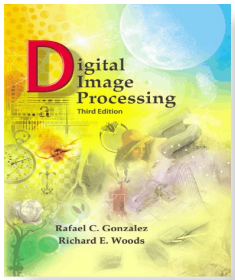
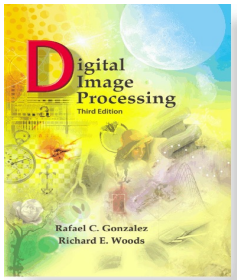


REVIEWS

Matrices
Probability and Statistics
Linear Systems



Matrices and Vector

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

- **A** is ***square*** if $m = n$.
- **A** is ***diagonal*** if all off-diagonal elements are 0, and not all diagonal elements are 0.
- **A** is the ***identity matrix*** (**I**) if it is diagonal and all diagonal elements are 1.
- **A** is the ***zero*** or ***null matrix*** (**0**) if all its elements are 0.
- The ***trace*** of **A** equals the sum of the elements along its main diagonal.
- Two matrices **A** and **B** are ***equal*** iff the have the same number of rows and columns, and $a_{ij} = b_{ij}$.

Definitions (Con't)

- The ***transpose*** $\mathbf{A}^T$ of an $m \times n$ matrix **A** is an $n \times m$ matrix obtained by interchanging the rows and columns of **A**.
- A square matrix for which $\mathbf{A}^T = \mathbf{A}$ is said to be ***symmetric***.
- Any matrix X for which **XA=I** and **AX=I** is called the ***inverse*** of **A**.
- Let c be a real or complex number (called a ***scalar***). The ***scalar multiple*** of c and matrix **A**, denoted $c\mathbf{A}$, is obtained by multiplying every elements of **A** by c. If $c = -1$, the scalar multiple is called the ***negative*** of **A**.

Definitions (Con't)

A ***column vector*** is an $m \times 1$ matrix:

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}$$

A ***row vector*** is a $1 \times n$ matrix:

$$\mathbf{b} = [b_1, b_2, \cdots b_n]$$

A column vector can be expressed as a row vector by using the transpose:

$$\mathbf{a}^T = [a_1, a_2, \cdots, a_m]$$

Some Basic Matrix Operations

- The ***sum*** of two matrices **A** and **B** (of equal dimension), denoted **A** + **B**, is the matrix with elements $a_{ij} + b_{ij}$.
- The ***difference*** of two matrices, **A**– **B**, has elements $a_{ij} - b_{ij}$.
- The ***product***, **AB**, of $m \times n$ matrix **A** and $p \times q$ matrix **B**, is an $m \times q$ matrix **C** whose (i,j)-th element is formed by multiplying the entries across the ith row of **A** times the entries down the jth column of **B**; that is,

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{pj}$$

Some Basic Matrix Operations (Con't)

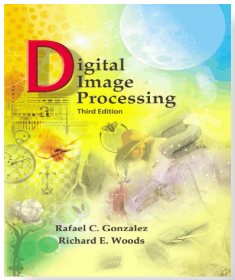
The ***inner product*** (also called ***dot product***) of two vectors

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

is defined as

$$\mathbf{a}^T\mathbf{b} = \mathbf{b}^T\mathbf{a} = a_1b_1 + a_2b_2 + \cdots + a_mb_m = \sum_{i=1}^{m} a_ib_i.$$

Note that the inner product is a scalar.



Vectors and Vector Spaces (Con't)

A ***linear combination*** of $\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n$ is an expression of the form

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_n\mathbf{v}_n$$

where the α's are scalars.

A vector $\mathbf{v}$ is said to be ***linearly dependent*** on a set, S, of vectors $\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n$ if and only if $\mathbf{v}$ can be written as a linear combination of these vectors. Otherwise, $\mathbf{v}$ is ***linearly independent*** of the set of vectors $\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n$.

Vector Norms

A ***vector norm*** on a vector space V is a function that assigns to each vector $\mathbf{v}$ in V a nonnegative real number, called the ***norm*** of $\mathbf{v}$, denoted by $\|\mathbf{v}\|$. By definition, the norm satisfies the following conditions:

(1) $\|\mathbf{v}\| > 0$ for $\mathbf{v} \neq \mathbf{0}$; $\|\mathbf{0}\| = 0$,

(2) $\|c\mathbf{v}\| = |c|\|\mathbf{v}\|$ for all scalars c and vectors $\mathbf{v}$, and

(3) $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$.

Vector Norms (Con't)

There are numerous norms that are used in practice. In our work, the norm most often used is the so-called ***2-norm***, which, for a vector $\mathbf{x}$ in real $\Re^m$, space is defined as

$$\|\mathbf{x}\| = [x_1^2 + x_2^2 + \cdots + x_m^2]^{1/2}$$

which is recognized as the *Euclidean distance* from the origin to point $\mathbf{x}$; this gives the expression the familiar name Euclidean norm. The expression also is recognized as the length of a vector $\mathbf{x}$, with origin at point $\mathbf{0}$. From earlier discussions, the norm also can be written as

$$\|\mathbf{x}\| = [\mathbf{x}^T\mathbf{x}]^{1/2}$$

Vector Norms (Con't)

The ***Cauchy-Schwartz*** inequality states that

$$|\mathbf{x}^T\mathbf{y}| \leq \|\mathbf{x}\|\,\|\mathbf{y}\|$$

Another well-known result used in the book is the expression

$$\cos\theta = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|\,\|\mathbf{y}\|}$$

where θ is the angle between vectors **x** and **y**. From these expressions it follows that the inner product of two vectors can be written as

$$\mathbf{x}^T\mathbf{y} = \|\mathbf{x}\|\,\|\mathbf{y}\|\cos\theta$$

Thus, the inner product can be expressed as a function of the norms of the vectors and the angle between the vectors.

Vector Norms (Con't)

From the preceding results, two vectors in $\Re^m$ are ***orthogonal*** if and only if their inner product is zero. Two vectors are ***orthonormal*** if, in addition to being orthogonal, the length of each vector is 1.

From the concepts just discussed, we see that an arbitrary vector $\mathbf{a}$ is turned into a vector $\mathbf{a}_n$ of unit length by performing the operation $\mathbf{a}_n = \mathbf{a}/\|\mathbf{a}\|$. Clearly, then, $\|\mathbf{a}_n\| = 1$.

A ***set of vectors*** is said to be an ***orthogonal*** set if every two vectors in the set are orthogonal. A ***set of vectors*** is ***orthonormal*** if every two vectors in the set are orthonormal.

