5.7 Inverse Filtering

Degraded image is given by $$G(u,v) = \underbrace{H(u,v)}_{\text{degradation function}} F(u,v) + \underbrace{N(u,v)}_{\text{noise}}$$

Estimate $\tilde{F}(u,v)$ by simply dividing $G(u,v)$ by $H(u,v)$ $\tilde{F}(u,v) = \dfrac{G(u,v)}{H(u,v)}$

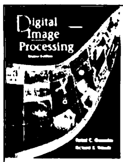
Then $$\tilde{F}(u,v) = \frac{H(u,v)\,F(u,v) + N(u,v)}{H(u,v)}$$

$$\tilde{F}(u,v) = F(u,v) + \frac{N(u,v)}{H(u,v)}$$

can never recover $F(u,v)$ exactly

1. $N(u,v)$ is not known since $n(x,y)$ is a random variable
2. If $H(u,v) \to 0$ then $\dfrac{N(u,v)}{H(u,v)}$ will dominate

   This can be somewhat overcome by restricting the analysis to values near the origin.



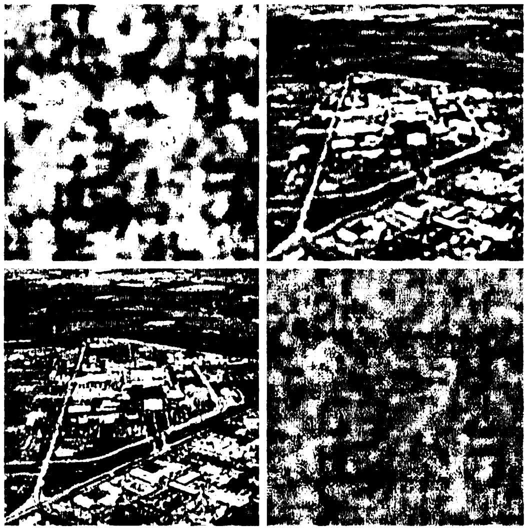
Digital Image Processing, 2nd ed.

www.imageprocessingbook.com

Chapter 5
Image Restoration

a b
c d

FIGURE 5.27 Restoring Fig. 5.25(b) with Eq. (5.7-1). (a) Result of using the full filter. (b) Result with H cut off outside a radius of 40; (c) outside a radius of 70; and (d) outside a radius of 85.



$D_0 = 40$

$D_0 = 70$

$D_0 = 85$

anything above this resembled (a)

This example shows the problems of small values of $H(u,v)$ in the inversion process.

For $H(u,v) = e^{-k\left[\left(u-\frac{M}{2}\right)^2+\left(v-\frac{N}{2}\right)^2\right]^{5/6}}$

This is never zero but can get small.

Using $\hat{F}(u,v) = \dfrac{G(u,v)}{H(u,v)}$ gives (a) above.

We can improve the result by cutting off values of $\dfrac{G(u,v)}{H(u,v)}$ outside a radius D_0.

The cutoff shown above was done using a Butterworth low-pass filter of order 10.

5.8 Minimum Mean Square Error (Wiener) Filtering

To generate the best estimate $\hat{f}$ of f we minimize

$$e^2 = \underset{\substack{\uparrow \\ \text{expected value}}}{E}\left\{ (f-\hat{f})^2 \right\}$$

Assumptions

1. f and n are uncorrelated
2. f and/or n is zero mean
3. gray levels in $\hat{f}$ are a linear function of gray levels in f

Then the best estimate $\hat{F}(u,v)$ is given by

$$\hat{F}(u,v) = \left[\frac{H^*(u,v)\, S_f(u,v)}{S_f(u,v)\,|H(u,v)|^2 + S_\eta(u,v)} \right] G(u,v)$$

$$\hat{F}(u,v) = \left[\frac{H^*(u,v)}{|H(u,v)|^2 + \frac{S_\eta(u,v)}{S_f(u,v)}} \right] G(u,v)$$

$$\hat{F}(u,v) = \left[\frac{1}{H(u,v)} \frac{|H(u,v)|^2}{|H(u,v)|^2 + \frac{S_\eta(u,v)}{S_f(u,v)}} \right] G(u,v)$$

where $H(u,v)$ = degradation function

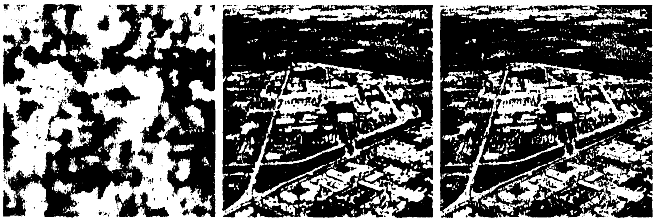
$H^*(u,v)$ = complex conjugate of $H(u,v)$

