A function where both U and W are real subsets (i.e., $U, W \subset R$) is called a *real function*. Examples of real functions include:

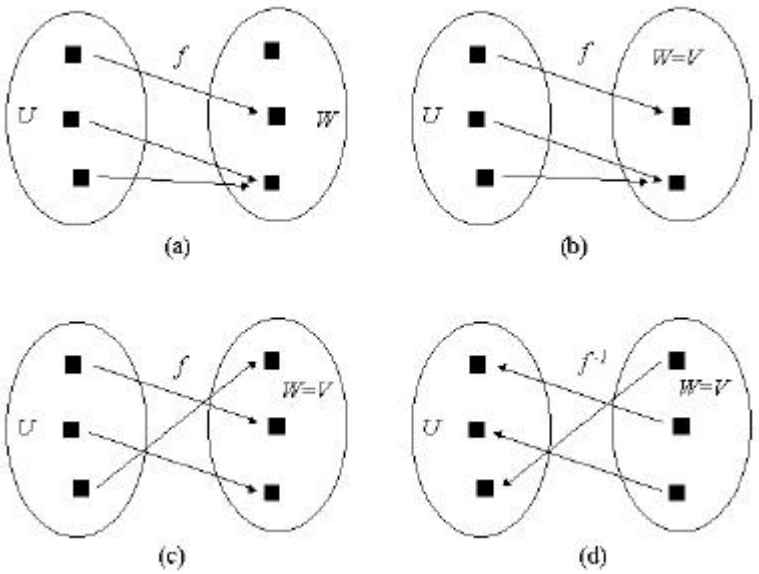


Figure 2.1: *Graphical representations of some types of functions (see text for explanation)*

$$f : R \to R \mid y = f(x) = x^2$$

and

$$g : R \to R \mid y = g(t) = \begin{cases} t & if \quad 0 \le t \le w, \quad w \in R, w > 0 \\ 0 & otherwise \end{cases}$$

The visual representation of real functions is always useful as a means to better understand their characteristics. Figure 2.2 depicts the graphical representations of the aforementioned functions.

Some special types of functions include:

Differentiable (analytical, or smooth) functions: These are functions for which all derivatives exist. For instance, the function $f(t) = t^2$ is differentiable in R, and its first derivative is $f'(t) = \frac{df}{dt} = 2t$. In case the derivatives exist only up to a maximum order k, the function is said to be

angle *y* with the real axis. Observe that this ray does not reach the plane origin. The above mapping can be graphically represented as in Figure 2.27.

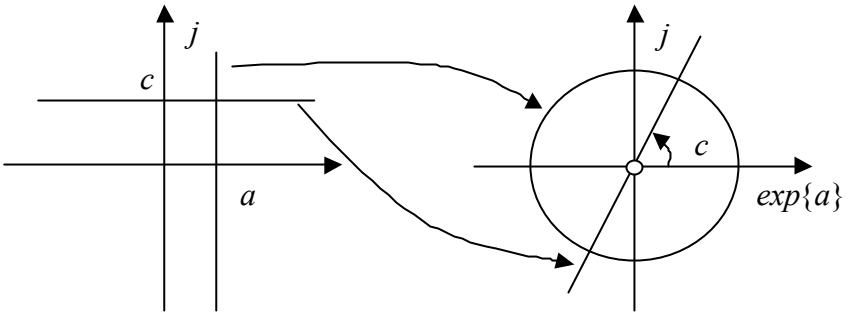


Figure 2.27: *The mapping of vertical and horizontal lines by* $g(z) = \exp(z)$.

Now, consider the rectangular region of the domain space defined as

$$a \le x \le b \quad \text{and} \quad c \le y \le d$$

The reader should have no difficulty verifing that this rectangular region is mapped by *g*(*z*) as illustrated in Figure 2.28.

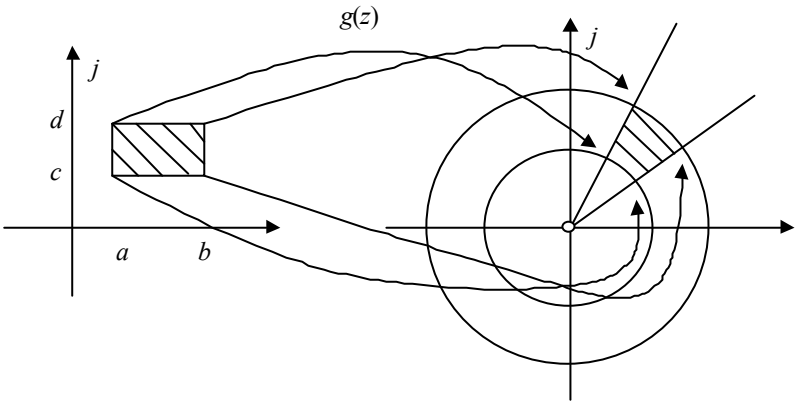


Figure 2.28: *The mapping of a rectangle by* $g(z) = \exp(z)$.

Figure 2.29 presents a more comprehensive illustration of the mapping implemented by $g(z) = \exp(z)$ with respect to an orthogonal grid in the domain space.

As is clear from the above example, it is not always easy to identify the more interesting behavior to be illustrated for each specifically considered complex function.

A non-empty subset X of a vector space S, such as the addition between any of its elements and the product of any of its elements by a scalar result a vector in X (i.e., closure with respect to addition and multiplication by a scalar), is called a *subspace of* S. It can be readily verified that the null vector $\vec{0}$ must be included in any subspace. For instance, the space R is a subspace of R^2. In addition, observe that a subspace is also a vector space.

Given M vectors $\vec{p}_i$; i = 1, 2, … , M; in the vector space S, the *linear combination* of such vectors, resulting a vector $\vec{q}$ also in S, is defined as

$$\vec{q} = a_1\vec{p}_1 + a_2\vec{p}_2 + \cdots + a_M\vec{p}_M$$

where a_i are any scalar values. The above M vectors are said to be *linearly independent* (*l.i.*) if and only

$$\vec{0} = a_1\vec{p}_1 + a_2\vec{p}_2 + \cdots + a_i\vec{p}_i \cdots + a_M\vec{p}_M \Leftrightarrow$$
$$\Leftrightarrow a_1 = a_2 = \cdots a_M = 0$$

In other words, it is not possible to express one of the vectors as a linear combination of the other vectors. Otherwise, the M vectors are said to be *linearly dependent* (*l.d.*). A practical way to determine whether a set of vectors is *l.i.* can be obtained by using the determinants or rank of matrices, as described in Sections 2.2.5.3 and 2.2.5.6, respectively.

For any vector space S, it is always possible to identify a minimal set, in the sense of involving the minimum number of elements, of linearly independent vectors in S whose linear combinations produce (or *span*) all the possible vectors in S. Such a set of elementary vectors is called a *basis* of S. For instance, both

$$B_1 = \left\{\begin{bmatrix}0\\1\end{bmatrix}, \begin{bmatrix}1\\0\end{bmatrix}\right\} \quad \text{and} \quad B_2 = \left\{\begin{bmatrix}0\\-1\end{bmatrix}, \begin{bmatrix}1\\1\end{bmatrix}\right\}$$

are valid bases for R^2. The *dimension* of a vector space is defined as the number of vectors in any of the bases spanning this space. It should be observed that the vector space containing only the null vector has dimension zero and not one. The above examples (*vii*) and (*viii*) of vector spaces have infinite dimension.

Let S be an N-dimensional vector space and $B = \left\{\vec{b}_1, \vec{b}_2, \cdots, \vec{b}_N\right\}$ be one of its possible bases. Then any vector $\vec{p}$ in this space can be represented as a unique linear combination of the vectors in B, i.e., $\vec{p} = a_1\vec{b}_1 + a_2\vec{b}_2 + \cdots + a_N\vec{b}_N$, and scalars $a_1, a_2, \cdots, a_N$ are called

coordinates of the vector $\vec{p}$ with respect to basis B, which are often represented as

$$\vec{p} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix}_B = \begin{bmatrix} a_1 & a_2 & \cdots & a_N \end{bmatrix}_B^T$$

Although a vector space can have an infinite number of alternative bases, it always has a special basis, in the sense of being the most elementary and simple one, which is called its respective *canonical basis*. For instance, the space R^N has the following canonical basis:

$$C_N = \{\vec{e}_1, \vec{e}_2, \cdots, \vec{e}_N\} = \left\{ \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}; \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}; \cdots; \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \right\}$$

These are the bases normally adopted as default for representing vectors. In such situations, the subscript indicating the basis is often ommited. For instance:

$$\begin{bmatrix} -1 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \\ 2 \end{bmatrix}_{C_3} = -1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 3 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 2 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

It makes no sense to think about the orientations of these vectors with respect to some absolute reference, since this is not known to exist in the universe. What does matter are the intrinsical properties of the canonical basis, such as having unit magnitude and being orthogonal (see Section 2.2.4). Observe that all the thus far presented examples in this chapter have considered canonical bases.

