Appendix A Detection and estimation in additive Gaussian noise

A.1 Gaussian random variables

A.1.1 Scalar real Gaussian random variables

A *standard Gaussian* random variable w takes values over the real line and has the probability density function

$$f(w) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{w^2}{2}\right), \qquad w \in \Re. \tag{A.1}$$

The mean of w is zero and the variance is 1. A (general) Gaussian random variable x is of the form

$$x = \sigma w + \mu. \tag{A.2}$$

The mean of x is μ and the variance is equal to σ^2. The random variable x is a one-to-one function of w and thus the probability density function follows from (A.1) as

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \qquad x \in \Re. \tag{A.3}$$

Since the random variable is completely characterized by its mean and variance, we denote x by $\mathcal{N}(\mu, \sigma^2)$. In particular, the standard Gaussian random variable is denoted by $\mathcal{N}(0, 1)$. The *tail* of the Gaussian random variable w

$$Q(a) := \mathbb{P}\{w > a\} \tag{A.4}$$

is plotted in Figure A.1. The plot and the computations $Q(1) = 0.159$ and $Q(3) = 0.00015$ give a sense of how rapidly the tail decays. The tail decays *exponentially* fast as evident by the following upper and lower bounds:

$$\frac{1}{\sqrt{2\pi}a}\left(1 - \frac{1}{a^2}\right) \mathrm{e}^{-a^2/2} < Q(a) < \mathrm{e}^{-a^2/2}, \qquad a > 1. \tag{A.5}$$

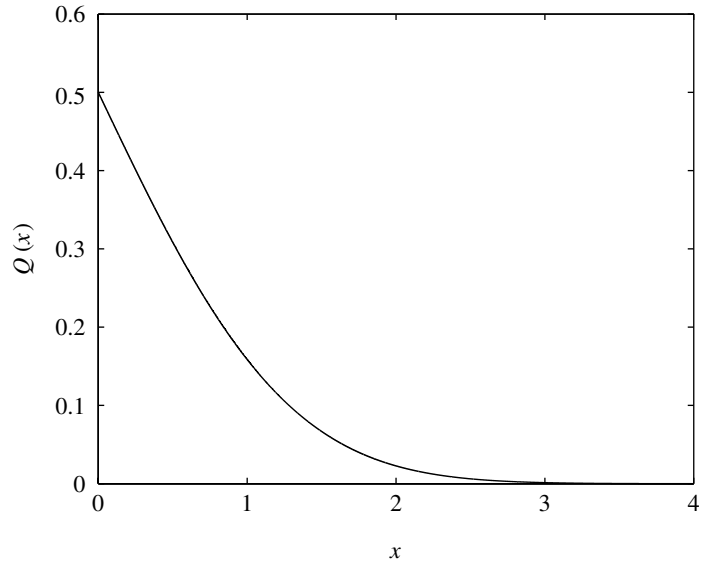

Figure A.1 The Q function.

An important property of Gaussianity is that it is preserved by linear transformations: linear combinations of independent Gaussian random variables are still Gaussian. If $x_1, \ldots, x_n$ are independent and $x_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ (where the $\sim$ notation represents the phrase "is distributed as"), then

$$\sum_{i=1}^{n} c_i x_i \sim \mathcal{N}\left(\sum_{i=1}^{n} c_i \mu_i, \sum_{i=1}^{n} c_i^2 \sigma_i^2\right). \tag{A.6}$$

A.1.2 Real Gaussian random vectors

A *standard Gaussian random vector* $\mathbf{w}$ is a collection of n independent and identically distributed (i.i.d.) standard Gaussian random variables $w_1, \ldots, w_n$. The vector $\mathbf{w} = (w_1, \ldots, w_n)^t$ takes values in the vector space $\Re^n$. The probability density function of $\mathbf{w}$ follows from (A.1):

$$f(\mathbf{w}) = \frac{1}{\left(\sqrt{2\pi}\right)^n} \exp\left(-\frac{\|\mathbf{w}\|^2}{2}\right), \qquad \mathbf{w} \in \Re^n. \tag{A.7}$$

Here $\|\mathbf{w}\| := \sqrt{\sum_{i=1}^{n} w_i^2}$, is the Euclidean distance from the origin to $\mathbf{w} := (w_1, \ldots, w_n)^t$. Note that the density depends only on the *magnitude* of the argument. Since an orthogonal transformation $\mathbf{O}$ (i.e., $\mathbf{O}^t\mathbf{O} = \mathbf{O}\mathbf{O}^t = \mathbf{I}$) preserves the magnitude of a vector, we can immediately conclude:

If $\mathbf{w}$ is standard Gaussian, then $\mathbf{O}\mathbf{w}$ is also standard Gaussian.	(A.8)

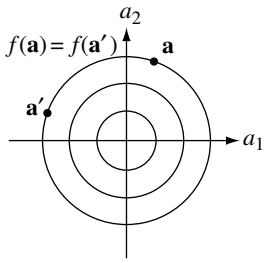

Figure A.2 The isobars, i.e., level sets for the density $f(\mathbf{w})$ of the standard Gaussian random vector, are circles for $n = 2$.

What this result says is that $\mathbf{w}$ has the same distribution in any orthonormal basis. Geometrically, the distribution of $\mathbf{w}$ is invariant to rotations and reflections and hence $\mathbf{w}$ does not prefer any specific direction. Figure A.2 illustrates this *isotropic* behavior of the density of the standard Gaussian random vector $\mathbf{w}$. Another conclusion from (A.8) comes from observing that the rows of matrix $\mathbf{O}$ are orthonormal: the projections of the standard Gaussian random vector in orthogonal directions are independent.

How is the squared magnitude $\|\mathbf{w}\|^2$ distributed? The squared magnitude is equal to the sum of the square of n i.i.d. zero-mean Gaussian random variables. In the literature this sum is called a χ-squared random variable with n degrees of freedom and denoted by χ_n^2. With $n = 2$, the squared magnitude has density

$$f(a) = \frac{1}{2}\exp\left(-\frac{a}{2}\right), \qquad a \geq 0, \tag{A.9}$$

and is said to be *exponentially* distributed. The density of the χ_n^2 random variable for general n is derived in Exercise A.1.

