

## Chapter 2

# Matrices and Linear Maps

### 2.1 Representation of Linear Maps by Matrices

Proposition 1.11 shows that given two vector spaces  $E$  and  $F$  and a basis  $(u_j)_{j \in J}$  of  $E$ , every linear map  $f: E \rightarrow F$  is uniquely determined by the family  $(f(u_j))_{j \in J}$  of the images under  $f$  of the vectors in the basis  $(u_j)_{j \in J}$ .

If we also have a basis  $(v_i)_{i \in I}$  of  $F$ , then every vector  $f(u_j)$  can be written in a unique way as

$$f(u_j) = \sum_{i \in I} a_{ij} v_i,$$

where  $j \in J$ , for a family of scalars  $(a_{ij})_{i \in I}$ .

Thus, with respect to the two bases  $(u_j)_{j \in J}$  of  $E$  and  $(v_i)_{i \in I}$  of  $F$ , the linear map  $f$  is completely determined by a “ $I \times J$ -matrix”

$$M(f) = (a_{ij})_{i \in I, j \in J}.$$

**Remark:** Note that we intentionally assigned the index set  $J$  to the basis  $(u_j)_{j \in J}$  of  $E$ , and the index  $I$  to the basis  $(v_i)_{i \in I}$  of  $F$ , so that the *rows* of the matrix  $M(f)$  associated with  $f: E \rightarrow F$  are indexed by  $I$ , and the *columns* of the matrix  $M(f)$  are indexed by  $J$ .

Obviously, this causes a mildly unpleasant reversal. If we had considered the bases  $(u_i)_{i \in I}$  of  $E$  and  $(v_j)_{j \in J}$  of  $F$ , we would obtain a  $J \times I$ -matrix  $M(f) = (a_{ji})_{j \in J, i \in I}$ .

No matter what we do, there will be a reversal! We decided to stick to the bases  $(u_j)_{j \in J}$  of  $E$  and  $(v_i)_{i \in I}$  of  $F$ , so that we get an  $I \times J$ -matrix  $M(f)$ , knowing that we may occasionally suffer from this decision!

When  $I$  and  $J$  are finite, and say, when  $|I| = m$  and  $|J| = n$ , the linear map  $f$  is determined by the matrix  $M(f)$  whose entries in the  $j$ -th column are the components of the vector  $f(u_j)$  over the basis  $(v_1, \dots, v_m)$ , that is, the matrix

$$M(f) = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

whose entry on row  $i$  and column  $j$  is  $a_{ij}$  ( $1 \leq i \leq m$ ,  $1 \leq j \leq n$ ).

We will now show that when  $E$  and  $F$  have finite dimension, linear maps can be very conveniently represented by matrices, and that composition of linear maps corresponds to matrix multiplication.

We will follow rather closely an elegant presentation method due to Emil Artin.

Let  $E$  and  $F$  be two vector spaces, and assume that  $E$  has a finite basis  $(u_1, \dots, u_n)$  and that  $F$  has a finite basis  $(v_1, \dots, v_m)$ . Recall that we have shown that every vector  $x \in E$  can be written in a unique way as

$$x = x_1u_1 + \cdots + x_nu_n,$$

and similarly every vector  $y \in F$  can be written in a unique way as

$$y = y_1v_1 + \cdots + y_mv_m.$$

Let  $f: E \rightarrow F$  be a linear map between  $E$  and  $F$ .

Then, for every  $x = x_1u_1 + \cdots + x_nu_n$  in  $E$ , by linearity, we have

$$f(x) = x_1f(u_1) + \cdots + x_nf(u_n).$$

Let

$$f(u_j) = a_{1j}v_1 + \cdots + a_{mj}v_m,$$

or more concisely,

$$f(u_j) = \sum_{i=1}^m a_{ij}v_i,$$

for every  $j$ ,  $1 \leq j \leq n$ .

This can be expressed by writing the coefficients  $a_{1j}, a_{2j}, \dots, a_{mj}$  of  $f(u_j)$  over the basis  $(v_1, \dots, v_m)$ , as the  $j$ th column of a matrix, as shown below:

$$\begin{array}{cccc} & f(u_1) & f(u_2) & \cdots & f(u_n) \\ v_1 & \left( a_{11} & a_{12} & \cdots & a_{1n} \right) \\ v_2 & a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ v_m & a_{m1} & a_{m2} & \cdots & a_{mn} \end{array}$$

Then, substituting the right-hand side of each  $f(u_j)$  into the expression for  $f(x)$ , we get

$$f(x) = x_1 \left( \sum_{i=1}^m a_{i1} v_i \right) + \cdots + x_n \left( \sum_{i=1}^m a_{in} v_i \right),$$

which, by regrouping terms to obtain a linear combination of the  $v_i$ , yields

$$f(x) = \left( \sum_{j=1}^n a_{1j} x_j \right) v_1 + \cdots + \left( \sum_{j=1}^n a_{mj} x_j \right) v_m.$$

Thus, letting  $f(x) = y = y_1 v_1 + \cdots + y_m v_m$ , we have

$$y_i = \sum_{j=1}^n a_{ij} x_j \tag{1}$$

for all  $i$ ,  $1 \leq i \leq m$ .

To make things more concrete, let us treat the case where  $n = 3$  and  $m = 2$ .

In this case,

$$\begin{aligned}f(u_1) &= a_{11}v_1 + a_{21}v_2 \\f(u_2) &= a_{12}v_1 + a_{22}v_2 \\f(u_3) &= a_{13}v_1 + a_{23}v_2,\end{aligned}$$

which in matrix form is expressed by

$$\begin{array}{c} f(u_1) \quad f(u_2) \quad f(u_3) \\ v_1 \left( \begin{array}{ccc} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{array} \right), \\ v_2 \end{array}$$

and for any  $x = x_1u_1 + x_2u_2 + x_3u_3$ , we have

$$\begin{aligned}f(x) &= f(x_1u_1 + x_2u_2 + x_3u_3) \\&= x_1f(u_1) + x_2f(u_2) + x_3f(u_3) \\&= x_1(a_{11}v_1 + a_{21}v_2) + x_2(a_{12}v_1 + a_{22}v_2) \\&\quad + x_3(a_{13}v_1 + a_{23}v_2) \\&= (a_{11}x_1 + a_{12}x_2 + a_{13}x_3)v_1 \\&\quad + (a_{21}x_1 + a_{22}x_2 + a_{23}x_3)v_2.\end{aligned}$$

Consequently, since

$$y = y_1v_1 + y_2v_2,$$

we have

$$\begin{aligned}y_1 &= a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \\y_2 &= a_{21}x_1 + a_{22}x_2 + a_{23}x_3.\end{aligned}$$

This agrees with the matrix equation

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

We now formalize the representation of linear maps by matrices.



**Definition 2.1.** Let  $E$  and  $F$  be two vector spaces, and let  $(u_1, \dots, u_n)$  be a basis for  $E$ , and  $(v_1, \dots, v_m)$  be a basis for  $F$ . Each vector  $x \in E$  expressed in the basis  $(u_1, \dots, u_n)$  as  $x = x_1u_1 + \dots + x_nu_n$  is represented by the column matrix

$$M(x) = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

and similarly for each vector  $y \in F$  expressed in the basis  $(v_1, \dots, v_m)$ . Every linear map  $f: E \rightarrow F$  is represented by the matrix  $M(f) = (a_{ij})$ , where  $a_{ij}$  is the  $i$ -th component of the vector  $f(u_j)$  over the basis  $(v_1, \dots, v_m)$ , i.e., where

$$f(u_j) = \sum_{i=1}^m a_{ij}v_i, \quad \text{for every } j, 1 \leq j \leq n.$$

The coefficients  $a_{1j}, a_{2j}, \dots, a_{mj}$  of  $f(u_j)$  over the basis  $(v_1, \dots, v_m)$  form the  $j$ th column of the matrix  $M(f)$  shown below:

$$\begin{array}{cccc} & f(u_1) & f(u_2) & \dots & f(u_n) \\ v_1 & \left( a_{11} & a_{12} & \dots & a_{1n} \right) \\ v_2 & \left( a_{21} & a_{22} & \dots & a_{2n} \right) \\ \vdots & \left( \vdots & \vdots & \ddots & \vdots \right) \\ v_m & \left( a_{m1} & a_{m2} & \dots & a_{mn} \right) \end{array} \cdot$$

The matrix  $M(f)$  associated with the linear map  $f: E \rightarrow F$  is called the *matrix of  $f$  with respect to the bases  $(u_1, \dots, u_n)$  and  $(v_1, \dots, v_m)$* .

When  $E = F$  and the basis  $(v_1, \dots, v_m)$  is identical to the basis  $(u_1, \dots, u_n)$  of  $E$ , the matrix  $M(f)$  associated with  $f: E \rightarrow E$  (as above) is called the *matrix of  $f$  with respect to the basis  $(u_1, \dots, u_n)$* .

**Remark:** As in the remark after Definition 1.6, there is no reason to assume that the vectors in the bases  $(u_1, \dots, u_n)$  and  $(v_1, \dots, v_m)$  are ordered in any particular way.

However, it is often convenient to assume the natural ordering. When this is so, authors sometimes refer to the matrix  $M(f)$  as the matrix of  $f$  with respect to the *ordered bases  $(u_1, \dots, u_n)$  and  $(v_1, \dots, v_m)$* .

Let us now consider how the composition of linear maps is expressed in terms of bases.

Let  $E$ ,  $F$ , and  $G$ , be three vectors spaces with respective bases  $(u_1, \dots, u_p)$  for  $E$ ,  $(v_1, \dots, v_n)$  for  $F$ , and  $(w_1, \dots, w_m)$  for  $G$ .

Let  $g: E \rightarrow F$  and  $f: F \rightarrow G$  be linear maps.

As explained earlier,  $g: E \rightarrow F$  is determined by the images of the basis vectors  $u_j$ , and  $f: F \rightarrow G$  is determined by the images of the basis vectors  $v_k$ .

We would like to understand how  $f \circ g: E \rightarrow G$  is determined by the images of the basis vectors  $u_j$ .

