3. Gaussian Noise

R.G. Gallager

The stochastic processes of almost exclusive interest in modeling channel noise are the Gaussian processes. Gaussian processes are stochastic processes for which the random variables $N(t_1), N(t_2), \ldots, N(t_k)$ are jointly Gaussian for all $t_1, \ldots, t_k$ and all $k > 0$. Today we start by giving a more complete discussion of jointly Gaussian random variables. We restrict our attention to zero mean jointly Gaussian rv's and zero mean Gaussian processes, both because a mean can be trivially added to a zero mean fluctuation, and because zero mean processes are used to model noise.

1 Zero mean jointly Gaussian random variables

We represent a set of k random variables (rv's) $Z_1, \ldots, Z_k$ as a random vector (rv) $\boldsymbol{Z} = (Z_1, \ldots, Z_k)^\intercal$. Such vectors satisfy the axioms of a vector space, but here we use vector notation primarily as a notational convenience. A k-vector $\boldsymbol{Z}$ is said to be zero mean jointly Gaussian if $\boldsymbol{Z} = A\boldsymbol{N}$ where $\boldsymbol{N} = (N_1, N_2, \ldots, N_m)^\intercal$ is a vector of iid normalized Gaussian random variables and A is a k by m real matrix. Each rv Z_j is zero mean Gaussian since it is a linear combination of zero mean iid Gaussian rv's. In addition, the vector rv $\boldsymbol{Z}$ has the special property that all linear combinations of its components are also Gaussian. In particular, if a rv $\boldsymbol{Y}$ is defined as $\boldsymbol{Y} = B\boldsymbol{Z}$, where B is a k by k real matrix, then

$$\boldsymbol{Y} = B(A\boldsymbol{N}) = (BA)\boldsymbol{N}.$$

Since BA is a k by m real matrix, $\boldsymbol{Y}$ is also jointly Gaussian.

The above definition of jointly Gaussian looks a little contrived at first, but is in fact very natural. Gaussian rv's often make excellent models for physical noise-like processes because noise is often the summation of many small effects, and the central limit theorem is a mathematically exact way of saying that the sum of a very large number of small random variables is approximately Gaussian. Even when these variables are statistically dependent, they are linear combinations of a common set of the small random variables. Thus the jointly Gaussian assumption is very closely linked to the assumption that the individual rv's are Gaussian.

The covariance of two rv's Z_1, Z_2 is by definition $\mathsf{E}[(Z_1 - \bar{Z}_1)(Z_2 - \bar{Z}_2)]$. For zero mean rv's (all of the rv's in this lecture will be zero mean), the covariance is $\mathsf{E}[Z_1 Z_2]$. For a rv $\boldsymbol{Z} = (Z_1, \ldots Z_k)^\intercal$ the covariance between each pair of random variables is very

conveniently represented by the covariance matrix,

$$\mathsf{K}_{\boldsymbol{Z}} = \mathsf{E}[\boldsymbol{Z}\boldsymbol{Z}^{\mathsf{T}}]$$

For a vector $\boldsymbol{N} = N_1, \ldots, N_m$ of normalized Gaussian rv's, $E[N_l N_j] = 0$ for $l \neq j$ and 1 for $l = j$. Thus

$$\mathsf{K}_{\boldsymbol{N}} = \mathsf{E}[\boldsymbol{N}\boldsymbol{N}^{\mathsf{T}}] = I_m$$

where I_m is the the m by m identity matrix. For a jointly Gaussian vector $\boldsymbol{Z} = A\boldsymbol{N}$, the covariance matrix is thus

$$\mathsf{K}_{\boldsymbol{Z}} = \mathsf{E}[A\boldsymbol{N}\boldsymbol{N}^{\mathsf{T}}A^{\mathsf{T}}] = A\mathsf{E}[\boldsymbol{N}\boldsymbol{N}^{\mathsf{T}}]A^{\mathsf{T}} = AA^{\mathsf{T}} \tag{1}$$

The *probability density*, $f_{\boldsymbol{Z}}(\boldsymbol{z})$, of a r**v** $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_k)^{\mathsf{T}}$ is simply the joint probability density of the components $Z_1, \ldots, Z_k$. An important example is the iid r**v** $\boldsymbol{N}$ where the components N_j, $1 \leq j \leq m$, are iid and normalized Gaussian, $N_j \sim \mathcal{N}(0,1)$. By taking the product of the m densities of the individual random variables, the density of $\boldsymbol{N} = (N_1, N_2, \ldots, N_m)^{\mathsf{T}}$ is

$$f_{\boldsymbol{N}}(\boldsymbol{n}) = \frac{1}{(2\pi)^{m/2}} \exp\left(\frac{-n_1^2 - n_2^2 - \cdots - n_m^2}{2}\right) = \frac{1}{(2\pi)^{m/2}} \exp\left(\frac{-\|\boldsymbol{n}\|^2}{2}\right). \tag{2}$$

The density of $\boldsymbol{N}$ at a sample value $\boldsymbol{n}$ depends only on the squared distance $\|\boldsymbol{n}\|^2$ of the sample value from the origin. That is, $f_{\boldsymbol{N}}(\boldsymbol{n})$ is spherically symmetric around the origin, and points of equal probability density lie on concentric spheres around the origin.

Consider the transformation $\boldsymbol{Z} = \mathsf{A}\boldsymbol{N}$ where $\boldsymbol{Z}$ and $\boldsymbol{N}$ each have k components and A is k by k. If we let $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k$ be the k columns of A, then this means that $\boldsymbol{Z} = \sum_j \boldsymbol{a}_j N_j$. That is, for any sample values $n_1, \ldots n_k$ for $\boldsymbol{N}$, the corresponding sample value for $\boldsymbol{Z}$ is $\boldsymbol{z} = \sum_j \boldsymbol{a}_j n_j$.