Now, let $\vec{v}$ be a vector originally represented in terms of its coordinates with respect to the basis $A = \{\vec{a}_1, \vec{a}_2, \cdots, \vec{a}_N\}$. What will the new coordinates of this vector be when it is expressed with respect to the new basis $B = \{\vec{b}_1, \vec{b}_2, \cdots, \vec{b}_N\}$? This important problem, known as *change of coordinates*, can be addressed as follows. We start by expressing the vectors in the new basis B in terms of the coordinates relative to the original basis A:

$$\vec{b}_1 = \alpha_{1,1}\vec{a}_1 + \alpha_{2,1}\vec{a}_2 + \cdots + \alpha_{N,1}\vec{a}_N$$

$$\vec{v}_A = \begin{bmatrix} \hat{v}_1 \\ \hat{v}_2 \\ \cdots \\ \hat{v}_N \end{bmatrix} = \begin{bmatrix} \alpha_{1,1} & \alpha_{1,2} & \cdots & \alpha_{1,N} \\ \alpha_{2,1} & \alpha_{2,2} & \cdots & \alpha_{2,N} \\ \cdots & & \cdots & \cdots \\ \alpha_{N,1} & \alpha_{N,2} & \cdots & \alpha_{N,N} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \cdots \\ v_N \end{bmatrix} \Leftrightarrow \vec{v}_A = C\vec{v}_B \tag{2.9}$$

Now, provided C is invertible (see Section 2.2.5.8), we have

$$\vec{v}_B = C^{-1}\vec{v}_A \tag{2.10}$$

which provides a practical method for changing coordinates. The above procedure is illustrated in the accompanying box.

Example: ***Change of Coordinates***

Find the coordinates of the vector $\vec{v}_A = (-1, 2)^T$ (represented with respect to the canonical basis) in the new basis defined by $B = \left\{\vec{b}_1 = (1, 1)^T ; \vec{b}_2 = (-2, 0)^T\right\}$.

Solution:

Since matrix C is readily obtained as $C = \begin{bmatrix} 1 & -2 \\ 1 & 0 \end{bmatrix}$, we have

$$\vec{v}_B = C^{-1}\vec{v}_A = \begin{bmatrix} 0 & 1 \\ -0.5 & 0.5 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1.5 \end{bmatrix}$$

Figure 2.32 shows the representation of the above vector with respect to both considered bases.

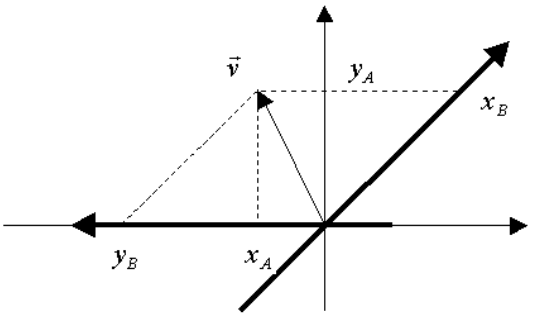


Figure 2.32: *The vector $\vec{v}$ represented with respect to both considered bases. The axes defined by bases A and B are represented by thin and thick arrows, respectively.*

Linear transforms taking vectors from an *N*-dimensional space into an *M*-dimensional space, *M*<*N*, (i.e., transformations which are not full rank) are said to be *degenerated*, and to find its inverse in this case is impossible. It should be observed at this early stage of the book that this type of transformation characterizes a large number of practical situations in shape analysis and vision. For instance, the 2D projections of the 3D world falling onto our retinas (or onto a digital camera) provide but a degenerate representation of the 3D imaged objects.

An important class of linear transformation is that implementing *rotations.* Figure 2.33 illustrates such a situation with respect to a vector $\vec{v}$ in the plane pointing at a point *P*, where the new and old coordinate systems are represented by full and dotted axes, respectively. It should be observed that rotating the old system by an angle $\boldsymbol{q}$ (counterclockwise, with respect to the x-axis) corresponds to rotating vector $\vec{v}$, with respect to the coordinate system, by an angle $-\boldsymbol{q}$. Consequently, both these problems can be treated in the same unified way.

The matrix representing the linear transformation, which rotates the coordinate system of the two-dimensional space R^2 by an angle $\boldsymbol{q}$ counterclockwise, is immediately obtained by using the coordinates exchange procedure discussed in Section 2.2.2. We start by expressing the basis vectors of the new space, i.e., $\tilde{B}=\{\tilde{i},\tilde{j}\}$, in terms of the vectors of the old basis $\hat{B}=\{\hat{i},\hat{j}\}$:

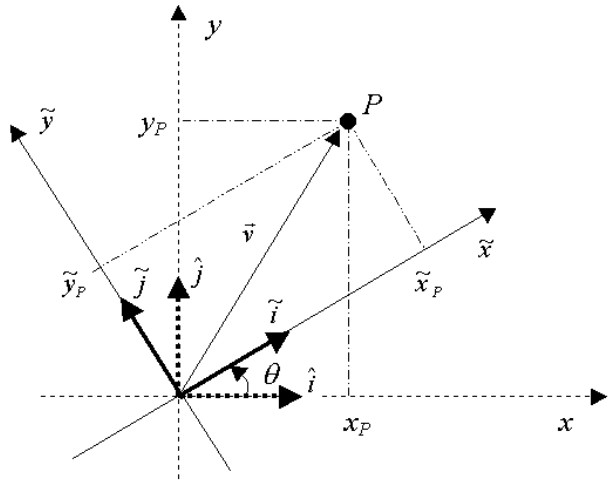


Figure 2.33: *Rotations of the coordinate system can be implemented by a specific class of linear transformations.*

$$\tilde{i}=\cos(q)\hat{i}+\sin(q)\hat{j}$$

It is easy to see that a space allowing an inner product is also a normed and metric space, since we can always define a norm in terms of the inner product by making $\|\vec{p}\| = +\sqrt{\langle \vec{p}, \vec{p} \rangle}$.

Although not every vector space is normed, metric or allows inner products, most developments in the present book deal with concepts related to vector spaces with inner products. The box *Metrics in C^N* exemplifies some of the valid norms, distances and inner products in those spaces.

Example: ***Metrics in C^N***

Let $\vec{p} = (p_1, p_2, \cdots, p_N)^T$ and $\vec{q} = (q_1, q_2, \cdots, q_N)^T$ be two generic vectors in the N-dimensional complex space C^N.

Norms of $\vec{p}$:

Euclidean: $\|\vec{p}\|_2 = \sqrt{p_1^2 + p_2^2 + \cdots + p_N^2}$

City-block: $\|\vec{p}\|_1 = |p_1| + |p_2| + \cdots + |p_N|$

Chessboard: $\|\vec{p}\|_\infty = Max\{|p_1|, |p_2|, \cdots, |p_N|\}$

Distances between $\vec{p}$ and $\vec{q}$:

Euclidean:

$$\|\vec{p} - \vec{q}\|_2 = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_N - q_N)^2}$$

City-block:

$$\|\vec{p} - \vec{q}\|_1 = |p_1 - q_1| + |p_2 - q_2| + \cdots + |p_N - q_N|$$

Chessboard:

$$\|\vec{p} - \vec{q}\|_\infty = Max\{|p_1 - q_1|, |p_2 - q_2|, \cdots, |p_N - q_N|\}$$

Inner product between $\vec{p}$ and $\vec{q}$:

$$\langle \vec{p}, \vec{q} \rangle = \vec{p}.\vec{q} = (\vec{p}^T)^* \vec{q} = p_1^* q_1 + p_2^* q_2 + \cdots p_N^* q_N$$

Example: ***Norms, Distances and Inner Products in Function Spaces***

Let $f(t)$ and $g(t)$ be two generic functions in the space of the continuous functions in the interval $[a, b]$, i.e., $C[a, b]$.

Norm of f(t):

$$\Rightarrow q_i = \begin{bmatrix} a_{i,1} & a_{i,2} & \cdots & a_{i,i} & \cdots & a_{i,N} \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_i \\ \vdots \\ p_N \end{bmatrix}$$

Consequently, the elements of the vector resulting from a linear transformation can be understood as a measure of similarity between the orientations of the input vector $\vec{p}$ and each of the respective vectors defined by the rows of the transformation matrix. This interpretation is essential for the full conceptual understanding of several properties in signal and image transforms, including the Fourier and Karhunen-Loève transforms.

2.2.5 More about Vectors and Matrices

We have thus far limited our discussion of vectors and matrices as elements of vector spaces, and as representations of linear transforms. This section provides additional concepts and properties including the more general cases of complex vectors and matrices, i.e., those vectors and matrices having complex numbers as elements, such as

$$\vec{v} = \begin{bmatrix} 2 \\ -j \\ 1+j \\ 3 \end{bmatrix} \qquad A = \begin{bmatrix} 1 & 0 & 3j \\ -j & 3 & -1-j \\ 0 & 2+j & 2 \end{bmatrix}$$

2.2.5.1 Some Basic Concepts

The *null* $N \times N$ matrix, henceforth represented as Φ, is a matrix having all elements equal to zero. A matrix A having dimension $N \times N$ is said to be a *square matrix*. Its *main diagonal* corresponds to the elements $a_{i,i}$, $i = 1, 2, \ldots, N$. A square matrix having all the elements below (above) its main diagonal equal to zero is said to be an *upper (lower) triangular matrix*, as illustrated in the following:

$$A = \begin{bmatrix} 9 & 0 & 2j \\ 0 & 3 & -1+j \\ 0 & 0 & 5 \end{bmatrix} \text{ is upper triangular,}$$

and

$$B = \begin{bmatrix} 9 & 0 & 0 \\ 0 & 3+2j & 0 \\ 2-j & 4 & 5 \end{bmatrix} \text{ is lower triangular.}$$

The *identity* matrix, represented as I, is a square matrix having ones along its main diagonal and zeroes elsewhere. For example, the 3×3 identity matrix is