**Remark:** Note that we are considering linear maps  $g: E \rightarrow F$  and  $f: F \rightarrow G$ , instead of  $f: E \rightarrow F$  and  $g: F \rightarrow G$ , which yields the composition  $f \circ g: E \rightarrow G$  instead of  $g \circ f: E \rightarrow G$ .

Our perhaps unusual choice is motivated by the fact that if  $f$  is represented by a matrix  $M(f) = (a_{ik})$  and  $g$  is represented by a matrix  $M(g) = (b_{kj})$ , then  $f \circ g: E \rightarrow G$  is represented by the product  $AB$  of the matrices  $A$  and  $B$ .

If we had adopted the other choice where  $f: E \rightarrow F$  and  $g: F \rightarrow G$ , then  $g \circ f: E \rightarrow G$  would be represented by the product  $BA$ .

Obviously, this is a matter of taste! We will have to live with our perhaps unorthodox choice.

Thus, let

$$f(v_k) = \sum_{i=1}^m a_{i k} w_i,$$

for every  $k$ ,  $1 \leq k \leq n$ , and let

$$g(u_j) = \sum_{k=1}^n b_{k j} v_k,$$

for every  $j$ ,  $1 \leq j \leq p$ .

Also if

$$x = x_1 u_1 + \cdots + x_p u_p,$$

let

$$y = g(x)$$

and

$$z = f(y) = (f \circ g)(x),$$

with

$$y = y_1 v_1 + \cdots + y_n v_n$$

and

$$z = z_1 w_1 + \cdots + z_m w_m.$$

After some calculations, we get

$$z_i = \sum_{j=1}^p \left( \sum_{k=1}^n a_{i k} b_{k j} \right) x_j.$$

Thus, defining  $c_{i j}$  such that

$$c_{i j} = \sum_{k=1}^n a_{i k} b_{k j},$$

for  $1 \leq i \leq m$ , and  $1 \leq j \leq p$ , we have

$$z_i = \sum_{j=1}^p c_{i j} x_j \tag{4}$$

Identity (4) shows that the composition of linear maps corresponds to the product of matrices.

Then, given a linear map  $f: E \rightarrow F$  represented by the matrix  $M(f) = (a_{ij})$  w.r.t. the bases  $(u_1, \dots, u_n)$  and  $(v_1, \dots, v_m)$ , by equations (1), namely

$$y_i = \sum_{j=1}^n a_{ij} x_j \quad 1 \leq i \leq m,$$

and the definition of matrix multiplication, *the equation  $y = f(x)$  corresponds to the matrix equation  $M(y) = M(f)M(x)$* , that is,

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Recall that

$$\begin{aligned} & \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \\ &= x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} + \cdots + x_n \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}. \end{aligned}$$

Sometimes, it is necessary to incorporate the bases  $(u_1, \dots, u_n)$  and  $(v_1, \dots, v_m)$  in the notation for the matrix  $M(f)$  expressing  $f$  with respect to these bases. This turns out to be a messy enterprise!

We propose the following course of action:

**Definition 2.2.** Write  $\mathcal{U} = (u_1, \dots, u_n)$  and  $\mathcal{V} = (v_1, \dots, v_m)$  for the bases of  $E$  and  $F$ , and denote by

$$M_{\mathcal{U}, \mathcal{V}}(f)$$

the *matrix of  $f$  with respect to the bases  $\mathcal{U}$  and  $\mathcal{V}$* .

Furthermore, write  $x_{\mathcal{U}}$  for the coordinates  $M(x) = (x_1, \dots, x_n)$  of  $x \in E$  w.r.t. the basis  $\mathcal{U}$  and write  $y_{\mathcal{V}}$  for the coordinates  $M(y) = (y_1, \dots, y_m)$  of  $y \in F$  w.r.t. the basis  $\mathcal{V}$ . Then,

$$y = f(x)$$

is expressed in matrix form by

$$y_{\mathcal{V}} = M_{\mathcal{U}, \mathcal{V}}(f) x_{\mathcal{U}}.$$

When  $\mathcal{U} = \mathcal{V}$ , we abbreviate  $M_{\mathcal{U}, \mathcal{V}}(f)$  as  $M_{\mathcal{U}}(f)$ .



The above notation seems reasonable, but it has the slight disadvantage that in the expression  $M_{\mathcal{U},\mathcal{V}}(f)x_{\mathcal{U}}$ , the input argument  $x_{\mathcal{U}}$  which is fed to the matrix  $M_{\mathcal{U},\mathcal{V}}(f)$  does not appear next to the subscript  $\mathcal{U}$  in  $M_{\mathcal{U},\mathcal{V}}(f)$ .

We could have used the notation  $M_{\mathcal{V},\mathcal{U}}(f)$ , and some people do that. But then, we find a bit confusing that  $\mathcal{V}$  comes before  $\mathcal{U}$  when  $f$  maps from the space  $E$  with the basis  $\mathcal{U}$  to the space  $F$  with the basis  $\mathcal{V}$ .

So, we prefer to use the notation  $M_{\mathcal{U},\mathcal{V}}(f)$ .

Be aware that other authors such as Meyer [27] use the notation  $[f]_{\mathcal{U},\mathcal{V}}$ , and others such as Dummit and Foote [13] use the notation  $M_{\mathcal{U}}^{\mathcal{V}}(f)$ , instead of  $M_{\mathcal{U},\mathcal{V}}(f)$ .

This gets worse! You may find the notation  $M_{\mathcal{V}}^{\mathcal{U}}(f)$  (as in Lang [23]), or  ${}_{\mathcal{U}}[f]_{\mathcal{V}}$ , or other strange notations.

Let us illustrate the representation of a linear map by a matrix in a concrete situation.

Let  $E$  be the vector space  $\mathbb{R}[X]_4$  of polynomials of degree at most 4, let  $F$  be the vector space  $\mathbb{R}[X]_3$  of polynomials of degree at most 3, and let the linear map be the derivative map  $d$ : that is,

$$\begin{aligned}d(P + Q) &= dP + dQ \\d(\lambda P) &= \lambda dP,\end{aligned}$$

with  $\lambda \in \mathbb{R}$ .

We choose  $(1, x, x^2, x^3, x^4)$  as a basis of  $E$  and  $(1, x, x^2, x^3)$  as a basis of  $F$ .

Then, the  $4 \times 5$  matrix  $D$  associated with  $d$  is obtained by expressing the derivative  $dx^i$  of each basis vector  $x^i$  for  $i = 0, 1, 2, 3, 4$  over the basis  $(1, x, x^2, x^3)$ .

We find

$$D = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{pmatrix}.$$

Then, if  $P$  denotes the polynomial

$$P = 3x^4 - 5x^3 + x^2 - 7x + 5,$$

we have

$$dP = 12x^3 - 15x^2 + 2x - 7,$$

the polynomial  $P$  is represented by the vector  $(5, -7, 1, -5, 3)$  and  $dP$  is represented by the vector  $(-7, 2, -15, 12)$ , and we have

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} 5 \\ -7 \\ 1 \\ -5 \\ 3 \end{pmatrix} = \begin{pmatrix} -7 \\ 2 \\ -15 \\ 12 \end{pmatrix},$$

as expected!

The kernel (nullspace) of  $d$  consists of the polynomials of degree 0, that is, the constant polynomials.

Therefore  $\dim(\text{Ker } d) = 1$ , and from

$$\dim(E) = \dim(\text{Ker } d) + \dim(\text{Im } d)$$

(see Theorem 3.6), we get  $\dim(\text{Im } d) = 4$   
(since  $\dim(E) = 5$ ).

For fun, let us figure out the linear map from the vector space  $\mathbb{R}[X]_3$  to the vector space  $\mathbb{R}[X]_4$  given by integration (finding the primitive, or anti-derivative) of  $x^i$ , for  $i = 0, 1, 2, 3$ ).

The  $5 \times 4$  matrix  $S$  representing  $\int$  with respect to the same bases as before is

$$S = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 \\ 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/4 \end{pmatrix}.$$

We verify that  $DS = I_4$ ,

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 \\ 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/4 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

as it should!

The equation  $DS = I_4$  show that  $S$  is injective and has  $D$  as a left inverse. However,  $SD \neq I_5$ , and instead

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 \\ 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/4 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

because constant polynomials (polynomials of degree 0) belong to the kernel of  $D$ .

The function that associates to a linear map  $f: E \rightarrow F$  the matrix  $M(f)$  w.r.t. the bases  $(u_1, \dots, u_n)$  and  $(v_1, \dots, v_m)$  has the property that matrix multiplication corresponds to composition of linear maps.

This allows us to transfer properties of linear maps to matrices.

**Proposition 2.1.** (1) *Given any matrices  $A \in M_{m,n}(K)$ ,  $B \in M_{n,p}(K)$ , and  $C \in M_{p,q}(K)$ , we have*

$$(AB)C = A(BC);$$

*that is, matrix multiplication is associative.*

(2) *Given any matrices  $A, B \in M_{m,n}(K)$ , and  $C, D \in M_{n,p}(K)$ , for all  $\lambda \in K$ , we have*

$$(A + B)C = AC + BC$$

$$A(C + D) = AC + AD$$

$$(\lambda A)C = \lambda(AC)$$

$$A(\lambda C) = \lambda(AC),$$

*so that matrix multiplication*

*$\cdot: M_{m,n}(K) \times M_{n,p}(K) \rightarrow M_{m,p}(K)$  is bilinear.*

Note that Proposition 2.1 implies that the vector space  $M_n(K)$  of square matrices is a (noncommutative) *ring* with unit  $I_n$ .

The following proposition states the main properties of the mapping  $f \mapsto M(f)$  between  $\text{Hom}(E, F)$  and  $M_{m,n}$ .

In short, it is an isomorphism of vector spaces.

