Chapter 3

Basis Functions

In Section 2.1.2, we introduced transform coding. We said that any discrete signal $\boldsymbol{v} \in \mathbb{R}^M$ can be expressed in a different basis via a basis transform:

$$\boldsymbol{v} = \boldsymbol{\Psi}\boldsymbol{w}$$

where $\boldsymbol{\Psi}$ is the $M \times M$ basis matrix and $\boldsymbol{w}\mathbb{R}^M$ is the representation of $\boldsymbol{v}$ in the $\boldsymbol{\Psi}$ basis.

The particular classes of signals $\boldsymbol{v}$ that we are interested in are digital images and digital video. The aim of this chapter is to construct a basis matrix $\boldsymbol{\Psi}$ that gives us a (near-) sparse representation of a wide range of such signals $\boldsymbol{v}$. Finding a set of basis functions $\boldsymbol{\Psi}$ that achieve such a transformation lies at the heart of many lossy compression techniques.

It is important to note here that the choice of basis functions $\boldsymbol{\Psi}$ typically has a significant effect on the performance of the reconstruction algorithms.

3.1 Discrete Cosine Transform

The first basis transform that we will use is the Discrete Cosine Transform (DCT), one of the most widely used transforms in signal processing. It underlies JPEG image compression and is used in various video compression algorithms such as MJPEG, MPEG, H.261 and H.263 [10]. [1]

A DCT decomposes a signal in terms of cosine functions with different frequencies. Its extensive use in lossy compression algorithms is due to the DCT's *energy*

[1] A related transform, known as the *Modified DCT* is used in many lossy audio compression formats such as MP3, AAC and Vorbis.

(a) Original signal $\boldsymbol{v}$ (b) DCT of $\boldsymbol{v}$

Fig. 3.1 Panel (a) shows the original signal $\boldsymbol{v}$, a 256×256 grayscale image known as "cameraman". Panel (b) illustrates the 2-D DCT of $\boldsymbol{v}$. The brightness of a an element increases with the absolute value of the corresponding DCT coefficient. (The high-frequency coefficients have been enhanced to show more detail).

compaction properties. The majority of a signal's energy is contained within relatively few coefficients - typically those corresponding to the lower frequency basis functions.

On a side note, the DCT comes in a various versions that have minor differences between them. In the following, we will describe the most widely used version, known as the *DCT-II*, as well as its inverse transform, the *DCT-III*. We will refer to them simply as "the DCT" and "the Inverse DCT (IDCT)", respectively.

3.1.1 One-Dimensional DCT

Formally, the DCT $\boldsymbol{w}$ of a one-dimensional signal $\boldsymbol{v}$ of length M is given by

$$w_k = c(k) \sum_{m=1}^{M} v_m \cos\left(\frac{\pi(2i-1)(k-1)}{2M}\right) \qquad k = 1, \cdots, M \tag{3.1}$$

where

$$c(k) = \begin{cases} \sqrt{\frac{1}{M}} & \text{if } k = 1 \\ \sqrt{\frac{2}{M}} & \text{otherwise} \end{cases}$$

This transforms a signal $\boldsymbol{v}$ in the original domain (time or space) into its representation $\boldsymbol{w}$ in the DCT domain.

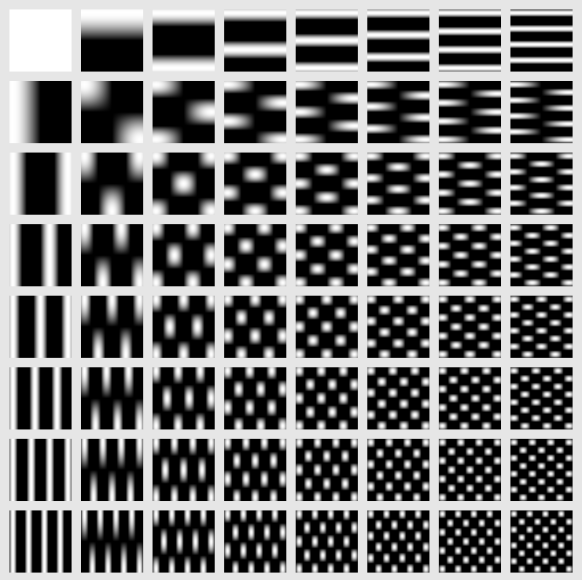


Fig. 3.2 The 2-D DCT basis functions that are used by the DCT to decompose a 8×8 image. The spatial frequency increases towards the bottom right corner.

