theorem 9.1. *A square matrix of the order $n \times n$ is diagonalizable if and only if A has n linearly independent eigenvectors.*

The diagonal matrix D in the definition 9.1 contains the eigenvalues of the matrix A in diagonal and the columns of A are the corresponding eigenvectors of the matrix A. We consider an example and try to diagonalize the matrix by using the above diagonalization

Example 9.1. Consider a 2×2 matrix

$$A = \begin{bmatrix} 7 & -4 \\ 2 & 1 \end{bmatrix}.$$

We need to find the invertible matrix P and the diagonal matrix D such that $A = PDP^{-1}$. First we find the eigenvalues of the matrix A, the associated equation of the matrix A is

$$\det(A - \lambda I) = 0,$$
$$\implies \lambda^2 - 8\lambda + 15 = 0.$$

By solving the above equation for λ gives $\lambda_1 = 5$ and $\lambda_2 = 3$. The corresponding eigenvector for $\lambda_1 = 5$ is

$$v_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix},$$

and the eigenvector for $\lambda_2 = 3$ is

$$v_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Since the eigenvectors form the columns of the matrix P, the matrix P is constructed and is given as

$$P = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}.$$

Now, we construct the diagonal matrix D from the corresponding eigenvalues and it becomes

$$D = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}.$$

By using the standard formula for the inverse of a 2×2 matrix the inverse of the matrix P is calculated and is given as

$$P^{-1} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}.$$

Finally, the diagonalization of the matrix A is

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}.$$

9.2 Diagonalization of Symmetric Matrices

In this section, we study the diagonalization of another type of matrices. Matrices which will be under discussion in this section are symmetric matrices. The theory of symmetric matrices is very rich and beautiful that arise often in application than another class of matrices. Before going into details we need to define orthogonal sets and orthonormal sets, which are defined as follows.

Theorem 9.2. *Any two eigenvectors from different eigen spaces of a symmetric matrix are orthogonal.*

If P is an orthogonal matrix (with $P^{-1} = P^T$) and D a diagonal matrix such that

$$A = PDP^T = PDP^{-1} \tag{9.1}$$

then the matrix A is said to be orthogonally diagonalizable. We consider an example to study the diagonalization of symmetric matrix.

Example 9.2. Consider the matrix

$$A = \begin{bmatrix} 6 & -2 & -1 \\ -2 & 6 & -1 \\ -1 & -1 & 5 \end{bmatrix}$$

the associated equation is

$$-\lambda^3 + 17\lambda^2 - 90\lambda + 144 = 0$$
$$-(\lambda - 8)(\lambda - 6)(\lambda - 3) = 0.$$

Solving the above equation for λ gives $\lambda_1 = 8$, $\lambda_2 = 6$, $\lambda_3 = 3$. The corresponding basis vectors are

$$u_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \quad u_2 = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}, \quad u_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

The vecors u_1, u_2 and u_3 can be used as columns of P as they form basis of $\mathbb{R}^3$. We see that these vectors are orthogonal and form an orthogonal set and it is easier to use P if its columns are orthogonal. Thus the normalized unit vectors are

$$v_1 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}, \quad v_2 = \begin{bmatrix} -1/\sqrt{6} \\ 1/\sqrt{6} \\ 2/\sqrt{6} \end{bmatrix}, \quad u_3 = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}.$$

The matrix P is formed as

$$P = \begin{bmatrix} -1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ 0 & 2/\sqrt{6} & 1/\sqrt{3}, \end{bmatrix}$$

and the diagonal matrix D is

$$D = \begin{bmatrix} 8 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

Hence the diagonalization of the matrix A is $A = PDP^{-1}$

Since the matrix P is a square matrix and its columns are orthonormal, P is an orthogonal matrix and simply $P^{-1} = P^T$. Theorem 9.2 shows how the eigenvectors in the example 9.2 are orthogonal. From the equation 9.1

$$A^T = (PDP^T)^T = PDP^T = A,$$

this implies that the matrix A is symmetric. The following theorem shows that every symmetric matrix is orthogonally diagonalizable.

Theorem 9.3. *An $n \times n$ square matrix A is orthogonally diagonalizable if and only if A is a symmetric matrix.*

9.3 Singular Value Decomposition

So far in this chapter, we have studied the diagonalization of a square matrix. We know that all the matrices can not be factorized as $A = PDP^{-1}$, where D is a diagonal matrix. Now in this section we introduce another important diagonalization called the singular value decomposition. This diagonalization has advantage that it can be used for any $m \times n$ matrix.