Gaussian random vectors are defined as linear transformations of a standard Gaussian random vector plus a constant vector, a natural generalization of the scalar case (cf. (A.2)):

$$\mathbf{x} = \mathbf{A}\mathbf{w} + \boldsymbol{\mu}. \tag{A.10}$$

Here $\mathbf{A}$ is a matrix representing a linear transformation from $\Re^n$ to $\Re^n$ and $\boldsymbol{\mu}$ is a fixed vector in $\Re^n$. Several implications follow:

1. A standard Gaussian random vector is also Gaussian (with $\mathbf{A} = \mathbf{I}$ and $\boldsymbol{\mu} = \mathbf{0}$).
2. For any $\mathbf{c}$, a vector in $\Re^n$, the random variable

$$\mathbf{c}^t\mathbf{x} \sim \mathcal{N}(\mathbf{c}^t\boldsymbol{\mu}, \mathbf{c}^t\mathbf{A}\mathbf{A}^t\mathbf{c}); \tag{A.11}$$

   this follows directly from (A.6). Thus any linear combination of the elements of a Gaussian random vector is a Gaussian random variable.[1] More generally, any linear transformation of a Gaussian random vector is also Gaussian.
3. If $\mathbf{A}$ is invertible, then the probability density function of $\mathbf{x}$ follows directly from (A.7) and (A.10):

$$f(\mathbf{x}) = \frac{1}{(\sqrt{2\pi})^n\sqrt{\det(\mathbf{A}\mathbf{A}^t)}}\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^t(\mathbf{A}\mathbf{A}^t)^{-1}(\mathbf{x}-\boldsymbol{\mu})\right), \quad \mathbf{x} \in \Re^n. \tag{A.12}$$

[1] This property can be used to define a Gaussian random vector; it is equivalent to our definition in (A.10).

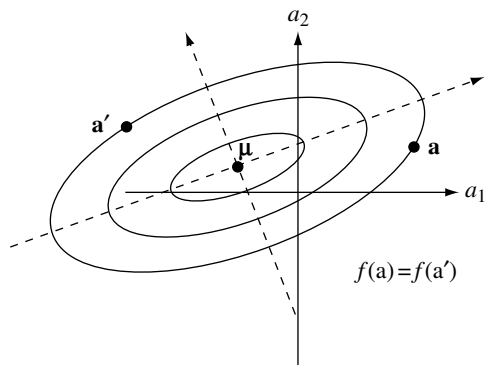

Figure A.3 The isobars of a general Gaussian random vector are ellipses. They corresponds to level sets $\{\mathbf{x} : \|\mathbf{A}^{-1}(\mathbf{x}-\boldsymbol{\mu})\|^2 = c\}$ for constants c.

The isobars of this density are ellipses; the circles of the standard Gaussian vectors being rotated and scaled by $\mathbf{A}$ (Figure A.3). The matrix $\mathbf{A}\mathbf{A}^t$ replaces σ^2 in the scalar Gaussian random variable (cf. (A.3)) and is equal to the *covariance matrix* of $\mathbf{x}$:

$$\mathbf{K} := \mathbb{E}[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^t] = \mathbf{A}\mathbf{A}^t. \tag{A.13}$$

For invertible $\mathbf{A}$, the Gaussian random vector is completely characterized by its mean vector $\boldsymbol{\mu}$ and its covariance matrix $\mathbf{K} = \mathbf{A}\mathbf{A}^t$, which is a symmetric and non-negative definite matrix. We make a few inferences from this observation:

(a) Even though the Gaussian random vector is defined via the matrix $\mathbf{A}$, only the covariance matrix $\mathbf{K} = \mathbf{A}\mathbf{A}^t$ is used to characterize the density of $\mathbf{x}$. Is this surprising? Consider two matrices $\mathbf{A}$ and $\mathbf{AO}$ used to define two Gaussian random vectors as in (A.10). When $\mathbf{O}$ is orthogonal, the covariance matrices of both these random vectors are the same, equal to $\mathbf{A}\mathbf{A}^t$; so the two random vectors must be distributed identically. We can see this directly using our earlier observation (see (A.8)) that $\mathbf{Ow}$ has the same distribution as $\mathbf{w}$ and thus $\mathbf{AOw}$ has the same distribution as $\mathbf{Aw}$.

(b) A Gaussian random vector is composed of independent Gaussian random variables exactly when the covariance matrix $\mathbf{K}$ is diagonal, i.e., the component random variables are *uncorrelated*. Such a random vector is also called a *white* Gaussian random vector.

(c) When the covariance matrix $\mathbf{K}$ is equal to identity, i.e., the component random variables are uncorrelated and have the same unit variance, then the Gaussian random vector reduces to the standard Gaussian random vector.

4. Now suppose that $\mathbf{A}$ is not invertible. Then $\mathbf{Aw}$ maps the standard Gaussian random vector $\mathbf{w}$ into a subspace of dimension less than n, and the density of $\mathbf{Aw}$ is equal to zero outside that subspace and impulsive inside. This means that some components of $\mathbf{Aw}$ can be expressed as linear

combinations of the others. To avoid messy notation, we can focus only on those components of $\mathbf{Aw}$ that are linearly independent and represent them as a lower dimensional vector $\tilde{\mathbf{x}}$, and represent the other components of $\mathbf{Aw}$ as (deterministic) linear combinations of the components of $\tilde{\mathbf{x}}$. By this strategem, we can always take the covariance $\mathbf{K}$ to be invertible.

In general, a Gaussian random vector is completely characterized by its mean $\boldsymbol{\mu}$ and by the covariance matrix $\mathbf{K}$; we denote the random vector by $\mathcal{N}(\boldsymbol{\mu}, \mathbf{K})$.

A.1.3 Complex Gaussian random vectors

So far we have considered real random vectors. In this book, we are primarily interested in *complex* random vectors; these are of the form $\mathbf{x} = \mathbf{x}_\mathrm{R} + \mathrm{j}\mathbf{x}_\mathrm{I}$ where $\mathbf{x}_\mathrm{R}, \mathbf{x}_\mathrm{I}$ are real random vectors. *Complex Gaussian* random vectors are ones in which $[\mathbf{x}_\mathrm{R}, \mathbf{x}_\mathrm{I}]^t$ is a real Gaussian random vector. The distribution is completely specified by the mean and covariance matrix of the real vector $[\mathbf{x}_\mathrm{R}, \mathbf{x}_\mathrm{I}]^t$. Exercise A.3 shows that the same information is contained in the mean $\boldsymbol{\mu}$, the covariance matrix $\mathbf{K}$, and the *pseudo-covariance* matrix $\mathbf{J}$ of the complex vector $\mathbf{x}$, where

$$\boldsymbol{\mu} := \mathbb{E}[\mathbf{x}], \tag{A.14}$$

$$\mathbf{K} := \mathbb{E}[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^*], \tag{A.15}$$

$$\mathbf{J} := \mathbb{E}[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^t]. \tag{A.16}$$

Here, $\mathbf{A}^*$ is the transpose of the matrix $\mathbf{A}$ with each element replaced by its complex conjugate, and $\mathbf{A}^t$ is just the transpose of $\mathbf{A}$. Note that in general the covariance matrix $\mathbf{K}$ of the complex random vector $\mathbf{x}$ by itself is not enough to specify the full second-order statistics of $\mathbf{x}$. Indeed, since $\mathbf{K}$ is Hermitian, i.e., $\mathbf{K} = \mathbf{K}^*$, the diagonal elements are real and the elements in the lower and upper triangles are complex conjugates of each other. Hence it is specified by n^2 real parameters, where n is the (complex) dimension of $\mathbf{x}$. On the other hand, the full second-order statistics of $\mathbf{x}$ are specified by the $n(2n+1)$ real parameters in the symmetric $2n \times 2n$ covariance matrix of $[\mathbf{x}_\mathrm{R}, \mathbf{x}_\mathrm{I}]^t$.