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The *complex conjugate* of a matrix (vector) is obtained by taking the complex conjugate of each of its elements. For instance, the complex conjugate of the above matrix B is

$$B^* = \begin{bmatrix} 9 & 0 & 0 \\ 0 & 3-2j & 0 \\ 2+j & 4 & 5 \end{bmatrix}$$

The *derivative of a matrix* is given by the derivatives of each of its components. For instance

$$\text{If } A = \begin{bmatrix} 1 & 3t^3+t & \cos(j2\pi t) \\ j & \sin(2t) & t \\ 2t+3 & 2-j & -t^2 \end{bmatrix},$$

$$\text{then } \frac{dA}{dt} = \begin{bmatrix} 0 & 9t^2+1 & -(j2\pi)\sin(j2\pi t) \\ 0 & 2\cos(2t) & 1 \\ 2 & 0 & -2t \end{bmatrix}$$

Given an $N \times M$ matrix A and an $M \times N$ matrix B, the *product* between A and B, indicated as $C = AB$, is defined as

$$c_{i,j} = \sum_{k=1}^{M} a_{i,k} b_{k,j} \tag{2.11}$$

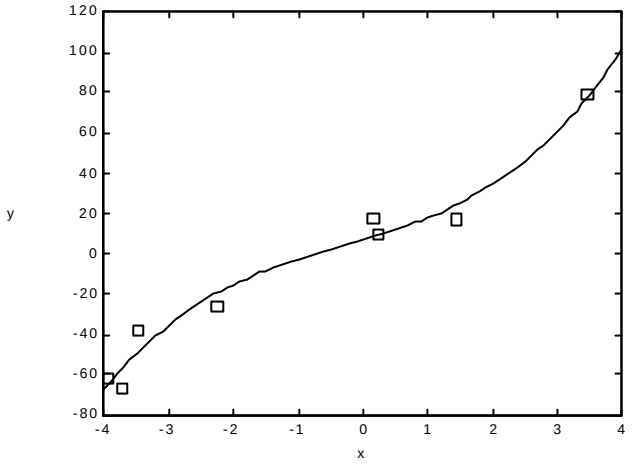


Figure 2.39: *The original points (squares) and the obtained cubic polynomial.*

The procedure illustrated in the above example can be generalized to any polynomial or function. As a matter of fact, in the particular case of straight line fitting, we have the general line equation:

$$y = a_0 + a_1 x$$

and, therefore:

$$C = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix}; \qquad \vec{b} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}; \qquad \text{and}$$

$$\vec{x} = \begin{bmatrix} a_0 \\ a_1 \end{bmatrix}$$

Applying Equation (2.15):

$$\left(C^T C\right)\vec{x} = C^T \vec{b} \Rightarrow \left(\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_N \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix} \right) \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} =$$

$$AV = V\Lambda \Leftrightarrow \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}\begin{bmatrix} a & b \\ -a & b \end{bmatrix} = \begin{bmatrix} a & b \\ -a & b \end{bmatrix}\begin{bmatrix} -2 & 0 \\ 0 & 4 \end{bmatrix}$$

Designing Matrices to Have Specific Eigenvectors: Section 2.2.5.9 has briefly addressed the situation where we wanted to identify the eigenvalues and eigenvectors of a specific square matrix A. Here we present how to build a matrix A having a specific $N\times1$ eigenvector $\vec{v}$ or a set of $k \le N$ orthogonal eigenvectors $\vec{v}_i$ with dimension $N\times1$. In the former case, the sought matrix is $A = \vec{v}\vec{v}^T$, since:

$$A = \vec{v}\vec{v}^T \Rightarrow A\vec{v} = \left(\vec{v}\vec{v}^T\right)\vec{v} = \vec{v}\left(\vec{v}^T\vec{v}\right) = r\vec{v}, r = \vec{v}^T\vec{v} \in R$$

Observe that the matrix product $\vec{v}\vec{v}^T$ can be understood in terms of the above concept of building a matrix by columns. This product implies that vector $\vec{v}$ is copied into a subsequent column j of A weighted by each of its respective coordinates v_j. This implies that matrix A columns are all equal except for a multiplicative constant, and therefore A necessarily has rank 1.

In the latter case, i.e., we want $k \le N$ orthogonal eigenvectors $\vec{v}_i$; i = 1, 2, …, k; the matrix A is also easily obtained as

$$A = \vec{v}_1\vec{v}_1^T + \vec{v}_2\vec{v}_2^T + \cdots + \vec{v}_k\vec{v}_k^T \Rightarrow A\vec{v}_i = \left(\vec{v}_1\vec{v}_1^T + \vec{v}_2\vec{v}_2^T + \cdots + \vec{v}_k\vec{v}_k^T\right)\vec{v}_i =$$
$$= \vec{v}_i\left(\vec{v}_i{}^T\vec{v}_i\right) = r\vec{v}_i, r = \vec{v}_i{}^T\vec{v}_i \in R$$

It can be verified that the so obtained matrix A has rank k.

To probe further: ***Functions, Matrices and Linear Algebra***

A good and relatively comprehensive introduction to many of the covered issues, including propositional logic, functions, linear algebra, matrices, calculus and complex numbers can be found in [James, 1996]. A more advanced reference on mathematical concepts include the outstanding textbook by [Kreyszig, 1993], which covers linear algebra, calculus, complex numbers, and much more. Other good general references are [Bell, 1990] and [Ma Fong 1997]. Interesting references covering complementary aspects related to mathematical physics, including variational calculus, are provided by [Boas, 1996] and [Dettman, 1988]. For those interested in probing further into function vector spaces (i.e., functional analysis), the books [Oden, 1979; Halmos, 1958; Michel and Herget, 1981; Kreyszig, 1993] provide excellent reading, also including good reviews of basic concepts. An interesting approach to complex number and analysis, based on visualization of the

$$\vec{n}(t) = \frac{\ddot{\vec{p}}(t)}{\left\|\ddot{\vec{p}}(t)\right\|}$$

Figure 2.48 illustrates the unit normal field for the curve in Figure 2.45. For an arc length parametrized curve $\vec{p}(t)$, the magnitude of its second derivative, i.e., $k(t) = \left\|\ddot{\vec{p}}(t)\right\| = \left\|a_R(t)\right\|$, is said to be the *curvature* of the curve at t. Observe that this expression always produces a non-negative curvature value. In the general case, i.e., the curve is not necessarily arc length parametrized, the curvature of a plane curve can be calculated as

$$k(t) = \frac{\dot{x}(t)\ddot{y}(t) - \dot{y}(t)\ddot{x}(t)}{\left(\dot{x}^2(t) + \dot{y}^2(t)\right)^{3/2}} \tag{2.21}$$

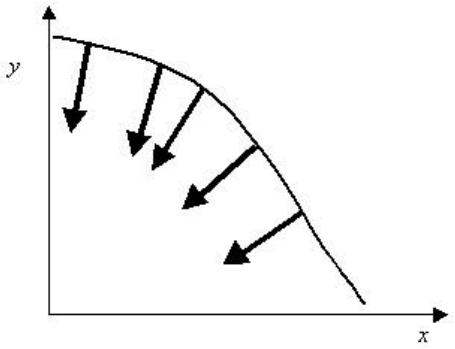


Figure 2.48: *Unitary normal field for the curve in Figure* 2.45*. Only a few of the infinite unit normal vectors are shown. All vectors have unit magnitude.*

Unlike in the previous equation, the signal of the curvature calculated by Equation (2.21) can be positive or negative, indicating the respective local concavity (see below). The curvature is an extremely important concept because it nicely expresses the local "geometric nature" of a curve. For instance, if zero curvature is observed along a portion of a curve, this portion will correspond to a straight line segment. On the other hand, a constant curvature value indicates a circle or an arc of circle. Generally, the curvature value is proportional to the local variation of the curve. More precisely, as defined above, it corresponds to the radial acceleration magnitude of the arclength parametrized version of the curve, therefore indicating how fast the tangent vector changes its orientation. Another interesting feature exhibited by

curvature is the fact that it is *invariant* to rotations, translations and reflections of the original curve (observe that it is not invariant to scaling). Moreover, the curvature is *information preserving* in the sense that it allows the original curve to be recovered, up to a rigid body transformation (i.e., combinations of translations, rotations and reflections that do not alter the size of the shape – see Section 4.9.3). Thus we have that, if $k(t)$ is a differentiable function expressing the curvature of a curve from t_0 to t, its reconstruction can be obtained as

$$\vec{p}(t) = \left(\int_{t_0}^{t} \cos(\alpha(r)dr) + c_1, \int_{t_0}^{t} \sin(\alpha(r)dr) + c_2 \right), \text{ where } \alpha(t) = \int_{t_0}^{t} k(r)dr + c_3$$

and (c_1, c_2) and c_3 represent the translation vector and the rotation angle, respectively.

Although it is clear from the above curvature definition $k(t) = \left\| \ddot{\vec{p}}(t) \right\| = \left\| a_R(t) \right\|$ that its values are non-negative, it is often interesting to consider an alternative definition allowing negative curvature values. This is done by considering the standard coordinate system $(\hat{i}, \hat{j}, \hat{k})$ of R^3 (i.e., $\hat{k} = \hat{i} \wedge \hat{j}$). The signed curvature $k_s(t)$, which can be calculated by Equation (2.21), can be defined as

$$k_s(t) = \operatorname{sgn}\left\{ \left\langle \dot{\vec{p}}(t) \wedge \vec{n}(t), \hat{i} \wedge \hat{j} \right\rangle \right\} k(t)$$

This means that positive curvature will be obtained whenever the sense of the vector $\dot{\vec{p}}(t) \wedge \vec{n}(t)$ agrees with that of the unit vector $\hat{i} \wedge \hat{j}$. Negative curvature is obtained otherwise.