We now assume that A is invertible, *i.e.*, that there is a matrix A^{-1} such that $\mathsf{A}\mathsf{A}^{-1} = \mathsf{A}^{-1}\mathsf{A} = \mathsf{I}_k$ where I_k is the k by k identity matrix. In terms of our earlier study of $\mathbb{R}^k$, this means that $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k$ form a basis for $\mathbb{R}^k$. The columns of A^{-1} then represent the unit vectors of $\mathbb{R}^k$ in terms of the basis $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k$. If A was not invertible, *i.e.*, if $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k$ did not form a basis for $\mathbb{R}^k$, then the possible sample values for $\boldsymbol{Z}$ would all lie in a proper subspace of $\mathbb{R}^k$, and $\boldsymbol{Z}$ would not have a probability density.

Consider a small cube, δ on a side, of sample values for $\boldsymbol{n}$. That is, consider the set B_δ of vectors for which $0 \leq n_j \leq \delta$ for $1 \leq j \leq k$. The set B'_δ of vectors $\boldsymbol{z} = A\boldsymbol{n}$ that $\boldsymbol{n} \in B_\delta$ map into is a parallepiped whose sides are the vectors $\boldsymbol{a}_1\delta, \ldots, \boldsymbol{a}_k\delta$. The determinant, $\det(A)$, of A has the remarkable geometric property that its magnitude, $|\det(A)|$, is equal to the volume of the parallelepiped with sides $\boldsymbol{a}_j$; $1 \leq j \leq k$. Thus the unit cube B_δ above, with volume δ^k, is mapped by A into a parallelepiped of volume $|\det A|\delta^k$.

Now let $\boldsymbol{z}$ be a sample value of $\boldsymbol{Z}$, and let $\boldsymbol{n} = \mathsf{A}^{-1}\boldsymbol{z}$ be the corresponding sample value of $\boldsymbol{N}$. Consider the cube B_δ around $\boldsymbol{n}$. The probability P that $\boldsymbol{N}$ lies in this cube is $f_{\boldsymbol{N}}(\boldsymbol{n})\delta^k$ plus terms that go to zero faster than δ^k as $\delta \to 0$. This cube around $\boldsymbol{n}$ maps

into a parallelepiped of volume $\delta^k|\det(\mathsf{A})|$ around $\boldsymbol{z}$, and no other sample value of $\boldsymbol{N}$ maps into this parallelepiped. Thus P is also equal to $f_{\boldsymbol{Z}}(\boldsymbol{z})\delta^k|\det(\mathsf{A})|$ plus negligible terms. Going to the limit $\delta \to 0$, we have

$$f_{\boldsymbol{Z}}(\boldsymbol{z})|\det(\mathsf{A})| = f_{\boldsymbol{N}}(\boldsymbol{n}). \tag{3}$$

Since $\boldsymbol{n} = \mathsf{A}^{-1}\boldsymbol{z}$, we get the explicit formula

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = \frac{f_{\boldsymbol{N}}(\mathsf{A}^{-1}\boldsymbol{z})}{|\det(\mathsf{A})|}. \tag{4}$$

This formula is valid for any random vector $\boldsymbol{N}$ with a density, but we are interested in the vector $\boldsymbol{N}$ of iid Gaussian random variables, $\mathcal{N}(0,1)$. Then, from (3), (4) becomes

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = \frac{1}{(2\pi)^{n/2}|\det(\mathsf{A})|}\exp\left(\frac{-\|A^{-1}\boldsymbol{z}\|^2}{2}\right) \tag{5}$$

$$= \frac{1}{(2\pi)^{n/2}|\det(\mathsf{A})|}\exp\left[-\frac{1}{2}\boldsymbol{z}^{\mathsf{T}}(\mathsf{A}^{-1})^{\mathsf{T}}\mathsf{A}^{-1}\boldsymbol{z}\right] \tag{6}$$

We can simplify this somewhat by recalling from (2) that the covariance matrix of $\boldsymbol{Z}$ is given by $\mathsf{K}_{\boldsymbol{Z}} = AA^{\mathsf{T}}$. Thus $\mathsf{K}_{\boldsymbol{Z}}^{-1} = (\mathsf{A}^{-1})^{\mathsf{T}}\mathsf{A}^{-1}$. We also have $\det(\mathsf{K}_{\boldsymbol{Z}}) = \det(AA^{\mathsf{T}}) = |\det A|^2$.

Substituting this into (6) and noting that $\det(\mathsf{K}_{\boldsymbol{Z}}) = [\det(\mathsf{A})]^2$,

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det(\mathsf{K}_{\boldsymbol{Z}})}}\exp\left[-\frac{1}{2}\boldsymbol{z}^{\mathsf{T}}\mathsf{K}_{\boldsymbol{Z}}^{-1}\boldsymbol{z}\right] \tag{7}$$

Note that this probability density depends only on the covariance matrix of $\boldsymbol{Z}$ and not directly on the matrix A. In particular, $\boldsymbol{z}^{\mathsf{T}}\mathsf{K}_{\boldsymbol{Z}}^{-1}\boldsymbol{z}$ is a quadratic expression in the components of $\boldsymbol{z}$. The set of $\boldsymbol{z}$ for which (7) is constant forms an ellipsoid centered at the origin, and the contours of equal probability density for $\boldsymbol{Z}$ form a set of concentric ellipsoids.