$|H(u,v)|^2 = H^*(u,v)\, H(u,v)$

$S_\eta(u,v) = |N(u,v)|^2$ = power spectrum of noise

$S_f(u,v) = |F(u,v)|^2$ = power spectrum of undegraded image

Chapter 5
Image Restoration



a b c

FIGURE 5.28 Comparison of inverse- and Wiener filtering. (a) Result of full inverse filtering of Fig. 5.25(b). (b) Radially limited inverse filter result. (c) Wiener filter result.

just computing $\dfrac{G(u,v)}{H(u,v)}$

$D_0 = 75$ radially limited $\dfrac{G(u,v)}{H(u,v)}$

Wiener filtering using interactive values of K

In practice $S_f(u,v) = |F(u,v)|^2$ of the undegraded image is not usually known.

So we simply replace $\dfrac{S_\eta(u,v)}{S_f(u,v)}$ by a constant K

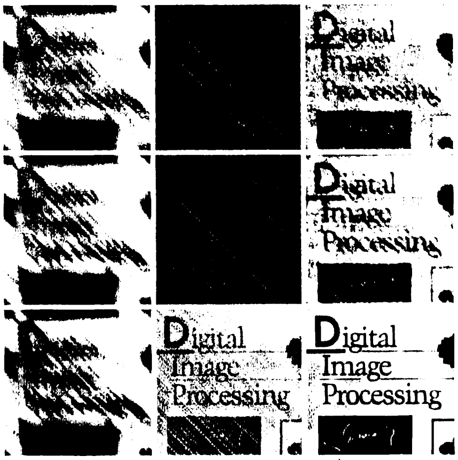
$$\hat{F}(u,v) = \frac{1}{H(u,v)} \frac{|H(u,v)|^2}{|H(u,v)| + K} G(u,v)$$

Chapter 5
Image Restoration

$\sigma^2 = 650$

$\sigma^2 = 65$

$\sigma^2 = 0.065$



inverse filtering Wiener filtering

a b c
d e f
g h i

FIGURE 5.29 (a) Image corrupted by motion blur and additive noise. (b) Result of inverse filtering. (c) Result of Wiener filtering. (d)–(f) Same sequence, but with noise variance one order of magnitude less. (g)–(i) Same sequence, but noise variance reduced by five orders of magnitude from (a). Note in (h) how the deblurred image is quite visible through a "curtain" of noise.

Another example of Wiener filtering

Degradation = motion + noise

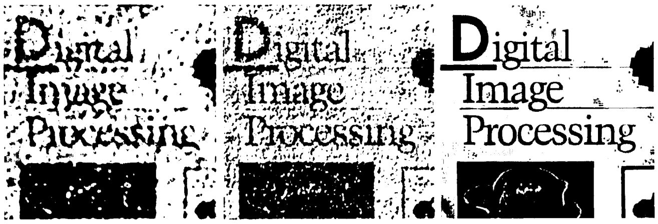
$x_0 = 0.1t$ Gaussian noise

$y_0 = 0.1t$ $\mu = 0$

$\sigma^2 = 650$

Inverse filtering (g) still shows a "curtain" of noise
but reasonable at removing noise

Chapter 5
Image Restoration



a b c

FIGURE 5.30 Results of constrained least squares filtering. Compare (a), (b), and (c) with the Wiener filtering results in Figs. 5.29(c), (f), and (i), respectively.

Additive Gaussian noise

$\sigma^2 = 650$ $\sigma^2 = 65$ $\sigma^2 = 0.065$

This shows the result of processing Fig. 5.29 using constrained least squares filter with γ manually adjusted.

5.9 Constrained Least Squares Filtering

There is an alternative to the Wiener statistical least squared error approach. It relies upon expressing the images and the degration in matrix form.

$$\underline{g} = \underline{H}\,\underline{f} + \underline{\eta} \qquad (1)$$

where

$$\underline{g} = [\; g^{(x,y)}_{row1} \quad g^{(x,y)}_{row2} \quad g^{(x,y)}_{row3} \quad \cdots \quad g^{(x,y)}_{row\,N} \;]$$ this is a very long vector.