**Proposition 2.2.** *Given three vector spaces  $E$ ,  $F$ ,  $G$ , with respective bases  $(u_1, \dots, u_p)$ ,  $(v_1, \dots, v_n)$ , and  $(w_1, \dots, w_m)$ , the mapping  $M: \text{Hom}(E, F) \rightarrow M_{n,p}$  that associates the matrix  $M(g)$  to a linear map  $g: E \rightarrow F$  satisfies the following properties for all  $x \in E$ , all  $g, h: E \rightarrow F$ , and all  $f: F \rightarrow G$ :*

$$\begin{aligned} M(g(x)) &= M(g)M(x) \\ M(g + h) &= M(g) + M(h) \\ M(\lambda g) &= \lambda M(g) \\ M(f \circ g) &= M(f)M(g). \end{aligned}$$

where  $M(x)$  is the column vector associated with the vector  $x$  and  $M(g(x))$  is the column vector associated with  $g(x)$ , as explained in Definition 2.1.

Thus,  $M: \text{Hom}(E, F) \rightarrow M_{n,p}$  is an isomorphism of vector spaces, and when  $p = n$  and the basis  $(v_1, \dots, v_n)$  is identical to the basis  $(u_1, \dots, u_p)$ ,  $M: \text{Hom}(E, E) \rightarrow M_n$  is an isomorphism of rings.



In view of Proposition 2.2, it seems preferable to represent vectors from a vector space of finite dimension as column vectors rather than row vectors.

Thus, from now on, we will denote vectors of  $\mathbb{R}^n$  (or more generally, of  $K^n$ ) as column vectors.

## 2.2 Change of Basis Matrix

It is important to observe that the isomorphism  $M: \text{Hom}(E, F) \rightarrow M_{n,p}$  given by Proposition 2.2 depends on the choice of the bases  $(u_1, \dots, u_p)$  and  $(v_1, \dots, v_n)$ , and similarly for the isomorphism  $M: \text{Hom}(E, E) \rightarrow M_n$ , which depends on the choice of the basis  $(u_1, \dots, u_n)$ .

Thus, it would be useful to know how a change of basis affects the representation of a linear map  $f: E \rightarrow F$  as a matrix.

**Proposition 2.3.** *Let  $E$  be a vector space, and let  $(u_1, \dots, u_n)$  be a basis of  $E$ . For every family  $(v_1, \dots, v_n)$ , let  $P = (a_{ij})$  be the matrix defined such that  $v_j = \sum_{i=1}^n a_{ij}u_i$ . The matrix  $P$  is invertible iff  $(v_1, \dots, v_n)$  is a basis of  $E$ .*

**Definition 2.3.** Given a vector space  $E$  of dimension  $n$ , for any two bases  $(u_1, \dots, u_n)$  and  $(v_1, \dots, v_n)$  of  $E$ , let  $P = (a_{ij})$  be the invertible matrix defined such that

$$v_j = \sum_{i=1}^n a_{ij} u_i,$$

which is also the matrix of the identity  $\text{id}: E \rightarrow E$  with respect to the bases  $(v_1, \dots, v_n)$  and  $(u_1, \dots, u_n)$ , *in that order*. Indeed, we express each  $\text{id}(v_j) = v_j$  over the basis  $(u_1, \dots, u_n)$ . The coefficients  $a_{1j}, a_{2j}, \dots, a_{nj}$  of  $v_j$  over the basis  $(u_1, \dots, u_n)$  form the  $j$ th column of the matrix  $P$  shown below:

$$\begin{array}{cccc} & v_1 & v_2 & \dots & v_n \\ \begin{array}{c} u_1 \\ u_2 \\ \vdots \\ u_n \end{array} & \left( \begin{array}{cccc} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{array} \right) & & \end{array}.$$

The matrix  $P$  is called the *change of basis matrix from  $(u_1, \dots, u_n)$  to  $(v_1, \dots, v_n)$* .

Clearly, the change of basis matrix from  $(v_1, \dots, v_n)$  to  $(u_1, \dots, u_n)$  is  $P^{-1}$ .

Since  $P = (a_{ij})$  is the matrix of the identity  $\text{id}: E \rightarrow E$  with respect to the bases  $(v_1, \dots, v_n)$  and  $(u_1, \dots, u_n)$ , given any vector  $x \in E$ , if  $x = x_1u_1 + \dots + x_nu_n$  over the basis  $(u_1, \dots, u_n)$  and  $x = x'_1v_1 + \dots + x'_nv_n$  over the basis  $(v_1, \dots, v_n)$ , from Proposition 2.2, we have

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix}$$

showing that the *old* coordinates  $(x_i)$  of  $x$  (over  $(u_1, \dots, u_n)$ ) are expressed in terms of the *new* coordinates  $(x'_i)$  of  $x$  (over  $(v_1, \dots, v_n)$ ).

Now we face the painful task of assigning a “good” notation incorporating the bases  $\mathcal{U} = (u_1, \dots, u_n)$  and  $\mathcal{V} = (v_1, \dots, v_n)$  into the notation for the change of basis matrix from  $\mathcal{U}$  to  $\mathcal{V}$ .

Because the change of basis matrix from  $\mathcal{U}$  to  $\mathcal{V}$  is the matrix of the identity map  $\text{id}_E$  *with respect to the bases  $\mathcal{V}$  and  $\mathcal{U}$  in that order*, we could denote it by  $M_{\mathcal{V},\mathcal{U}}(\text{id})$  (Meyer [27] uses the notation  $[I]_{\mathcal{V},\mathcal{U}}$ ).

We prefer to use an abbreviation for  $M_{\mathcal{V},\mathcal{U}}(\text{id})$  and we will use the notation

$$P_{\mathcal{V},\mathcal{U}}$$

for the *change of basis matrix from  $\mathcal{U}$  to  $\mathcal{V}$* .

Note that

$$P_{\mathcal{U},\mathcal{V}} = P_{\mathcal{V},\mathcal{U}}^{-1}.$$

Then, if we write  $x_{\mathcal{U}} = (x_1, \dots, x_n)$  for the *old* coordinates of  $x$  with respect to the basis  $\mathcal{U}$  and  $x_{\mathcal{V}} = (x'_1, \dots, x'_n)$  for the *new* coordinates of  $x$  with respect to the basis  $\mathcal{V}$ , we have

$$x_{\mathcal{U}} = P_{\mathcal{V},\mathcal{U}} x_{\mathcal{V}}, \quad x_{\mathcal{V}} = P_{\mathcal{V},\mathcal{U}}^{-1} x_{\mathcal{U}}.$$

The above may look backward, but remember that the matrix  $M_{\mathcal{U},\mathcal{V}}(f)$  takes input expressed over the basis  $\mathcal{U}$  to output expressed over the basis  $\mathcal{V}$ .

Consequently,  $P_{\mathcal{V},\mathcal{U}}$  takes input expressed over the basis  $\mathcal{V}$  to output expressed over the basis  $\mathcal{U}$ , and  $x_{\mathcal{U}} = P_{\mathcal{V},\mathcal{U}} x_{\mathcal{V}}$  matches this point of view!



Beware that some authors (such as Artin [1]) define the change of basis matrix from  $\mathcal{U}$  to  $\mathcal{V}$  as  $P_{\mathcal{U},\mathcal{V}} = P_{\mathcal{V},\mathcal{U}}^{-1}$ . Under this point of view, the old basis  $\mathcal{U}$  is expressed in terms of the new basis  $\mathcal{V}$ . We find this a bit unnatural.

Also, in practice, it seems that the new basis is often expressed in terms of the old basis, rather than the other way around.

Since the matrix  $P = P_{\mathcal{V},\mathcal{U}}$  expresses the *new* basis  $(v_1, \dots, v_n)$  in terms of the *old* basis  $(u_1, \dots, u_n)$ , we observe that the coordinates  $(x_i)$  of a vector  $x$  vary in the *opposite direction* of the change of basis.

For this reason, vectors are sometimes said to be *contravariant*.

However, this expression does not make sense! Indeed, a vector is an intrinsic quantity that does not depend on a specific basis.

What makes sense is that the *coordinates* of a vector vary in a contravariant fashion.

Let us consider some concrete examples of change of bases.

**Example 2.1.** Let  $E = F = \mathbb{R}^2$ , with  $u_1 = (1, 0)$ ,  $u_2 = (0, 1)$ ,  $v_1 = (1, 1)$  and  $v_2 = (-1, 1)$ .

The change of basis matrix  $P$  from the basis  $\mathcal{U} = (u_1, u_2)$  to the basis  $\mathcal{V} = (v_1, v_2)$  is

$$P = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

and its inverse is

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

The old coordinates  $(x_1, x_2)$  with respect to  $(u_1, u_2)$  are expressed in terms of the new coordinates  $(x'_1, x'_2)$  with respect to  $(v_1, v_2)$  by

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x'_1 \\ x'_2 \end{pmatrix},$$

and the new coordinates  $(x'_1, x'_2)$  with respect to  $(v_1, v_2)$  are expressed in terms of the old coordinates  $(x_1, x_2)$  with respect to  $(u_1, u_2)$  by

$$\begin{pmatrix} x'_1 \\ x'_2 \end{pmatrix} = \begin{pmatrix} 1/2 & 1/2 \\ -1/2 & 1/2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$



**Example 2.2.** Let  $E = F = \mathbb{R}[X]_3$  be the set of polynomials of degree at most 3, and consider the bases  $\mathcal{U} = (1, x, x^2, x^3)$  and  $\mathcal{V} = (B_0^3(x), B_1^3(x), B_2^3(x), B_3^3(x))$ , where  $B_0^3(x), B_1^3(x), B_2^3(x), B_3^3(x)$  are the *Bernstein polynomials* of degree 3, given by

$$\begin{aligned} B_0^3(x) &= (1-x)^3 & B_1^3(x) &= 3(1-x)^2x \\ B_2^3(x) &= 3(1-x)x^2 & B_3^3(x) &= x^3. \end{aligned}$$

By expanding the Bernstein polynomials, we find that the change of basis matrix  $P_{\mathcal{V},\mathcal{U}}$  is given by

$$P_{\mathcal{V},\mathcal{U}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -3 & 3 & 0 & 0 \\ 3 & -6 & 3 & 0 \\ -1 & 3 & -3 & 1 \end{pmatrix}.$$

We also find that the inverse of  $P_{\mathcal{V},\mathcal{U}}$  is

$$P_{\mathcal{V},\mathcal{U}}^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1/3 & 0 & 0 \\ 1 & 2/3 & 1/3 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}.$$

Therefore, the coordinates of the polynomial  $2x^3 - x + 1$  over the basis  $\mathcal{V}$  are

$$\begin{pmatrix} 1 \\ 2/3 \\ 1/3 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1/3 & 0 & 0 \\ 1 & 2/3 & 1/3 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ 0 \\ 2 \end{pmatrix},$$

and so

$$2x^3 - x + 1 = B_0^3(x) + \frac{2}{3}B_1^3(x) + \frac{1}{3}B_2^3(x) + 2B_3^3(x).$$

Our next example is the Haar wavelets, a fundamental tool in signal processing.