Example 1 Before proceeding, we want to look at the important special case of (1) and (7) where $AA^{\mathsf{T}} = \mathsf{I}_k$. In this case, from (1), $\mathsf{K}_{\boldsymbol{Z}} = \mathsf{K}_{\boldsymbol{Z}}^{-1} = \mathsf{I}_k$ and (7) becomes

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = \frac{\exp\left[-\frac{1}{2}\boldsymbol{z}^{\mathsf{T}}\boldsymbol{z}\right]}{(2\pi)^{n/2}} = \prod_{i=1}^{n}\frac{\exp[-z_i^2/2]}{2\pi}. \tag{8}$$

Thus, the transformation A is such that the r**v** $\boldsymbol{Z}$ has iid normalized Gaussian components. To understand this better, note that $AA^{\mathsf{T}} = \mathsf{I}_k$ means that A^{T} is the inverse of A and thus that $A^{\mathsf{T}}A = \mathsf{I}_k$. Letting $\boldsymbol{a}_i$ be the i^{th} column of A, the equation $A^{\mathsf{T}}A = \mathsf{I}_k$ means that $\boldsymbol{a}_i^{\mathsf{T}}\boldsymbol{a}_j = \delta_{ij}$, i.e., that the columns of A are orthonormal. Thus, in this example, the unit vectors $\boldsymbol{e}_1, \boldsymbol{e}_2$ are mapped into orthonormal vectors $\boldsymbol{a}_1, \boldsymbol{a}_2$, so that the transformation simply rotates the points in the plane. Although it is difficult to visualize such a transformation in higher dimensional space, it is still called a rotation, and has the property that $||\mathsf{A}\boldsymbol{n}||^2 = \boldsymbol{n}^{\mathsf{T}}\mathsf{A}^{\mathsf{T}}\mathsf{A}\boldsymbol{n}$, which is just $\boldsymbol{n}^{\mathsf{T}}\boldsymbol{n} = ||\boldsymbol{n}||^2$. Thus, each point $\boldsymbol{n}$ maps

into a point $\mathsf{A}\boldsymbol{n}$ at the same distance from the origin as itself. Because of the circular symmetry of the iid Gaussian distribution, it is not too surprising that $\mathsf{A}\boldsymbol{N}$ is still iid.

A matrix A with the property that $AA^\intercal = \mathsf{I}_k$ is called an *orthogonal matrix* or *orthonormal matrix* (orthonormal is more appropriate, but orthogonal is more common). Not only the columns of A are orthonormal, but the rows, say $\boldsymbol{b}_i$; $1 \leq i \leq n$ are also orthonormal (as is seen directly from $AA^\intercal = \mathsf{I}_k$). Since $Z_i = \boldsymbol{b}_i \boldsymbol{n}$, this means that any set of orthonormal vectors $\boldsymbol{b}_i$ leads to the property that the random variables $Z_i = \boldsymbol{b}_i \boldsymbol{n}$ are iid and $\mathcal{N}(0,1)$

Example 2 For a 2-dimensional zero mean Gaussian r$\boldsymbol{v}$ $\boldsymbol{Z}$, let $\mathsf{E}[Z_1^2] = \sigma_1^2$, $\mathsf{E}[Z_2^2] = \sigma_2^2$ and $\mathsf{E}[Z_1 Z_2] = k_{12}$. Let ρ be the *normalized covariance* $\rho = k_{12}/(\sigma_1\sigma_2)$. Then $\det(\mathsf{K}_{\boldsymbol{Z}}) = \sigma_1^2\sigma_2^2 - k_{12}^2 = \sigma_1^2\sigma_2^2(1-\rho^2)$. For A to be non-singular, we need $\det(\mathsf{K}_{\boldsymbol{Z}}) = [\det(\mathsf{A})]^2 > 0$, so we need $|\rho| < 1$ (note that for any rv's, $\rho \leq 1$). We then have

$$\mathsf{K}_{\boldsymbol{Z}}^{-1} = \frac{1}{\sigma_1^2\sigma_2^2 - k_{12}^2} \begin{bmatrix} \sigma_2^2 & -k_{12} \\ -k_{12} & \sigma_1^2 \end{bmatrix} = \frac{1}{1-\rho^2} \begin{bmatrix} 1/\sigma_1^2 & -\rho/(\sigma_1\sigma_2) \\ -\rho/(\sigma_1\sigma_2) & 1/\sigma_2^2 \end{bmatrix}$$

$$\begin{aligned} f_{\boldsymbol{Z}}(\boldsymbol{z}) &= \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2 - k_{12}^2}} \exp\left(\frac{-z_1^2\sigma_2^2 + 2z_1z_2k_{12} - z_2^2\sigma_1^2}{2(\sigma_1^2\sigma_2^2 - k_{12}^2)}\right) \\ &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(\frac{-(z_1/\sigma_1)^2 + 2\rho(z_1/\sigma_1)(z_2/\sigma_2) - (z_2/\sigma_2)^2}{2(1-\rho^2)}\right) \end{aligned} \tag{9}$$

Note that curves of equal probability density in the plane correspond to points where the argument of the exponential function in (9) are equal, and this argument is quadratic. For the zero covariance case, this corresponds to ellipses centered on the axes. If $\sigma_1 > \sigma_2$, these ellipses are elongated in the z_1 direction. If $\rho > 0$, this increases the density in the first and third quadrants at the expense of the second and fourth, and thus the ellipses are elongated in the first and third quadrants. The main point to be learned from this, however, is that vector notation is highly beneficial. Hand calculation, even for two dimensions, is messy and gets messier in higher dimension. Thus, we must learn to reason directly from the vector equations and use standard computer programs to do calculations.