Conversely, given a signal $\boldsymbol{w}$ in the DCT domain, we can transform it back to the orignal (time or space) domain via the IDCT defined by

$$v_n = \sum_{k=1}^{M} c(k) w_k \cos\left(\frac{\pi(2i-1)(k-1)}{2M}\right) \qquad n = 1, \cdots, M \tag{3.2}$$

We can express equation (3.2) in the desired form $\boldsymbol{v} = \boldsymbol{P}\boldsymbol{w}$. The entries of the basis matrix $\boldsymbol{P}$ are given by

$$P_{n,k} = c(k) \cos\left(\frac{\pi(2i-1)(k-1)}{2M}\right). \tag{3.3}$$

Note that the basis matrix $\boldsymbol{P}$ is orthogonal, $\boldsymbol{P}^T\boldsymbol{P} = \boldsymbol{I}_M$.

3.1.2 Multi-Dimensional DCT

Once we know how to perform the DCT on a one-dimensional signal, we can easily extend the transform to multi-dimensional signals (images, video, etc). To do so, we simply perform successive 1-D transforms along each dimension of the signal. This property is known as *seperability*.

Suppose the signal of interest is a digital image. That means that $\boldsymbol{v}$ is a $M_1 \times M_2$ matrix where $M_1 \times M_2$ is the resolution of the image. To transform the signal, we first

perform the DCT on every row of the matrix. Following that, we perform the DCT on every column of the resulting matrix to get the final transformed signal.

Figure 3.1 shows an example of a 2-D signal $\boldsymbol{v}$ and its transform $\boldsymbol{w}$. Note that the majority of the energy of the transformed signal is concentrated in the top left corner. Most of the DCT coefficients are zero or very close to zero.

In Figure 3.2, we show the 2-D basis functions that would used by the DCT to decompose a signal of size 8×8. Each basis function is characterised by a horizontal and vertical spatial frequency. Typically, natural images are mostly made up of low-frequency components and the corresponding coefficients are therefore relatively large. The highest-frequency components are usually only needed to describe very fine details.

The DCT can be used to decompose video signals with 3-D basis functions. Besides the spatial frequencies, the 3-D basis functions have an additional temporal frequency component. To perform the DCT on a video, we could first perform the 2-D DCT on every frame of the video followed by a 1-D DCT across the temporal axis for each pixel.

For a discussion on the properties of the DCT, see [3]

3.2 Discrete Wavelet Transform

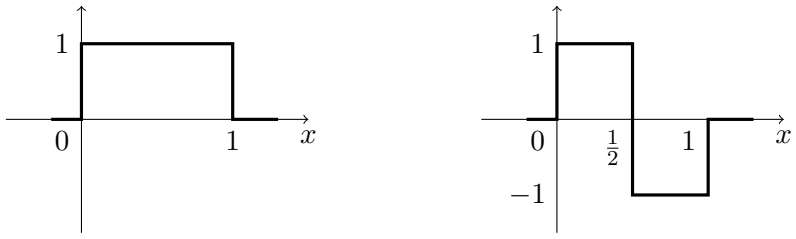
Wavelets have become a very popular tool in signal processing. Their energy compaction properties are on par and often superior to those of the DCT for a wide range of signal classes. In 2000, the JPEG committee released a new image coding standard, JPEG2000, that is gradually replacing the original JPEG standard. The new format moved away from the DCT and uses a Discrete Wavelet Transform (DWT) instead.