$\underline{f}$ & $\underline{\eta}$ have the same form, and dimensions $MN + 1$

$\underline{H}$ has dimensions $MN \times MN$ which is VERY big

Pose the restoration as finding the minimum of $\nabla^2 f$, i.e., smoothness, constrained by (1),

$$\text{minimize} \quad C = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \left[\nabla^2 f(x,y)\right]^2$$

subject to the constraint

$$\|\underline{g} - \underline{H}\,\hat{\underline{f}}\| = \|\underline{\eta}\|^2 \leftarrow \|\underline{w}\|^2 = \underline{w}^T\underline{w}$$

where $\|\underline{\eta}\|^2 = \underline{\eta}^T\underline{\eta}$, $\hat{f}$ is the estimate of the degraded image.

See Castleman [1996]

In the frequency domain the solution is given by.

$$\hat{F}(u,v) = \left[\frac{H^*(u,v)}{|H(u,v)|^2 + \gamma\,|P(u,v)|^2}\right] G(u,v)$$

using $P(u,v) = \mathfrak{F}[p(x,y)]$ where $p(x,y) = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}$ (a positive Laplacian)

γ is adjusted to satisfy the constraint.

γ can be computed. Define

$$\underline{r} = \underline{g} - \underline{H}\,\underline{f}$$

We want to find γ such that

$$\|\underline{r}\|^2 = \|\underline{\eta}\|^2 \pm a$$

accuracy factor

It can be shown that $\|\underline{r}\|^2$ is a monotonically increasing function of γ

Find γ by

1. Specifying an initial value of γ
2. Compute $\|\underline{r}\|^2$
3. Stop if $\|\underline{r}\|^2 = \|\underline{\eta}\|^2 \pm a$. Otherwise

$$\left[\begin{array}{l} \text{increase } \gamma \text{ if } \|\underline{r}\|^2 < \|\underline{\eta}\|^2 + a \\ \text{decrease } \gamma \text{ if } \|r\|^2 > \|\eta\|^2 + a \end{array}\right.$$

Recompute $\hat{F}(u,v)$ using this new value of γ

Go to 2.

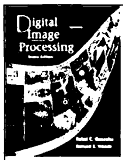
$$\sigma_n^2 = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \left[\eta(x,y) - m_\eta\right]^2$$ is used to estimate the variance of the noise

where $m_\eta = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \eta(x,y)$, i.e. the mean of the noise.

But $\frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \left[\eta(x,y) - m_\eta\right]^2 = \|\eta\|^2$

$$\Rightarrow \|\eta\|^2 = MN\left[\sigma_\eta^2 - m_\eta\right]$$

so we only need the mean and variance of the noise to compute $\|\eta\|^2$ and $\hat{F}(u,v)$



Digital Image Processing, 2nd ed.

www.imageprocessingbook.com

Chapter 5
Image Restoration

a b
FIGURE 5.31 (a) Iteratively determined constrained least squares restoration of Fig. 5.16(b), using correct noise parameters. (b) Result obtained with wrong noise parameters.

Result of restoring image using

γ based on correct noise parameters	γ based on incorrect noise parameters

Random Variables

Random variables often are a source of confusion when first encountered. This need not be so, as the concept of a random variable is in principle quite simple. 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. That is it. A simple, straightforward definition.

Random Variables (Con't)

Part of the confusion often found in connection with random variables is the fact that they are ***functions***. The notation also is partly responsible for the problem. In other words, although typically the notation used to denote a random variable is as we have shown it here, x, or some other appropriate variable, to be strictly formal, a random variable should be written as a function $x(\cdot)$ where the argument is a specific event being considered. However, this is seldom done, and, in our experience, trying to be formal by using function notation complicates the issue more than the clarity it introduces. Thus, we will opt for the less formal notation, with the warning that it must be keep clearly in mind that random variables are functions.

Random Variables (Con't)

Example: Consider again the experiment of drawing a single card from a standard deck of 52 cards. Suppose that we define the following events. A: a heart; B: a spade; C: a club; and D: a diamond, so that $S = \{A, B, C, D\}$. A random variable is easily defined by letting $x = 1$ represent event A, $x = 2$ represent event B, and so on.

As a second illustration, consider the experiment of throwing a single die and observing the value of the up-face. We can define a random variable as the numerical outcome of the experiment (i.e., 1 through 6), but there are many other possibilities. For example, a binary random variable could be defined simply by letting $x = 0$ represent the event that the outcome of throw is an even number and $x = 1$ otherwise.

Random Variables (Con't)

Note the important fact in the examples just given that the probability of the events have not changed; all a random variable does is map events onto the real line.