2 Stationary Gaussian Processes

A stochastic process $\{N(t)\}$ is a collection of rv's, one rv for each real number t; these rv's are defined on a common sample space Ω. The process is zero mean if $\mathsf{E}[N(t)] = 0$ for each t. The covariance function for a zero mean stochastic process is defined as $\mathsf{K}_{\boldsymbol{N}}(t,\tau) = \mathsf{E}[N(t)N(\tau)]$. A stochastic process is specified by its joint probability assignments on $N(t_1), \ldots, N(t_k)$ over all finite sets of epochs $t_1, \ldots, t_k$. A *zero mean Gaussian process* $\{N(t)\}$ is a zero mean stochastic process for which, for any integer $k > 0$ and any k epochs, $t_1, \ldots, t_k$, the random variables (rv's) $N(t_1), \ldots, N(t_k)$ are jointly Gaussian (and, of course, zero mean). For any k tuple of epochs, $t_1, \ldots, t_k$, the random vector $\boldsymbol{Y} = (N(t_1), \ldots, N(t_k))^\intercal$ is then jointly Gaussian. The l, j component of $\mathsf{K}_{\boldsymbol{Y}}$ is then

$\mathsf{E}[N(t_l)N(t_j)] = \mathsf{K}_{\boldsymbol{N}}(t_l, t_j)$, so that $\mathsf{K}_{\boldsymbol{Y}}$ is determined by the function $\mathsf{K}_{\boldsymbol{N}}(t, \tau)$. Thus $\mathsf{K}_{\boldsymbol{N}}(t, \tau)$ specifies the covariance matrix for each k-tuple of epochs of $\{N(t)\}$, and thus specifies the corresponding joint probability assignment. It follows that for a zero mean Gaussian stochastic process, the covariance function completely specifies the process.

A stochastic process is *stationary* if the joint probability assignment for any set of epochs, $t_1, \ldots, t_k$ is the same, for any τ, as the probability assignment for $t_1+\tau, t_2+\tau, \ldots, t_k+\tau$. For a zero mean Gaussian process $\{N(t)\}$, this condition is satisfied if and only if the covariance function $\mathsf{K}_{\boldsymbol{N}}(t, \tau) = \mathsf{K}_{\boldsymbol{N}}(t-\tau, 0)$. We abbreviate the covariance function for a stationary process by $\mathsf{K}_{\boldsymbol{N}}(t) = \mathsf{E}[N(t)N(0)]$.

As a model for any real engineering situation, we are never interested in more than some finite interval of interest, and defining how a stochastic process behaves outside of that interval is done solely for simplicity. Here, stationarity creates a kind of simplicity, since the probabilistic description does not change with shifts in time. It also creates a complication, in the sense that the sample functions of the process are no longer finite energy functions (*i.e.*, waveforms in $\mathcal{L}_2$). We solve this problem by defining a stochastic process to be *effectively stationary over* $(-T_0, T_0)$ if the joint probability assignment over $t_1, \ldots, t_k$ is the same as that over $t_1+\tau, t_2+\tau, \ldots, t_k+\tau$ whenever $t_1, \ldots, t_k$ and $t_1+\tau, t_2+\tau, \ldots, t_k+\tau$ are all contained in the interval $(-T_0, T_0)$. A zero mean Gaussian process is then effectively stationary over $(-T_0, T_0)$ if $\mathsf{K}_{\boldsymbol{N}}(t, \tau) = \mathsf{K}_{\boldsymbol{N}}(t-\tau)$ whenever t, τ, and $t-\tau$ lie within the interval $(-T_0, T_0)$. The sample waveforms of an effectively stationary stochastic process can then die out gradually outside of $(-T_0, T_0)$, and thus can be $\mathcal{L}_2$ waveforms.

We should recognize that the difference between stationary and effectively stationary for a large interval is only a difference in the model and not in the situation being modeled. If the two models behave differently in some significant way, it is a sign that the entire modeling issue should be rethought. In more concrete terms, this says that if the noise in the distant past or the distant future turns out to have a critical effect on a communication process, then the model should be rethought.

3 Linear functionals for Gaussian processes

Suppose the sample functions $N(t, \omega)$ of a zero mean stochastic process are real waveforms (*i.e.*, real finite energy functions). The process might be effectively stationary or might be non-stationary. Then these sample functions can be viewed as vectors over $\mathbb{R}$ in the $\mathcal{L}_2$ space of real waveforms. For any given real waveform $g(t)$, there is an inner product,

$$\langle N(t, \omega), g(t) \rangle = \int_{-\infty}^{\infty} N(t, \omega) g(t)\, dt.$$

Since this has a real value for each $\omega \in \Omega$, it maps sample points ω into real numbers and thus is a random variable, denoted $Z = \int_{-\infty}^{\infty} N(t) g(t)\, dt$. This random variable Z is known as a *linear functional* of the process $\{N(t)\}$. For a zero mean Gaussian process, we would certainly expect Z to be a zero mean Gaussian rv, but it would require a fairly intricate mathematical argument to show this.

As a simpler approach, which is instructive on its own merits, suppose the sample functions of $\{N(t)\}$ are baseband limited to $(-W_0, W_0)$. We view W_0 as much larger than any frequency of interest in whatever communication problem we consider. Thus W_0 exceeds any passband of interest, any band from which any given interference might occur, any higher frequencies which might be desirable for use, etc. We argue that any noise waveform is essentially limited in this way for large enough W_0, especially after it is actually received as a waveform from the channel medium.