For reasons explained in Chapter 2, in wireless communication we are almost exclusively interested in complex random vectors that have the *circular symmetry* property:

$\mathbf{x}$ is circular symmetric if $\mathrm{e}^{\mathrm{j}\theta}\mathbf{x}$ has the same distribution of $\mathbf{x}$ for any θ. (A.17)

For a circular symmetric complex random vector $\mathbf{x}$,

$$\mathbb{E}[\mathbf{x}] = \mathbb{E}[\mathrm{e}^{\mathrm{j}\theta}\mathbf{x}] = \mathrm{e}^{\mathrm{j}\theta}\mathbb{E}[\mathbf{x}] \tag{A.18}$$

for any θ; hence the mean $\boldsymbol{\mu} = \mathbf{0}$. Moreover

$$\mathbb{E}[\mathbf{x}\mathbf{x}^t] = \mathbb{E}[\mathrm{e}^{\mathrm{j}\theta}\mathbf{x}(\mathrm{e}^{\mathrm{j}\theta}\mathbf{x})^t] = \mathrm{e}^{\mathrm{j}2\theta}\mathbb{E}[\mathbf{x}\mathbf{x}^t] \tag{A.19}$$

for any θ; hence the pseudo-covariance matrix $\mathbf{J}$ is also zero. Thus, the covariance matrix $\mathbf{K}$ fully specifies the first- and second-order statistics of a circular symmetric random vector. And if the complex random vector is also Gaussian, $\mathbf{K}$ in fact specifies its entire statistics. A circular symmetric Gaussian random vector with covariance matrix $\mathbf{K}$ is denoted as $\mathcal{CN}(0,\mathbf{K})$.

Some special cases:

1. A complex Gaussian random variable $w = w_\mathrm{R} + \mathrm{j}w_\mathrm{I}$ with i.i.d. zero-mean Gaussian real and imaginary components is circular symmetric. The circular symmetry of w is in fact a restatement of the rotational invariance of the real Gaussian random vector $[w_\mathrm{R}, w_\mathrm{I}]^t$ already observed (cf. (A.8)). In fact, a circular symmetric Gaussian random variable *must* have i.i.d. zero-mean real and imaginary components (Exercise A.5). The statistics are fully specified by the variance $\sigma^2 := \mathbb{E}[|w|^2]$, and the complex random variable is denoted as $\mathcal{CN}(0, \sigma^2)$. (Note that, in contrast, the statistics of a general complex Gaussian random variable are specified by five real parameters: the means and the variances of the real and imaginary components and their correlation.) The phase of w is *uniform* over the range $[0, 2\pi]$ and independent of the magnitude $\|w\|$, which has a density given by

$$f(r) = \frac{r}{\sigma^2}\exp\left\{\frac{-r^2}{2\sigma^2}\right\}, \qquad r \geq 0 \tag{A.20}$$

and is known as a *Rayleigh* random variable. The square of the magnitude, i.e., $w_1^2 + w_2^2$, is χ_2^2, i.e., exponentially distributed, cf. (A.9). A random variable distributed as $\mathcal{CN}(0, 1)$ is said to be *standard*, with the real and imaginary parts each having variance 1/2.

2. A collection of n i.i.d. $\mathcal{CN}(0, 1)$ random variables forms a standard circular symmetric Gaussian random vector $\mathbf{w}$ and is denoted by $\mathcal{CN}(0, \mathbf{I})$. The density function of $\mathbf{w}$ can be explicitly written as, following from (A.7),

$$f(\mathbf{w}) = \frac{1}{\pi^n}\exp(-\|\mathbf{w}\|^2), \qquad \mathbf{w} \in \mathcal{C}^n. \tag{A.21}$$

As in the case of a real Gaussian random vector $\mathcal{N}(0, \mathbf{I})$ (cf. (A.8)), we have the property that

$$\boxed{\mathbf{U}\mathbf{w} \text{ has the same distribution as } \mathbf{w},} \tag{A.22}$$

for any complex orthogonal matrix $\mathbf{U}$ (such a matrix is called a *unitary* matrix and is characterized by the property $\mathbf{U}^*\mathbf{U} = \mathbf{I}$). The property (A.22) is the complex extension of the isotropic property of the real standard Gaussian random vector (cf. (A.8)). Note the distinction between the *circular*

symmetry (A.17) and the *isotropic* (A.22) properties: the latter is in general much stronger than the former except that they coincide when $\mathbf{w}$ is scalar.

The square of the magnitude of $\mathbf{w}$, as in the real case, is a χ^2_{2n} random variable.

3. If $\mathbf{w}$ is $\mathcal{CN}(0, \mathbf{I})$ and $\mathbf{A}$ is a complex matrix, then $\mathbf{x} = \mathbf{A}\mathbf{w}$ is also circular symmetric Gaussian, with covariance matrix $\mathbf{K} = \mathbf{A}\mathbf{A}^*$, i.e., $\mathcal{CN}(0, \mathbf{K})$. Conversely, any circular symmetric Gaussian random vector with covariance matrix $\mathbf{K}$ can be written as a linearly transformed version of a standard circular symmetric random vector. If $\mathbf{A}$ is invertible, the density function of $\mathbf{x}$ can be explicitly calculated via (A.21), as in (A.12),

$$f(\mathbf{x}) = \frac{1}{\pi^n \det \mathbf{K}} \exp\left(-\mathbf{x}^* \mathbf{K}^{-1} \mathbf{x}\right), \qquad \mathbf{x} \in \mathcal{C}^n. \tag{A.23}$$

When $\mathbf{A}$ is not invertible, the earlier discussion for real random vectors applies here as well: we focus only on the linearly independent components of $\mathbf{x}$, and treat the other components as deterministic linear combinations of these. This allows us to work with a compact notation.