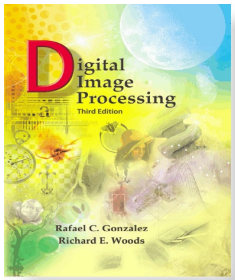
Some Important Aspects of Orthogonality

Let $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be an orthogonal or orthonormal basis in the sense defined in the previous section. Then, an important result in vector analysis is that any vector v can be represented with respect to the orthogonal basis B as

$$\mathbf{v} = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_n \mathbf{v}_n$$

where the coefficients are given by

$$\alpha_i = \frac{\mathbf{v}^T \mathbf{v}_i}{\mathbf{v}_i^T \mathbf{v}_i} = \frac{\mathbf{v}^T \mathbf{v}_i}{\|\mathbf{v}_i\|^2}$$



Eigenvalues & Eigenvectors

Definition: The ***eigenvalues*** of a real matrix **M** are the real numbers λ for which there is a nonzero vector **e** such that

$$\mathbf{Me} = \lambda\,\mathbf{e}.$$

The ***eigenvectors*** of **M** are the nonzero vectors **e** for which there is a real number λ such that $\mathbf{Me} = \lambda\,\mathbf{e}$.

If $\mathbf{Me} = \lambda\,\mathbf{e}$ for $\mathbf{e} \neq 0$, then **e** is an ***eigenvector*** of **M** associated with ***eigenvalue*** λ, and vice versa.

Eigenvalues & Eigenvectors (Con't)

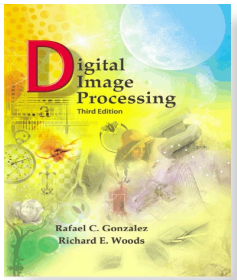
Example: Consider the matrix

$$\mathbf{M} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

It is easy to verify that $\mathbf{M}\mathbf{e}_1 = \lambda_1\mathbf{e}_1$ and $\mathbf{M}\mathbf{e}_2 = \lambda_2\mathbf{e}_2$ for $\lambda_1 = 1$, $\lambda_2 = 2$ and

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

In other words, $\mathbf{e}_1$ is an eigenvector of $\mathbf{M}$ with associated eigenvalue λ_1, and similarly for $\mathbf{e}_2$ and λ_2.



Eigenvalues & Eigenvectors (Con't)

The following properties, which we give without proof, are essential background in the use of vectors and matrices in digital image processing. In each case, we assume a real matrix of order $m \times m$ although, as stated earlier, these results are equally applicable to complex numbers.

1. If $\{\lambda_1, \lambda_2, \ldots, \lambda_q, q \leq m$, is set of distinct eigenvalues of $\mathbf{M}$, and $\mathbf{e}_i$ is an eigenvector of $\mathbf{M}$ with corresponding eigenvalue λ_i, $i = 1,2,\ldots,q$, then $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_q\}$ is a linearly independent set of vectors. An important implication of this property: If an $m \times m$ matrix $\mathbf{M}$ has m distinct eigenvalues, its eigenvectors will constitute an orthogonal (orthonormal) set, which means that any m-dimensional vector can be expressed as a linear combination of the eigenvectors of $\mathbf{M}$.

Eigenvalues & Eigenvectors (Con't)

2. The numbers along the main diagonal of a diagonal matrix are equal to its eigenvalues. It is not difficult to show using the definition $\mathbf{Me} = \lambda\,\mathbf{e}$ that the eigenvectors can be written by inspection when **M** is diagonal.

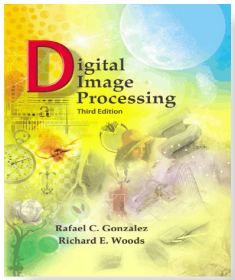
3. A real, symmetric $m \times m$ matrix **M** has a set of m linearly independent eigenvectors that may be chosen to form an orthonormal set. This property is of particular importance when dealing with covariance matrices (e.g., see Section 11.4 and our review of probability) which are real and symmetric.

Eigenvalues & Eigenvectors (Con't)

4. A corollary of Property 3 is that the eigenvalues of an $m \times m$ real symmetric matrix are real, and the associated eigenvectors may be chosen to form an orthonormal set of m vectors.

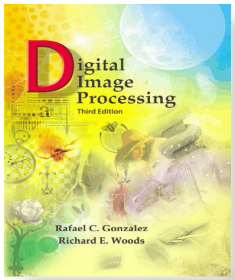
5. Suppose that $\mathbf{M}$ is a real, symmetric $m \times m$ matrix, and that we form a matrix $\mathbf{A}$ whose rows are the m orthonormal eigenvectors of $\mathbf{M}$. Then, the product $\mathbf{A}\mathbf{A}^T = \mathbf{I}$ because the rows of $\mathbf{A}$ are orthonormal vectors. Thus, we see that $\mathbf{A}^{-1} = \mathbf{A}^T$ when matrix $\mathbf{A}$ is formed in the manner just described.

6. Consider matrices $\mathbf{M}$ and $\mathbf{A}$ in 5. The product $\mathbf{D} = \mathbf{A}\mathbf{M}\mathbf{A}^{-1} = \mathbf{A}\mathbf{M}\mathbf{A}^T$ is a diagonal matrix whose elements along the main diagonal are the eigenvalues of $\mathbf{M}$. The eigenvectors of $\mathbf{D}$ are the same as the eigenvectors of $\mathbf{M}$.