An immediate advantage allowed by the signed curvature is that its sign provides indication about the *concavity* at each of the curve points. It should however be taken into account that the sign of $k_s(t)$ depends on the sense of the curve, and will change with the sense in which the curve is followed and with the sign of t. Figure 2.49 illustrates the change of curvature sign considering two senses along a closed curve, and the respective concavity criteria. A point where $k_s(t) = 0$ and $\dot{k}_s(t) \neq 0$ is said to be an *ordinary inflection point*. Such a point corresponds to a change of concavity along the curve.

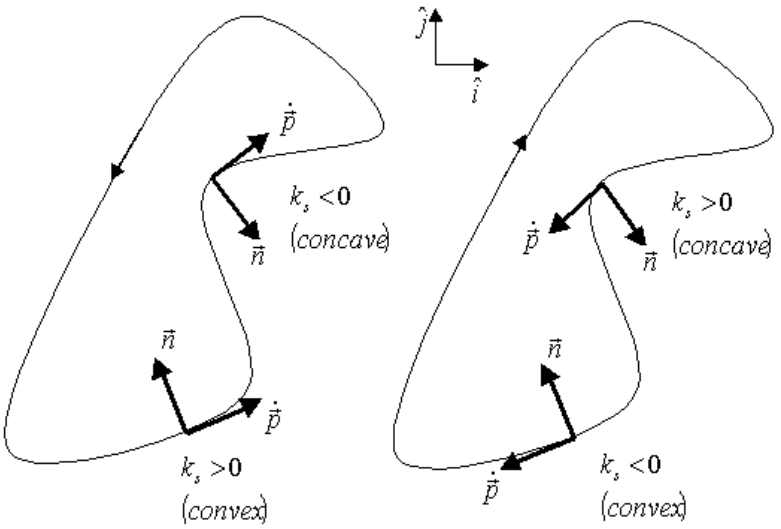


Figure 2.49: *The sign of the signed curvature k_s changes as the sense of the curve is inverted. The concavity criterion also depends on the adopted curve sense.*

The curvature can also be geometrically understood in terms of osculating circles and radius of curvature. Consider Figure 2.50 and assume that the curvature is never zero. The circle having radius $r(t) = 1/k(t)$, called *radius of curvature*, and centered at $\vec{u}(t) + \vec{n}(t)/k(t)$, where $\vec{u}(t)$ is an arc length parametrized curve and $\vec{n}(t)$ is the unit normal field to $\vec{u}(t)$, is called the *osculating circle* at t.

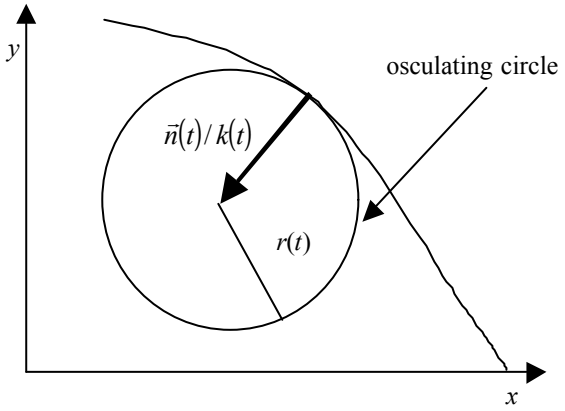
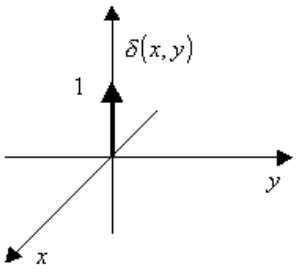
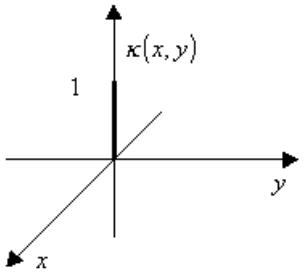
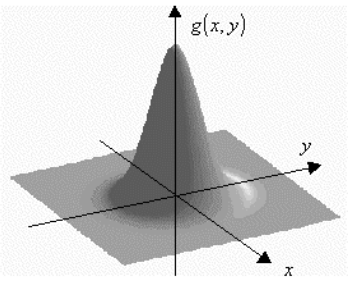


Figure 2.50: *Osculating circle and radius of curvature at a point t.*

Table 2.4: *Some particularly important bivariate real scalar fields.*

2D Dirac Delta:
$\boldsymbol{d}(x,y)=\begin{cases}\text{not defined} & \text{if } x=0 \text{ and } y=0\\ 0 & \text{otherwise}\end{cases}$ and $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\boldsymbol{d}(x,y)dxdy=1$ 
2D Kronecker Delta:
$\boldsymbol{k}(x,y)=\begin{cases}1 & \text{if } x=0 \text{ and } y=0\\ 0 & \text{otherwise}\end{cases}$ 
A circularly symmetric Gaussian:
$g(x,y)=\exp\left(-a\left(x^2+y^2\right)\right);$ *where* $a>0$ 

A real vector field $\vec{g}:R^N\rightarrow R^M \mid \vec{q}=\vec{g}(\vec{v})$ is said to be *continuous at a point* $\vec{q}_0 \in R^M$ if for each open ball B_e with radius $\boldsymbol{e}$ centered at $\vec{q}_0$ (i.e., the vectors $\vec{q}\in R^M$ such as $\|\vec{q}-\vec{q}_0\|<\boldsymbol{e}$), it is always possible to find an open ball B_d with radius $\boldsymbol{d}$ centered at $\vec{v}_0 \in R^N$ (i.e., the vectors $\vec{v}\in R^N$ such as $\|\vec{v}-\vec{v}_0\|<\boldsymbol{d}$), such as the mapping of this ball by the vector field $\vec{g}$, i.e., $\vec{g}(B_d)$, falls complely inside B_e. A vector field that is continuous at all the points of its domain is simply said to be *continuous.* The continuity of a

vector field can be understood as a particular case of the above definition in the case M=1.

Given a bivariate function $z = g(x,y)$, we can think about this function in terms of unidimensional functions by taking slices of $g(x,y)$ along planes perpendicular to the (x,y) plane. Observe that any of such planes is completely specified by the straight line L defined by the intersection of this perpendicular plane with the plane (x,y). It is particularly useful to define such lines in a parametric fashion (see Section 2.3.1), which can be done by imposing that these lines are parallel to a vector $\vec{v} = (a, b)$ and contain a specific point, identified by the vector $\vec{p}_0 = (c, d)$. Therefore the general form of these straight lines is $L : \vec{p} = \vec{v}t + \vec{p}_0 = (at + c, bt + d)$.

Since this line will be used to follow the line along the slice, defining a function of a single variable, unit speed (see Section 2.3.2), and therefore arc length parametrization, is required. This can be easily achieved by imposing that $a^2 + b^2 = 1$. Now, the values of the function g along the slice can easily be obtained by substituting the x and y coordinates of the positions defining the line L into the function $z = g(x,y)$ to yield $z = g(at+c, bt+d)$, which is a function on the single variable t. The Box *Slicing a Circularly Symmetric Bivariate Gaussian* provides an example about scalar fields slicing.

Example: ***Slicing a Circularly Symmetric Bivariate Gaussian***

Consider the Gaussian scalar field $F(x,y) = \exp\{-(x^2 + y^2)/4\}$. Obtain the univariate function defined by slicing this field along the plane that is orthogonal to the (x,y) plane and contains the line L, which is parallel to the vector $\vec{v} = (1, 0.5)$ and passes onto the point $\vec{b} = (0, -2)$, which defines the origin along the slice.

Solution:

First, we obtain the equation of the line L. In order to have arc length parametrization, we impose $\|\vec{v}\| = 1 \Rightarrow \tilde{\vec{v}} = \frac{\vec{v}}{\|\vec{v}\|} = \left(\frac{1}{\sqrt{1.25}}, \frac{0.5}{\sqrt{1.25}}\right)$. Now, the sought line equation can be expressed as

$$\vec{p}(s) = (x(s), y(s)) = \vec{b} + \tilde{\vec{v}}s = (0, -2) + \left(\frac{1}{\sqrt{1.25}}, \frac{0.5}{\sqrt{1.25}}\right)s \Leftrightarrow \begin{cases} x(s) = \frac{s}{\sqrt{1.25}} \\ y(s) = -2 + \frac{0.5s}{\sqrt{1.25}} \end{cases}$$

Substituting these coordinates into the scalar field we obtain the following:

speaking, these two operations provide a means for "combining" or "mixing" the two functions as to allow important properties, such as the convolution and correlation theorems to be presented in Sections 2.7.3.7 and 2.7.3.8. In addition, convolution provides the basis for several filters, and correlation provides a means for comparing two functions. These operations are presented in the following, first with respect to continuous domains, then to discrete domains.

