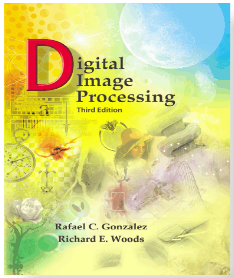
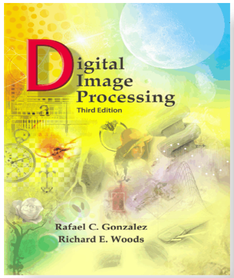


Lectures 2
REVIEWS

Matrices
Probability and Statistics
Linear Systems



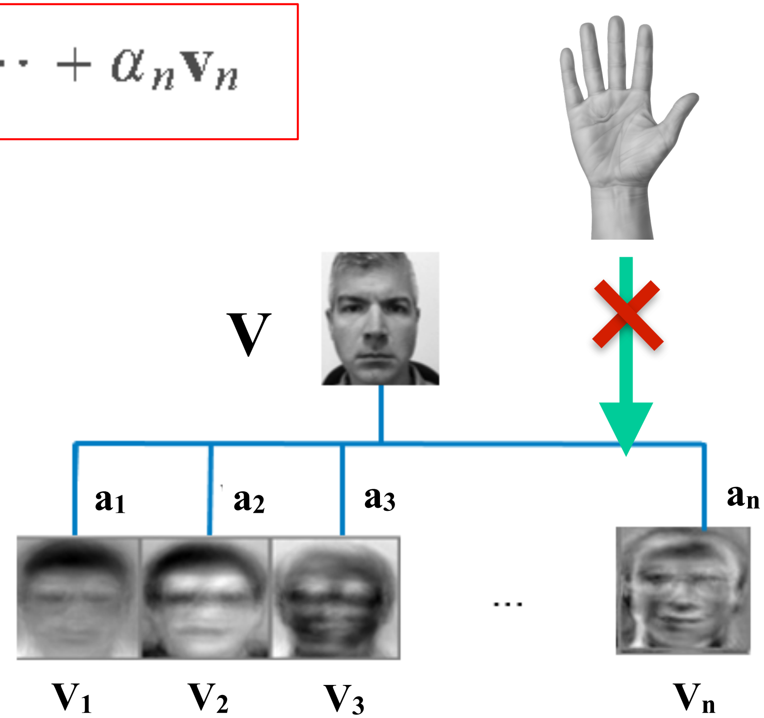
Vectors and Vector Spaces (Con't)

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

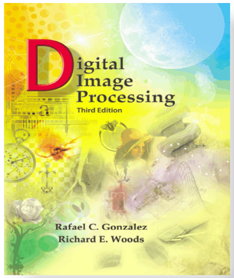
$$\boxed{\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_n \mathbf{v}_n}$$

where the α's are scalars.

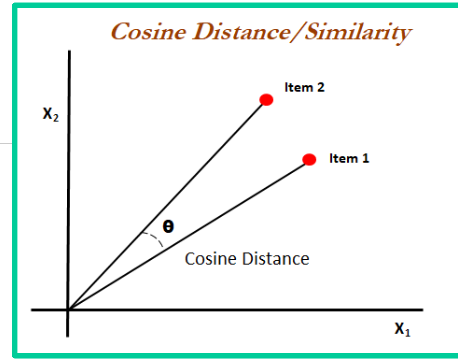
A vector $\mathbf{v}$ is said to be ***linearly dependent*** on a set, S, of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ if and only if $\mathbf{v}$ can be written as a linear combination of these vectors. Otherwise, $\mathbf{v}$ is ***linearly independent*** of the set of vectors $\mathbf{v}_1,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



https://blogs.sas.com/content/subconsciousmusings/2015/10/26/principal-component-analysis-for-dimensionality-reduction/



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

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

similarity measure

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

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

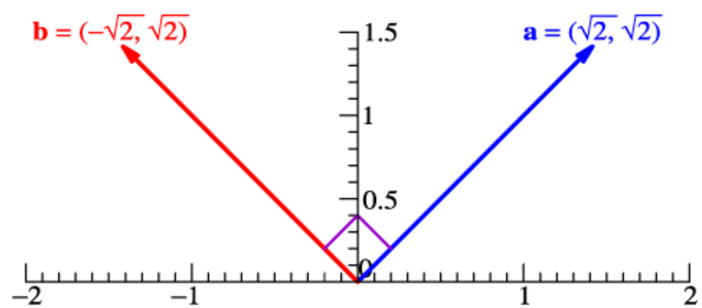
Thus, the inner product can be expressed as a function of the norms of the vectors and the angle between the vectors.