Review

Probability & Random Variables

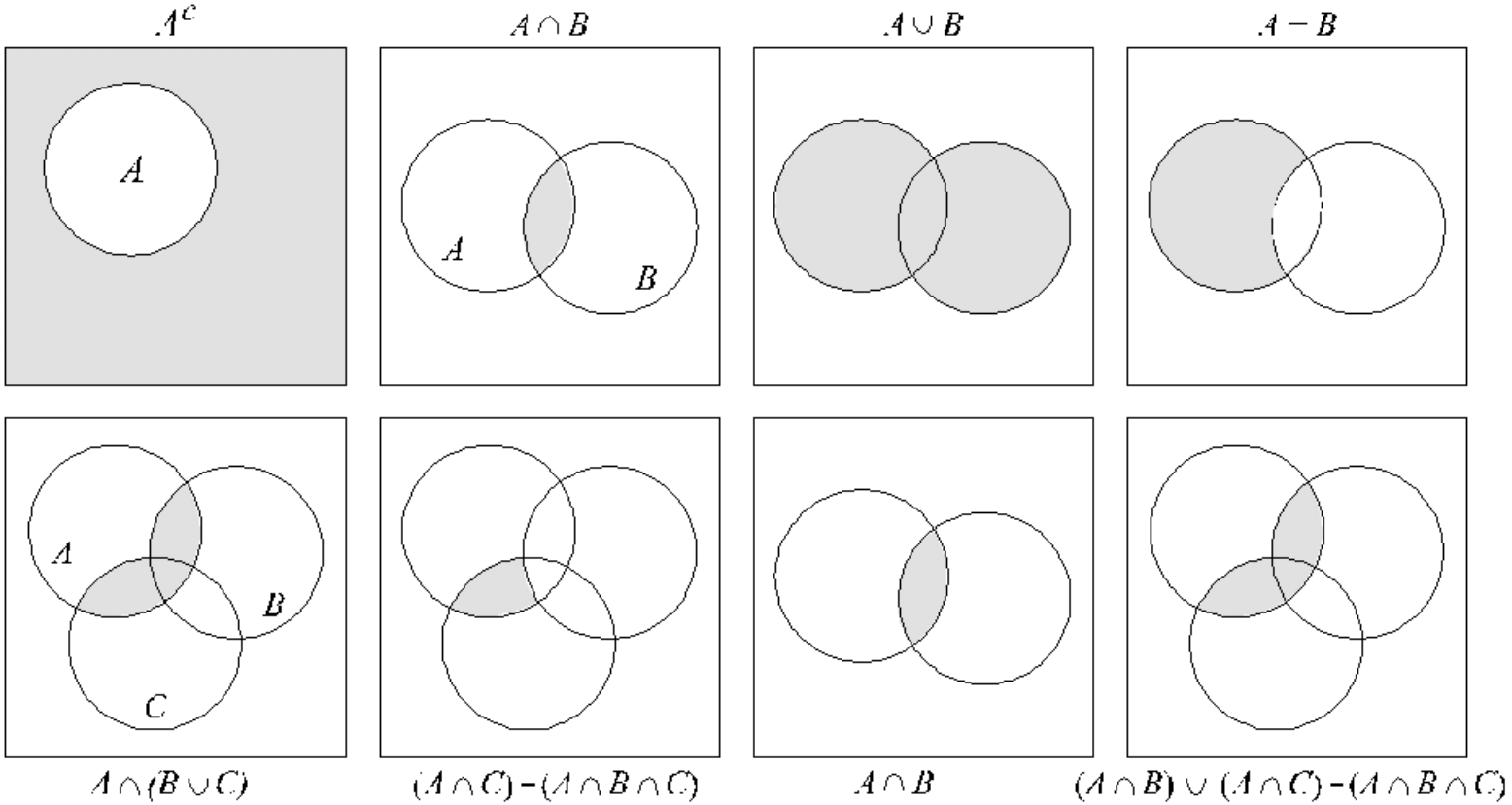


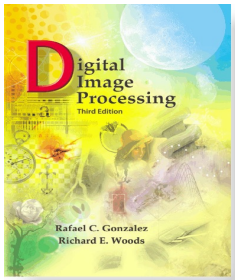
Sets and Set Operations

Probability events are modeled as sets, so it is customary to begin a study of probability by defining sets and some simple operations among sets.

A ***set*** is a collection of objects, with each object in a set often referred to as an ***element*** or ***member*** of the set. Familiar examples include the set of all image processing books in the world, the set of prime numbers, and the set of planets circling the sun. Typically, sets are represented by uppercase letters, such as A, B, and C, and members of sets by lowercase letters, such as a, b, and c.

Set Operations (Con't)

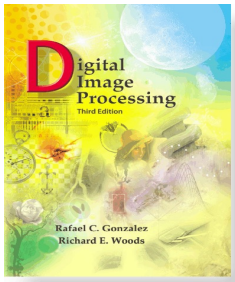




Relative Frequency & Probability

A ***random experiment*** is an experiment in which it is not possible to predict the outcome. Perhaps the best known random experiment is the tossing of a coin. Assuming that the coin is not biased, we are used to the concept that, on average, half the tosses will produce heads (H) and the others will produce tails (T). This is intuitive and we do not question it. In fact, few of us have taken the time to verify that this is true. If we did, we would make use of the concept of relative frequency. Let n denote the total number of tosses, n_H the number of heads that turn up, and n_T the number of tails. Clearly,

$$n_H + n_T = n.$$

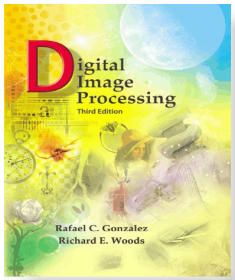


Relative Frequency & Prob. (Con't)

Dividing both sides by n gives

$$\frac{n_H}{n} + \frac{n_T}{n} = 1.$$

The term n_H/n is called the ***relative frequency*** of the event we have denoted by H, and similarly for n_T/n. If we performed the tossing experiment a large number of times, we would find that each of these relative frequencies tends toward a stable, limiting value. We call this value the ***probability of the event***, and denoted it by P(event).



Relative Frequency & Prob. (Con't)

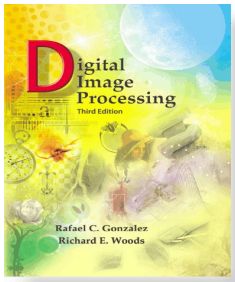
In the current discussion the probabilities of interest are $P(H)$ and $P(T)$. We know in this case that $P(H) = P(T) = 1/2$. Note that the event of an experiment need not signify a single outcome. For example, in the tossing experiment we could let D denote the event "heads or tails," (note that the event is now a set) and the event E, "neither heads nor tails." Then, $P(D) = 1$ and $P(E) = 0$.

The first important property of P is that, for an event A,

$$0 \leq P(A) \leq 1.$$

That is, the probability of an event is a positive number bounded by 0 and 1. For the certain event, S,

$$P(S) = 1.$$

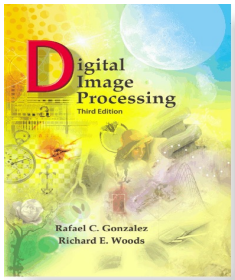


Relative Frequency & Prob. (Con't)

The relative frequency of event A occurring, ***given that*** event B has occurred, is given by

$$\frac{n_{A/B}}{n} = \frac{\frac{n_{AB}}{n}}{\frac{n_B}{n}} = \frac{n_3}{n_2 + n_3}.$$

This ***conditional probability*** is denoted by $P(A/B)$, where we note the use of the symbol " / " to denote conditional occurrence. It is common terminology to refer to $P(A/B)$ as the ***probability of*** A ***given*** B.