Summary A.1 Complex Gaussian random vectors

- An n-dimensional complex Gaussian random vector $\mathbf{x}$ has real and imaginary components which form a $2n$-dimensional real Gaussian random vector.
- $\mathbf{x}$ is *circular symmetric* if for any θ,

$$\mathrm{e}^{\mathrm{j}\theta}\mathbf{x} \sim \mathbf{x}. \tag{A.24}$$

- A circular symmetric Gaussian $\mathbf{x}$ has zero mean and its statistics are fully specified by the covariance matrix $\mathbf{K} := \mathbb{E}[\mathbf{x}\mathbf{x}^*]$. It is denoted by $\mathcal{CN}(0, \mathbf{K})$.
- The scalar complex random variable $w \sim \mathcal{CN}(0, 1)$ has i.i.d. real and imaginary components each distributed as $\mathcal{N}(0, 1/2)$. The phase of w is uniformly distributed in $[0, 2\pi]$ and independent of its magnitude $|w|$, which is Rayleigh distributed:

$$f(r) = r \exp\left(-\frac{r^2}{2}\right), \qquad r \geq 0. \tag{A.25}$$

  $|w|^2$ is exponentially distributed.
- If the random vector $\mathbf{w} \sim \mathcal{CN}(0, \mathbf{I})$, then its real and imaginary components are all i.i.d., and $\mathbf{w}$ is *isotropic*, i.e., for any unitary matrix $\mathbf{U}$,

$$\mathbf{U}\mathbf{w} \sim \mathbf{w}. \tag{A.26}$$

Equivalently, the projections of $\mathbf{w}$ onto orthogonal directions are i.i.d. $\mathcal{CN}(0,1)$. The squared magnitude $\|\mathbf{w}\|^2$ is distributed as χ^2_{2n} with mean n.

- If $\mathbf{x} \sim \mathcal{CN}(0, \mathbf{K})$ and $\mathbf{K}$ is invertible, then the density of $\mathbf{x}$ is

$$f(\mathbf{x}) = \frac{1}{\pi^n \det \mathbf{K}} \exp(-\mathbf{x}^* \mathbf{K}^{-1} \mathbf{x}), \qquad \mathbf{x} \in \mathcal{C}^n. \tag{A.27}$$

A.2 Detection in Gaussian noise

A.2.1 Scalar detection

Consider the real additive Gaussian noise channel:

$$y = u + w, \tag{A.28}$$

where the transmit symbol u is equally likely to be u_A or u_B and $w \sim \mathcal{N}(0, N_0/2)$ is real Gaussian noise. The *detection* problem involves making a decision on whether u_A or u_B was transmitted based on the observation y. The optimal detector, with the least probability of making an erroneous decision, chooses the symbol that is most likely to have been transmitted given the received signal y, i.e., u_A is chosen if

$$\mathbb{P}\{u = u_A | y\} \geq \mathbb{P}\{u = u_B | y\}. \tag{A.29}$$

Since the two symbols u_A, u_B are equally likely to have been transmitted, Bayes' rule lets us simplify this to the *maximum likelihood* (ML) receiver, which chooses the transmit symbol that makes the observation y most likely. Conditioned on $u = u_i$, the received signal $y \sim \mathcal{N}(u_i, N_0/2)$, $i = A, B$, and the decision rule is to choose u_A if

$$\frac{1}{\sqrt{\pi N_0}} \exp\left(-\frac{(y-u_A)^2}{N_0}\right) \geq \frac{1}{\sqrt{\pi N_0}} \exp\left(-\frac{(y-u_B)^2}{N_0}\right), \tag{A.30}$$

and u_B otherwise. The ML rule in (A.30) further simplifies: choose u_A when

$$|y - u_A| < |y - u_B|. \tag{A.31}$$

The rule is illustrated in Figure A.4 and can be interpreted as corresponding to choosing the *nearest neighboring* transmit symbol. The probability of making an error, the same whether the symbol u_A or u_B was transmitted, is equal to

$$\mathbb{P}\left\{y < \frac{u_A + u_B}{2} | u = u_A\right\} = \mathbb{P}\left\{w > \frac{|u_A - u_B|}{2}\right\} = Q\left(\frac{|u_A - u_B|}{2\sqrt{N_0/2}}\right). \tag{A.32}$$

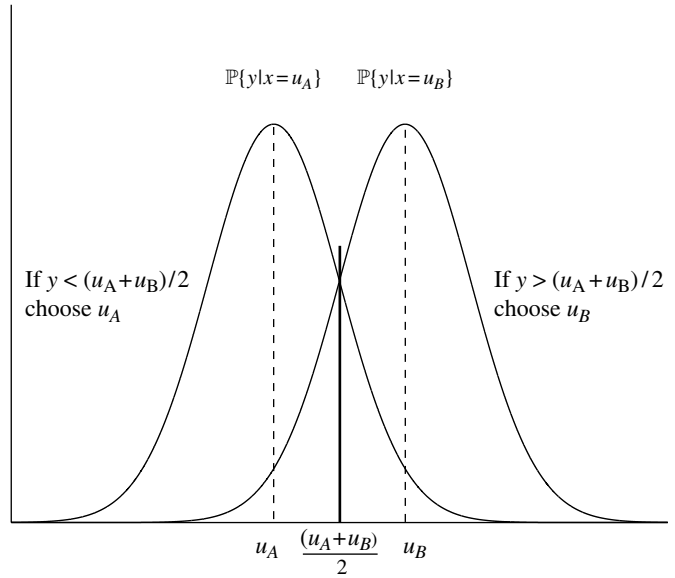

Figure A.4 The ML rule is to choose the symbol that is *closest* to the received symbol.

Thus, the error probability only depends on the distance between the two transmit symbols u_A, u_B.

A.2.2 Detection in a vector space

Now consider detecting the transmit *vector* $\mathbf{u}$ equally likely to be $\mathbf{u}_A$ or $\mathbf{u}_B$ (both elements of $\Re^n$). The received vector is

$$\mathbf{y} = \mathbf{u} + \mathbf{w}, \tag{A.33}$$

and $\mathbf{w} \sim \mathcal{N}(0, (N_0/2)\mathbf{I})$. Analogous to (A.30), the ML decision rule is to choose $\mathbf{u}_A$ if

$$\frac{1}{(\pi N_0)^{n/2}} \exp\left(-\frac{\|\mathbf{y}-\mathbf{u}_A\|^2}{N_0}\right) \geq \frac{1}{(\pi N_0)^{n/2}} \exp\left(-\frac{\|\mathbf{y}-\mathbf{u}_B\|^2}{N_0}\right), \tag{A.34}$$

which simplifies to, analogous to (A.31),

$$\|\mathbf{y}-\mathbf{u}_A\| < \|\mathbf{y}-\mathbf{u}_B\|, \tag{A.35}$$

the same *nearest neighbor* rule. By the isotropic property of the Gaussian noise, we expect the error probability to be the same for both the transmit symbols $\mathbf{u}_A, \mathbf{u}_B$. Suppose $\mathbf{u}_A$ is transmitted, so $\mathbf{y} = \mathbf{u}_A + \mathbf{w}$. Then an error occurs when the event in (A.35) does not occur, i.e., $\|\mathbf{w}\| > \|\mathbf{w}+\mathbf{u}_A-\mathbf{u}_B\|$. So, the error probability is equal to

$$\mathbb{P}\{\|\mathbf{w}\|^2 > \|\mathbf{w}+\mathbf{u}_A-\mathbf{u}_B\|^2\} = \mathbb{P}\left\{(\mathbf{u}_A-\mathbf{u}_B)^t\mathbf{w} < -\frac{\|\mathbf{u}_A-\mathbf{u}_B\|^2}{2}\right\}. \tag{A.36}$$