2.5.1 Continuous Convolution and Correlation

Let $g(t)$ and $h(t)$ be two real or complex functions. The *convolution* between these functions is the univariate function resulting from the operation defined as

$$q(\tau) = g(\tau) * h(\tau) = (g * h)(\tau) = \int_{-\infty}^{\infty} g(t)h(\tau - t)dt \tag{2.22}$$

The *correlation* between two real or complex functions $g(t)$ and $h(t)$ is the function defined as

$$q(\boldsymbol{t}) = g(\boldsymbol{t}) \circ h(\boldsymbol{t}) = (g \circ h)(\boldsymbol{t}) = \int_{-\infty}^{\infty} g^*(t)h(\boldsymbol{t} + t)dt \tag{2.23}$$

As is clear from the above equations, the correlation and convolution operations are similar, except that in the latter the first function is conjugated and the signal of the free variable t in the argument of $h(t)$ is inverted. As a consequence, while the convolution can be verified to be commutative, i.e.,

$$(g * h)(\boldsymbol{t}) = \int_{-\infty}^{\infty} g(t)h(\boldsymbol{t} - t)dt \overset{a=t-t}{=} \int_{-\infty}^{\infty} g(\boldsymbol{t} - a)h(a)da = (h * g)(\boldsymbol{t})$$

we have that the correlation is not, i.e.

$$(g \circ h)(\boldsymbol{t}) = \int_{-\infty}^{\infty} g^*(t)h(\boldsymbol{t} + t)dt \overset{a=t+t}{=} \int_{-\infty}^{\infty} g^*(a - \boldsymbol{t})h(a)da \neq (h \circ g)(\boldsymbol{t})$$

However, in case both $g(t)$ and $h(t)$ are real, we have

$$(g \circ h)(\boldsymbol{t}) = \int_{-\infty}^{\infty} g(t)h(\boldsymbol{t} + t)dt \overset{a=t+t}{=} \int_{-\infty}^{\infty} g(a - \boldsymbol{t})h(a)da = (h \circ g)(-\boldsymbol{t})$$

In other words, although the correlation of two real functions is not commutative, we still have $(g \circ h)(\tau) = (h \circ g)(-\tau)$. In case both $g(t)$ and $h(t)$ are real and even, then $(g \circ h)(\tau) = (h \circ g)(\tau)$. For real functions, the convolution and correlation are related as

$$g(\tau) * h(-\tau) = \int_{-\infty}^{\infty} g(t)h(t-\tau)dt \overset{a=t-t}{=} \int_{-\infty}^{\infty} g(a+\tau)h(a)da = h(\tau) \circ g(\tau)$$

If, in addition, $h(t)$ is even, we have

$$g(\tau) * h(\tau) = \int_{-\infty}^{\infty} g(t)h(t-\tau)dt = h(\tau) \circ g(\tau) = (g \circ h)(-\tau)$$

An interesting property is that the convolution of any function $g(t)$ with the Dirac delta reproduces the function $g(t)$, i.e.

$$(g * \delta)(\tau) = \int_{-\infty}^{\infty} g(t)\delta(\tau - t)dt = \int_{-\infty}^{\infty} g(\tau)\delta(\tau - t)dt = g(\tau)\int_{-\infty}^{\infty} \delta(\tau - t)\,dt = g(\tau)$$

An effective way to achieve a sound conceptual understanding of the convolution and correlation operations is through graphical developments, which is done in the following with respect to the convolution. Let $g(t)$ and $h(t)$ be given by Equations (2.24) and (2.25), as illustrated in Figure 2.53.

$$g(t) = \begin{cases} 1.5 & if \quad -1 < t \le 0 \\ 0 & otherwise \end{cases} \tag{2.24}$$

and

$$h(t) = \begin{cases} 2 & if \quad 0 < t \le 2 \\ 0 & otherwise \end{cases} \tag{2.25}$$

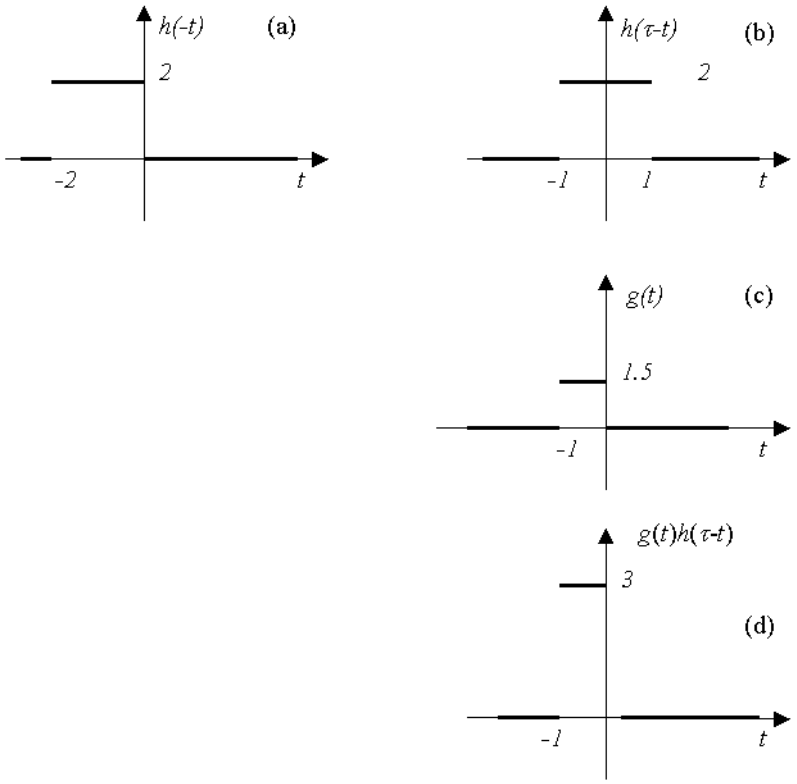


Figure 2.54: *Illustration of the basic operations involved in the convolution of the functions g(t) and h(t). See text for explanation.*

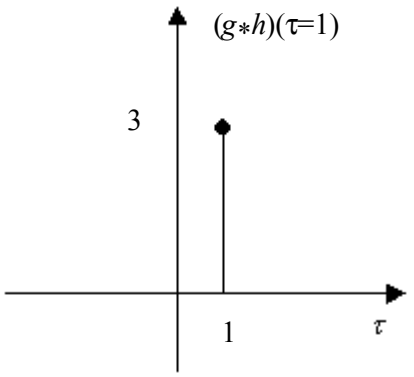


Figure 2.55: *The convolution (g* h)(**t**) for **t**=**1**.

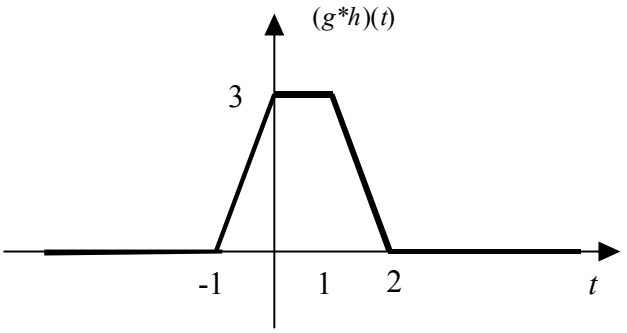


Figure 2.56: *The complete convolution* $(g * h)(t)$.

The correlation can be understood in a similar manner, except for the fact that the second function is not reflected and, for complex functions, by the conjugation of the first function. Figure 2.57 shows the correlation of the above real functions, i.e., $g(t) \circ h(t)$.

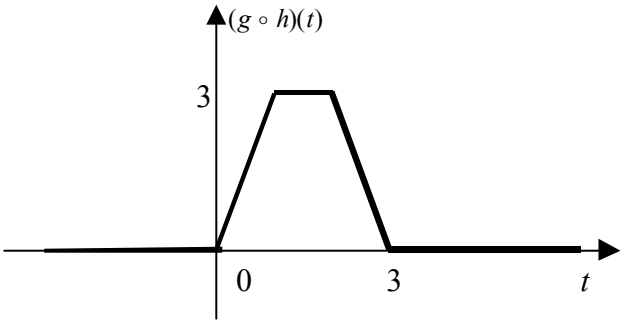


Figure 2.57: *The correlation* $g(t) \circ h(t)$.

Let us now consider that both $g(t)$ and $h(t)$ have finite extension along the domain, i.e., $g(t), h(t) = 0$ for $t < r$ and $t > s$. Recall from Section 2.2.4 that the inner product between two functions $g(t)$ and $h(t)$ with respect to the interval $[a, b]$ is given by

$$\langle g, h \rangle = \int_a^b g^*(t) h(t) dt$$

Observe that this equation is similar to the correlation equation, except that the latter includes the parameter $\boldsymbol{t}$ in the argument of the second function, which allows the second function to be shifted along the x-axis with respect to the first function. As a matter of fact, for each fixed value of $\boldsymbol{t}$, the correlation equation becomes an internal product between the first function and the

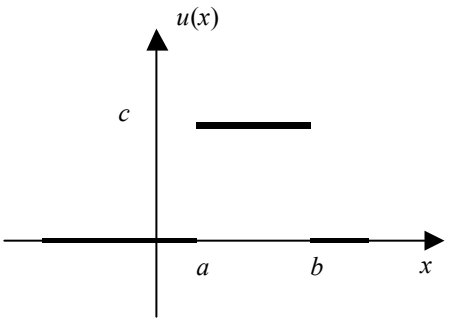
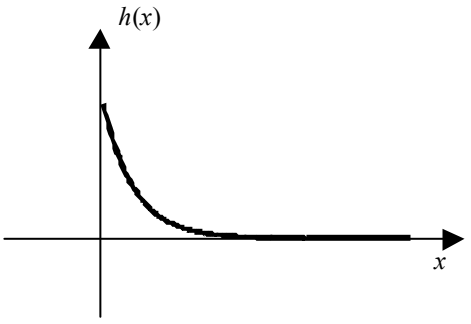
2.6.2.5 Random Variables Transformations, Conditional Density Functions and Discrete Random Variables

Given a random variable X, it is possible to transform it in order to obtain a new random variable Y. A particularly useful *random variable transformation*, called *normal transformation*, is obtained by applying the following equation:

$$Y = \frac{X - E[X]}{\sigma_X}$$

It can be verified that the new random variable Y has zero mean and unit standard deviation.