Relative Frequency & Prob. (Con't)

A little manipulation of the preceding results yields the following important relationships

$$P(A/B) = \frac{P(B/A)P(A)}{P(B)}$$

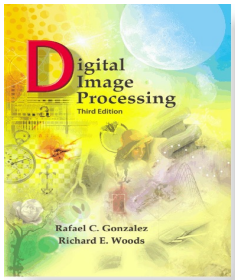
and

$$P(AB) = P(A)P(B/A) = P(B)P(A/B).$$

The second expression may be written as

$$P(B/A) = \frac{P(A/B)P(B)}{P(A)}$$

which is known as ***Bayes' theorem***, so named after the 18th century mathematician Thomas Bayes.



Relative Frequency & Prob. (Con't)

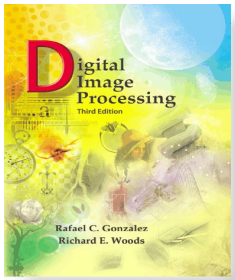
If A and B are ***statistically independent***, then $P(B/A) = P(B)$ and it follows that

$$P(A/B) = P(A)$$
$$P(B/A) = P(B)$$

and

$$P(AB) = P(A)P(B).$$

As was just shown, the two sets are statistically independent if $P(AB)=P(A)P(B)$, which we assume to be nonzero in general.



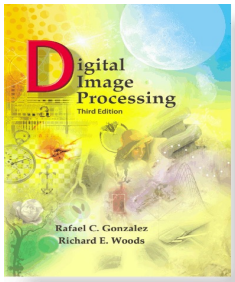
Relative Frequency & Prob. (Con't)

For three events A, B, and C to be independent, it must be true that

$$P(AB) = P(A)P(B)$$
$$P(AC) = P(A)P(C)$$
$$P(BC) = P(B)P(C)$$

and

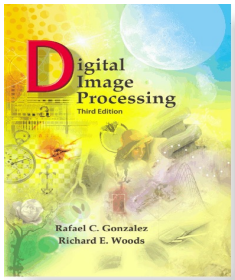
$$P(ABC) = P(A)P(B)P(C).$$



Relative Frequency & Prob. (Con't)

In general, for N events to be statistically independent, it must be true that, for all combinations $1 \leq i \leq j \leq k \leq \ldots \leq N$

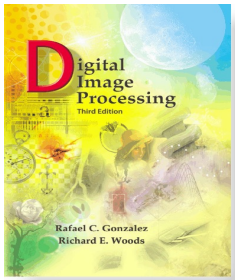
$$
\begin{aligned}
P(A_iA_j) &= P(A_i)P(A_j) \\
P(A_iA_jA_k) &= P(A_i)P(A_j)P(A_k) \\
&\vdots \\
P(A_1A_2\cdots A_N) &= P(A_1)P(A_2)\cdots P(A_N).
\end{aligned}
$$



Random Variables

A ***random variable***, x, is a real-valued function ***defined*** on the events of the sample space, S. In words, for each event in S, there is a real number that is the corresponding value of the random variable. Viewed yet another way, a random variable maps each event in S onto the real line.

We can have continuous or discrete random variables.

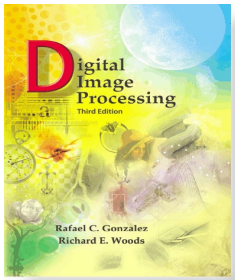


Random Variables (Con't)

Thus, instead of talking about the probability of a specific value, we talk about the probability that the value of the random variable lies in a specified ***range***. In particular, we are interested in the probability that the random variable is less than or equal to (or, similarly, greater than or equal to) a specified constant a. We write this as

$$F(a) = P(x \leq a).$$

If this function is given for all values of a (i.e., $-\infty < a < \infty$), then the values of random variable x have been defined. Function F is called the ***cumulative probability distribution function*** or simply the ***cumulative distribution function*** (cdf). The shortened term ***distribution function*** also is used.

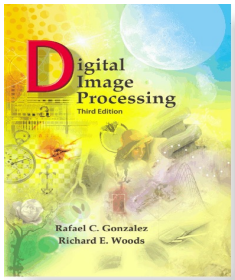


Random Variables (Con't)

Due to the fact that it is a probability, the cdf has the following properties:

1. $F(-\infty) = 0$
2. $F(\infty) = 1$
3. $0 \leq F(x) \leq 1$
4. $F(x_1) \leq F(x_2)$ if $x_1 < x_2$
5. $P(x_1 < x \leq x_2) = F(x_2) - F(x_1)$
6. $F(x^+) = F(x)$,

where $x^+ = x + \varepsilon$, with ε being a positive, infinitesimally small number.



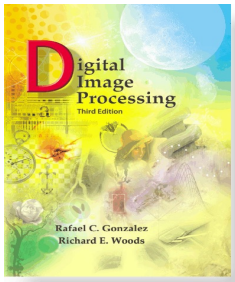
Random Variables (Con't)

The ***probability density function*** (pdf) of random variable x is defined as the derivative of the cdf:

$$p(x) = \frac{dF(x)}{dx}.$$

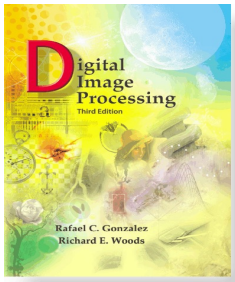
The term ***density function*** is commonly used also. The pdf satisfies the following properties:

1. $p(x) \geq 0$ for all x
2. $\int_{-\infty}^{\infty} p(x)dx = 1$
3. $F(x) = \int_{-\infty}^{x} p(\alpha)d\alpha$, where α is a dummy variable
4. $P(x_1 < x \leq x_2) = \int_{x_1}^{x_2} p(x)dx.$



Random Variables (Con't)

The preceding concepts are applicable to discrete random variables. In this case, there is a finite no. of events and we talk about ***probabilities***, rather than probability density functions. Integrals are replaced by summations and, sometimes, the random variables are subscripted. For example, in the case of a discrete variable with N possible values we would denote the probabilities by $P(x_i)$, i=1, 2,…, N.