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.



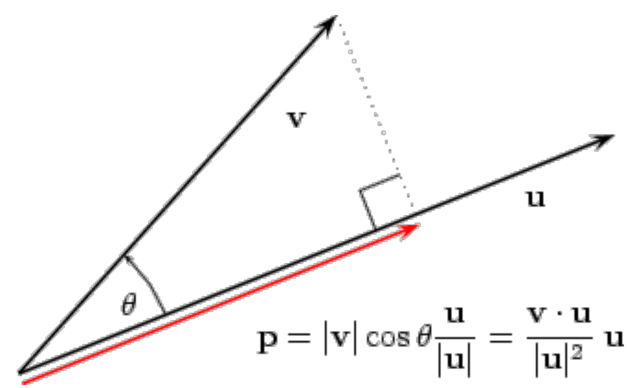
Some Important Aspects of Orthogonality

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

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

where the coefficients are given by

$$\boxed{\begin{aligned} \alpha_i &= \frac{\mathbf{v}^T \mathbf{v}_i}{\mathbf{v}_i^T \mathbf{v}_i} \\ &= \frac{\mathbf{v}^T \mathbf{v}_i}{\|\mathbf{v}_i\|^2} \end{aligned}}$$



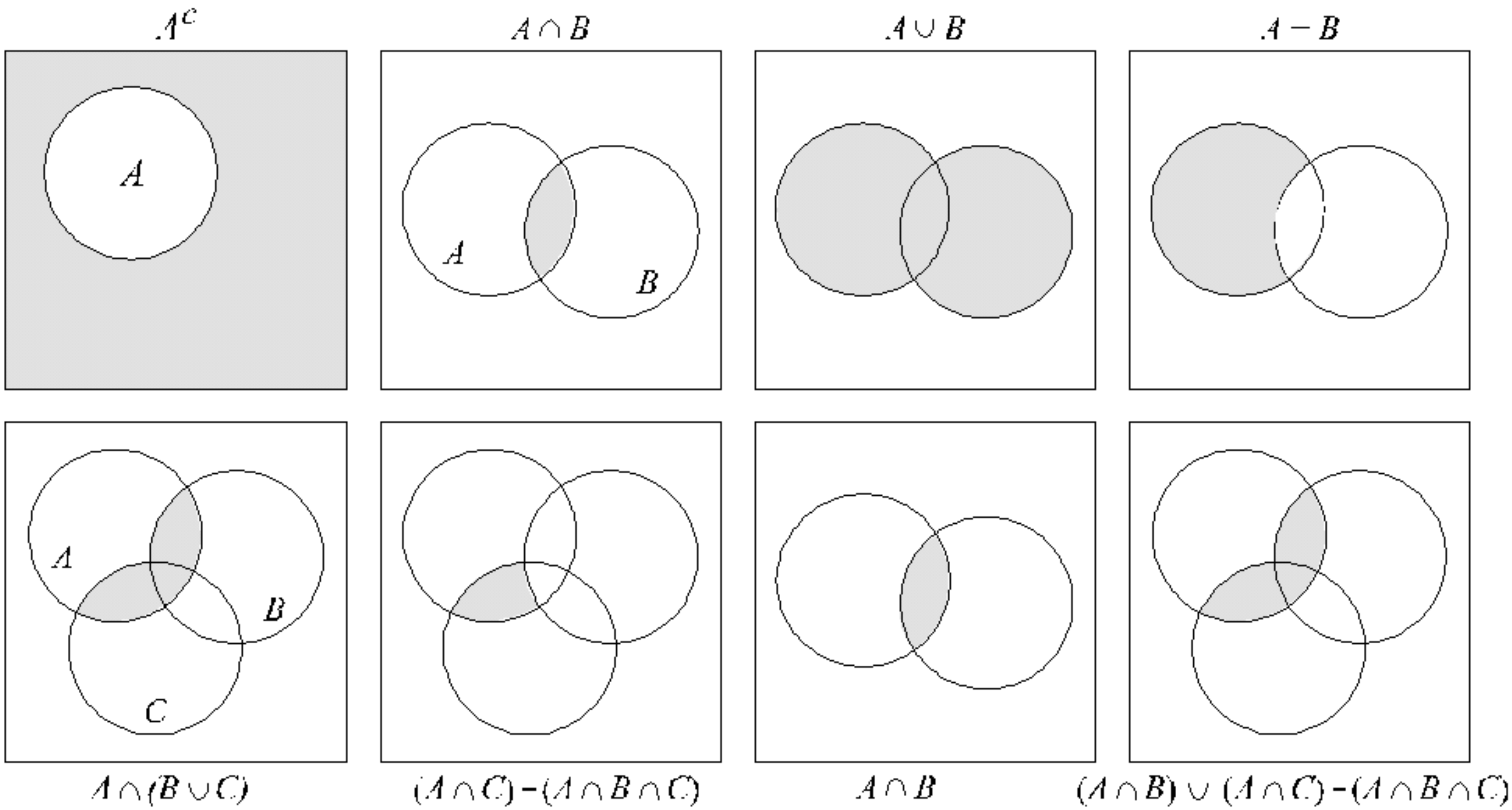
Sets and Set Operations

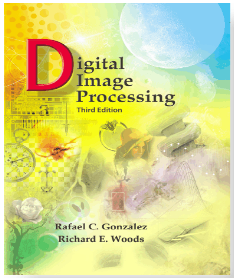
Probability events are modeled as sets, so it is customary to begin a study of probability by defining sets and some simple operations among sets.

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 & Prob. (Con't)

The probability of event A occurring, ***given that*** event B has occurred, is expressed by the ***conditional probability***, which 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)

What is the relationship between events A and B?

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