For any $m \times n$ matrix A, A^TA and AA^T are symmetric and we know that a symmetric matrix can be orthogonally diagonalized. We define some basic things like the singular value and the rank of the matrix before defining the singular value decomposition.

Definition 9.2. The Singular Values

The positive square root of the eigenvalues of A^TA and AA^T are called the singular values and are denoted by $\sigma_1, \sigma_2, \cdots \sigma_n$ arranged in decreasing order i.e. $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0$.

Theorem 9.4. *If $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k \geq 0$ are the k non zero singular values of a $m \times n$ matrix with $\sigma_{k+1} = \cdots = \sigma_n = 0$, then the rank of A is k i.e rank A = k.*

Now the singular value decomposition can be constructed and is given as

$$A = U\Sigma V^T.$$

Where U is a $m \times m$ orthogonal matrix that satisfies $U^TU = 1$. The columns of U are the eigenvectors AA^T and these eigenvectors are called the left singular vectors. Similarly, V is a $n \times n$ matrix that satisfy $V^TV = 1$. The columns of V are the eigenvectors of A^TA and are called right singular vectors. Also, σ is a $m \times n$ diagonal matrix whose diagonal elements satisfies $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0$.

Now we state the following theorem to define the singular value decomposition.

Theorem 9.5. ***The Singular Value Decomposition*** *Let A be an arbitrary $m \times n$ matrix. Then $A = U\Sigma V^T$, where U is an orthogonal $m \times m$ matrix, V is an $n \times n$ orthogonal matrix and Σ is $m \times n$ diagonal matrix.*

In simple words the theorem 9.5 mean as follows. For an arbitrary $m \times n$ matrix A, consider a mapping $A : X \to Y$ where X and Y are Hilbert spaces with $\dim X = n$ and $\dim Y = m$. We can choose an orthonormal basis V in X and an orthonormal basis U in Y such that A maps a vector $x \in X$ to a vector $y = Ax \in Y$.

Example 9.3. Consider a 2×2 matrix

$$A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix},$$

we find a singular value decomposition of the matrix A. First, we find $A^T A$ which is given as

$$A^T A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 10 & 6 \\ 6 & 10 \end{bmatrix}.$$

The associated equation is

$$\lambda^2 - 20\lambda + 64 = 0,$$

solving the above equation for λ gives $\lambda_1 = 16$ and $\lambda_2 = 4$. The normalized eigenvector for $\lambda_1 = 16$ is

$$v_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

and the normalized eigenvector for $\lambda_2 = 4$ is

$$v_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

The singular values are $\sigma_1 = \sqrt{16} = 4$ and $\sigma_2 = \sqrt{4} = 2$. Now we find u_1 and u_2

$$u_1 = \frac{1}{\sigma_1} A v_1 = \frac{1}{4} \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

similarly,

$$u_2 = \frac{1}{\sigma_2} A v_2 = \frac{1}{2} \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}.$$

Finally

$$V = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix},$$

$$U = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix},$$

and

$$\sigma = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}.$$

Now we have all the information to built the singular value decomposition, hence the singular value decomposition of the matrix A is

$$A = U\Sigma V^T = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}^T$$

9.4 Application of Singular Value Decomposition

To see the application of the singular value decomposition, we consider an example of the image compression. In this example, we will try to show how the singular value decomposition works and some picture will be shown to illustrate. For this purpose we use the image of panda again with same resolution of 512×512. The image will be treated as a matrix and we will form its singular value decomposition.

First, we import the image of the Panda by using the command, in MATLAB

```
I=imread('panda.png');
I=double(I);
```

To find the singular value decomposition if the image matrix we use the command

```
[U,S,V]=svd(I);
```

Here I is the variable in which the image is stored. With the help of the above command, we can decompose the image matrix into three matrices, U, S and V. The matrices U, V are orthogonal matrices and the matrix S is a diagonal matrix. We consider an example to describe the matrices U, S and V.