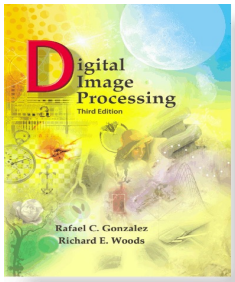


Random Variables (Con't)

If a random variable x is ***transformed*** by a monotonic transformation function $T(x)$ to produce a new random variable y, the probability density function of y can be obtained from knowledge of $T(x)$ and the probability density function of x, as follows:

$$p_y(y) = p_x(x)\left|\frac{dx}{dy}\right|$$

where the subscripts on the p's are used to denote the fact that they are different functions, and the vertical bars signify the absolute value. A function $T(x)$ is ***monotonically increasing*** if $T(x_1) < T(x_2)$ for $x_1 < x_2$, and ***monotonically decreasing*** if $T(x_1) > T(x_2)$ for $x_1 < x_2$. The preceding equation is valid if $T(x)$ is an increasing or decreasing monotonic function.



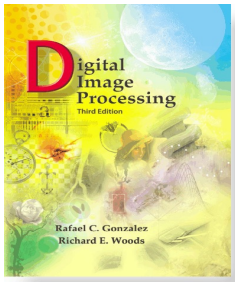
Expected Value and Moments

The ***expected value*** of a function $g(x)$ of a ***continuos*** random variable is defined as

$$E[g(x)] = \int_{-\infty}^{\infty} g(x)p(x)dx.$$

If the random variable is ***discrete*** the definition becomes

$$E[g(x)] = \sum_{i=1}^{N} g(x_i)P(x_i).$$



Expected Value & Moments (Con't)

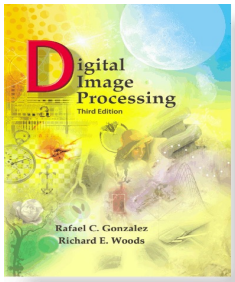
The expected value is one of the operations used most frequently when working with random variables. For example, the expected value of random variable x is obtained by letting $g(x) = x$:

$$E[x] = \overline{x} = m = \int_{-\infty}^{\infty} xp(x)dx$$

when x is continuos and

$$E[x] = \overline{x} = m = \sum_{i=1}^{N} x_i P(x_i)$$

when x is discrete. The expected value of x is equal to its ***average*** (or ***mean***) ***value***, hence the use of the equivalent notation $\overline{x}$ and m.



Expected Value & Moments (Con't)

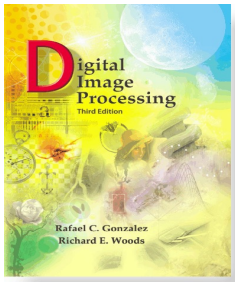
The ***mean-squared value*** of a random variable is obtained by letting $g(x) = x^2$ which gives

$$\sigma^2 = E[x^2] = \int_{-\infty}^{\infty} x^2 p(x) dx$$

for continuous random variables and

$$\sigma^2 = E[x^2] = \sum_{i=1}^{N} x_i^2 P(x_i)$$

for discrete variables.



Expected Value & Moments (Con't)

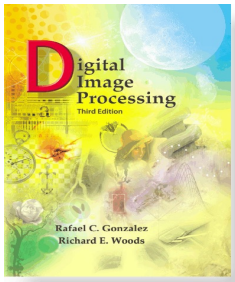
Of particular importance is the variance of random variables that have been ***normalized*** by subtracting their mean. In this case, the variance is

$$\sigma^2 = E[(x-m)^2] = \int_{-\infty}^{\infty}(x-m)^2 p(x)dx$$

and

$$\sigma^2 = E[(x-m)^2] = \sum_{i=1}^{N}(x_i-m)^2 P(x_i)$$

for continuous and discrete random variables, respectively. The square root of the variance is called the ***standard deviation***, and is denoted by σ.



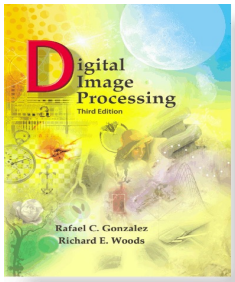
The Gaussian Probability Density Function

Because of its importance, we will focus in this tutorial on the ***Gaussian probability density function*** to illustrate many of the preceding concepts, and also as the basis for generalization to more than one random variable. The reader is referred to Section 5.2.2 of the book for examples of other density functions.

A random variable is called ***Gaussian*** if it has a probability density of the form

$$p(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-m)^2/\sigma^2}$$

where m and σ are as defined in the previous section. The term ***normal*** also is used to refer to the Gaussian density. A plot and properties of this density function are given in Section 5.2.2 of the book.

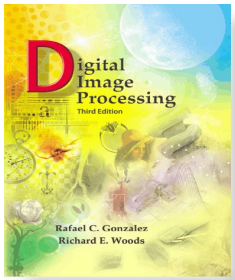


The Gaussian PDF (Con't)

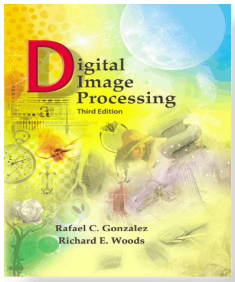
The cumulative distribution function corresponding to the Gaussian density is

$$F(x) = \int_{-\infty}^{x} p(x)dx$$
$$= \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{x} e^{-(x-m)^2/\sigma^2 2} dx.$$

which, as before, we interpret as the probability that the random variable lies between minus infinite and an arbitrary value x. This integral has no known closed-form solution, and it must be solved by numerical or other approximation methods. Extensive tables exist for the Gaussian cdf.



Several Random Variables



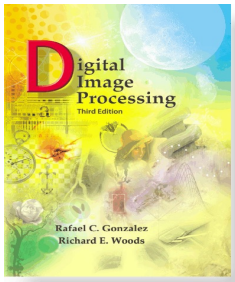
Several Random Variables (Con't)

It is convenient to use vector notation when dealing with several random variables. Thus, we represent a ***vector random variable*** **x** as

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

Then, for example, the cumulative distribution function introduced earlier becomes

$$\begin{aligned} F(\mathbf{a}) &= F(a_1, a_2, \ldots, a_n) \\ &= P\{x_1 \leq a_1, x_2 \leq a_2, \ldots, x_n \leq a_n\} \end{aligned}$$

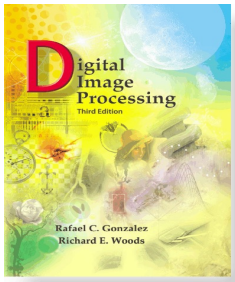


Several Random Variables (Con't)

when using vectors. As before, when confusion is not likely, the ***cdf of a random variable vector*** often is written simply as $F(\mathbf{x})$. This notation will be used in the following discussion when speaking generally about the cdf of a random variable vector.