## 2.3 Haar Basis Vectors and a Glimpse at Wavelets

We begin by considering *Haar wavelets* in  $\mathbb{R}^4$ .

Wavelets play an important role in audio and video signal processing, especially for *compressing* long signals into much smaller ones that still retain enough information so that when they are played, we can't see or hear any difference.

Consider the four vectors  $w_1, w_2, w_3, w_4$  given by

$$w_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \quad w_2 = \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix} \quad w_3 = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix} \quad w_4 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix}.$$

Note that these vectors are pairwise orthogonal, so they are indeed linearly independent (we will see this in a later chapter).

Let  $\mathcal{W} = \{w_1, w_2, w_3, w_4\}$  be the *Haar basis*, and let  $\mathcal{U} = \{e_1, e_2, e_3, e_4\}$  be the canonical basis of  $\mathbb{R}^4$ .

The change of basis matrix  $W = P_{\mathcal{W}, \mathcal{U}}$  from  $\mathcal{U}$  to  $\mathcal{W}$  is given by

$$W = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & -1 & 0 \\ 1 & -1 & 0 & 1 \\ 1 & -1 & 0 & -1 \end{pmatrix},$$

and we easily find that the inverse of  $W$  is given by

$$W^{-1} = \begin{pmatrix} 1/4 & 0 & 0 & 0 \\ 0 & 1/4 & 0 & 0 \\ 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1/2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}.$$

So, the vector  $v = (6, 4, 5, 1)$  over the basis  $\mathcal{U}$  becomes  $c = (c_1, c_2, c_3, c_4) = (4, 1, 1, 2)$  over the Haar basis  $\mathcal{W}$ , with

$$\begin{pmatrix} 4 \\ 1 \\ 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1/4 & 0 & 0 & 0 \\ 0 & 1/4 & 0 & 0 \\ 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1/2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} 6 \\ 4 \\ 5 \\ 1 \end{pmatrix}.$$

Given a signal  $v = (v_1, v_2, v_3, v_4)$ , we first *transform*  $v$  into its coefficients  $c = (c_1, c_2, c_3, c_4)$  over the Haar basis by computing  $c = W^{-1}v$ . Observe that

$$c_1 = \frac{v_1 + v_2 + v_3 + v_4}{4}$$

is the overall *average* value of the signal  $v$ . The coefficient  $c_1$  corresponds to the background of the image (or of the sound).

Then,  $c_2$  gives the coarse details of  $v$ , whereas,  $c_3$  gives the details in the first part of  $v$ , and  $c_4$  gives the details in the second half of  $v$ .

*Reconstruction* of the signal consists in computing  $v = Wc$ .

The trick for good *compression* is to throw away some of the coefficients of  $c$  (set them to zero), obtaining a *compressed signal*  $\hat{c}$ , and still retain enough crucial information so that the reconstructed signal  $\hat{v} = W\hat{c}$  looks almost as good as the original signal  $v$ .

Thus, the steps are:

$$\begin{aligned} \text{input } v &\longrightarrow \text{coefficients } c = W^{-1}v \longrightarrow \text{compressed } \hat{c} \\ &\longrightarrow \text{compressed } \hat{v} = W\hat{c}. \end{aligned}$$

This kind of compression scheme makes modern video conferencing possible.

It turns out that there is a faster way to find  $c = W^{-1}v$ , without actually using  $W^{-1}$ . This has to do with the multiscale nature of Haar wavelets.

Given the original signal  $v = (6, 4, 5, 1)$  shown in Figure 2.1, we compute averages and half differences obtaining

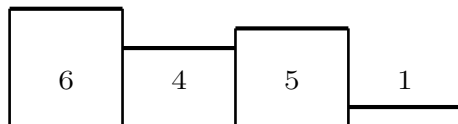


Figure 2.1: The original signal  $v$

Figure 2.2: We get the coefficients  $c_3 = 1$  and  $c_4 = 2$ .



Figure 2.2: First averages and first half differences

Note that the original signal  $v$  can be reconstructed from the two signals in Figure 2.2.

Then, again we compute averages and half differences obtaining Figure 2.3.

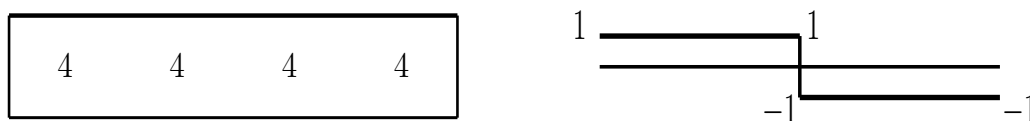


Figure 2.3: Second averages and second half differences

We get the coefficients  $c_1 = 4$  and  $c_2 = 1$ .

Again, the signal on the left of Figure 2.2 can be reconstructed from the two signals in Figure 2.3.

This method can be generalized to signals of any length  $2^n$ . The previous case corresponds to  $n = 2$ .

Let us consider the case  $n = 3$ .

The Haar basis  $(w_1, w_2, w_3, w_4, w_5, w_6, w_7, w_8)$  is given by the matrix

$$W = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & -1 & 0 & 0 & 0 \\ 1 & 1 & -1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & -1 & 0 & 0 & -1 & 0 & 0 \\ 1 & -1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & -1 & 0 & 1 & 0 & 0 & -1 & 0 \\ 1 & -1 & 0 & -1 & 0 & 0 & 0 & 1 \\ 1 & -1 & 0 & -1 & 0 & 0 & 0 & -1 \end{pmatrix}$$



The columns of this matrix are orthogonal and it is easy to see that

$$W^{-1} = \text{diag}(1/8, 1/8, 1/4, 1/4, 1/2, 1/2, 1/2, 1/2)W^{\top}.$$

A pattern is beginning to emerge. It looks like the second Haar basis vector  $w_2$  is the “mother” of all the other basis vectors, except the first, whose purpose is to perform averaging.

Indeed, in general, given

$$w_2 = \underbrace{(1, \dots, 1, -1, \dots, -1)}_{2^n},$$

the other Haar basis vectors are obtained by a “scaling and shifting process.”

Starting from  $w_2$ , the scaling process generates the vectors

$$w_3, w_5, w_9, \dots, w_{2^j+1}, \dots, w_{2^{n-1}+1},$$

such that  $w_{2^{j+1}+1}$  is obtained from  $w_{2^j+1}$  by forming two consecutive blocks of 1 and  $-1$  of half the size of the blocks in  $w_{2^j+1}$ , and setting all other entries to zero. Observe that  $w_{2^j+1}$  has  $2^j$  blocks of  $2^{n-j}$  elements.

The shifting process, consists in shifting the blocks of 1 and  $-1$  in  $w_{2^j+1}$  to the right by inserting a block of  $(k-1)2^{n-j}$  zeros from the left, with  $0 \leq j \leq n-1$  and  $1 \leq k \leq 2^j$ .

Thus, we obtain the following formula for  $w_{2^j+k}$ :

$$w_{2^j+k}(i) = \begin{cases} 0 & 1 \leq i \leq (k-1)2^{n-j} \\ 1 & (k-1)2^{n-j} + 1 \leq i \leq (k-1)2^{n-j} + 2^{n-j-1} \\ -1 & (k-1)2^{n-j} + 2^{n-j-1} + 1 \leq i \leq k2^{n-j} \\ 0 & k2^{n-j} + 1 \leq i \leq 2^n, \end{cases}$$

with  $0 \leq j \leq n-1$  and  $1 \leq k \leq 2^j$ .

Of course

$$w_1 = \underbrace{(1, \dots, 1)}_{2^n}.$$

The above formulae look a little better if we change our indexing slightly by letting  $k$  vary from 0 to  $2^j - 1$  and using the index  $j$  instead of  $2^j$ .

In this case, the Haar basis is denoted by

$$w_1, h_0^0, h_0^1, h_1^1, h_0^2, h_1^2, h_2^2, h_3^2, \dots, h_k^j, \dots, h_{2^{n-1}-1}^{n-1},$$

and

$$h_k^j(i) = \begin{cases} 0 & 1 \leq i \leq k2^{n-j} \\ 1 & k2^{n-j} + 1 \leq i \leq k2^{n-j} + 2^{n-j-1} \\ -1 & k2^{n-j} + 2^{n-j-1} + 1 \leq i \leq (k+1)2^{n-j} \\ 0 & (k+1)2^{n-j} + 1 \leq i \leq 2^n, \end{cases}$$

with  $0 \leq j \leq n - 1$  and  $0 \leq k \leq 2^j - 1$ .

It turns out that there is a way to understand these formulae better if we interpret a vector  $u = (u_1, \dots, u_m)$  as a piecewise linear function over the interval  $[0, 1)$ .

We define the function  $\text{plf}(u)$  such that

$$\text{plf}(u)(x) = u_i, \quad \frac{i-1}{m} \leq x < \frac{i}{m}, \quad 1 \leq i \leq m.$$

In words, the function  $\text{plf}(u)$  has the value  $u_1$  on the interval  $[0, 1/m)$ , the value  $u_2$  on  $[1/m, 2/m)$ , etc., and the value  $u_m$  on the interval  $[(m-1)/m, 1)$ .

For example, the piecewise linear function associated with the vector

$$u = (2.4, 2.2, 2.15, 2.05, 6.8, 2.8, -1.1, -1.3)$$

is shown in Figure 2.4.