Table 2.6: *Important density probability functions.*

Uniform: $u(x) = \begin{cases} c = \dfrac{1}{b-a} & a < x \le b \\ 0 & otherwise \end{cases}$ $E[x] = \dfrac{a+b}{2}$ $Var[x] = \dfrac{(b-a)^2}{12}$	
Exponential: $h(x) = \alpha \exp\{-\alpha x\}, \quad x > 0$ $E[x] = \dfrac{1}{\alpha}$ $Var[x] = \dfrac{1}{\alpha^2}$	
Gaussian: $g(x) =$ $= \dfrac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\dfrac{1}{2}\left(\dfrac{x-\mu}{\sigma}\right)^2\right\}$	

and

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} p(x_1, x_2, \cdots, x_N)dx_1 dx_2 \cdots dx_N = 1$$

Given a density function $g(\vec{x}) = g(x_1, x_2, \cdots, x_N)$, it is possible to define *marginal density function*s for any of its random variables x_i, integrating along all the other variables, i.e.,

$$p(x_i) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(x_1, x_2, \cdots x_i, \cdots, x_N)\, dx_1 dx_2 \cdots dx_{i-1} dx_{i+1} \cdots dx_N$$

where the fixed variables are represented by a tilde. An example of joint density function is the multivariate Gaussian, given by the following:

$$p(\vec{x}) = \frac{1}{(2\pi)^{N/2}\sqrt{Det\{K\}}} \exp\left\{-\frac{1}{2}(\vec{x} - \vec{m}_{\vec{X}})^T K^{-1}(\vec{x} - \vec{m}_{\vec{X}})\right\}$$

This density function is completely specified by the mean vector $\vec{m}_{\vec{X}}$ and the covariance matrix K (see below).

The moments and central moments of an $N\times1$ random vector $\vec{X}$, modelled by the joint density function $p(\vec{x})$, are defined as

$(n_1, n_2, \ldots, n_N)$-th moment:

$$M_{(n_1,n_2,\cdots,n_N)}[\vec{X}] = E\left[X_1^{n_1} X_2^{n_2} \cdots X_N^{n_N}\right] =$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x_1^{n_1} x_2^{n_2} \cdots x_N^{n_N}\, p(\vec{x})dx_1 dx_2 \cdots dx_N$$

$(n_1, n_2, \ldots, n_N)$-th central moment:

$$M_{(n_1,n_2,\cdots,n_N)}[\vec{X}] = E\left[(X_1 - E[X_1])^{n_1}(X_2 - E[X_2])^{n_2} \cdots (X_N - E[X_N])^{n_N}\right] =$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}(x_1 - E[X_1])^{n_1}(x_2 - E[X_2])^{n_2} \cdots (x_N - E[X_N])^{n_N}\, p(\vec{x})dx_1 dx_2 \cdots dx_N$$

where $n_1, n_2, \cdots, n_N \in \{0, 1, 2, \cdots\}$. As with scalar random variables, such moments provide *global* information about the behaviour of the random

In case Cov(X, Y) = 0, we say that the random variables X and Y are *uncorrelated*. It is important to note that the fact that two random variables X and Y are uncorrelated does not imply that they are independent (see Section 2.6.1), but independence implies uncorrelation. The covariance can be alternatively expressed as

$$Cov(X_i, X_j) = E[X_i X_j] - E[X_i]E[X_j]$$

In addition, it is interesting to observe that

$$\sigma_{X_i}^2 = Var(X_i) = Cov(X_i, X_i)$$

and, consequently, the standard deviation of the random variable X_i can be alternatively expressed as

$$\sigma_{X_i} = +\sqrt{Cov(X_i, X_i)}$$

Since the covariance between two random variables is not dimensionless, it becomes interesting to define the *correlation coefficient*, which provides a dimensionless and relative measure of the correlation between the two variables. The correlation coefficient $CorrCoef(X_i, X_j)$ is defined as

$$CorrCoef(X_i, X_j) = E\left[\left(\frac{X_i - \mu_{X_i}}{\sigma_{X_i}}\right)\left(\frac{X_j - \mu_{X_j}}{\sigma_{X_j}}\right)\right] = \frac{Cov(X_i, X_j)}{\sigma_{X_i}\sigma_{X_j}}$$

An important property of the correlation coefficient is that $\left| CorrCoef(X_i, X_j) \right| \leq 1$.

It should be observed that when the means of all the involved random variables are zero, the correlation between two variables becomes equal to the respective covariance. Similarly, the covariance becomes equal to the correlation coefficient when the standard deviations of all involved random variables have unit value. When all means are zero and all standard deviations are one, the correlations are equal to the covariances, which in turn are equal to the correlation coefficients. In other words, the covariance and correlation coefficient between random variables can be understood in terms of correlations between versions of those random variables transformed in order to have zero means and unit standard deviations (i.e., a normal transformation), respectively. These important and practical properties are summarized in the following:

$$P_g\{f\} = |G(f)|^2 = G(f)^* G(f)$$

An important property of the power spectrum is that it does not change as the original function is shifted along its domain, which is explored by the so-called Fourier descriptors for shape analysis (see Section 6.5).

Consider the following example:

Example: ***Fourier Transform I***

Calculate the Fourier transform and power spectrum of the function:

$$g(t) = \exp\{-t\},\ 0 \le t < \infty$$

First, we apply Equation (2.48):

$$\Im\{g(t)\} = \int_0^\infty \exp\{-t\}\exp\{-j2\pi ft\}dt = \int_0^\infty \exp\{-t(j2\pi f + 1)\}dt =$$

$$= -\frac{1}{1+j2\pi f}\left[\exp\{-t(j2\pi f + 1)\}\right]_0^\infty =$$

$$= -\left(\frac{1}{1+j2\pi f}\right)[0-1] = \frac{1}{1+j2\pi f}\left(\frac{1-j2\pi f}{1-j2\pi f}\right) = \frac{1-j2\pi f}{1+(2\pi f)^2} = G(f)$$

Thus, the Fourier transform $G(f)$ of $g(t)$ is a complex function with the following real and imaginary parts, shown in Figure 2.75.

$$\mathrm{Re}\{G(f)\} = \frac{1}{1+(2\pi f)^2} \quad \text{and} \quad \mathrm{Im}\{G(f)\} = \frac{-2\pi f}{1+(2\pi f)^2}$$

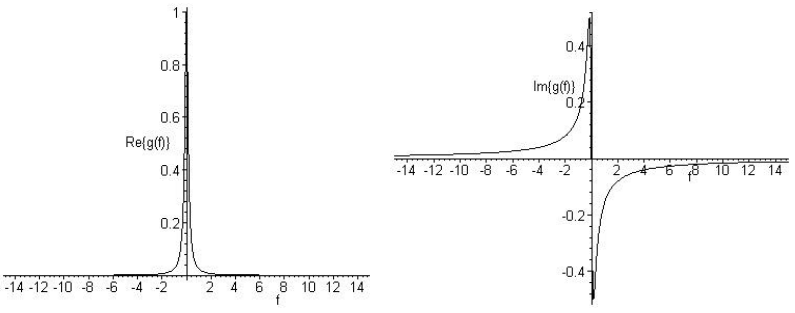


Figure 2.75: *The real and imaginary parts of G(f).*

Alternatively, in the magnitude and phase representation, we have

$$|G(f)| = \sqrt{\frac{1+4\pi^2 f^2}{\left[1+(2\pi f)^2\right]^2}} \quad \text{and} \quad \Phi\{G(f)\} = atg\{-2\pi f\}$$

The power spectrum can be calculated as

$$P_g\{f\} = G(f)^* G(f) = \frac{1-j2\pi f}{1+(2\pi f)^2}\frac{1+j2\pi f}{1+(2\pi f)^2} = \frac{1+(2\pi f)^2}{\left(1+(2\pi f)^2\right)^2} = |G(f)|^2$$

and is illustrated in Figure 2.76.

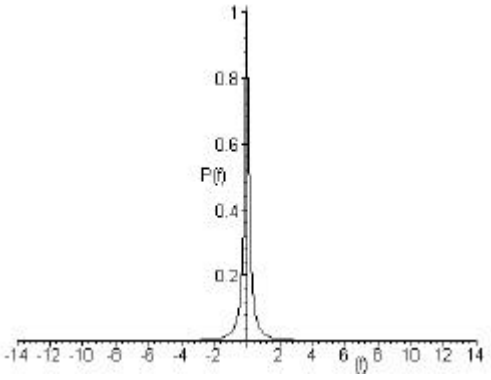


Figure 2.76: *The power spectrum of the function g(t).*

Although the Fourier transform of a complex function is usually (as in the above example) a complex function, it can also be a purely real (or imaginary) function. On the other hand, observe that the power spectrum is always a real function of the frequency. Consider the following example:

Example: ***Fourier Transform II***

Calculate the Fourier transform of the function:

$$g(t) = \begin{cases} 1 & if\ -a \le t < a \\ 0 & otherwise \end{cases}$$

Applying Equation 2.48:

$$\sin(2\pi f_0 t) \quad \leftrightarrow \quad \frac{j}{2}\delta(f+f_0) - \frac{j}{2}\delta(f-f_0)$$

In a similar fashion

$$\Im\{\cos(2\pi f_0 t)\} = \Im\left\{\frac{\exp\{j2\pi f_0 t\} + \exp\{-j2\pi f_0 t\}}{2}\right\} =$$

$$= \frac{1}{2}\Im\{\exp\{j2\pi f_0 t\}\} + \frac{1}{2}\Im\{\exp\{-j2\pi f_0 t\}\} = \left(\frac{1}{2}\delta(f-f_0) + \frac{1}{2}\delta(f+f_0)\right)$$

and , therefore $\cos(2\pi f_0 t) \quad \leftrightarrow \quad \frac{1}{2}\delta(f+f_0) + \frac{1}{2}\delta(f-f_0)$.