[CAVEAT and INTRO]

3.2.1 Introduction to Wavelets

To motivate wavelets, consider again the one-dimensional signal $\boldsymbol{v}$ of length M. Suppose, for simplicity, that M is a power of 2, $M = 2^q$ say. We can view the $\boldsymbol{v}$ as a piecewise-constant function $v(x)$ on the half-open interval $[0, 1)$, where $v(x) = v_i$ if $x \in [\frac{i-1}{M}, \frac{i}{M})$.

Let V^j denote the vector space containing all piecewise-constant functions f defined on the interval $[0, 1)$ that consist of 2^j pieces, each of which is constant across a sub-interval of size 2^{-j}. Thus, V^0 consists of all functions that are constant on $[0, 1)$, while V^1 consists of all functions that have two constant pieces, one over $[0, 1/2)$ and one over $[1/2, 1)$. In particular, our signal $v(x)$ resides in the space V^q.



(a) Haar scaling function $\phi(x)$ (b) Haar mother wavelet $\psi(x)$

Fig. 3.3 The scaling function and wavelet function for the Haar wavelets.

Note that if $f \in V^j$, then $f \in V^{j+1}$. Thus, the vector spaces V^j are nested: $V^0 \subset V^1 \subset V^2 \subset \cdots$.

Next, we need to choose a basis for each vector space V^j. To do so, we introduce a *scaling function* (also known as *scalet*, or *father wavelet*) that is usually denoted $\phi(x)$. The form of the scaling function depends on the particular choice wavelet decomposition.

For example, for the *Haar wavelets* the scaling function is given by

$$\phi(x) = \begin{cases} 1 & \text{if } 0 \leq x < 1 \\ 0 & \text{otherwise} \end{cases} \tag{3.4}$$

See Figure 3.3a for a plot of $\phi(x)$.

Given the scaling function $\phi(x)$, we can define the following basis for V^j:

$$\phi_k^j(x) := 2^{j/2}\phi(2^j x - k) \qquad k = 0, \cdots, 2^j - 1$$

Using this basis, we can decompose our signal $v(x) \in V^q$ as

$$v(x) = \sum_{k=0}^{2^q-1} c_k^q \phi_k^q(x)$$

For the scaling function defined in equation (3.4), we have that $c_k^q = v_{k+1}$.

To obtain *wavelets*, consider the *orthogonal complement* of V^j in V^{j+1} and denote it W^j. That is, $W^j = \{f \in V^{j+1} : \quad \langle f, g\rangle = 0 \quad \forall g \in V^j\}$ where the inner product $\langle f, g\rangle$ is given by

$$\langle f, g\rangle = \int_0^1 f(x)g(x)dx.$$

By forming a basis for W^j, we obtain a set of *wavelet functions* $\{\psi_k^j,\ k = 0, \cdots, 2^j - 1\}$. Wavelet functions can be constructed by scaling and shifting a so-called *mother wavelet*

$\psi(x)$ as follows:

$$\psi_k^j(x) = 2^{j/2}\psi(2^j x - k) \qquad k = 0, \cdots, 2^j - 1$$

For the Haar wavelets, the mother wavelet is given by:

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

The Haar mother wavelet is shown in Figure 3.3b.

Note that, since the scaling functions ϕ_k^j form a basis of V^j and the wavelet functions ψ_k^j form a basis of W^k, and since W^j is the orthogonal complement to V^j in V^{j+1}, it follows that the set $\{\phi_k^j, \psi_k^j : k = 0, \cdots, 2^j - 1\}$ forms a basis of the vector space V^{j+1}.

This allows us to express our signal $v \in V^q$ as

$$v(x) = \sum_{k=0}^{2^{q-1}-1} d_k^{q-1}\psi_k^{q-1}(x) + \sum_{k=0}^{2^{q-1}-1} c_k^{q-1}\phi_k^{q-1}(x)$$

This gives us the first level of the discrete wavelet transform of v. The coefficients c_k and d_k are sometimes referred to as "approximation" coefficients and "detail" coefficients, respectively.