As in the single variable case, the ***probability density function of a random variable vector*** is defined in terms of derivatives of the cdf; that is,

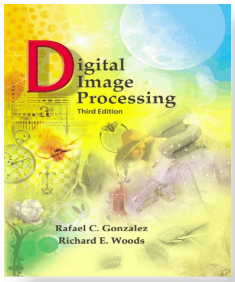
$$p(\mathbf{x}) = p(x_1, x_2, \ldots, x_n) = \frac{\partial^n F(x_1, x_2, \ldots, x_n)}{\partial x_1 \partial x_2 \cdots \partial x_n}.$$



Several Random Variables (Con't)

The ***expected value*** of a function of **x** is defined basically as before:

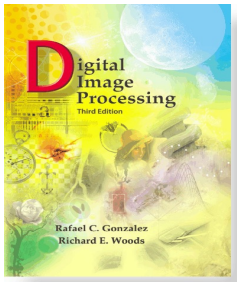
$$E[g(\mathbf{x})] = E[g(x_1, x_2, \ldots, x_n)]$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, x_2, \ldots, x_n) p(x_1, x_2, \ldots, x_n) dx_1 dx_2 \cdots dx_n.$$



Several Random Variables (Con't)

The moment $\eta_{11} = E[xy]$ is called the ***correlation*** of x and y. As discussed in Chapters 4 and 12 of the book, correlation is an important concept in image processing. In fact, it is important in most areas of signal processing, where typically it is given a special symbol, such as R_{xy}:

$$R_{xy} = \eta_{11} = E[xy] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyp(x,y)dxdy.$$



Several Random Variables (Con't)

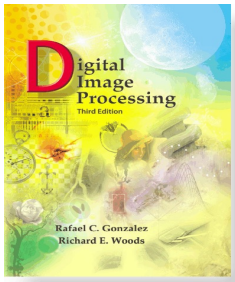
If the condition

$$R_{xy} = E[x]E[y]$$

holds, then the two random variables are said to be ***uncorrelated***. From our earlier discussion, we know that if x and y are ***statistically independent***, then $p(x, y) = p(x)p(y)$, in which case we write

$$R_{xy} = \int_{-\infty}^{\infty} xp(x)dx \int_{-\infty}^{\infty} yp(y)dy = E[x]E[y].$$

Thus, we see that ***if two random variables are statistically independent then they are also uncorrelated***. The converse of this statement is ***not*** true in general (except for Gaussians).

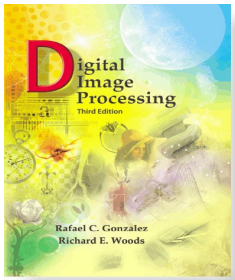


Several Random Variables (Con't)

The moment μ_{11}

$$\mu_{11} = E[(x - m_x)(y - m_y)]$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - m_x)(y - m_y)p(x,y)dxdy$$

is called the ***covariance*** of x and y. As in the case of correlation, the covariance is an important concept, usually given a special symbol such as C_{xy}.

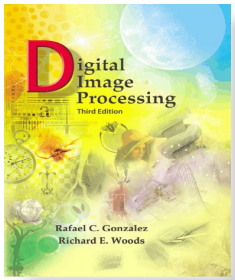


Several Random Variables (Con't)

By direct expansion of the terms inside the expected value brackets, and recalling the $m_x = E[x]$ and $m_y = E[y]$, it is straightforward to show that

$$\begin{aligned} C_{xy} &= E[xy] - m_y E[x] - m_x E[y] + m_x my \\ &= E[xy] - E[x]E[y] \\ &= R_{xy} - E[x]E[y]. \end{aligned}$$

From our discussion on correlation, we see that the covariance is zero if the random variables are either uncorrelated ***or*** statistically independent. This is an important result worth remembering.



Several Random Variables (Con't)

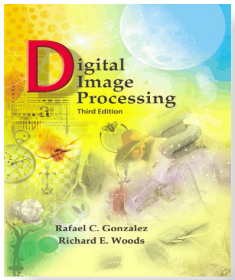
If we divide the covariance by the square root of the product of the variances we obtain

$$\gamma = \frac{\mu_{11}}{\sqrt{\mu_{20}\mu_{02}}}$$

$$= \frac{C_{xy}}{\sigma_x \sigma_y}$$

$$= E\left[\frac{(x - m_x)}{\sigma_x}\frac{(y - m_y)}{\sigma_y}\right].$$

The quantity γ is called the ***correlation coefficient*** of random variables x and y. It can be shown that γ is in the range $-1 \leq \gamma \leq 1$

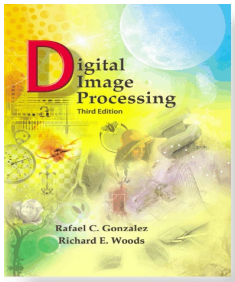


The Multivariate Gaussian Density

As an illustration of a probability density function of more than one random variable, we consider the ***multivariate Gaussian probability density function***, defined as

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\mathbf{C}|^{1/2}} e^{-\frac{1}{2}\left[(\mathbf{x}-\mathbf{m})^T\mathbf{C}^{-1}(\mathbf{x}-\mathbf{m})\right]}$$

where n is the ***dimensionality*** (number of components) of the random vector $\mathbf{x}$, $\mathbf{C}$ is the ***covariance matrix*** (to be defined below), $|\mathbf{C}|$ is the determinant of matrix $\mathbf{C}$, $\mathbf{m}$ is the ***mean vector*** (also to be defined below) and T indicates transposition (see the review of matrices and vectors).



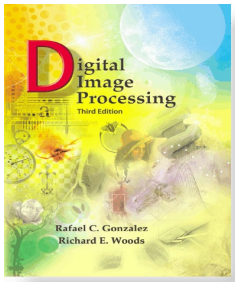
The Multivariate Gaussian Density (Con't)

The ***mean vector*** is defined as

$$\mathbf{m} = E[\mathbf{x}] = \begin{bmatrix} E[x_1] \\ E[x_2] \\ \vdots \\ E[x_n] \end{bmatrix}$$

and the ***covariance matrix*** is defined as

$$\mathbf{C} = E[(\mathbf{x} - \mathbf{m})(\mathbf{x} - \mathbf{m})^T].$$

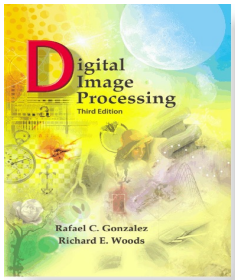


The Multivariate Gaussian Density (Con't)

The element of **C** are the covariances of the elements of **x**, such that

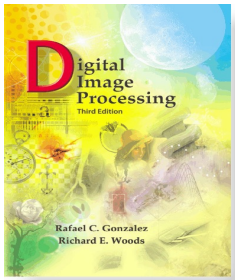
$$c_{ij} = C_{x_i x_j} = E[(x_i - m_i)(x_j - m_j)]$$

where, for example, x_i is the ith component of **x** and m_i is the ith component of **m**.