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.

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

$$\boxed{F(a) = P(x \leq a).}$$

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. $F(x_1) \leq F(x_2) \quad \text{if} \quad x_1 < x_2$
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:

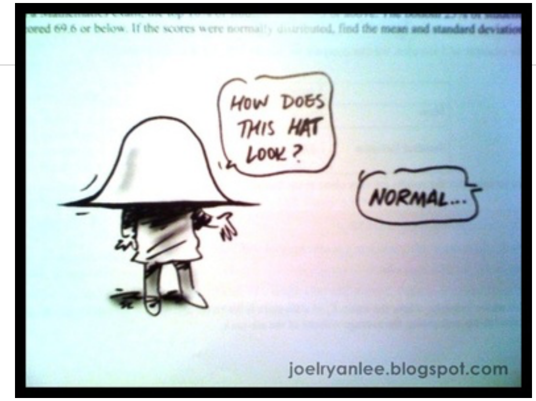
$$\boxed{p(x) = \frac{dF(x)}{dx}.}$$

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. $P(x_1 < x \leq x_2) = \int_{x_1}^{x_2} p(x)dx.$

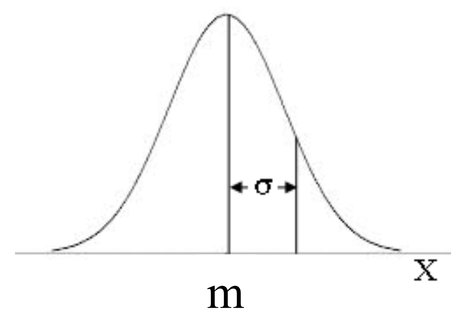
The Gaussian Probability Density Function

Because of its importance, we will focus in this lecture 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 (3rd ed.) for examples of other density functions.



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

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



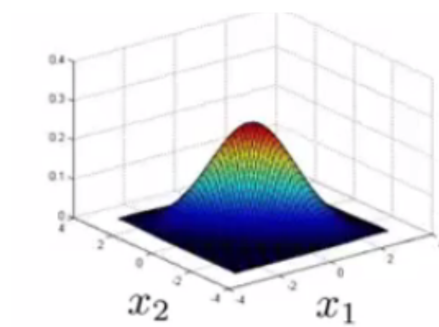
where m and σ are as defined in the previous section. The term ***normal*** also is used to refer to the Gaussian density. A plot and properties of this density function are given in Section 5.2.2 of the book.