We can continue the decomposition by splitting the basis for V^{q-1} into the bases for V^{q-2} and W^{q-2} to get the next level of the transform:

$$v(x) = \sum_{k=0}^{2^{q-1}-1} d_k^{q-1}\psi_k^{q-1}(x) + \sum_{k=0}^{2^{q-2}-1} d_k^{q-2}\psi_k^{q-2}(x) + \sum_{k=0}^{2^{q-2}-1} c_k^{q-2}\phi_k^{q-2}(x)$$

To get the full decomposition, we continue in this fashion up to the qth level:

$$v(x) = \sum_{j=0}^{q-1}\sum_{k=0}^{2^j-1} d_k^j\psi_k^j(x) + c_0^0\phi(x)$$

The full DWT of $\boldsymbol{v}$ consists of the coefficients $\{c_0^0, d_k^j : j = 0, \cdots, q-1,\ k = 0, \cdots, 2^j - 1\}$.

3.2.2 Filter Banks

In practice, we can compute one level of the DWT coefficients by passing the signal $\boldsymbol{v}$ through a *low-pass filter* $\boldsymbol{h}$ and a high-pass filter $\boldsymbol{g}$, respectively, and then downsampling the results by a factor of two. The procedure is depicted in Figure 3.4.

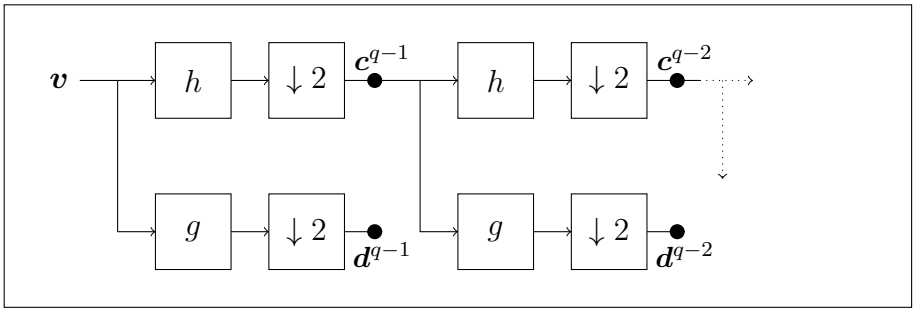


Fig. 3.4 Filter Banks

References

[1] Baraniuk, R. G. (2007). Compressive sensing. *IEEE signal processing magazine*, 24(4).

[2] Candès, E. J. and Wakin, M. B. (2008). An introduction to compressive sampling. *Signal Processing Magazine, IEEE*, 25(2):21–30.

[3] Khayam, S. A. (2003). The discrete cosine transform (dct): theory and application. *Michigan State University*, 114.

[4] Pilikos, G. (2014). Signal reconstruction using compressive sensing. MPhil thesis, University of Cambridge.

[5] Shannon, C. E. (1949). Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21.

[6] Taubman, D. and Marcellin, M. (2012). *JPEG2000 Image Compression Fundamentals, Standards and Practice: Image Compression Fundamentals, Standards and Practice*. The Springer International Series in Engineering and Computer Science. Springer US.

[7] Tipping, A. and Faul, A. (2002). Analysis of sparse bayesian learning. *Advances in neural information processing systems*, 14:383–389.

[8] Tipping, M. E. (2001). Sparse bayesian learning and the relevance vector machine. *The journal of machine learning research*, 1:211–244.

[9] Tipping, M. E., Faul, A. C., et al. (2003). Fast marginal likelihood maximisation for sparse bayesian models. In *AISTATS*.

[10] Zeng, J., Au, O. C., Dai, W., Kong, Y., Jia, L., and Zhu, W. (2013). A tutorial on image/video coding standards. In *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2013 Asia-Pacific*, pages 1–7. IEEE.