Example 9.4. For simplicity we consider rather a small matrix then the actual size of image matrix. In this example, the size of the matrix is 8×8. The matrix I represents the gray scale image matrix. If we are working with colored pictures the image matrix will be stored into three different matrices. For gray image we work only with a single matrix and for colored image we have to do the same procedure with all three matrices.

$$I = \begin{bmatrix} 90 & 84 & 76 & 67 & 59 & 55 & 60 & 65 \\ 92 & 83 & 78 & 68 & 61 & 54 & 59 & 68 \\ 92 & 80 & 73 & 68 & 62 & 56 & 62 & 71 \\ 91 & 81 & 73 & 66 & 62 & 59 & 66 & 68 \\ 89 & 80 & 72 & 68 & 63 & 65 & 68 & 71 \\ 87 & 77 & 71 & 68 & 64 & 70 & 68 & 76 \\ 84 & 79 & 72 & 70 & 66 & 71 & 73 & 77 \\ 85 & 81 & 77 & 76 & 73 & 76 & 73 & 75 \end{bmatrix}.$$

To factor the image matrix we use the command $[U,S,V] = svd(I)$ in MATLAB. Where U consists of left singular vectors, the diagonal matrix S consists of singular values and the matrix V consists of right singular vectors. These matrices are given below.

$$U = \begin{bmatrix} -0.3426 & -0.4229 & 0.2954 & -0.2997 & 0.4768 & 0.0752 & -0.4228 & 0.3389 \\ -0.3471 & -0.4634 & 0.2265 & 0.3834 & 0.1915 & -0.0650 & 0.5210 & -0.3926 \\ -0.3470 & -0.2719 & -0.2706 & 0.4698 & -0.4857 & -0.2401 & -0.4445 & 0.1433 \\ -0.3478 & -0.1759 & -0.1855 & -0.5374 & -0.4260 & 0.0086 & 0.4890 & 0.3228 \\ -0.3528 & 0.0667 & -0.2036 & -0.3609 & -0.0893 & 0.2663 & -0.3034 & -0.7270 \\ -0.3548 & 0.2990 & -0.4132 & 0.3241 & 0.3061 & 0.5697 & 0.1330 & 0.2701 \\ -0.3607 & 0.4204 & -0.1916 & -0.0896 & 0.3482 & -0.7240 & 0.0513 & -0.0176 \\ -0.3746 & 0.4805 & 0.7105 & 0.1019 & -0.3093 & 0.1127 & -0.0277 & 0.0671 \end{bmatrix}$$

$$S = \begin{bmatrix} 580.7522 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 28.2405 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 8.7345 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 5.3787 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3.7269 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3.1732 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1.4564 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.3215 \end{bmatrix}$$

$$V = \begin{bmatrix} -0.4318 & -0.4821 & -0.4720 & -0.0327 & -0.3435 & 0.4719 & -0.0047 & -0.1285 \\ -0.3924 & -0.3364 & 0.1430 & -0.3799 & 0.3926 & -0.1566 & -0.3152 & 0.5386 \\ -0.3603 & -0.2720 & 0.4281 & 0.1135 & 0.3158 & -0.0290 & 0.6178 & -0.3432 \\ -0.3357 & 0.0306 & 0.3656 & 0.2669 & -0.6294 & -0.0657 & 0.2368 & 0.5141 \\ -0.3089 & 0.6282 & 0.0705 & -0.1059 & 0.2872 & 0.6397 & -0.0323 & 0.0316 \\ -0.3225 & 0.3346 & -0.2285 & -0.6154 & -0.1988 & -0.4497 & 0.1438 & -0.3012 \\ -0.3479 & 0.1982 & -0.5208 & 0.5857 & 0.2947 & -0.3449 & 0.0446 & 0.1328 \end{bmatrix}$$

To compress the result we need to find the best rank. By using the command in MATLB

$$\begin{aligned} U1 &= U(:,1:k); \\ S1 &= S(1:k,1:k); \\ V1 &= V(:,1:k)'; \end{aligned}$$

we can find $U1$, $S1$ and $V1$ for different values of k, where k is the rank of the matrix S. For instance, we choose $K = 5$ and for this value of k we have the following compressed matrices. The size of $U1$ is 8×5

$$U1 = \begin{bmatrix} -0.3426 & -0.4229 & 0.2954 & -0.2997 & 0.4768 \\ -0.3471 & -0.4634 & 0.2265 & 0.3834 & 0.1915 \\ -0.3470 & -0.2719 & -0.2706 & 0.4698 & -0.4857 \\ -0.3478 & -0.1759 & -0.1855 & -0.5374 & -0.4260 \\ -0.3528 & 0.0667 & -0.2036 & -0.3609 & -0.0893 \\ -0.3548 & 0.2990 & -0.4132 & 0.3241 & 0.3061 \\ -0.3607 & 0.4204 & -0.1916 & -0.0896 & 0.3482 \\ -0.3746 & 0.4805 & 0.7105 & 0.1019 & -0.3093 \end{bmatrix},$$