Assuming that the waveform $g(t)$ is similarly limited, we can expand both a sample function of the noise $N(t,\omega)$ and $g(t)$ by the sampling theorem. By Parseval, then

$$Z(\omega) = \int_{-\infty}^{\infty} N(t,\omega)g(t)\ dt = T \sum_k N(kT,\omega)\, g(kT)$$

where $T = 1/(2W_0)$. We then have

$$Z = \int_{-\infty}^{\infty} N(t)g(t)\ dt = T \sum_k N(kT)\, g(kT) \tag{10}$$

Since $\{N(t)\}$ is a Gaussian process, any finite set of the variables $\{N(kT)\}$ is jointly Gaussian, so any finite sum in (10) is Gaussian. Since g is $\mathcal{L}_2$, the coefficients $g(kT)$ decrease with k, and thus we assert without proof that the sum in (10) converges to a zero-mean Gaussian rv.

Next, suppose we look at a finite set of waveforms $\{g_j(t)\};\ 1 \le j \le j_0$, each limited to the band $|f| \le W_0$. Let $Z_j = \int_{-\infty}^{\infty} N(t)g_j(t)\,dt$. By the same argument as above,

$$Z_j = \int_{-\infty}^{\infty} N(t)g_j(t)\ dt = T \sum_k N(kT)\, g_j(kT) \tag{11}$$

What is more, each Z_j is a linear combination of the same underlying set of rv's $\{N(kT)\}$. Thus, we can assert as above that $\{Z_j\}$ is a set of jointly Gaussian rv's.

We can use the same argument to look at the output of a linear filter for which the input is the Gaussian process $\{N(t)\}$ above.

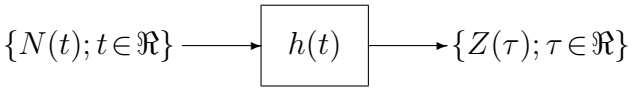

Figure 1: Filtered Stochastic Process

Assume that the impulse response $h(t)$ of the filter is a real $\mathcal{L}_2$ waveform limited to the band $|f| \le W_0$. Then for any given sample function $\{N(t,\omega)\}$ of the input, the filter output at any epoch τ is given by

$$Z(\tau,\omega) = \int_{-\infty}^{\infty} N(t,\omega)h(\tau{-}t)\ dt = T \sum_k N(kT,\,\omega)\, h(\tau{-}kT). \tag{12}$$

The sample function output $\{Z(\tau,\omega)\}$ is also limited to the band $|f| \le W_0$. For any given epoch τ, (12) maps sample points to real values and is thus a random variable $Z(\tau)$; in fact $Z(\tau)$ for the given τ is simply the linear functional corresponding to $h(\tau-t)$ (as a function of t) analyzed above. The random variables $Z(\tau)$ over all τ form a stochastic process,

$$Z(\tau) = \int_{-\infty}^{\infty} N(t)h(\tau-t)\, dt = T \sum_k N(kT)\, h(\tau-kT). \tag{13}$$

This is a zero-mean Gaussian rv for each epoch τ. For any set of epochs, $\tau_1, \ldots, \tau_{j_0}$, we see that $Z(\tau_1), \ldots, Z(\tau_{j_0})$ are jointly Gaussian. Thus $\{Z(\tau)\}$ is a zero mean Gaussian stochastic process.

We summarize these results in the following theorem.

Theorem 3.1 *Let $\{N(t)\}$ be a zero-mean Gaussian process whose sample functions are real waveforms in $\mathcal{L}_2$ limited to the band $|f| \le W_0$. Let $T = 1/(2W_0)$. Then*

- *For any set of $\mathcal{L}_2$ waveforms $g_1(t), \ldots, g_{j_0}(t)$ all limited to the band $|f| \le W_0$, the linear functionals $Z_1, \ldots, Z_{j_0}$ given by (11) are jointly Gaussian.*
- *For any filter with real $\mathcal{L}_2$ impulse response $h(t)$ limited to the band $|f| \le W_0$, the filter output $\{Z(\tau)\}$ given by (13) is a zero mean Gaussian process.*

These are important results. The first, concerning sets of linear functionals, is important when we represent the input to the channel in terms of an orthonormal expansion; this result will let us show that the noise can be expanded in the same orthonormal expansion. The second, concerning linear filtering, shows that when the received signal and noise are passed through a linear filter, the noise at the filter output is simply another zero mean Gaussian process. This theorem is often summarized by saying that linear operations preserve Gaussianity.

We now return briefly to the question whether the limitation to a finite band of frequencies is a serious restriction. Since W_0 can be arbitrarily large, the restriction is at most a mathematical irritant; the irritant is that waveforms cannot be both strictly time-limited and strictly band-limited. From an engineering standpoint, we rarely either know or care what happens at frequencies very far from those of interest, so any modeling assumption that is self consistent is essentially equally good. We return to this issue later after discussing covariance functions and spectral density.

Covariance for linear functionals and filters

We assume again that $\{N(t)\}$ is a zero mean Gaussian process limited to some band $|f| \le W_0$. We also assume that $g_1(t), \ldots, g_{j_0}(t)$ are real $\mathcal{L}_2$ waveforms limited to the band $|f| \le W_0$. We have seen that the linear functionals $Z_1, \ldots, Z_{j_0}$ given by $Z_j = \int_{-\infty}^{\infty} N(t)g_j(t)\, dt$ are jointly Gaussian for $1 \le j \le j_0$. We now want to find the covariance for each pair Z_i, Z_j of these random variables. The computation is quite simple, although we omit questions of limits, interchanges of order of expectation and integration, etc. A

more careful derivation could be made by returning to the sampling theorem arguments before, but this would somewhat obscure the ideas. The results do not really depend on $\{N(t)\}$ being Gaussian, but only on the mathematical details of integration and limits, etc.

$$\begin{aligned}
\mathsf{E}[Z_i Z_j] &= \mathsf{E}\left[\int_{-\infty}^{\infty} N(t) g_i(t)\, dt \int_{-\infty}^{\infty} N(\tau) g_j(\tau)\, d\tau\right] && (14)\\
&= \int_{t=-\infty}^{\infty}\int_{\tau=-\infty}^{\infty} g_i(t)\mathsf{E}[N(t)N(\tau)]g_j(\tau)\, dt\, d\tau && (15)\\
&= \int_{t=-\infty}^{\infty}\int_{\tau=-\infty}^{\infty} g_i(t)\mathsf{K}_{\boldsymbol{N}}(t,\tau)g_j(\tau)\, dt\, d\tau && (16)
\end{aligned}$$

We see that each covariance term (and also $E[Z_j^2]$ for each j) is simply a function of the covariance function of the process and of the waveforms $\{\boldsymbol{g}_j\}$.