The Multivariate Gaussian Density (Con't)

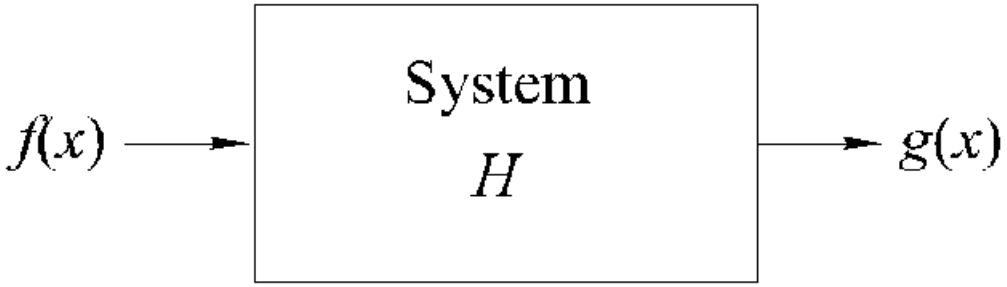
Covariance matrices are ***real*** and ***symmetric*** (see the review of matrices and vectors). The elements along the main diagonal of **C** are the variances of the elements **x**, such that $c_{ii} = \sigma_{x_i}^2$. When all the elements of **x** are uncorrelated or statistically independent, $c_{ij} = 0$, and the covariance matrix becomes a ***diagonal matrix***. If all the variances are equal, then the covariance matrix becomes proportional to the ***identity matrix***, with the constant of proportionality being the variance of the elements of **x**.



Review
Linear Systems

Some Definitions

A ***system*** converts an input function $f(x)$ into an output (or response) function $g(x)$, where x is an independent variable, such as time or, as in the case of images, spatial position. We assume for simplicity that x is a continuous variable, but the results that will be derived are equally applicable to discrete variables.



Some Definitions (Con't)

It is required that the system output be determined completely by the input, the system properties, and a set of initial conditions. From the figure in the previous page, we write

$$g(x) = H[f(x)]$$

where H is the ***system operator***, defined as a mapping or assignment of a member of the set of possible outputs $\{g(x)\}$ to each member of the set of possible inputs $\{f(x)\}$. In other words, the system operator completely characterizes the system response for a given set of inputs $\{f(x)\}$.

Some Definitions (Con't)

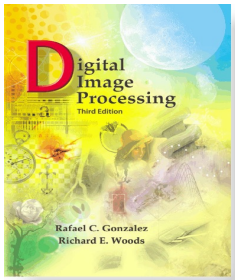
An operator H is called a ***linear operator*** for a class of inputs $\{f(x)\}$ if

$$H[a_i f_i(x) + a_j f_j(x)] = a_i H[f_i(x)] + a_j H[f_{ji}(x)] = a_i g_i(x) + a_j g_j(x)$$

for all $f_i(x)$ and $f_j(x)$ belonging to $\{f(x)\}$, where the a's are arbitrary constants and

$$g_i(x) = H[f_i(x)]$$

is the output for an arbitrary input $f_i(x) \in \{f(x)\}$.



Some Definitions (Con't)

The system described by a linear operator is called a ***linear system*** (with respect to the same class of inputs as the operator). The property that performing a linear process on the sum of inputs is the same that performing the operations individually and then summing the results is called the property of ***additivity***. The property that the response of a linear system to a constant times an input is the same as the response to the original input multiplied by a constant is called the property of ***homogeneity***.

Some Definitions (Con't)

An operator H is called ***time invariant*** (if x represents time), ***spatially invariant*** (if x is a spatial variable), or simply ***fixed parameter***, for some class of inputs $\{f(x)\}$ if

$$g_i(x) = H[f_i(x)] \text{ implies that } g_i(x + x_0) = H[f_i(x + x_0)]$$

for all $f_i(x) \in \{f(x)\}$ and for all x_0. A system described by a fixed-parameter operator is said to be a ***fixed-parameter system***. Basically all this means is that offsetting the independent variable of the input by x_0 causes the same offset in the independent variable of the output. Hence, the input-output relationship remains the same.

Some Definitions (Con't)

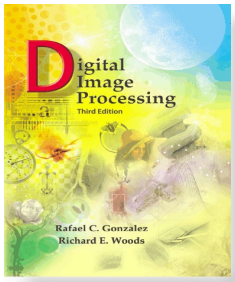
An operator H is said to be ***causal***, and hence the system described by H is a ***causal system***, if there is no output before there is an input. In other words,

$$f(x) = 0 \text{ for } x < x_0 \text{ implies that } g(x) = H[f(x)] = 0 \text{ for } x < x_0.$$

Finally, a linear system H is said to be ***stable*** if its response to any bounded input is bounded. That is, if

$$|f(x)| < K \text{ implies that } |g(x)| < cK$$

where K and c are constants.