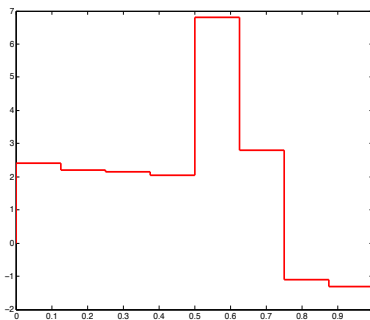


Figure 2.4: The piecewise linear function  $\text{plf}(u)$

Then, each basis vector  $h_k^j$  corresponds to the function

$$\psi_k^j = \text{plf}(h_k^j).$$

In particular, for all  $n$ , the Haar basis vectors

$$h_0^0 = w_2 = \underbrace{(1, \dots, 1, -1, \dots, -1)}_{2^n}$$

yield the same piecewise linear function  $\psi$  given by

$$\psi(x) = \begin{cases} 1 & \text{if } 0 \leq x < 1/2 \\ -1 & \text{if } 1/2 \leq x < 1 \\ 0 & \text{otherwise,} \end{cases}$$

whose graph is shown in Figure 2.5.

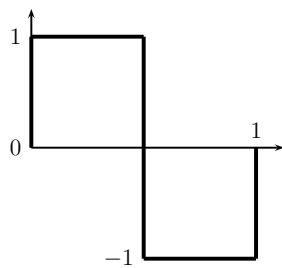


Figure 2.5: The Haar wavelet  $\psi$

Then, it is easy to see that  $\psi_k^j$  is given by the simple expression

$$\psi_k^j(x) = \psi(2^j x - k), \quad 0 \leq j \leq n - 1, \quad 0 \leq k \leq 2^j - 1.$$

The above formula makes it clear that  $\psi_k^j$  is obtained from  $\psi$  by scaling and shifting.

The function  $\phi_0^0 = \text{plf}(w_1)$  is the piecewise linear function with the constant value 1 on  $[0, 1)$ , and the functions  $\psi_k^j$  together with  $\phi_0^0$  are known as the *Haar wavelets*.

Rather than using  $W^{-1}$  to convert a vector  $u$  to a vector  $c$  of coefficients over the Haar basis, and the matrix  $W$  to reconstruct the vector  $u$  from its Haar coefficients  $c$ , we can use faster algorithms that use averaging and differencing.

If  $c$  is a vector of Haar coefficients of dimension  $2^n$ , we compute the sequence of vectors  $u^0, u^1, \dots, u^n$  as follows:

$$\begin{aligned} u^0 &= c \\ u^{j+1} &= u^j \\ u^{j+1}(2i-1) &= u^j(i) + u^j(2^j + i) \\ u^{j+1}(2i) &= u^j(i) - u^j(2^j + i), \end{aligned}$$

for  $j = 0, \dots, n-1$  and  $i = 1, \dots, 2^j$ .

The reconstructed vector (signal) is  $u = u^n$ .

If  $u$  is a vector of dimension  $2^n$ , we compute the sequence of vectors  $c^n, c^{n-1}, \dots, c^0$  as follows:

$$\begin{aligned} c^n &= u \\ c^j &= c^{j+1} \\ c^j(i) &= (c^{j+1}(2i-1) + c^{j+1}(2i))/2 \\ c^j(2^j + i) &= (c^{j+1}(2i-1) - c^{j+1}(2i))/2, \end{aligned}$$

for  $j = n-1, \dots, 0$  and  $i = 1, \dots, 2^j$ .

The vector over the Haar basis is  $c = c^0$ .

Here is an example of the conversion of a vector to its Haar coefficients for  $n = 3$ .

Given the sequence  $u = (31, 29, 23, 17, -6, -8, -2, -4)$ , we get the sequence

$$c^3 = (31, 29, 23, 17, -6, -8, -2, -4)$$

$$c^2 = (30, 20, -7, -3, 1, 3, 1, 1)$$

$$c^1 = (25, -5, 5, -2, 1, 3, 1, 1)$$

$$c^0 = (10, 15, 5, -2, 1, 3, 1, 1).$$

Conversely, given  $c = (10, 15, 5, -2, 1, 3, 1, 1)$ , we get the sequence

$$u^0 = (10, 15, 5, -2, 1, 3, 1, 1),$$

$$u^1 = (25, -5, 5, -2, 1, 3, 1, 1)$$

$$u^2 = (30, 20, -7, -3, 1, 3, 1, 1)$$

$$u^3 = (31, 29, 23, 17, -6, -8, -2, -4),$$

which gives back  $u = (31, 29, 23, 17, -6, -8, -2, -4)$ .



An important and attractive feature of the Haar basis is that it provides a *multiresolution analysis* of a signal.

Indeed, given a signal  $u$ , if  $c = (c_1, \dots, c_{2^n})$  is the vector of its Haar coefficients, the coefficients with low index give coarse information about  $u$ , and the coefficients with high index represent fine information.

This multiresolution feature of wavelets can be exploited to compress a signal, that is, to use fewer coefficients to represent it. Here is an example.

Consider the signal

$$u = (2.4, 2.2, 2.15, 2.05, 6.8, 2.8, -1.1, -1.3),$$

whose Haar transform is

$$c = (2, 0.2, 0.1, 3, 0.1, 0.05, 2, 0.1).$$

The piecewise-linear curves corresponding to  $u$  and  $c$  are shown in Figure 2.6.

Since some of the coefficients in  $c$  are small (smaller than or equal to 0.2) we can compress  $c$  by replacing them by 0.

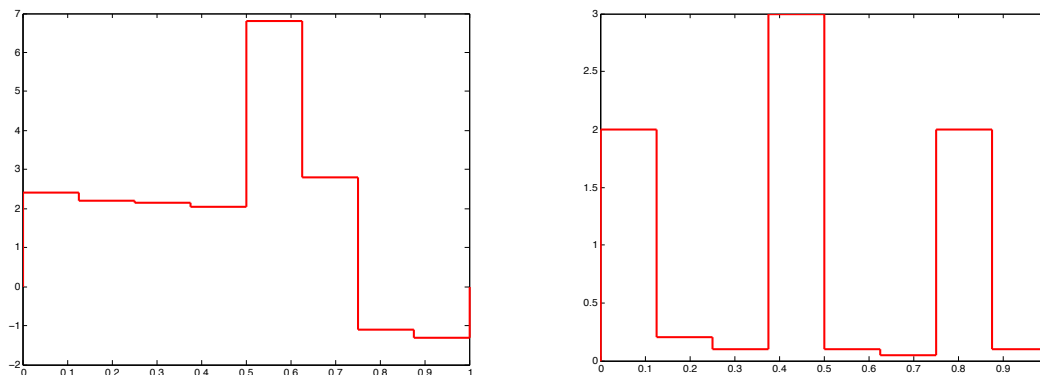


Figure 2.6: A signal and its Haar transform

We get

$$c_2 = (2, 0, 0, 3, 0, 0, 2, 0),$$

and the reconstructed signal is

$$u_2 = (2, 2, 2, 2, 7, 3, -1, -1).$$

The piecewise-linear curves corresponding to  $u_2$  and  $c_2$  are shown in Figure 2.7.

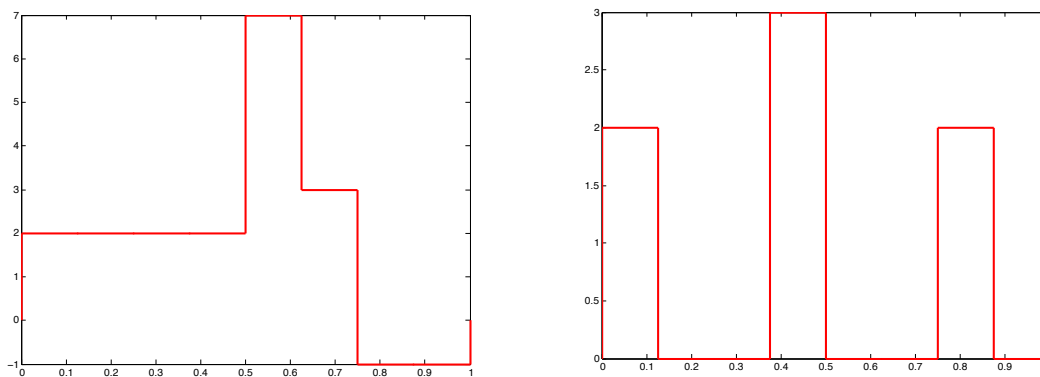


Figure 2.7: A compressed signal and its compressed Haar transform

An interesting (and amusing) application of the Haar wavelets is to the compression of audio signals.

It turns out that if your type `load handel` in `Matlab` an audio file will be loaded in a vector denoted by  $y$ , and if you type `sound(y)`, the computer will play this piece of music.

You can convert  $y$  to its vector of Haar coefficients,  $c$ . The length of  $y$  is 73113, so first tuncate the tail of  $y$  to get a vector of length  $65536 = 2^{16}$ .

A plot of the signals corresponding to  $y$  and  $c$  is shown in Figure 2.8.

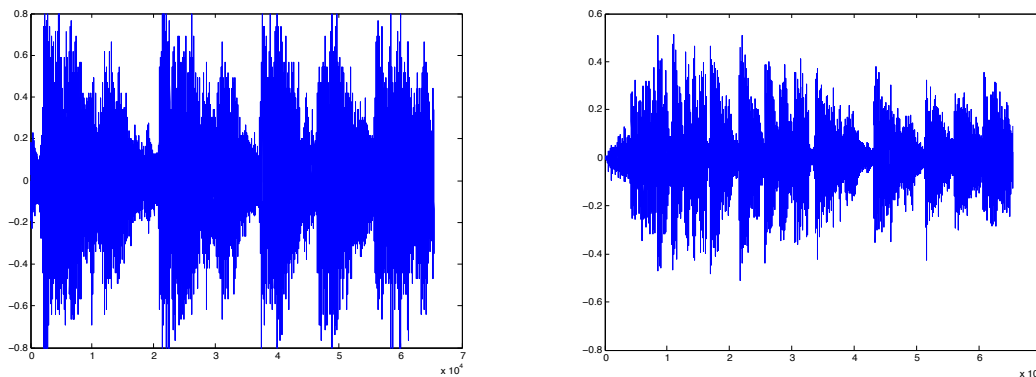


Figure 2.8: The signal “handel” and its Haar transform

Then, run a program that sets all coefficients of  $c$  whose absolute value is less than 0.05 to zero. This sets 37272 coefficients to 0.