Next, suppose the process $\{N(t)\}$ is effectively stationary over $(-T_0, T_0)$ and that the waveforms $\boldsymbol{g}_i$ and $\boldsymbol{g}_j$ are zero outside of $(-T_0, T_0)$. Then (16) simplifies somewhat to

$$E[Z_i Z_j] = \int_{t=-\infty}^{\infty}\int_{\tau=-\infty}^{\infty} g_i(t)\mathsf{K}_{\boldsymbol{N}}(t-\tau)g_j(\tau)\, d\tau\, dt \qquad (17)$$

Consider the covariance for the filtered output of a zero-mean Gaussian process $\{N(t)\}$. The output $\{Z(t)\}$ for a filter with a real $\mathcal{L}_2$ impulse response $\boldsymbol{h}$ is given by (13), so the covariance of the output can be found as

$$\begin{aligned}
\mathsf{K}_{\boldsymbol{Z}}(u,v) &= \mathsf{E}[Z(u)Z(v)]\\
&= E\left[\int_{-\infty}^{\infty} N(t)h(u-t)dt \int_{-\infty}^{\infty} N(\tau)h(v-\tau)d\tau\right]\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(u-t)\mathsf{K}_{\boldsymbol{N}}(t,\,\tau)h(v-\tau)dtd\tau && (18)
\end{aligned}$$

This can be simplified if $\{N(t)\}$ is effectively stationary over some interval $(-T_0, T_0)$. Then $\mathsf{K}_{\boldsymbol{N}}(t,\tau)$ can be replaced by $\mathsf{K}_{\boldsymbol{N}}(t-\tau)$. Replacing $t-\tau$ by ϕ (*i.e.*, t by by $\phi+\tau$), (18) becomes

$$\mathsf{K}_{\boldsymbol{Z}}(u,v) = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} h(u-\tau-\phi)\mathsf{K}_{\boldsymbol{N}}(\phi)\, d\phi\right] h(v-\tau)\, d\tau$$

Replacing τ by $\mu+v$,

$$\mathsf{K}_{\boldsymbol{Z}}(u,v) = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} h(u-v-\mu-\phi)\mathsf{K}_{\boldsymbol{N}}(\phi)\, d\phi\right] h(-\mu)\, d\mu$$

We see from this that $\mathsf{K}_{\boldsymbol{Z}}(u,v)$ is a function only of $u-v$. Replacing $u-v$ by τ, this then becomes

$$\mathsf{K}_{\boldsymbol{Z}}(\tau) = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} h(\tau-\mu-\phi)\mathsf{K}_{\boldsymbol{N}}(\phi)\, d\phi\right] h(-\mu)\, d\mu \qquad (19)$$

Since $\{Z(t)\}$ is effectively Gaussian, this means that $\{Z(t)\}$ is stationary. This is not surprising – putting a stationary Gaussian process through a linear time-invariant filter results in another effectively stationary Gaussian process.

We were somewhat careless above about the interval $(-T_0, T_0)$. If $\{N(t)\}$ is effectively stationary over $(-T_0, T_0)$, and the filter response $\boldsymbol{h}$ has a duration at most T_1 around 0, then it can be seen from (18) that $\{Z(t)\}$ is effectively stationary over the smaller region, $(-T_0+T_1,\ T_0-T_1)$. This is insignificant if T_0 is very much larger than any time durations of interest. Thus, in practice, we use the word *stationary* to refer to effectively stationary processes and ignore the interval $(-T_0, T_0)$ over which it is effectively stationary. This makes sense in modeling since we usually have no idea of the range in time over which stationarity is a reasonable assumption; we simply assume that a process is stationary over a much larger interval than any other intervals of interest.

4 Stationary Processes in the Frequency Domain

Stationary zero-mean Gaussian processes are often viewed more insightfully in the frequency domain than in the time domain. First we look at linear functionals, and then look at linear filters. As indicated above, we are implicitly assuming effective stationarity, and thus assuming that the sample waveforms are $\mathcal{L}_2$. We will ignore the interval $(-T_0, T_0)$ however.

First, for a zero mean WSS process $\{N(t)\}$, consider the linear functionals $Z_j = \int g_j(t)N(t)\,dt$ where $\boldsymbol{g}_1, \ldots \boldsymbol{g}_{j_0}$ are real waveforms. From (16),

$$\mathsf{E}[Z_i Z_j] = \int_{t=-\infty}^{\infty} g_i(t) \left[\int_{\tau=-\infty}^{\infty} \mathsf{K}_{\boldsymbol{N}}(t-\tau) g_j(\tau)\,d\tau \right] dt \tag{20}$$

$$= \int_{t=-\infty}^{\infty} g_i(t) \left[\mathsf{K}_{\boldsymbol{N}} * \boldsymbol{g}_j \right](t)\,dt \tag{21}$$

where $\mathsf{K}_{\boldsymbol{N}} * \boldsymbol{g}_j$ denotes the convolution of the waveforms $\mathsf{K}_{\boldsymbol{N}}(t)$ and $g_j(t)$. The *spectral density* $S_{\boldsymbol{N}}(f)$ of a WSS process $\{N(t)\}$ is defined as the Fourier transform of $\mathsf{K}_{\boldsymbol{N}}(t)$. Since $\mathsf{K}_{\boldsymbol{N}}(t)$ is real and symmetric, $S_{\boldsymbol{N}}$ is also real and symmetric..