Linear System Characterization-Convolution

A ***unit impulse function***, denoted $\delta(x - a)$, is ***defined*** by the expression

$$\int_{-\infty}^{\infty} f(\alpha)\delta(x - \alpha)d\alpha = f(x).$$

From the previous sections, the output of a system is given by $g(x) = H[f(x)]$. But, we can express $f(x)$ in terms of the impulse function just defined, so

$$g(x) = H\left[\int_{-\infty}^{\infty} f(\alpha)\delta(x - \alpha)d\alpha\right].$$

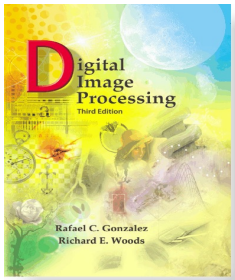
System Characterization (Con't)

Extending the property of addivity to integrals (recall that an integral can be approximated by limiting summations) allows us to write

$$g(x) = \int_{-\infty}^{\infty} H[f(\alpha)\delta(x-\alpha)]d\alpha.$$

Because $f(\alpha)$ is independent of x, and using the homogeneity property, it follows that

$$\begin{aligned} g(x) &= \int_{-\infty}^{\infty} f(\alpha)H[\delta(x-\alpha)]d\alpha \\ &= \int_{-\infty}^{\infty} f(\alpha)h(x,\alpha)d\alpha. \end{aligned}$$



System Characterization (Con't)

The term

$$h(x, \alpha) = H[\delta(x - \alpha)]$$

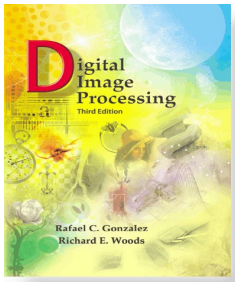
is called the ***impulse response*** of H. In other words, $h(x, \alpha)$ is the response of the linear system to a unit impulse located at coordinate x (the origin of the impulse is the value of α that produces $\delta(0)$; in this case, this happens when $\alpha = x$).

System Characterization (Con't)

The expression

$$g(x) = \int_{-\infty}^{\infty} f(\alpha)h(x,\alpha)d\alpha$$

is called the ***superposition*** (or ***Fredholm***) ***integral of the first kind***. This expression is a fundamental result that is at the core of linear system theory. It states that, if the response of H to a unit impulse [i.e., $h(x, \alpha)$], is known, then response to ***any*** input f can be computed using the preceding integral. ***In other words, the response of a linear system is characterized completely by its impulse response.***



System Characterization (Con't)

If H is a fixed-parameter operator, then

$$H[\delta(x-\alpha)] = h(x-\alpha)$$

and the superposition integral becomes

$$g(x) = \int_{-\infty}^{\infty} f(\alpha)h(x-\alpha)d\alpha.$$

This expression is called the ***convolution integral***. It states that the response of a linear, fixed-parameter system is completely characterized by the convolution of the input with the system impulse response. As will be seen shortly, this is a powerful and most practical result.

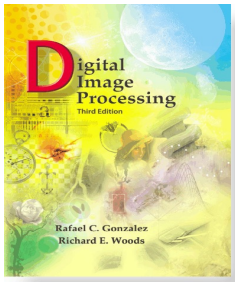
System Characterization (Con't)

Because the variable α in the preceding equation is integrated out, it is customary to write the convolution of f and h (both of which are functions of x) as

$$g(x) = f(x) * h(x).$$

In other words,

$$f(x) * h(x) \triangleq \int_{-\infty}^{\infty} f(\alpha) g(x - \alpha) d\alpha.$$



System Characterization (Con't)

The Fourier transform of the preceding expression is

$$\Im[f(x) * h(x)] = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} f(\alpha)h(x-\alpha)d\alpha\right]e^{-j2\pi ux}dx$$
$$= \int_{-\infty}^{\infty} f(\alpha)\left[\int_{-\infty}^{\infty} h(x-\alpha)e^{-j2\pi ux}dx\right]d\alpha.$$

The term inside the inner brackets is the Fourier transform of the term $h(x - \alpha)$. But,

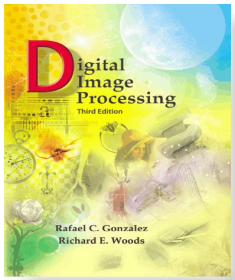
$$\Im[h(x-\alpha)] = H(u)e^{-j2\pi u\alpha}$$

System Characterization (Con't)

so,

$$\begin{aligned}\Im[f(x) * h(x)] &= \int_{-\infty}^{\infty} f(\alpha)[H(u)e^{-j2\pi u\alpha}]d\alpha \\ &= H(u)\int_{-\infty}^{\infty} f(\alpha)e^{-j2\pi u\alpha}d\alpha \\ &= H(u)F(u).\end{aligned}$$

We have succeeded in proving the important result that the Fourier transform of the convolution of two functions is the product of their Fourier transforms. As noted below, this result is the foundation for linear filtering



System Characterization (Con't)

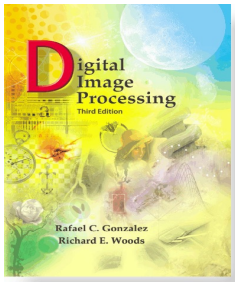
Following a similar development, it is not difficult to show that the inverse Fourier transform of the convolution of $H(u)$ and $F(u)$ [i.e., $H(u)*F(u)$] is the product $f(x)g(x)$. This result is known as the ***convolution theorem***, typically written as

$$f(x) * h(x) \Leftrightarrow H(u)F(u)$$

and

$$f(x)g(x) \Leftrightarrow H(u) * F(u)$$

where " $\Leftrightarrow$ " is used to indicate that the quantity on the right is obtained by taking the Fourier transform of the quantity on the left, and, conversely, the quantity on the left is obtained by taking the inverse Fourier transform of the quantity on the right.



System Characterization (Con't)

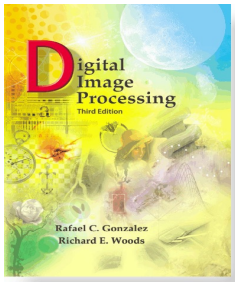
The mechanics of convolution are explained in detail in the book. We have just filled in the details of the proof of validity in the preceding paragraphs.

Because the output of our linear, fixed-parameter system is

$$g(x) = f(x) * h(x)$$

if we take the Fourier transform of both sides of this expression, it follows from the convolution theorem that

$$G(u) = H(u)F(u).$$

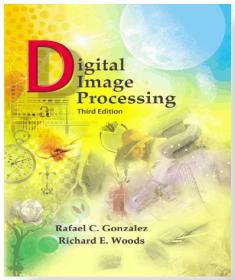


System Characterization (Con't)

The key importance of the result $G(u)=H(u)F(u)$ is that, instead of performing a convolution to obtain the output of the system, we computer the Fourier transform of the impulse response and the input, multiply them and then take the inverse Fourier transform of the product to obtain $g(x)$; that is,

$$\begin{aligned} g(x) &= \Im^{-1}[G(u)] \\ &= \Im^{-1}[H(u)F(u)]. \end{aligned}$$

These results are the basis for all the filtering work done in Chapter 4, and some of the work in Chapter 5 of *Digital Image Processing*. Those chapters extend the results to two dimensions, and illustrate their application in considerable detail.



Reading

- Chapters 1 and 2 of book
- Review Signal Processing material