Random Variables (Con't)

Thus far we have been concerned with random variables whose values are discrete. To handle ***continuous random variables*** we need some additional tools. In the discrete case, the probabilities of events are numbers between 0 and 1. When dealing with continuous quantities (which are not denumerable) we can no longer talk about the "probability of an event" because that probability is zero. This is not as unfamiliar as it may seem. For example, given a continuous function we know that the area of the function between two limits a and b is the integral from a to b of the function. However, the area ***at a point*** is zero because the integral from,say, a to a is zero. We are dealing with the same concept in the case of continuous 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)

Observe that the notation we have used makes no distinction between a random variable and the values it assumes. If confusion is likely to arise, we can use more formal notation in which we let capital letters denote the random variable and lowercase letters denote its values. For example, the cdf using this notation is written as

$$\boxed{F_X(x) = P(X \leq x).}$$

When confusion is not likely, the cdf often is written simply as $F(x)$. This notation will be used in the following discussion when speaking generally about the cdf of a random variable.

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:

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

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

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)

In Sec. 3.3 of the book we used the notation $p(r_k)$, $k = 0,1,\ldots, L - 1$, to denote the ***histogram*** of an image with L possible gray levels, r_k, $k = 0,1,\ldots, L - 1$, where $p(r_k)$ is the probability of the kth gray level (random event) occurring. The discrete random variables in this case are gray levels. It generally is clear from the context whether one is working with continuous or discrete random variables, and whether the use of subscripting is necessary for clarity. Also, uppercase letters (e.g., P) are frequently used to distinguish between probabilities and probability density functions (e.g., p) when they are used together in the same discussion.

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:

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

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.

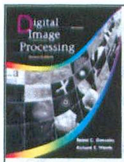
Expected Value and Moments

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

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

If the random variable is *discrete* the definition becomes

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



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 $g(x) = x$:

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

when x is continuous and

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

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 ***variance*** of a random variable, denoted by σ^2, is obtained by letting $g(x) = x^2$ which gives

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

for continuous random variables and

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

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.

Expected Value & Moments (Con't)

We can continue along this line of thought and define the *n*th ***central moment*** of a continuous random variable by letting $g(x) = (x - m)^n$:

$$\boxed{\mu_n = E[(x-m)^n] = \int_{-\infty}^{\infty} (x-m)^n p(x)dx}$$

and

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

for discrete variables, where we assume that $n \geq 0$. Clearly, $\mu_0=1$, $\mu_1=0$, and $\mu_2=\sigma^2$. The term ***central*** when referring to moments indicates that the mean of the random variables has been subtracted out. The moments defined above in which the mean is not subtracted out sometimes are called ***moments about the origin***.

Expected Value & Moments (Con't)

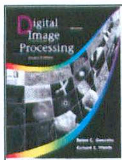
In image processing, moments are used for a variety of purposes, including histogram processing, segmentation, and description. In general, moments are used to characterize the probability density function of a random variable. For example, the second, third, and fourth central moments are intimately related to the ***shape*** of the probability density function of a random variable. The second central moment (the centralized variance) is a measure of ***spread*** of values of a random variable about its mean value, the third central moment is a measure of ***skewness*** (bias to the left or right) of the values of x about the mean value, and the fourth moment is a relative measure of ***flatness***. In general, knowing all the moments of a density specifies that density.

Expected Value & Moments (Con't)

Example: Consider an experiment consisting of repeatedly firing a rifle at a target, and suppose that we wish to characterize the behavior of bullet impacts on the target in terms of whether we are shooting high or low.. We divide the target into an upper and lower region by passing a horizontal line through the bull's-eye. The events of interest are the vertical distances from the center of an impact hole to the horizontal line just described. Distances above the line are considered positive and distances below the line are considered negative. The distance is zero when a bullet hits the line.

Expected Value & Moments (Con't)

In this case, we define a random variable directly as the value of the distances in our sample set. Computing the mean of the random variable indicates whether, ***on average***, we are shooting high or low. If the mean is zero, we know that the average of our shots are on the line. However, the mean does not tell us how far our shots deviated from the horizontal. The variance (or standard deviation) will give us an idea of the ***spread of the shots***. A small variance indicates a tight grouping (with respect to the mean, and in the vertical position); a large variance indicates the opposite. Finally, a third moment of zero would tell us that the spread of the shots is symmetric about the mean value, a positive third moment would indicate a high bias, and a negative third moment would tell us that we are shooting low more than we are shooting high with respect to the mean location.