Let $v(t) = [\mathsf{K}_{\boldsymbol{N}}(t) * \boldsymbol{g}_j](t)$ be the convolution of $\mathsf{K}_{\boldsymbol{N}}$ and $\boldsymbol{g}_j$. Thus the Fourier transform of $v(t)$ satisfies $\hat{v}(f) = S_{\boldsymbol{N}}(f)\hat{g}_j(f)$. Since $\int g_i(t)v^*(t)\,dt = \int \hat{g}_i(f)\hat{v}^*(f)\,df$, we have

$$\mathsf{E}[Z_i Z_j] = \int \hat{g}_i(f)\hat{v}^*(f)\,df = \int \hat{g}_i(f) S_{\boldsymbol{N}}(f) \hat{g}_j^*(f)\,df \tag{22}$$

Two important conclusions follow from this equation. First, if the waveforms $\boldsymbol{g}_i$ and $\boldsymbol{g}_j$ have non-overlapping transforms, then the linear functionals Z_i and Z_j are uncorrelated (and thus independent if $\{N(t)\}$ is Gaussian). This says essentially that a Gaussian random process at any given frequency is independent of itself at any other frequency. Second, for $i = j$, $\mathsf{E}[Z_j^2] = \int |\hat{g}_j(f)|^2 S_{\boldsymbol{N}}(f) \geq 0$. Since $\boldsymbol{g}$ can be made arbitrarily narrow band at any given frequency, this shows that the spectral density must be non-negative.

Next consider the frequency domain interpretation of filtering a WSS process. We can rewrite (19), focussing on viewing it as a convolution.

$$\begin{aligned}\mathsf{K}_{\boldsymbol{Z}}(\tau) &= \int \left[\int h(\tau-\mu-\phi)\mathsf{K}_{\boldsymbol{N}}(\phi)\, d\phi \right] h(-\mu)\, d\mu \\ &= \int [\boldsymbol{h} * \mathsf{K}_{\boldsymbol{N}}](\tau-\mu)\; h(-\mu)\; d\mu \\ &= \boldsymbol{h} * \mathsf{K}_{\boldsymbol{N}} * \boldsymbol{h}_b \end{aligned} \tag{23}$$

where $\boldsymbol{h}_b$ is the matched filter to $\boldsymbol{h}$. That is $\boldsymbol{h}_b$ is $\boldsymbol{h}$ turned around backward, *i.e.*, $h_b(\mu) = h(-\mu)$. Recognizing that the Fourier transform of $h(-\mu)$ is $\hat{h}^*(f)$, we can take the Fourier transform of (23) to get

$$S_{\boldsymbol{Z}}(f) = \hat{h}(f) S_{\boldsymbol{N}}(f) \hat{h}^*(f) = |\hat{h}(f)|^2 S_{\boldsymbol{N}}(f) \tag{24}$$

Since $S_{\boldsymbol{Z}}(f)$ is the spectral density of a WSS stochastic process $\{Z(t)\}$, it must be real and symmetric; this also follows from (24) since $|\hat{h}(f)|^2$ is real and symmetric and $S_{\boldsymbol{N}}$ is also.

Example 1: Since (24) is valid for any real $\mathcal{L}_2$ filter response $h(t)$, we can consider the response of a brickwall filter around an arbitrary frequency f_c:

$$\hat{h}(f) = \begin{cases} 1 & \text{for } \; f_c - W/2 \le |f| \le f_c + W/2 \\ 0 & \text{otherwise} \end{cases}$$

If W is so small that $S_{\boldsymbol{N}}(f)$ does not vary appreciably over the band around f_c, then $S_{\boldsymbol{Z}}(f) \approx S_{\boldsymbol{N}}(f_c)$ over $f_c - W/2 \le |f| \le f_c + W/2$ and is 0 elsewhere. Since $S_{\boldsymbol{Z}}(f)$ is non-zero over a band of width W around f_c and another band of width W around $-f_c$, the mean square value of $\{Z(t)\}$, *i.e.*, $\mathsf{E}[Z^2] = \mathsf{K}_{\boldsymbol{N}}(0)$ is given by

$$\mathsf{K}_{\boldsymbol{Z}}(0) = \int S_{\boldsymbol{Z}}(f)\, df \approx 2W S_{\boldsymbol{N}}(f_c).$$

If $\{N(t)\}$ is WGN of spectral density $N_0/2$, then $\mathsf{K}_{\boldsymbol{Z}}(0) = N_0 W$. In other words, N_0 *is the noise power per unit positive frequency of WGN of spectral density* $N_0/2$.

5 White Gaussian noise

5.1 Time Domain View

Physical noise processes are very often reasonably modeled as zero mean, effectively stationary over $(-T_0, T_0)$, and Gaussian. There is one further simplification that is often reasonable. This is that the covariance between the noise at two epochs dies out very rapidly as the interval between those epochs increases. The interval over which this covariance is significantly non-zero is often very small relative to the intervals over which

the signal varies appreciably. What this means is that the covariance function $\mathsf{K}_{\boldsymbol{N}}(\tau)$ looks like a short duration pulse around $\tau = 0$.