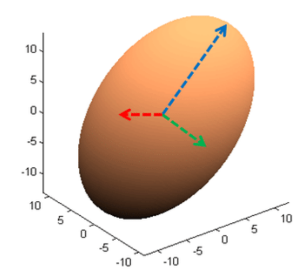
The Multivariate Gaussian Density

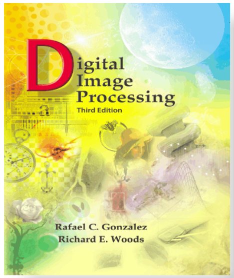
As an illustration of a probability density function of more than one random variable, we consider the ***multivariate Gaussian probability density function***, defined as

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

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





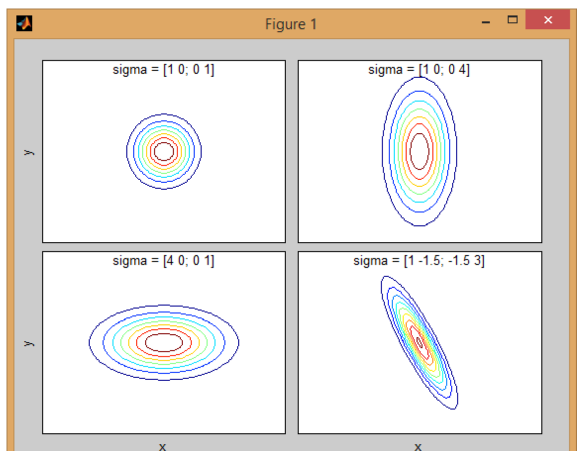


The Multivariate Gaussian Density (Con't)

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***.

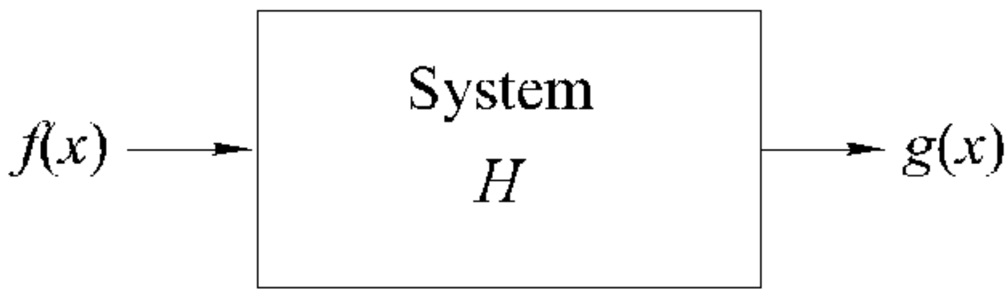
If all the variances are equal, then the covariance matrix becomes proportional to the ***identity matrix***, with the constant of proportionality being the variance of the elements of **x**.



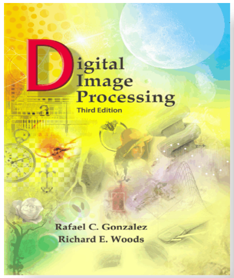
https://stackoverflow.com/questions/26019584/understanding-concept-of-gaussian-mixture-models

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.



Filter function



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

$$\boxed{g(x) = H[f(x)]}$$

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

$$\boxed{\begin{aligned} H[a_i f_i(x) + a_j f_j(x)] &= a_i H[f_i(x)] + a_j H[f_{ji}(x)] \\ &= a_i g_i(x) + a_j g_j(x) \end{aligned}}$$

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

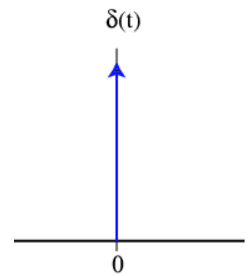
$$\boxed{g_i(x) = H[f_i(x)]}$$

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

Linear System Characterization-Convolution

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

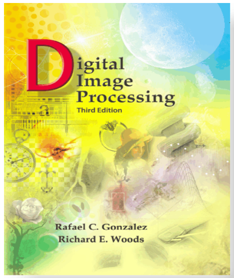
$$\boxed{\int_{-\infty}^{\infty} f(\alpha)\delta(x-\alpha)d\alpha = f(x).}$$



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

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

What does this mean?



System Characterization (Con't)

The term

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

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)

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

$$\boxed{g(x) = f(x) * h(x).}$$

In other words,

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

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

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

and

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

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

$$\boxed{g(x) = f(x) * h(x)}$$

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

$$\boxed{G(u) = H(u)F(u).}$$

Summary

- Linear algebra: Matrix, vectors, eigenvalues/vectors
- Probability theory: random variables, conditional probability, CDF, Gaussian distribution
- Linear system: definition & property, convolution

Reading

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

Additional reading on Eigenface

https://towardsdatascience.com/eigenfaces-recovering-humans-from-ghosts-17606c328184

Monty Hall Problem:

https://www.youtube.com/watch?v=i7OY7CZUdIY