Geometrically, this says that the decision regions are the two sides of the hyperplane perpendicular to the vector $\mathbf{u}_B - \mathbf{u}_A$, and an error occurs when the received vector lies on the side of the hyperplane opposite to the transmit vector (Figure A.5). We know from (A.11) that $(\mathbf{u}_A - \mathbf{u}_B)^t\mathbf{w} \sim \mathcal{N}(0, \|\mathbf{u}_A - \mathbf{u}_B\|^2 N_0/2)$. Thus the error probability in (A.36) can be written in compact notation as

$$Q\left(\frac{\|\mathbf{u}_A - \mathbf{u}_B\|}{2\sqrt{N_0/2}}\right). \tag{A.37}$$

The quantity $\|\mathbf{u}_A - \mathbf{u}_B\|/2$ is the distance from each of the vectors $\mathbf{u}_A, \mathbf{u}_B$ to the decision boundary. Comparing the error probability in (A.37) with that in the scalar case (cf. (A.32)), we see that the the error probability depends only on the Euclidean distance between $\mathbf{u}_A$ and $\mathbf{u}_B$ and not on the specific orientations and magnitudes of $\mathbf{u}_A$ and $\mathbf{u}_B$.

An alternative view

To see how we could have reduced the vector detection problem to the scalar one, consider a small change in the way we think of the transmit vector $\mathbf{u} \in \{\mathbf{u}_A, \mathbf{u}_B\}$. We can write the transmit vector $\mathbf{u}$ as

$$\mathbf{u} = x(\mathbf{u}_A - \mathbf{u}_B) + \frac{1}{2}(\mathbf{u}_A + \mathbf{u}_B), \tag{A.38}$$

where the information is in the *scalar* x, which is equally likely to be $\pm 1/2$. Substituting (A.38) in (A.33), we can subtract the constant vector $(\mathbf{u}_A + \mathbf{u}_B)/2$ from the received signal $\mathbf{y}$ to arrive at

$$\mathbf{y} - \frac{1}{2}(\mathbf{u}_A + \mathbf{u}_B) = x(\mathbf{u}_A - \mathbf{u}_B) + \mathbf{w}. \tag{A.39}$$

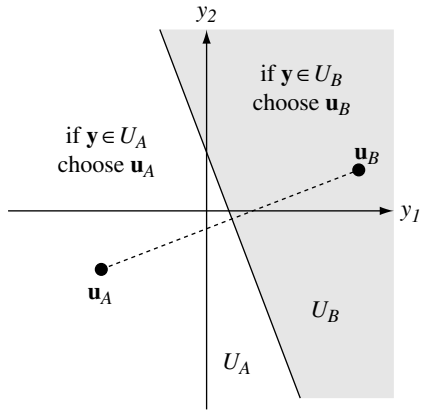

Figure A.5 The decision region for the nearest neighbor rule is partitioned by the hyperplane perpendicular to $\mathbf{u}_B - \mathbf{u}_A$ and halfway between $\mathbf{u}_A$ and $\mathbf{u}_B$.

We observe that the transmit symbol (a scalar x) is only in a specific direction:

$$\mathbf{v} := (\mathbf{u}_A - \mathbf{u}_B)/\|\mathbf{u}_A - \mathbf{u}_B\|. \tag{A.40}$$

The components of the received vector $\mathbf{y}$ in the directions orthogonal to $\mathbf{v}$ contain purely noise, and, due to the isotropic property of $\mathbf{w}$, the noise in these directions is also independent of the noise in the signal direction. This means that the components of the received vector in these directions are *irrrelevant* for detection. Therefore projecting the received vector along the signal direction $\mathbf{v}$ provides all the necessary information for detection:

$$\tilde{y} := \mathbf{v}^t \left(\mathbf{y} - \frac{1}{2}(\mathbf{u}_A + \mathbf{u}_B) \right). \tag{A.41}$$

We have thus reduced the vector detection problem to the scalar one. Figure A.6 summarizes the situation.

More formally, we are viewing the received vector in a different orthonormal basis: the first direction is that given by $\mathbf{v}$, and the other directions are orthogonal to each other and to the first one. In other words, we form an orthogonal matrix $\mathbf{O}$ whose first row is $\mathbf{v}$, and the other rows are orthogonal to each other and to the first one and have unit norm. Then

$$\mathbf{O}\left(\mathbf{y} - \frac{1}{2}(\mathbf{u}_A + \mathbf{u}_B) \right) = \begin{bmatrix} x\|\mathbf{u}_A - \mathbf{u}_B\| \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \mathbf{O}\mathbf{w}. \tag{A.42}$$

Since $\mathbf{O}\mathbf{w} \sim \mathcal{N}(0, (N_0/2)\mathbf{I})$ (cf. (A.8)), this means that all but the first component of the vector $\mathbf{O}(\mathbf{y} - \frac{1}{2}(\mathbf{u}_A + \mathbf{u}_B))$ are independent of the transmit symbol x and the noise in the first component. Thus it suffices to make a decision on the transmit symbol x, using only the first component, which is precisely (A.41).

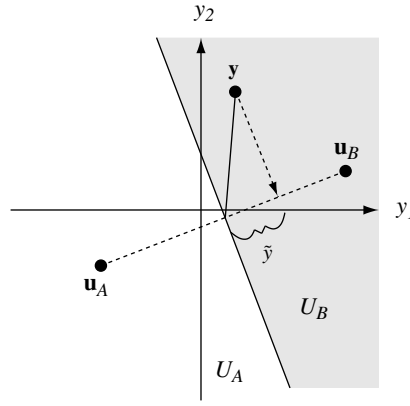

Figure A.6 Projecting the received vector $\mathbf{y}$ onto the signal direction $\mathbf{v}$ reduces the vector detection problem to the scalar one.

This important observation can be summarized:

1. In technical jargon, the scalar $\tilde{y}$ in (A.41) is called a *sufficient statistic* of the received vector $\mathbf{y}$ to detect the transmit symbol u.
2. The sufficient statistic $\tilde{y}$ is a projection of the received signal in the signal direction $\mathbf{v}$: in the literature on communication theory, this operation is called a *matched filter*; the linear filter at the receiver is "matched" to the direction of the transmit signal.
3. This argument explains why the error probability depends on $\mathbf{u}_A$ and $\mathbf{u}_B$ only through the distance between them: the noise is isotropic and the entire detection problem is rotationally invariant.