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

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

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 Gaussian PDF (Con't)

The cumulative distribution function corresponding to the Gaussian density is

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

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

In the previous example, we used a single random variable to describe the behavior of rifle shots with respect to a horizontal line passing through the bull's-eye in the target. Although this is useful information, it certainly leaves a lot to be desired in terms of telling us how well we are shooting with respect to the center of the target. In order to do this we need two random variables that will map our events onto the *xy*-plane. It is not difficult to see how if we wanted to describe events in 3-D space we would need three random variables. In general, we consider in this section the case of n random variables, which we denote by x_1, $x_2, \ldots, x_n$ (the use of n here is not related to our use of the same symbol to denote the nth moment of a random variable).

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*** $\mathbf{x}$ as

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

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

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

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,

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

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)

Cases dealing with expectation operations involving pairs of elements of **x** are particularly important. For example, the joint moment (about the origin) of order kq between variables x_i and x_j

$$\eta_{kq}(i,j) = E[x_i^k x_j^q] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_i^k x_j^q p(x_i, x_j)\,dx_i\,dx_j.$$

Several Random Variables (Con't)

When working with any two random variables (any two elements of **x**) it is common practice to simplify the notation by using x and y to denote the random variables. In this case the joint moment just defined becomes

$$\eta_{kq} = E[x^k y^q] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^k y^q p(x,y)\,dx\,dy.$$

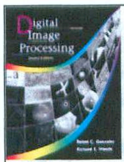
It is easy to see that η_{k0} is the kth moment of x and η_{0q} is the qth moment of y, as defined earlier.

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}:

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

This is often called the cross correlation



Several Random Variables (Con't)

If the condition

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

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

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

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.

Several Random Variables (Con't)

The joint central moment of order kq involving random variables x and y is defined as

$$\boxed{\begin{aligned}\mu_{kq} &= E[(x-m_x)^k(y-m_y)^q] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-m_x)^k(y-m_y)^q p(x,y)\,dx\,dy\end{aligned}}$$

where $m_x = E[x]$ and $m_y = E[y]$ are the means of x and y, as defined earlier. We note that

$$\boxed{\mu_{20} = E[(x-m_x)^2]} \quad \text{and} \quad \boxed{\mu_{02} = E[(y-m_y)^2]}$$

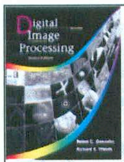
are the variances of x and y, respectively.

Several Random Variables (Con't)

The moment μ_{11}

$$\boxed{\begin{aligned}\mu_{11} &= E[(x-m_x)(y-m_y)] \\ C_{xy} = \quad &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-m_x)(y-m_y)p(x,y)\,dx\,dy\end{aligned}}$$

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

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

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.

$$C_{xy} = E[(x-m_x)(y-m_y)] = E[xy - m_x y - m_y x + m_x m_y]$$
$$= E[xy] - m_x E[y] - m_y E[x] + m_x m_y$$

If $R_{xy} = E[x]E[y]$ The variables are uncorrelated.

and $C_{xy} = R_{xy} - E[x]E[y] = 0$

Several Random Variables (Con't)

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

$$\boxed{\begin{aligned}\gamma &= \frac{\mu_{11}}{\sqrt{\mu_{20}\mu_{02}}} \\ &= \frac{C_{xy}}{\sigma_x \sigma_y} \\ &= E\left[\frac{(x-m_x)}{\sigma_x}\frac{(y-m_y)}{\sigma_y}\right].\end{aligned}}$$

The quantity γ is called the *correlation coefficient* of random variables x and y. It can be shown that γ is in the range $-1 \leq \gamma \leq 1$ (see Problem 12.5). As discussed in Section 12.2.1, the correlation coefficient is used in image processing for matching.

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 $\mathbf{x}$, $\mathbf{C}$ 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

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

and the *covariance matrix* is defined as

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

The Multivariate Gaussian Density (Con't)

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

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

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

Digital Image Processing, 2nd ed.

www.imageprocessingbook.com

Review: Probability and Random Variables

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**.

$$\begin{bmatrix} .3 & -.5 & -.2 \\ -.5 & .2 & 0 \\ -.2 & 0 & .6 \end{bmatrix}$$

element are real

symmetric about the diagonal.

the diagonal elements are the variances of the variables.

$$\begin{bmatrix} .4 & 0 & 0 \\ 0 & .5 & 0 \\ 0 & 0 & .2 \end{bmatrix}$$

in this case the variables are statistically independent