The resulting vector  $c_2$  is converted to a signal  $y_2$ . A plot of the signals corresponding to  $y_2$  and  $c_2$  is shown in Figure 2.9.

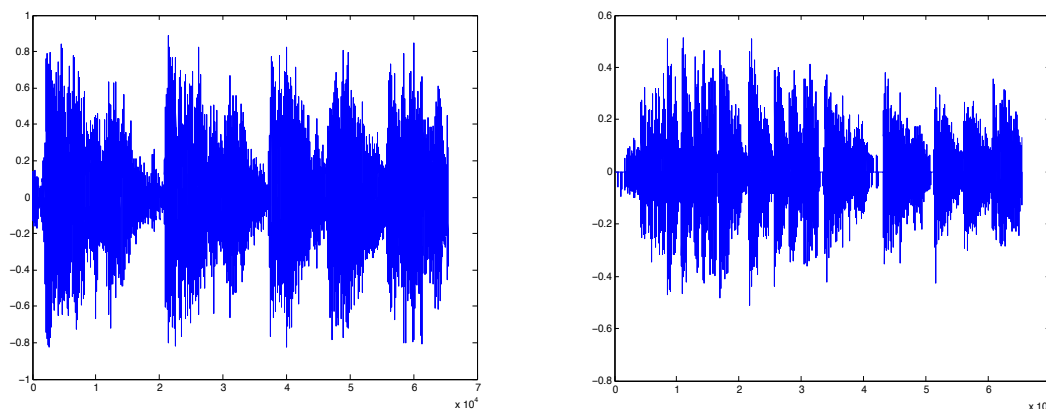


Figure 2.9: The compressed signal “handel” and its Haar transform

When you type `sound(y2)`, you find that the music doesn't differ much from the original, although it sounds less crisp.

Another neat property of the Haar transform is that it can be instantly generalized to matrices (even rectangular) without any extra effort!

This allows for the compression of digital images. But first, we address the issue of normalization of the Haar coefficients.

As we observed earlier, the  $2^n \times 2^n$  matrix  $W_n$  of Haar basis vectors has orthogonal columns, but its columns do not have unit length.

As a consequence,  $W_n^\top$  is not the inverse of  $W_n$ , but rather the matrix

$$W_n^{-1} = D_n W_n^\top$$

with

$$D_n = \text{diag} \left( 2^{-n}, \underbrace{2^{-n}}_{2^0}, \underbrace{2^{-(n-1)}, 2^{-(n-1)}}_{2^1}, \underbrace{2^{-(n-2)}, \dots, 2^{-(n-2)}}_{2^2}, \dots, \underbrace{2^{-1}, \dots, 2^{-1}}_{2^{n-1}} \right).$$

Therefore, we define the orthogonal matrix

$$H_n = W_n D_n^{\frac{1}{2}}$$

whose columns are the normalized Haar basis vectors, with

$$D_n^{\frac{1}{2}} = \text{diag} \left( 2^{-\frac{n}{2}}, \underbrace{2^{-\frac{n}{2}}}_{2^0}, \underbrace{2^{-\frac{n-1}{2}}, 2^{-\frac{n-1}{2}}}_{2^1}, \right. \\ \left. \underbrace{2^{-\frac{n-2}{2}}, \dots, 2^{-\frac{n-2}{2}}}_{2^2}, \dots, \underbrace{2^{-\frac{1}{2}}, \dots, 2^{-\frac{1}{2}}}_{2^{n-1}} \right).$$

We call  $H_n$  the *normalized Haar transform matrix*.

Because  $H_n$  is orthogonal,  $H_n^{-1} = H_n^\top$ .

Given a vector (signal)  $u$ , we call  $c = H_n^\top u$  the *normalized Haar coefficients* of  $u$ .

When computing the sequence of  $u^j$ s, use

$$\begin{aligned}u^{j+1}(2i - 1) &= (u^j(i) + u^j(2^j + i))/\sqrt{2} \\ u^{j+1}(2i) &= (u^j(i) - u^j(2^j + i))/\sqrt{2},\end{aligned}$$

and when computing the sequence of  $c^j$ s, use

$$\begin{aligned}c^j(i) &= (c^{j+1}(2i - 1) + c^{j+1}(2i))/\sqrt{2} \\ c^j(2^j + i) &= (c^{j+1}(2i - 1) - c^{j+1}(2i))/\sqrt{2}.\end{aligned}$$

Note that things are now more symmetric, at the expense of a division by  $\sqrt{2}$ . However, for long vectors, it turns out that these algorithms are numerically more stable.

Let us now explain the 2D version of the Haar transform.

We describe the version using the matrix  $W_n$ , the method using  $H_n$  being identical (except that  $H_n^{-1} = H_n^\top$ , but this does not hold for  $W_n^{-1}$ ).

Given a  $2^m \times 2^n$  matrix  $A$ , we can first convert the *rows* of  $A$  to their Haar coefficients using the Haar transform  $W_n^{-1}$ , obtaining a matrix  $B$ , and then convert the *columns* of  $B$  to their Haar coefficients, using the matrix  $W_m^{-1}$ .

Because columns and rows are exchanged in the first step,

$$B = A(W_n^{-1})^\top,$$

and in the second step  $C = W_m^{-1}B$ , thus, we have

$$C = W_m^{-1}A(W_n^{-1})^\top = D_m W_m^\top A W_n D_n.$$



In the other direction, given a matrix  $C$  of Haar coefficients, we reconstruct the matrix  $A$  (the image) by first applying  $W_m$  to the columns of  $C$ , obtaining  $B$ , and then  $W_n^\top$  to the rows of  $B$ . Therefore

$$A = W_m C W_n^\top.$$

Of course, we don't actually have to invert  $W_m$  and  $W_n$  and perform matrix multiplications. We just have to use our algorithms using averaging and differencing.

Here is an example. If the data matrix (the image) is the  $8 \times 8$  matrix

$$A = \begin{pmatrix} 64 & 2 & 3 & 61 & 60 & 6 & 7 & 57 \\ 9 & 55 & 54 & 12 & 13 & 51 & 50 & 16 \\ 17 & 47 & 46 & 20 & 21 & 43 & 42 & 24 \\ 40 & 26 & 27 & 37 & 36 & 30 & 31 & 33 \\ 32 & 34 & 35 & 29 & 28 & 38 & 39 & 25 \\ 41 & 23 & 22 & 44 & 45 & 19 & 18 & 48 \\ 49 & 15 & 14 & 52 & 53 & 11 & 10 & 56 \\ 8 & 58 & 59 & 5 & 4 & 62 & 63 & 1 \end{pmatrix},$$

then applying our algorithms, we find that

$$C = \begin{pmatrix} 32.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4 & -4 & 4 & -4 \\ 0 & 0 & 0 & 0 & 4 & -4 & 4 & -4 \\ 0 & 0 & 0.5 & 0.5 & 27 & -25 & 23 & -21 \\ 0 & 0 & -0.5 & -0.5 & -11 & 9 & -7 & 5 \\ 0 & 0 & 0.5 & 0.5 & -5 & 7 & -9 & 11 \\ 0 & 0 & -0.5 & -0.5 & 21 & -23 & 25 & -27 \end{pmatrix}.$$

As we can see,  $C$  has more zero entries than  $A$ ; it is a compressed version of  $A$ . We can further compress  $C$  by setting to 0 all entries of absolute value at most 0.5. Then, we get

$$C_2 = \begin{pmatrix} 32.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4 & -4 & 4 & -4 \\ 0 & 0 & 0 & 0 & 4 & -4 & 4 & -4 \\ 0 & 0 & 0 & 0 & 27 & -25 & 23 & -21 \\ 0 & 0 & 0 & 0 & -11 & 9 & -7 & 5 \\ 0 & 0 & 0 & 0 & -5 & 7 & -9 & 11 \\ 0 & 0 & 0 & 0 & 21 & -23 & 25 & -27 \end{pmatrix}.$$

We find that the reconstructed image is

$$A_2 = \begin{pmatrix} 63.5 & 1.5 & 3.5 & 61.5 & 59.5 & 5.5 & 7.5 & 57.5 \\ 9.5 & 55.5 & 53.5 & 11.5 & 13.5 & 51.5 & 49.5 & 15.5 \\ 17.5 & 47.5 & 45.5 & 19.5 & 21.5 & 43.5 & 41.5 & 23.5 \\ 39.5 & 25.5 & 27.5 & 37.5 & 35.5 & 29.5 & 31.5 & 33.5 \\ 31.5 & 33.5 & 35.5 & 29.5 & 27.5 & 37.5 & 39.5 & 25.5 \\ 41.5 & 23.5 & 21.5 & 43.5 & 45.5 & 19.5 & 17.5 & 47.5 \\ 49.5 & 15.5 & 13.5 & 51.5 & 53.5 & 11.5 & 9.5 & 55.5 \\ 7.5 & 57.5 & 59.5 & 5.5 & 3.5 & 61.5 & 63.5 & 1.5 \end{pmatrix},$$

which is pretty close to the original image matrix  $A$ .

It turns out that **Matlab** has a wonderful command, **image(X)**, which displays the matrix  $X$  as an image.

The images corresponding to  $A$  and  $C$  are shown in Figure 2.10. The compressed images corresponding to  $A_2$  and  $C_2$  are shown in Figure 2.11.

The compressed versions appear to be indistinguishable from the originals!