We now arrive at a scalar detection problem:

$$\tilde{y} = x\|\mathbf{u}_A - \mathbf{u}_B\| + w, \tag{A.43}$$

where w, the first component of $\mathbf{Ow}$ is $\mathcal{N}(0, N_0/2)$ and independent of the transmit symbol u. The effective distance between the two constellation points is $\|\mathbf{u}_A - \mathbf{u}_B\|$. The error probability is, from (A.32),

$$Q\left(\frac{\|\mathbf{u}_A - \mathbf{u}_B\|}{2\sqrt{N_0/2}}\right), \tag{A.44}$$

the same as that arrived at in (A.37), via a direct calculation.

The above argument for binary detection generalizes naturally to the case when the transmit vector can be one of M vectors $\mathbf{u}_1, \ldots, \mathbf{u}_M$. The projection of $\mathbf{y}$ onto the subspace spanned by $\mathbf{u}_1, \ldots, \mathbf{u}_M$ is a sufficient statistic for the detection problem. In the special case when the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_M$ are collinear, i.e., $\mathbf{u}_i = \mathbf{h}x_i$ for some vector $\mathbf{h}$ (for example, when we are transmitting from a PAM constellation), then a projection onto the direction $\mathbf{h}$ provides a sufficient statistic.

A.2.3 Detection in a complex vector space

Consider detecting the transmit symbol $\mathbf{u}$, equally likely to be one of two complex vectors $\mathbf{u}_A, \mathbf{u}_B$ in additive standard complex Gaussian noise. The received complex vector is

$$\mathbf{y} = \mathbf{u} + \mathbf{w}, \tag{A.45}$$

where $\mathbf{w} \sim \mathcal{CN}(0, N_0\mathbf{I})$. We can proceed as in the real case. Write

$$\mathbf{u} = x(\mathbf{u}_A - \mathbf{u}_B) + \frac{1}{2}(\mathbf{u}_A + \mathbf{u}_B). \tag{A.46}$$

The signal is in the direction

$$\mathbf{v} := (\mathbf{u}_A - \mathbf{u}_B)/\|\mathbf{u}_A - \mathbf{u}_B\|. \tag{A.47}$$

Projection of the received vector $\mathbf{y}$ onto $\mathbf{v}$ provides a (complex) scalar sufficient statistic:

$$\tilde{y} := \mathbf{v}^* \left(\mathbf{y} - \frac{1}{2}(\mathbf{u}_A + \mathbf{u}_B) \right) = x\|\mathbf{u}_A - \mathbf{u}_B\| + w, \tag{A.48}$$

where $w \sim \mathcal{CN}(0, N_0)$. Note that since x is real ($\pm 1/2$), we can further extract a sufficient statistic by looking only at the real component of $\tilde{\mathbf{y}}$:

$$\Re[\tilde{\mathbf{y}}] = x\|\mathbf{u}_A - \mathbf{u}_B\| + \Re[w], \tag{A.49}$$

where $\Re[w] \sim N(0, N_0/2)$. The error probability is exactly as in (A.44):

$$Q\left(\frac{\|\mathbf{u}_A - \mathbf{u}_B\|}{2\sqrt{N_0/2}} \right), \tag{A.50}$$

Note that although $\mathbf{u}_A$ and $\mathbf{u}_B$ are *complex* vectors, the transmit vectors

$$x(\mathbf{u}_A - \mathbf{u}_B) + \frac{1}{2}(\mathbf{u}_A + \mathbf{u}_B), \qquad x = \pm 1, \tag{A.51}$$

lie in a subspace of one *real* dimension and hence we can extract a *real* sufficient statistic. If there are more than two possible transmit vectors and they are of the form $\mathbf{h}x_i$, where x_i is *complex* valued, $\mathbf{h}^*\mathbf{y}$ is still a sufficient statistic but $\Re[\mathbf{h}^*\mathbf{y}]$ is sufficient only if x is real (for example, when we are transmitting a PAM constellation).

The main results of our discussion are summarized below.

Summary A.2 Vector detection in complex Gaussian noise

Binary signals

The transmit vector $\mathbf{u}$ is either $\mathbf{u}_A$ or $\mathbf{u}_B$ and we wish to detect $\mathbf{u}$ from received vector

$$\mathbf{y} = \mathbf{u} + \mathbf{w}, \tag{A.52}$$

where $\mathbf{w} \sim \mathcal{CN}(0, N_0\mathbf{I})$. The ML detector picks the transmit vector closest to $\mathbf{y}$ and the error probability is

$$Q\left(\frac{\|\mathbf{u}_A - \mathbf{u}_B\|}{2\sqrt{N_0/2}} \right). \tag{A.53}$$

Collinear signals
The transmit symbol x is equally likely to take one of a finite set of values in $\mathcal{C}$ (the *constellation* points) and the received vector is

$$\mathbf{y} = \mathbf{h}x + \mathbf{w}, \tag{A.54}$$

where $\mathbf{h}$ is a fixed vector.

Projecting $\mathbf{y}$ onto the unit vector $\mathbf{v} := \mathbf{h}/\|\mathbf{h}\|$ yields a scalar sufficient statistic:

$$\mathbf{v}^*\mathbf{y} = \|\mathbf{h}\|x + w. \tag{A.55}$$

Here $w \sim \mathcal{CN}(0, N_0)$.

If further the constellation is real-valued, then

$$\Re[\mathbf{v}^*\mathbf{y}] = \|\mathbf{h}\|x + \Re[w] \tag{A.56}$$

is sufficient. Here $\Re[w] \sim \mathcal{N}(0, N_0/2)$.

With antipodal signalling, $x = \pm a$, the ML error probability is simply

$$Q\left(\frac{a\|\mathbf{h}\|}{\sqrt{N_0/2}}\right). \tag{A.57}$$

Via a translation, the binary signal detection problem in the first part of the summary can be reduced to this antipodal signalling scenario.

A.3 Estimation in Gaussian noise

A.3.1 Scalar estimation

Consider a zero-mean real signal x embedded in independent additive real Gaussian noise ($w \sim \mathcal{N}(0, N_0/2)$):

$$y = x + w. \tag{A.58}$$

Suppose we wish to come up with an estimate $\hat{x}$ of x and we use the mean squared error (MSE) to evaluate the performance:

$$\text{MSE} := \mathbb{E}[(x - \hat{x})^2], \tag{A.59}$$

where the averaging is over the randomness of both the signal x and the noise w. This problem is quite different from the detection problem studied in Section A.2. The estimate that yields the smallest mean squared error is the classical *conditional mean*:

$$\hat{x} = \mathbb{E}[x|y], \tag{A.60}$$

which has the important *orthogonality* property: the error is independent of the observation. In particular, this implies that

$$\mathbb{E}[(\hat{x} - x)y] = 0. \tag{A.61}$$

The *orthogonality principle* is a classical result and all standard textbooks dealing with probability theory and random variables treat this material.

In general, the conditional mean $\mathbb{E}[x|y]$ is some complicated non-linear function of y. To simplify the analysis, one studies the restricted class of linear estimates that minimize the MSE. This restriction is without loss of generality in the important case when x is a Gaussian random variable because, in this case, the conditional mean operator is actually *linear*.