2.7.3.7 The Convolution Theorem

This important property of the Fourier transform is expressed as follows:

Let $g(t) \leftrightarrow G(f)$ and $h(t) \leftrightarrow H(f)$
Then $(g * h)(t) \leftrightarrow G(f)H(f)$
and $g(t)h(t) \leftrightarrow (G * H)(f)$

where $g(t)$ and $h(t)$ are generic complex functions. See Sections 2.7.4 and 7.2 for applications of this theorem.

2.7.3.8 The Correlation Theorem

Let $g(t)$ and $h(t)$ be real functions defining the Fourier pairs $g(t) \leftrightarrow G(f)$ and $h(t) \leftrightarrow H(f)$. Then $(g \circ h)(t) \leftrightarrow G^*(f)H(f)$.

2.7.3.9 The Derivative Property

Let the generic Fourier pair $g(t) \leftrightarrow G(f)$ and a be any non-negative real value. Then

$$\frac{d^a g(t)}{dt^a} \leftrightarrow D_a(f)G(f) \tag{2.51}$$

where $D_a(f) = (j2\pi f)^a$. This interesting property, which is used extensively in the present book (see Section 7.2), allows not only the calculation of many derivatives in terms of the respective Fourier transforms,

$$H(f) = G(f)[\Psi_{1/2L}(f)] = \sum_{i=-\infty}^{\infty} G\left(f - i\frac{1}{2L}\right)\delta\left(f - i\frac{1}{2L}\right)$$

The periodical function $h(t)$ and its respective Fourier transform $H(f)$ are shown in Figure 2.81(b) and (d), respectively, considering $a = 1$ and $L = 2$.

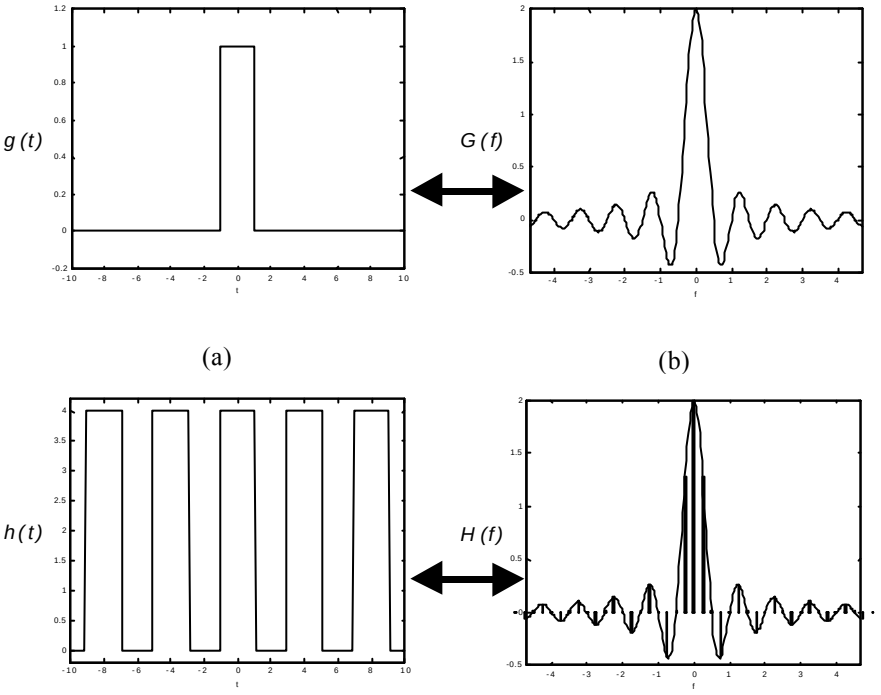
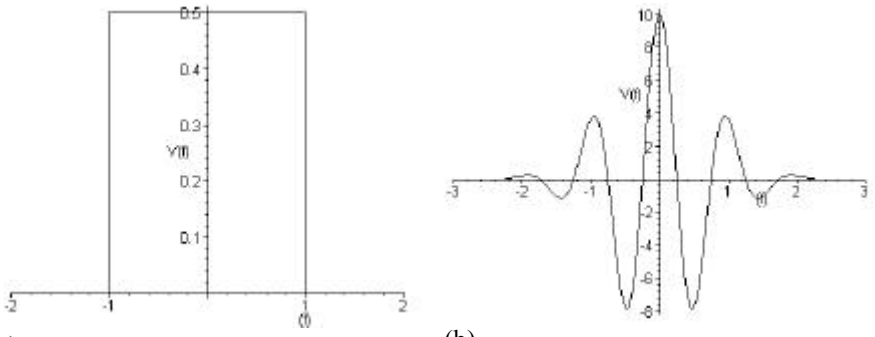


Figure 2.81: *The function g(t) (a) and its Fourier transform G(f) (b). The periodical version* $h(t) = g(t) * \Psi_{2L}(t)$ *of g(t), for a = 1 and L = 2 (c), and its respective Fourier transform H(f) (d).*

A completely similar effect is observed by sampling the function $g(t)$, implying the respective Fourier transform to be periodical. The above results are summarized below:

$g(t)$	$G(f)$
Periodical	Discrete
Discrete	Periodical

Observe that the low-pass filter in Figure 2.82(a) attenuates all frequencies, but the attenuation is smaller for the lower frequencies. Low-pass filtering tends to produce functions that are *smoother* and *more intensely correlated* than the original function $h(t)$ – see Section 2.6.5.



(a) (b)

Figure 2.82: *Two possible low-pass filtering functions.*

A typical low-pass filtering function is the zero-mean Gaussian (see Section 2.1.4). It is interesting to relate the Gaussian filtering function to its inverse Fourier transform, since this allows us to understand the filtering effect in terms of the standard deviation σ of the Gaussian respectively defined in the time domain (the higher this value, the more intense the low-pass filtering effect). Recall from Section 2.7.3 and Equation (2.50) that the Gaussian in the frequency domain has as parameter $\sigma_f = 1/(2\pi\sigma)$. The henceforth adopted Gaussian filtering function $V(f)$ and its respective inverse Fourier transform (which is a Gaussian in the strict sense), are given in terms of the following Fourier transform pair:

$$g_\sigma(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{t}{\sigma}\right)^2\right\} \leftrightarrow V(f) = \exp\left\{-2(\pi\sigma f)^2\right\}$$

Observe that the above Gaussian filter function $V(f)$ always varies between 0 and 1. Figure 2.83 illustrates the process of Gaussian low-pass filtering. The Fourier transform $H(f)$ (b) of the function $h(t)$ to be filtered (a) is multiplied by the filtering function $V(f)$ (c), which in this case is the Gaussian $V(f) = \exp\left\{-2(\pi\sigma f)^2\right\}$ with $\sigma = 0.1$, and the filtered function (d) is obtained by taking the inverse Fourier transform of $H(f)V(f)$. The effect of this filtering process over the original function, a cosine function corrupted by additive uniform noise, is clear in the sense that the higher frequency components of $h(t)$, i.e., the sharp oscillations along the cosine function, have been substantially attenuated, although at the expense of a substantial change

in the amplitude of $h(t)$. An additional discussion about Gaussian filtering, in the context of contour processing, is presented in Section 7.2.3.

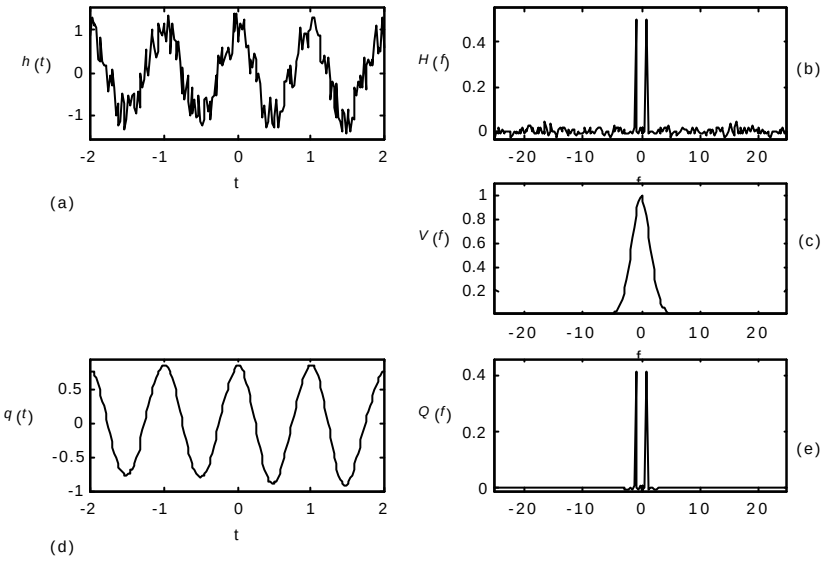


Figure 2.83: *The function h(t) to be low-pass filtered (a), its respective Fourier transform (b), the filtering function (in Fourier domain) (c), the filtered function q(t) (d), and its respective Fourier transform (e).*

The second class of filters, known as *high-pass filters*, act conversely to the low-pass filters, i.e., by attenuating the magnitude of the low frequency components of the signal, while the higher frequency components are allowed to pass. Such an attenuation should again be understood in *relative* terms. An example of high-pass filter is the *complemented Gaussian V(f)*, defined as

$$g_s(t) = \delta(t) - \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{t}{\sigma}\right)^2\right\} \leftrightarrow V(f) = 1 - \exp\left\{-2(\pi\sigma f)^2\right\}$$

It is interesting to observe that the complemented Gaussian filter function always varies between 0 and 1. This function is illustrated in Figure 2.84 for $\sigma = 0.25$.