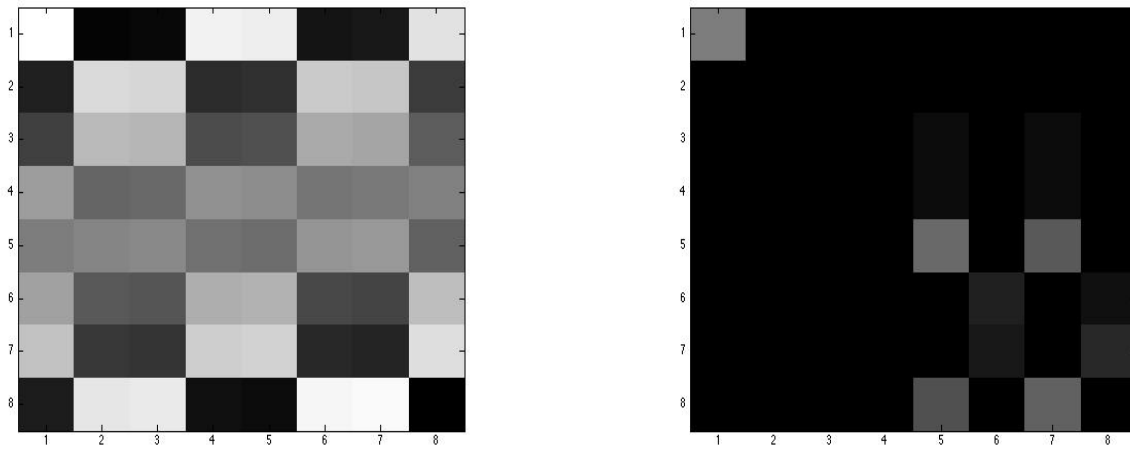


Figure 2.10: An image and its Haar transform

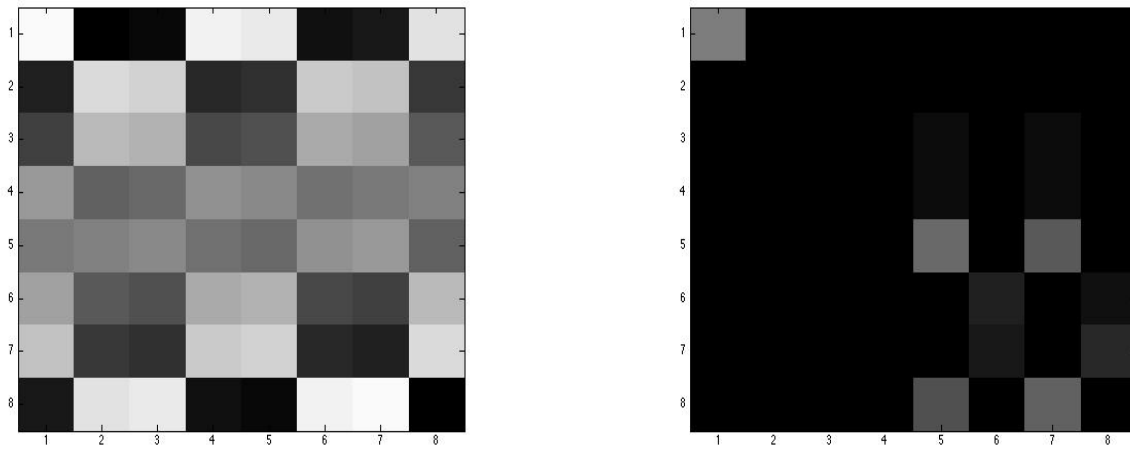


Figure 2.11: Compressed image and its Haar transform

If we use the normalized matrices  $H_m$  and  $H_n$ , then the equations relating the image matrix  $A$  and its normalized Haar transform  $C$  are

$$\begin{aligned}C &= H_m^\top A H_n \\A &= H_m C H_n^\top.\end{aligned}$$

The Haar transform can also be used to send large images progressively over the internet.

Observe that instead of performing all rounds of averaging and differencing on each row and each column, we can perform partial encoding (and decoding).

For example, we can perform a single round of averaging and differencing for each row and each column.

The result is an image consisting of four subimages, where the top left quarter is a coarser version of the original, and the rest (consisting of three pieces) contain the finest detail coefficients.

We can also perform two rounds of averaging and differencing, or three rounds, *etc.* This process is illustrated on the image shown in Figure 2.12. The result of performing

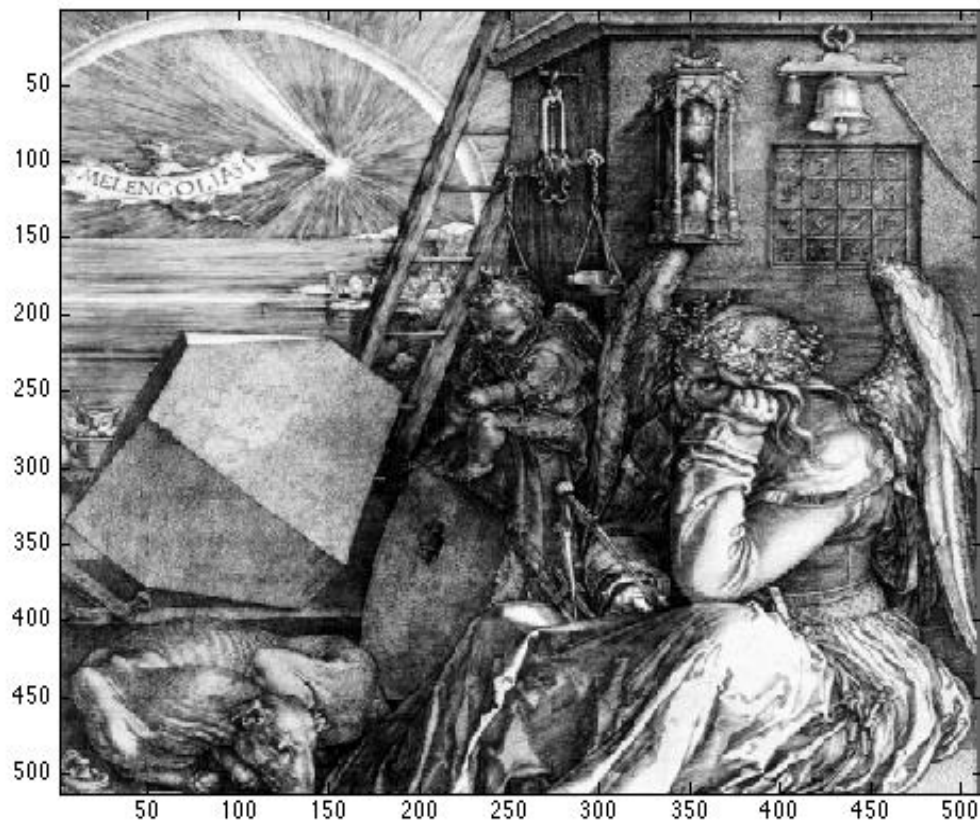


Figure 2.12: Original drawing by Durer

one round, two rounds, three rounds, and nine rounds of averaging is shown in Figure 2.13.

Since our images have size  $512 \times 512$ , nine rounds of averaging yields the Haar transform, displayed as the image on the bottom right. The original image has completely disappeared!

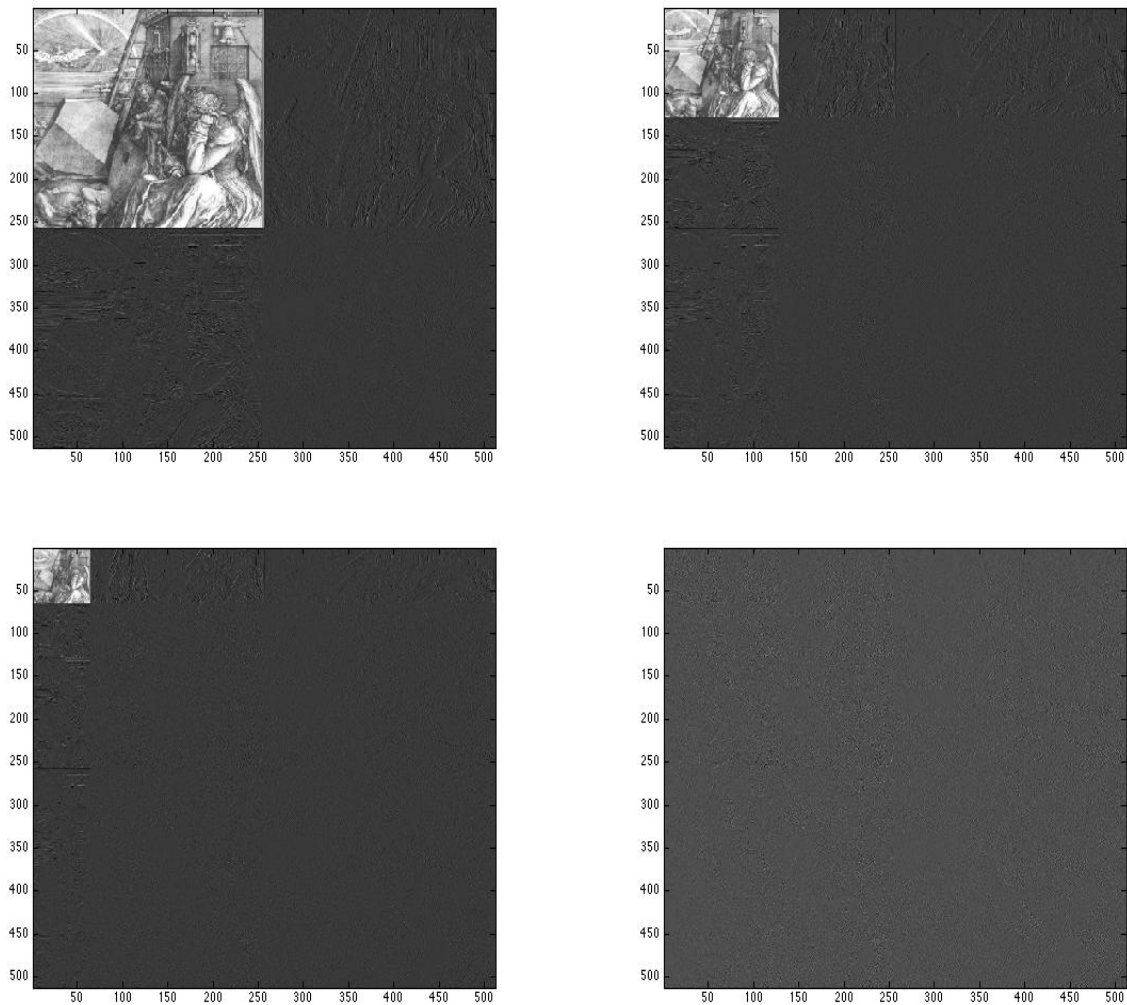


Figure 2.13: Haar tranforms after one, two, three, and nine rounds of averaging

We can find easily a basis of  $2^n \times 2^n = 2^{2n}$  vectors  $w_{ij}$  ( $2^n \times 2^n$  matrices) for the linear map that reconstructs an image from its Haar coefficients, in the sense that for any matrix  $C$  of Haar coefficients, the image matrix  $A$  is given by