Since x is zero mean, linear estimates are of the form $\hat{x} = cy$ for some real number c. What is the best coefficient c? This can be derived directly or via using the orthogonality principle (cf. (A.61)):

$$c = \frac{\mathbb{E}[x^2]}{\mathbb{E}[x^2] + N_0/2}. \tag{A.62}$$

Intuitively, we are weighting the received signal y by the transmitted signal energy as a fraction of the received signal energy. The corresponding minimum mean squared error (MMSE) is

$$\text{MMSE} = \frac{\mathbb{E}[x^2]N_0/2}{\mathbb{E}[x^2] + N_0/2}. \tag{A.63}$$

A.3.2 Estimation in a vector space

Now consider estimating x in a vector space:

$$\mathbf{y} = \mathbf{h}x + \mathbf{w}. \tag{A.64}$$

Here x and $\mathbf{w} \sim \mathcal{N}(0, (N_0/2)\mathbf{I})$ are independent and $\mathbf{h}$ is a fixed vector in $\Re^n$. We have seen that the projection of $\mathbf{y}$ in the direction of $\mathbf{h}$,

$$\tilde{y} = \frac{\mathbf{h}^t\mathbf{y}}{\|\mathbf{h}\|^2} = x + w, \tag{A.65}$$

is a sufficient statistic: the projections of $\mathbf{y}$ in directions orthogonal to $\mathbf{h}$ are independent of both the signal x and w, the noise in the direction

of $\mathbf{h}$. Thus we can convert this problem to a scalar one: estimate x from $\tilde{y}$, with $w \sim \mathcal{N}(0, N_0/(2\|\mathbf{h}\|^2))$. Now this problem is identical to the scalar estimation problem in (A.58) with the energy of the noise w suppressed by a factor of $\|\mathbf{h}\|^2$. The best linear estimate of x is thus, as in (A.62),

$$\frac{\mathbb{E}[x^2]\|\mathbf{h}\|^2}{\mathbb{E}[x^2]\|\mathbf{h}\|^2 + N_0/2}\,\tilde{y}. \tag{A.66}$$

We can combine the sufficient statistic calculation in (A.65) and the scalar linear estimate in (A.66) to arrive at the best linear estimate $\hat{x} = \mathbf{c}^t\mathbf{y}$ of x from $\mathbf{y}$:

$$\mathbf{c} = \frac{\mathbb{E}[x^2]}{\mathbb{E}[x^2]\|\mathbf{h}\|^2 + N_0/2}\,\mathbf{h}. \tag{A.67}$$

The corresponding minimum mean squared error is

$$\text{MMSE} = \frac{\mathbb{E}[x^2]N_0/2}{\mathbb{E}[x^2]\|\mathbf{h}\|^2 + N_0/2}. \tag{A.68}$$

An alternative performance measure to evaluate linear estimators is the *signal-to-noise ratio* (SNR) defined as the ratio of the signal energy in the estimate to the noise energy:

$$\text{SNR} := \frac{(\mathbf{c}^t\mathbf{h})^2\mathbb{E}[x^2]}{\|\mathbf{c}\|^2 N_0/2}. \tag{A.69}$$

That the matched filter ($\mathbf{c} = \mathbf{h}$) yields the maximal SNR at the output of any linear filter is a classical result in communication theory (and is studied in all standard textbooks on the topic). It follows directly from the Cauchy–Schwartz inequality:

$$(\mathbf{c}^t\mathbf{h})^2 \le \|\mathbf{c}\|^2\,\|\mathbf{h}\|^2, \tag{A.70}$$

with equality exactly when $\mathbf{c} = \mathbf{h}$. The fact that the matched filter maximizes the SNR and when appropriately scaled yields the MMSE is not coincidental; this is studied in greater detail in Exercise A.8.

A.3.3 Estimation in a complex vector space

The extension of our discussion to the complex field is natural. Let us first consider scalar complex estimation, an extension of the basic real setup in (A.58):

$$y = x + w, \tag{A.71}$$

where $w \sim \mathcal{CN}(0, N_0)$ is independent of the complex zero-mean transmitted signal x. We are interested in a linear estimate $\hat{x} = c^* y$, for some complex constant c. The performance metric is

$$\text{MSE} = \mathbb{E}[|x - \hat{x}|^2]. \tag{A.72}$$

The best linear estimate $\hat{x} = c^* y$ can be directly calculated to be, as an extension of (A.62),

$$c = \frac{\mathbb{E}[|x|^2]}{\mathbb{E}[|x|^2] + N_0}. \tag{A.73}$$

The corresponding minimum MSE is

$$\text{MMSE} = \frac{\mathbb{E}[|x|^2] N_0}{\mathbb{E}[|x|^2] + N_0}. \tag{A.74}$$

The orthogonality principle (cf. (A.61)) for the complex case is extended to:

$$\mathbb{E}[(\hat{x} - x) y^*] = 0. \tag{A.75}$$

The linear estimate in (A.73) is easily seen to satisfy (A.75).

Now let us consider estimating the scalar complex zero mean x in a complex vector space:

$$\mathbf{y} = \mathbf{h} x + \mathbf{w}, \tag{A.76}$$

with $\mathbf{w} \sim \mathcal{CN}(0, N_0 \mathbf{I})$ independent of x and $\mathbf{h}$ a fixed vector in $\mathcal{C}^n$. The projection of $\mathbf{y}$ in the direction of $\mathbf{h}$ is a sufficient statistic and we can reduce the vector estimation problem to a scalar one: estimate x from

$$\tilde{y} = \frac{\mathbf{h}^* \mathbf{y}}{\|\mathbf{h}\|^2} = x + w, \tag{A.77}$$

where $w \sim \mathcal{CN}(0, N_0 / \|\mathbf{h}\|^2)$.

Thus the best linear estimator is, as an extension of (A.67),

$$\mathbf{c} = \frac{\mathbb{E}[|x|^2]}{\mathbb{E}[|x|^2] \|\mathbf{h}\|^2 + N_0} \mathbf{h}. \tag{A.78}$$

The corresponding minimum MSE is, as an extension of (A.68),

$$\text{MMSE} = \frac{\mathbb{E}[x^2] N_0}{\mathbb{E}[x^2] \|\mathbf{h}\|^2 + N_0}. \tag{A.79}$$

Summary A.3 Mean square estimation in a complex vector space

The linear estimate with the smallest mean squared error of x from

$$y = x + w, \tag{A.80}$$

with $w \sim \mathcal{CN}(0, N_0)$, is

$$\hat{x} = \frac{\mathbb{E}[|x|^2]}{\mathbb{E}[|x|^2] + N_0} y. \tag{A.81}$$

To estimate x from

$$\mathbf{y} = \mathbf{h}x + \mathbf{w}, \tag{A.82}$$

where $\mathbf{w} \sim \mathcal{CN}(0, N_0\mathbf{I})$,

$$\mathbf{h}^*\mathbf{y} \tag{A.83}$$

is a sufficient statistic, reducing the vector estimation problem to the scalar one.