We know from linear system theory that $\int \mathsf{K}_{\boldsymbol{N}}(t-\tau)g(\tau)d\tau$ is equal to $g(t)$ if $\mathsf{K}_{\boldsymbol{N}}(t)$ is a unit impulse. We also know that this integral is approximately equal to $g(t)$ if $\mathsf{K}_{\boldsymbol{N}}(t)$ has unit area and is a narrow pulse relative to changes in $g(t)$. It follows that under the same circumstances, (17) becomes

$$\mathsf{E}[Z_1 Z_2] = \int_t \int_\tau g_1(t)\mathsf{K}_{\boldsymbol{N}}(t-\tau)g_2(\tau)\,d\tau\,dt \approx \int g_1(t)g_2(t)\,dt \tag{25}$$

What this means is that, if the covariance function is very narrow relative to the functions of interest, then the only thing we need to know about it is its area. We refer to a zero mean WSS Gaussian stochastic process with such a narrow covariance function as *White Gaussian Noise (WGN)*. The area under the covariance function is called the *intensity* or the *spectral density* of the WGN and is denoted by the symbol $N_0/2$. Thus, for finite energy functions $g_1(t), g_2(t), \ldots$ of interest, and for WGN (denoted by W(t)) of intensity $N_0/2$, the random variable $Z_j = \int W(t)g_j(t)\,dt$ has the variance

$$\mathsf{E}[Z_j^2] = (N_0/2)\int g_j^2(t)\,dt \tag{26}$$

Similarly, the random variables Z_i and Z_j have the covariance

$$\mathsf{E}[Z_i Z_j] = (N_0/2)\int g_i(t)g_j(t)\,dt \tag{27}$$

Also $Z_1, Z_2, \ldots$ are jointly Gaussian.

The most important special case of (26) and (27) is to let $\phi_j(t)$ be a set of orthonormal functions and let $W(t)$ be WGN of intensity $N_0/2$. Let $Z_j = \int \phi_j(t)W(t)\,dt$. Then, from (26) and (27),

$$\mathsf{E}[Z_i Z_j] = (N_0/2)\delta_{ij} \tag{28}$$

This is an important equation. It says that if the noise can be modeled as WGN, then when the noise is represented in terms of *any* orthonormal expansion, the resulting random variables are iid. Thus, we can represent signals in terms of orthonormal expansions, and represent WGN in terms of the same expansions, and the result is iid Gaussian random variables.

The careful reader will observe that we haven't really defined WGN. What we have said, in essence, is that if a stationary zero mean Gaussian process has a covariance function that is very narrow relative to the variation of all functions of interest, then we can pretend that the covariance function is an impulse times $N_0/2$, where $N_0/2$ is the area of the actual covariance function. Unfortunately, according to our definition of stochastic process, there cannot be any Gaussian stochastic process $N(t)$ whose covariance function is $\mathsf{K}(t) = (N_0/2)\delta(t)$. The reason for this dilemma is that $\mathsf{E}[N^2(t)] = \mathsf{K}_{\boldsymbol{N}}(0)$. We could interpret $\mathsf{K}_{\boldsymbol{N}}(0)$ to be either undefined or ∞, but either way, $N(t)$ cannot be a random variable (although we could think of it taking on only the values plus or minus ∞).

Mathematicians view WGN as a generalized stochastic process, in the same sense as the unit impulse $\delta(t)$ is viewed as a generalized function. That is, the impulse function $\delta(t)$ is not viewed as an ordinary function taking the value 0 for $t \neq 0$ and the value ∞ at $t = 0$. Rather, it is viewed in terms of its effect on other, better behaved, functions $g(t)$, where $\int_{-\infty}^{\infty} g(t)\delta(t)\,dt = g(0)$.

In the same way, WGN is not viewed in terms of random variables at each epoch of time. Rather it is viewed as a generalized zero mean stochastic process for which linear functionals are jointly Gaussian, for which variances and covariances are given by (26) and (27), and for which the covariance is formally taken to be $(N_0/2)\delta(t)$.

We have viewed both the delta function and WGN in terms of limiting the interval of stationarity and bandwidth to T_0 and W_0. Within this range, then the results are perfectly natural and well defined.

Finally, if we look at the noise over some frequency band with bandwidth W, and recall that the real noise waveform has both negative and positive frequency terms, the noise power in the bandwidth W is then N_0W.

5.2 Frequency Domain View

Previously, we characterized white Gaussian noise (WGN) as a zero mean Gaussian process $\{N(t)\}$ for which the covariance function is a very narrow pulse of area $N_0/2$ and width small relative to other time intervals of interest.

Here, in the frequency domain, we see how to characterize it in a much more sensible way: *WGN is a zero mean Gaussian process* $\{N(t)\}$ *for which* $S_{\boldsymbol{N}}(f) = N_0/2$ *over all frequencies of interest.* We say that it is WGN *relative to the given band* Most communication takes place over a bandwidth W that is narrow relative to the carrier frequency f_c. Thus the noise is viewed as WGN if its spectral density is equal to $N_0/2$ over that band. As we have seen above, this means that the noise power is N_0W_1 over any positive band W_1 within that bandwidth of interest. The noise thus has power N_0 per unit positive bandwidth over the entire range of frequencies of interest.

Note that we also saw that for any stationary Gaussian process, any linear functional for a function constrained to one frequency band is independent of that constrained to any other band. Thus for WGN relative to some given band, any linear functional constrained to a bandwidth outside of that given band is independent of all linear functionals within the band. Thus the noise has a constant spectral density within the band of interest, but is also independent of the out-of-band noise, regardless of the spectral density of the out-of-band noise.