$$A = \sum_{i=1}^{2^n} \sum_{j=1}^{2^n} c_{ij} w_{ij}.$$

Indeed, the matrix  $w_{ij}$  is given by the so-called outer product

$$w_{ij} = w_i (w_j)^\top.$$

Similarly, there is a basis of  $2^n \times 2^n = 2^{2n}$  vectors  $h_{ij}$  ( $2^n \times 2^n$  matrices) for the 2D Haar transform, in the sense that for any matrix  $A$ , its matrix  $C$  of Haar coefficients is given by

$$C = \sum_{i=1}^{2^n} \sum_{j=1}^{2^n} a_{ij} h_{ij}.$$

If the columns of  $W^{-1}$  are  $w'_1, \dots, w'_{2^n}$ , then

$$h_{ij} = w'_i (w'_j)^\top.$$



## 2.4 The Effect of a Change of Bases on Matrices

The effect of a change of bases on the representation of a linear map is described in the following proposition.

**Proposition 2.4.** *Let  $E$  and  $F$  be vector spaces, let  $\mathcal{U} = (u_1, \dots, u_n)$  and  $\mathcal{U}' = (u'_1, \dots, u'_n)$  be two bases of  $E$ , and let  $\mathcal{V} = (v_1, \dots, v_m)$  and  $\mathcal{V}' = (v'_1, \dots, v'_m)$  be two bases of  $F$ . Let  $P = P_{\mathcal{U}', \mathcal{U}}$  be the change of basis matrix from  $\mathcal{U}$  to  $\mathcal{U}'$ , and let  $Q = P_{\mathcal{V}', \mathcal{V}}$  be the change of basis matrix from  $\mathcal{V}$  to  $\mathcal{V}'$ . For any linear map  $f: E \rightarrow F$ , let  $M(f) = M_{\mathcal{U}, \mathcal{V}}(f)$  be the matrix associated to  $f$  w.r.t. the bases  $\mathcal{U}$  and  $\mathcal{V}$ , and let  $M'(f) = M_{\mathcal{U}', \mathcal{V}'}(f)$  be the matrix associated to  $f$  w.r.t. the bases  $\mathcal{U}'$  and  $\mathcal{V}'$ . We have*

$$M'(f) = Q^{-1}M(f)P,$$

or more explicitly

$$M_{\mathcal{U}', \mathcal{V}'}(f) = P_{\mathcal{V}', \mathcal{V}}^{-1} M_{\mathcal{U}, \mathcal{V}}(f) P_{\mathcal{U}', \mathcal{U}} = P_{\mathcal{V}, \mathcal{V}'} M_{\mathcal{U}, \mathcal{V}}(f) P_{\mathcal{U}', \mathcal{U}}.$$

As a corollary, we get the following result.

**Corollary 2.5.** *Let  $E$  be a vector space, and let  $\mathcal{U} = (u_1, \dots, u_n)$  and  $\mathcal{U}' = (u'_1, \dots, u'_n)$  be two bases of  $E$ . Let  $P = P_{\mathcal{U}', \mathcal{U}}$  be the change of basis matrix from  $\mathcal{U}$  to  $\mathcal{U}'$ . For any linear map  $f: E \rightarrow E$ , let  $M(f) = M_{\mathcal{U}}(f)$  be the matrix associated to  $f$  w.r.t. the basis  $\mathcal{U}$ , and let  $M'(f) = M_{\mathcal{U}'}(f)$  be the matrix associated to  $f$  w.r.t. the basis  $\mathcal{U}'$ . We have*

$$M'(f) = P^{-1}M(f)P,$$

or more explicitly,

$$M_{\mathcal{U}'}(f) = P_{\mathcal{U}', \mathcal{U}}^{-1} M_{\mathcal{U}}(f) P_{\mathcal{U}', \mathcal{U}} = P_{\mathcal{U}, \mathcal{U}'} M_{\mathcal{U}}(f) P_{\mathcal{U}', \mathcal{U}}.$$

**Example 2.3.** Let  $E = \mathbb{R}^2$ ,  $\mathcal{U} = (e_1, e_2)$  where  $e_1 = (1, 0)$  and  $e_2 = (0, 1)$  are the canonical basis vectors, let  $\mathcal{V} = (v_1, v_2) = (e_1, e_1 - e_2)$ , and let

$$A = \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}.$$

The change of basis matrix  $P = P_{\mathcal{V}, \mathcal{U}}$  from  $\mathcal{U}$  to  $\mathcal{V}$  is

$$P = \begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix},$$

and we check that  $P^{-1} = P$ .

Therefore, in the basis  $\mathcal{V}$ , the matrix representing the linear map  $f$  defined by  $A$  is

$$\begin{aligned} A' = P^{-1}AP &= PAP = \begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix} \\ &= \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} = D, \end{aligned}$$

a diagonal matrix.

Therefore, in the basis  $\mathcal{V}$ , it is clear what the action of  $f$  is: it is a stretch by a factor of 2 in the  $v_1$  direction and it is the identity in the  $v_2$  direction.

Observe that  $v_1$  and  $v_2$  are not orthogonal.

What happened is that we *diagonalized* the matrix  $A$ .

The diagonal entries 2 and 1 are the *eigenvalues* of  $A$  (and  $f$ ) and  $v_1$  and  $v_2$  are corresponding *eigenvectors*.

The above example showed that the same linear map can be represented by different matrices. This suggests making the following definition:

**Definition 2.4.** Two  $n \times n$  matrices  $A$  and  $B$  are said to be *similar* iff there is some invertible matrix  $P$  such that

$$B = P^{-1}AP.$$

It is easily checked that similarity is an equivalence relation.

From our previous considerations, two  $n \times n$  matrices  $A$  and  $B$  are similar iff they represent the same linear map with respect to two different bases.

The following surprising fact can be shown: *Every square matrix  $A$  is similar to its transpose  $A^T$ .*

The proof requires advanced concepts than we will not discuss in these notes (the Jordan form, or similarity invariants).

If  $\mathcal{U} = (u_1, \dots, u_n)$  and  $\mathcal{V} = (v_1, \dots, v_n)$  are two bases of  $E$ , the change of basis matrix

$$P = P_{\mathcal{V}, \mathcal{U}} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

from  $(u_1, \dots, u_n)$  to  $(v_1, \dots, v_n)$  is the matrix whose  $j$ th column consists of the coordinates of  $v_j$  over the basis  $(u_1, \dots, u_n)$ , which means that

$$v_j = \sum_{i=1}^n a_{ij} u_i.$$

It is natural to extend the matrix notation and to express the vector  $(v_1, \dots, v_n)$  in  $E^n$  as the product of a matrix times the vector  $(u_1, \dots, u_n)$  in  $E^n$ , namely as

$$\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix},$$

but notice that *the matrix involved is not  $P$ , but its transpose  $P^\top$ .*

This observation has the following consequence: if  $\mathcal{U} = (u_1, \dots, u_n)$  and  $\mathcal{V} = (v_1, \dots, v_n)$  are two bases of  $E$  and if

$$\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = A \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix},$$

that is,

$$v_i = \sum_{j=1}^n a_{ij} u_j,$$

for any vector  $w \in E$ , if

$$w = \sum_{i=1}^n x_i u_i = \sum_{k=1}^n y_k v_k,$$

then

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = A^\top \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix},$$

and so

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = (A^\top)^{-1} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

It is easy to see that  $(A^\top)^{-1} = (A^{-1})^\top$ .



Also, if  $\mathcal{U} = (u_1, \dots, u_n)$ ,  $\mathcal{V} = (v_1, \dots, v_n)$ , and  $\mathcal{W} = (w_1, \dots, w_n)$  are three bases of  $E$ , and if the change of basis matrix from  $\mathcal{U}$  to  $\mathcal{V}$  is  $P = P_{\mathcal{V}, \mathcal{U}}$  and the change of basis matrix from  $\mathcal{V}$  to  $\mathcal{W}$  is  $Q = P_{\mathcal{W}, \mathcal{V}}$ , then

$$\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = P^\top \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}, \quad \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = Q^\top \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix},$$

so

$$\begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = Q^\top P^\top \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} = (PQ)^\top \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix},$$

which means that the change of basis matrix  $P_{\mathcal{W}, \mathcal{U}}$  from  $\mathcal{U}$  to  $\mathcal{W}$  is  $PQ$ .

This proves that

$$P_{\mathcal{W}, \mathcal{U}} = P_{\mathcal{V}, \mathcal{U}} P_{\mathcal{W}, \mathcal{V}}.$$

Even though matrices are indispensable since they are *the major tool* in applications of linear algebra, one should not lose track of the fact that

*linear maps are more fundamental, because they are intrinsic objects that do not depend on the choice of bases. Consequently, we advise the reader to try to think in terms of linear maps rather than reduce everything to matrices.*

In our experience, this is particularly effective when it comes to proving results about linear maps and matrices, where proofs involving linear maps are often more “conceptual.”

Also, instead of thinking of a matrix decomposition, as a purely algebraic operation, it is often illuminating to view it as a *geometric decomposition*.

After all, a

*a matrix is a representation of a linear map*

and most decompositions of a matrix reflect the fact that with a *suitable choice of a basis (or bases)*, the linear map is represented by a matrix having a special shape.

The problem is then to find such bases.

Also, always try to keep in mind that

*linear maps are geometric in nature; they act on space.*