The best *linear* estimator is

$$\hat{x} = \frac{\mathbb{E}[|x|^2]}{\mathbb{E}[|x|^2]\|\mathbf{h}\|^2 + N_0} \mathbf{h}^*\mathbf{y}. \tag{A.84}$$

The corresponding minimum mean squared error (MMSE) is:

$$\text{MMSE} = \frac{\mathbb{E}[|x|^2]N_0}{\mathbb{E}[|x|^2]\|\mathbf{h}\|^2 + N_0}. \tag{A.85}$$

In the special case when $x \sim \mathcal{CN}(\mu, \sigma^2)$, this estimator yields the minimum mean squared error among *all* estimators, linear or non-linear.

A.4 Exercises

Exercise A.1 Consider the n-dimensional standard Gaussian random vector $\mathbf{w} \sim \mathcal{N}(0, \mathbf{I}_n)$ and its squared magnitude $\|\mathbf{w}\|^2$.

1. With $n = 1$, show that the density of $\|\mathbf{w}\|^2$ is

$$f_1(a) = \frac{1}{\sqrt{2\pi a}} \exp\left(-\frac{a}{2}\right), \qquad a \geq 0. \tag{A.86}$$

2. For any n, show that the density of $\|\mathbf{w}\|^2$ (denoted by $f_n(\cdot)$) satisfies the recursive relation:

$$f_{n+2}(a) = \frac{a}{n} f_n(a), \qquad a \geq 0. \tag{A.87}$$

3. Using the formulas for the densities for $n = 1$ and 2 ((A.86) and (A.9), respectively) and the recurisve relation in (A.87) determine the density of $\|\mathbf{w}\|^2$ for $n \geq 3$.

Exercise A.2 Let $\{w(t)\}$ be white Gaussian noise with power spectral density $N_0/2$. Let $\mathbf{s_1}, \ldots, \mathbf{s_M}$ be a set of finite orthonormal waveforms (i.e., orthogonal and unit energy), and define $z_i = \int_{-\infty}^{\infty} w(t) s_i(t) \mathrm{d}t$. Find the joint distribution of $\mathbf{z}$. *Hint*: Recall the isotropic property of the normalized Gaussian random vector (see (A.8)).

Exercise A.3 Consider a complex random vector $\mathbf{x}$.

1. Verify that the second-order statistics of $\mathbf{x}$ (i.e., the covariance matrix of the real representation $[\Re[\mathbf{x}], \Im[\mathbf{x}]]^t$) can be completely specified by the covariance and pseudo-covariance matrices of $\mathbf{x}$, defined in (A.15) and (A.16) respectively.
2. In the case where $\mathbf{x}$ is circular symmetric, express the covariance matrix $[\Re[\mathbf{x}], \Im[\mathbf{x}]]^t$ in terms of the covariance matrix of the complex vector $\mathbf{x}$ only.

Exercise A.4 Consider a complex Gaussian random vector $\mathbf{x}$.

1. Show that a necessary and sufficient condition for $\mathbf{x}$ to be circular symmetric is that the mean $\boldsymbol{\mu}$ and the pseudo-covariance matrix $\mathbf{J}$ are zero.
2. Now suppose the relationship between the covariance matrix of $[\Re[\mathbf{x}], \Im[\mathbf{x}]]^t$ and the covariance matrix of $\mathbf{x}$ in part (2) of Exercise A.3 holds. Can we conclude that $\mathbf{x}$ is circular symmetric?

Exercise A.5 Show that a circular symmetric complex Gaussian random variable must have i.i.d. real and imaginary components.

Exercise A.6 Let $\mathbf{x}$ be an n-dimensional i.i.d. complex Gaussian random vector, with the real and imaginary parts distributed as $\mathcal{N}(0, \mathbf{K}_x)$ where $\mathbf{K}_x$ is a 2×2 covariance matrix. Suppose $\mathbf{U}$ is a unitary matrix (i.e., $\mathbf{U}^*\mathbf{U} = \mathbf{I}$). Identify the conditions on $\mathbf{K}_x$ under which $\mathbf{U}\mathbf{x}$ has the same distribution as $\mathbf{x}$.

Exercise A.7 Let $\mathbf{z}$ be an n-dimensional i.i.d. complex Gaussian random vector, with the real and imaginary parts distributed as $\mathcal{N}(0, \mathbf{K}_x)$ where $\mathbf{K}_x$ is a 2×2 covariance matrix. We wish to detect a scalar x, equally likely to be ± 1 from

$$\mathbf{y} = \mathbf{h}x + \mathbf{z}, \tag{A.88}$$

where x and $\mathbf{z}$ are independent and $\mathbf{h}$ is a fixed vector in $\mathcal{C}^n$. Identify the conditions on $\mathbf{K}_x$ under which the scalar $\mathbf{h}^*\mathbf{y}$ is a sufficient statistic to detect x from $\mathbf{y}$.

Exercise A.8 Consider estimating the real zero-mean scalar x from:

$$\mathbf{y} = \mathbf{h}x + \mathbf{w}, \tag{A.89}$$

where $\mathbf{w} \sim \mathcal{N}(0, N_0/2\mathbf{I})$ is uncorrelated with x and $\mathbf{h}$ is a fixed vector in $\Re^n$.

1. Consider the scaled linear estimate $\mathbf{c}^t\mathbf{y}$ (with the normalization $\|\mathbf{c}\| = 1$):

$$\hat{x} := a\mathbf{c}^t\mathbf{y} = (a\mathbf{c}^t\mathbf{h})\, x + a\mathbf{c}^t\mathbf{z}. \tag{A.90}$$

   Show that the constant a that minimizes the mean square error ($\mathbb{E}[(x-\hat{x})^2]$) is equal to

$$\frac{\mathbb{E}[x^2]|\mathbf{c}^t\mathbf{h}|^2}{\mathbb{E}[x^2]|\mathbf{c}^t\mathbf{h}|^2 + N_0/2}. \tag{A.91}$$

2. Calculate the minimal mean square error (denoted by MMSE) of the linear estimate in (A.90) (by using the value of a in (A.91). Show that

$$\frac{\mathbb{E}[x^2]}{\text{MMSE}} = 1 + \text{SNR} := 1 + \frac{\mathbb{E}[x^2]|\mathbf{c}^t\mathbf{h}|^2}{N_0/2}. \tag{A.92}$$

For every fixed linear estimator $\mathbf{c}$, this shows the relationship between the corresponding SNR and MMSE (of an appropriately scaled estimate). In particular, this relation holds when we optimize over all $\mathbf{c}$ leading to the best linear estimator.