the size of the $S1$ is 5×5

$$S1 = \begin{bmatrix} 580.7522 & 0 & 0 & 0 & 0 \\ 0 & 28.2405 & 0 & 0 & 0 \\ 0 & 0 & 8.7345 & 0 & 0 \\ 0 & 0 & 0 & 5.3787 & 0 \\ 0 & 0 & 0 & 0 & 3.7269 \end{bmatrix},$$

and the size of the $V1$ is 8×5

$$V1 = \begin{bmatrix} -0.4318 & -0.3924 & -0.3603 & -0.3357 & -0.3109 & -0.3089 & -0.3225 & -0.3479 \\ -0.4821 & -0.3364 & -0.2720 & 0.0306 & 0.1833 & 0.6282 & 0.3346 & 0.1982 \\ -0.4720 & 0.1430 & 0.4281 & 0.3656 & 0.3338 & 0.0705 & -0.2285 & -0.5208 \\ -0.0327 & -0.3799 & 0.1135 & 0.2669 & 0.1936 & -0.1059 & -0.6154 & 0.5857 \\ -0.3435 & 0.3926 & 0.3158 & -0.1520 & -0.6294 & 0.2872 & -0.1988 & 0.2947 \end{bmatrix},$$

The following command is used to combining $U1$, $S1$ and $V1$ into a single matrix L.

$$L = U1 * S1 * V1;$$

The matrix L is compressed image matrix. We can observe that the pixels of this matrix are almost equal to original matrix but L matrix provides the less information than the original image matrix. The matrix L is given as below:

$$L = \begin{bmatrix} 89.8986 & 83.7846 & 76.4248 & 66.6722 & 59.1054 & 54.8241 & 60.2286 & 65.0953 \\ 92.0847 & 83.2748 & 77.4819 & 68.4190 & 60.8717 & 54.1605 & 58.7601 & 67.9117 \\ 92.3625 & 79.6519 & 73.3937 & 67.4917 & 62.0795 & 56.4651 & 61.7643 & 70.7599 \\ 91.0038 & 81.1728 & 72.5964 & 66.5222 & 61.7798 & 59.0022 & 65.9411 & 67.9639 \\ 88.5692 & 80.1189 & 72.2173 & 67.7130 & 63.2803 & 64.4527 & 68.3731 & 71.3422 \\ 86.1589 & 77.2973 & 70.9625 & 68.4046 & 64.0282 & 68.8471 & 68.8112 & 76.6034 \\ 85.0839 & 78.6669 & 71.8854 & 69.7454 & 65.8343 & 72.4723 & 71.9545 & 76.2050 \\ 84.8338 & 81.0316 & 77.0427 & 76.0299 & 73.0220 & 75.7692 & 73.1731 & 75.1223 \end{bmatrix}.$$

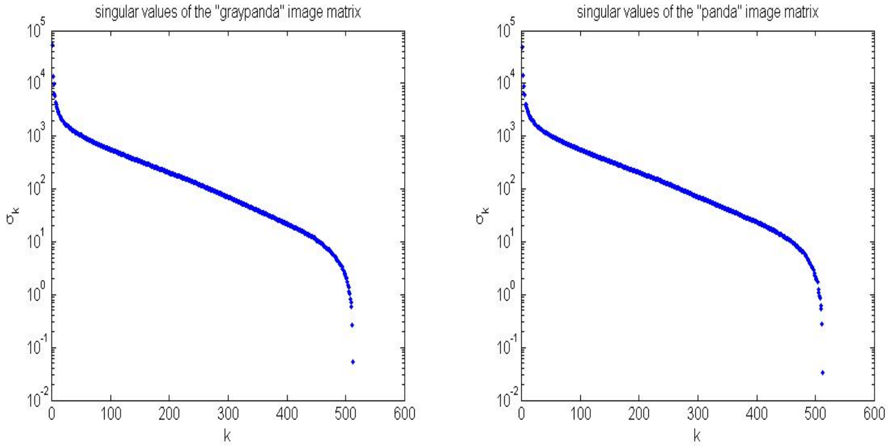
By choosing the small k we can observe that the difference between the matrices I and L is increasing, Therefore the distortion in image will also increase. So we can find the best approximation which provide best visualization.

9.5 Compressed Results

In this section we present compressed results using SVD. We compress the gray scale image and colored image, the results are shown below.

semilogy(diag(S))

First we plot the graph of the singular values of the image matrix.