As illustrated in Figure 2.85, a high-pass filter tends to accentuate the most abrupt variations in the function being filtered, i.e., the regions where the derivative magnitude is high (in image processing and analysis, such abrupt variations are related to the image contrast). In other words, high-pass filtering reduces the correlation and redundancy degree in the original signal.

called *deconvolution* (this follows from the fact that the filtering process can be alternatively understood as a convolution in the time space). If the original function $h(t)$ was filtered by a function $V(f)$, yielding $q(t)$, we may attempt to recover the original function by dividing the Fourier transform $Q(f)$ of the filtered function by the filter function $V(f)$ and taking the inverse Fourier transform as the result. Thus, the sought recovered function would be obtained as $h(t) = \Im^{-1}\{Q(f)/V(f)\}$. However, this process is not possible whenever $V(f)$ assumes zero value. In practice, the situation is complicated by the presence of noise in the signal and numeric calculation. Consequently, effective deconvolution involves more sophisticated procedures such as Wiener filtering (see, for instance, [Castleman, 1992]).

2.7.5 The Discrete One-Dimensional Fourier Transform

In order to be numerically processed by digital computers, and to be compatible with the discrete signals produced by digital measuring systems, the Fourier transform has to be reformulated into a suitable discrete version, henceforth called *discrete Fourier transform – DFT*.

First, the function g_i to be Fourier transformed is assumed to be a uniformly sampled (spaced by Δt) series of measures along time, which can be modelled in terms of multiplication of the original, continuous function $\tilde{g}(t)$ with the sampling function $\Psi_{\Delta t}(t) = \sum_{i=-\infty}^{\infty} \delta(t - i\Delta t)$. Second, by being the result of some measuring process (such as the recording of a sound signal) the function g_i is assumed to have *finite duration* along the time domain, let us say from time $a = i_a \Delta t$ to $b = i_b \Delta t$. The function g_i is henceforth represented as

$$g_i = \tilde{g}(i\Delta t)$$

Observe that the discrete function g_i can be conveniently represented in terms of the vector $\vec{g} = (g_{i_a}, g_{i_a+1}, \cdots, g_{i-1}, g_i, g_{i+1}, \cdots, g_{i_b-1}, g_{i_b})$. Figure 2.89 illustrates the generic appearance (i.e., sampled) of the function g_i.

$$N = \frac{1/\Delta f}{\Delta t} = \frac{1}{\Delta t \Delta f}$$

Observe that we have $N = 1/(\Delta t \Delta f)$ instead of $N = 1/(\Delta t \Delta f) + 1$ because we want to avoid repetition at the extremity of the period, i.e., the function is sampled along the interval [*a*, *b*). The number M of sampling points in any period of the output function $H(f)$ is similarly given by

$$M = \frac{1/\Delta t}{\Delta f} = \frac{1}{\Delta t \Delta f}$$

By considering $N = M$, i.e., the number of sampling points representing the input and output DFT functions are the same (which implies vectors of equal sizes in the DFT), we have

$$N = M = \frac{1}{\Delta t \Delta f} \tag{2.53}$$

Since the input function is always periodical, the DFT can be numerically approximated in terms of the Fourier series, which can be calculated by considering any full period of the input function $h(t)$. In order to be numerically processed, the Fourier series given by Equation (2.47) can be rewritten as follows. First, the integral is replaced by the sum symbol and the continuous functions are replaced by the above sampled input and output functions. In addition, this sum is multiplied by Δt because of the numerical integration and the relationship $n = 2Lf$ is taken into account, yielding

$$\begin{aligned} G_k = G(k\Delta f) = c_{n=2Lf} &= \frac{1}{2L}\int_0^{2L} h(t)\exp\left\{-\frac{j\pi nt}{L}\right\} = \\ &= \Delta t \frac{1}{2L}\sum_{i=0}^{N-1} h(i\Delta t)\exp\left\{-\frac{j\pi(2Lf)(i\Delta t)}{L}\right\} = \\ &= \Delta t \frac{1}{2L}\sum_{i=0}^{N-1} h(i\Delta t)\exp\{-j2\pi(k\Delta f)(i\Delta t)\} \end{aligned}$$

Observe that we have considered the time interval [0, 2*L*), in order to avoid redundancies. By considering the input function as having period $2L = \frac{1}{\Delta f}$, we obtain

$$H_k = H(k\Delta f) = \Delta t \Delta f \sum_{i=0}^{N-1} h(i\Delta t)\exp\{-j2\pi(k\Delta f)(i\Delta t)\}$$

Now, from Equation (2.53) we have $\Delta t \Delta f = \frac{1}{N}$, which implies that

$$H_k = H(k\Delta f) = \frac{1}{N}\sum_{i=0}^{N-1} h(i\Delta t)\exp\left\{-\frac{j2\pi ik}{N}\right\} \tag{2.54}$$

This equation, which is commonly known as the *discrete Fourier transform equation*, allows us to numerically estimate the Fourier series of the periodical function $h(t)$. It is easily verified that the computational execution of Equation (2.54) for each specific value of k demands N basic steps, being therefore an algorithm of complexity order $O(N)$. Since the complete Fourier series involves N calculations of this equation (i.e., k = 0, 1, … , N-1), the overall number of basic operations in the DFT algorithm is of $O(N^2)$.

2.7.6 Matrix Formulation of the DFT

Equation (2.54) can be compactly represented in matrix form, which is developed in the following. By defining the abbreviations

$$w_{k,i} = \exp\left\{-\frac{j2\pi ik}{N}\right\},\ h_i = h(i\Delta t), \text{ and } H_k = H(k\Delta f),$$

Equation (2.54) can be rewritten as

$$H_k = \frac{1}{N}\sum_{i=0}^{N-1} w_{k,i} h_i \tag{2.55}$$

Before proceeding with the derivation of the matrix form of the DFT, it is interesting to have a closer look at the discretized kernel function $w_{k,i} = \exp\left\{-\frac{j2\pi ik}{N}\right\}$. Let us introduce $w_{ki} = w_{k,i}$ and observe that $w_{k,i} = w_{i,k}$; for instance $w_4 = w_{1,4} = w_{4,1} = w_{2,2}$. From Section 2.1, it is easy to see that the complex exponential kernel function $w_{i,k}$ in the above equation can be understood as the sequence of complex points uniformly distributed

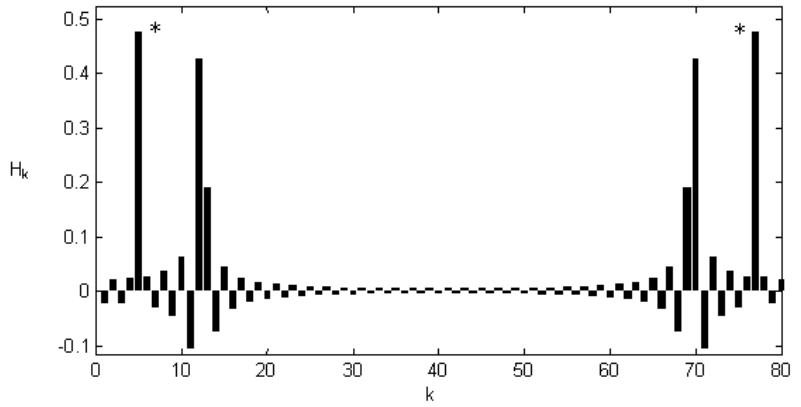


Figure 2.98: *The DFT output function H(kΔf). The pairof lower frequency peaks for* $f_1 = 1$ *Hz are indicated by asterisks.*

A more careful analysis of the Dirac delta approximations obtained in Figure 2.98 indicates that the lower frequency peaks (marked with asterisks), with respect to $f_1 = 1$ Hz, have been better represented (there is less ripple around it and the amplitude is exactly as expected, i.e., 0.5) than the deltas for $f_2 = 2\sqrt{2}$ Hz. This is because f_1 is an integer multiple of $\Delta f = 1/(N\Delta t) = 0.25$ Hz, while f_2 is not. Indeed, a complete cancellation of the ripple effect is observed in such a multiple situation, because the zero crossings of $\widetilde{G}(f)$ can be verified to coincide with the sampling points. However, this welcomed effect cannot usually be guaranteed in practical situations, and the unwanted rippling effect has to be somehow alleviated, for instance by using a smoother window function. Figure 2.99 illustrates this possibility considering as windowing function $f(t) = \exp\{-t^2\} \cdot r(t)$, i.e., the product of a Gaussian with the rectangular function (a truncated Gaussian function).

The ripple attenuation is readily verified, though at the expense of a decrease in the amplitude of the coefficients related to $f_2 = 2\sqrt{2}$ Hz. The reader is referred to the literature (e.g., Papoulis, 1977; Ingle and Proakis, 1997; Kamen and Heck, 1997]) for a more detailed discussion about several windowing functions and their respective properties.