(a) Singular Values of Gray Image (b) Singular Values of Colored Image

Figure 9.1: Singular Values of the Image Matrix

The figure 9.1 shows that the singular values are decline in magnitude. To calculate the compression degree, we used $1 - \frac{(2^j+2^j)\times k}{2^j \times 2^j}$, $j = 9$ and k is the rank. Again, we consider the picture of panda for compression using SVD method. In the pictures 9.2a and 9.2b, we use $k = 256$ largest singular values. We can observe, gray and color images, the compressed image is closer to the original image which is indistinguishable from the original image but it does not save the storage space.

(a) Gray Image (b) Colored Image

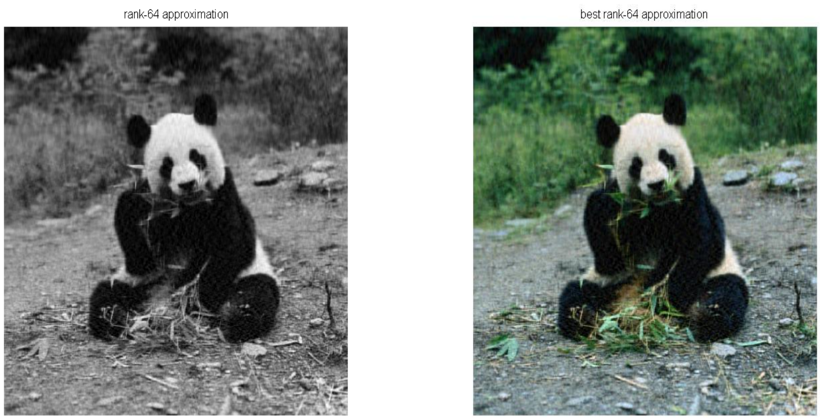
Figure 9.2: rank - 256 approximation

Now, we make a new matrix by choosing $k = 128$ largest singular values. We use this approximation to get a new compressed pictures, the pictures 9.3a and 9.3b are the pictures after other approximation. We can observe that the figure 9.3 has little distortion and human eye can visualize it, it requires 50% less storage space.

(a) Gray Image (b) Colored Image

Figure 9.3: rank - 128 approximation

Now we consider $K = 64$ The pictures 9.4a and 9.4b are the pictures after the approximation and we can observe that it has much distortion and it requires 25% less storage space.



(a) Gray Image (b) Colored Image

Figure 9.4: rank - 64 approximation

We did some experiments on the images and made some different approximations. In the first approximation we used only 256 largest singular values and then used 128 and 64 largest singular values and showed the results.

10 Conclusion

In this thesis we studied two different methods to compress the images. They are Haar wavelet and singular value decomposition. We explained the concepts of multiresolution analysis to construct the Haar wavelet basis, which is orthonormal. We used the Averaging and differencing technique as the application of Haar wavelet to compress the image. We discussed how the averaging and differencing technique is connected to the Haar wavelet. In the application of Haar wavelet we compressed the grey and colored images by considering the normalized and non-normalized Haar wavelet transform. The Haar wavelet transform is the simplest wavelet, which increases the speed of transferring the images and other electronic files. It also provides the better compressed results. We came to the conclusion that the normalized transformation is faster than the non normalized transformation because the rebuilt image is more closer to the original image and we have less distortion. We also concluded that for different pictures, the same threshold value gives us the different level of compression. The singular value decomposition (SVD) is also used to compress the images and it has many other applications too. The SVD can be good choice to get the better image result but it does not save much storage space. The Haar Wavelet transformation gives better results and the compressed data takes less storage space than the other methods.

References

[1] Greg Ames. Image compression. *Project Report Barcelona University*.

[2] James W. Demmel. *Applied Numerical linear Algebra*. Society for Industrial and Applied Mathematics, 1997.

[3] Ronald A. Devore and Brandley J. Lucier. Wavelets.

[4] Michael W. Frazier. *An Introduction To Wavelets Through Linear Algebra*. Springer, Third edition, 1999.

[5] Serge Lang. *Linear Algebra*. Springer, third edition, 1989.

[6] David C. Lay. *Linear Algebra and Its Applications*. Addison-Wesley Publishing Company, 1994.

[7] Kamrul Hasan Talukder and Koichi Harada. Haar wavelet based approach for image compression and quality assessment of compressed image. *IAENG International Journal of Applied Mathematics, 36:1, IJAM*$3 6_{19}$.

[8] David F. Walnut. *An Introduction To Wavelet Analysis*.

[9] Guido Wiess. *A First course On Wavelets*.

Linnæus University
School of Computer Science, Physics and Mathematics

SE-351 95 Växjö / SE-391 82 Kalmar
Tel +46-772-28 80 00
dfm@lnu.se
Lnu.se