Chapter 1

Signals in Radio Astronomy

Rajaram Nityananda

1.1 Introduction

The record of the electric field $E(t)$, received at a point on earth from a source of radio waves can be called a "signal", so long as we do not take this to imply intelligence at the transmitting end. Emanating as it does from a large object with many independently radiating parts, at different distances from our point, and containing many frequencies, this signal is naturally random in character. In fact, this randomness is of an extreme form. All measured statistical properties are consistent with a model in which different frequencies have completely unrelated phases, and each of these phases can vary randomly from 0 to 2π. A sketch of such a signal is given in Fig. 1.1. The strength (squared amplitude or power) of the different frequencies ω has a systematic variation which we call the "power spectrum" $S(\omega)$. This chapter covers the basic properties of such signals, which go by the name of "time-stationary gaussian noise". Both the signal from the source of interest, as well as the noise added to this cosmic signal by the radio telescope recievers can be described as time-stationary gaussian noise. The word noise of course refers to the random character. "Noise" also evokes unwanted disturbance, but this of course does not apply to the signal from the source (but does apply to what our receivers unavoidably add to it). The whole goal of radio astronomy is to receive, process, and interpret these cosmic signals, (which were, ironically enough, first discovered as a "noise" which affected trans-atlantic radio communication). "Time–Stationary" means that the signal in one time interval is statistically indistinguishable from that in another equal duration but time shifted interval. Like all probabilistic statements, this can never be precisely checked but its validity can be made more probable (circularity intended!) by repeated experiments. For example, we could look at the probability distribution of the signal amplitude. An experimenter could take a stretch of the signal say, from times 0 to T, select N equally spaced values $E(t_i), i$ going from 1 to N, and make a histogram of them. The property of time stationarity says that this histogram will turn out to be (statistically) the same — with calculable errors decreasing as N increases! — if one had chosen instead the stretch from t to $t + T$, for any t. The second important characteristic property of our random phase superposition of many frequencies is that this histogram will tend to a gaussian, with zero mean as N tends to infinity.

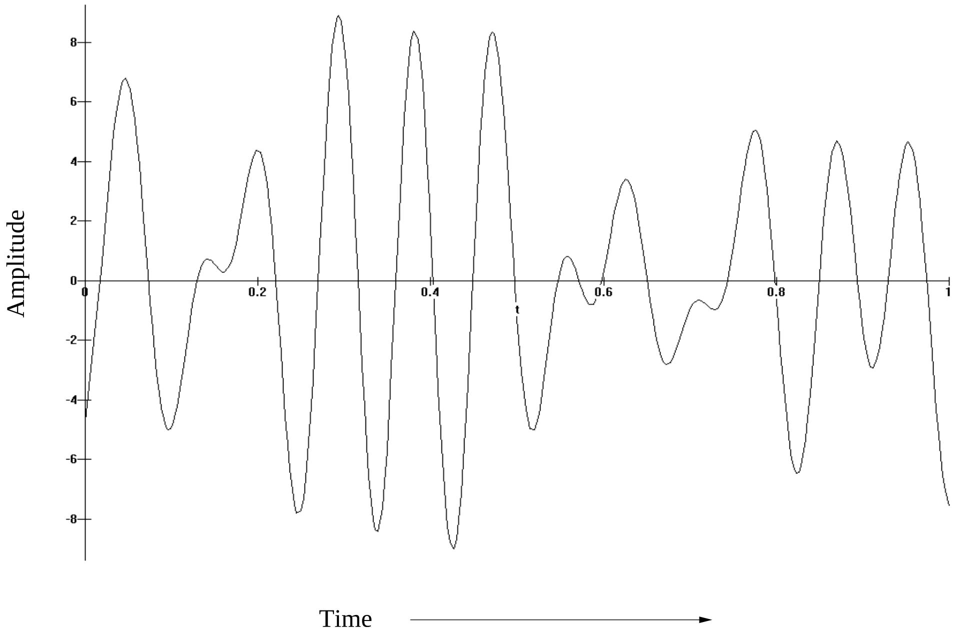

Figure 1.1: A signal made by superposition of many frequencies with random phases

1.2 Properties of the Gaussian

The general statement of gaussianity is that we look at the **joint** distribution of N amplitudes $x_1 = E(t_1), x_2 = E(t_2), \ldots$ etc. This is of the form

$$P(x_1 \ldots x_k) = const \times \exp\left(-Q(x_1, x_2, \ldots x_k)\right)$$

Q is a quadratic expression which clearly has to increase to $+\infty$ in any direction in the k dimensional space of the x's. For just one amplitude,

$$P(x_1) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x_1^2/2\sigma^2}$$

does the job and has one parameter, the "Variance"σ, the mean being zero. This variance is a measure of the power in the signal. For two variables, x_1 and x_2, the general mathematical form is the "bivariate gaussian"

$$P(x_1, x_2) = const \times \exp\left(-\frac{1}{2}(a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2)\right)$$

.

Such a distribution can be visualised as a cloud of points in $x_1 - x_2$ space, whose density is constant along ellipses Q =constant (see Fig. 1.2).

The following basic properties are worth noting (and even checking!).

1. We need a_{11}, a_{22}, and $a_{11}a_{22} - a_{12}^2$ all > 0 to have ellipses for the contours of constant P (hyperbolas or parabolas would be a disaster, since P would not fall off at infinity).

2. The constant in front is

$$(1/2\pi) \times \sqrt{det \begin{vmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{vmatrix}}$$

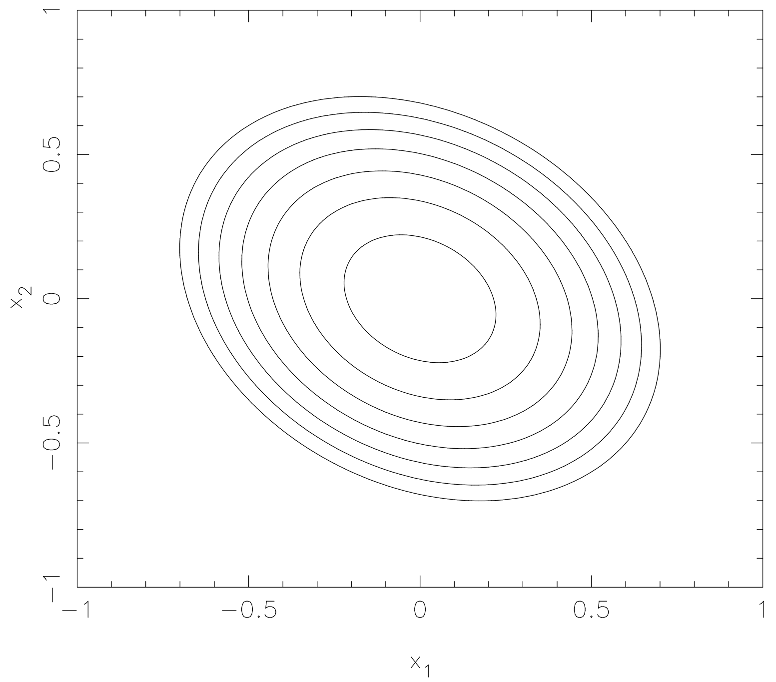

Figure 1.2: Contour lines of a bivariate gaussian distribution

3. The average values of x_1^2, x_2^2 and x_1x_2, when arranged as a matrix (the so called covariance matrix) are the **inverse** of the matrix of a's. For example,

$$\langle x_1^2 \rangle = a_{22}/detA$$

$$\langle x_1x_2 \rangle = a_{12}/detA$$

etc.

4. By time stationarity,

$$< x_1^2 >=< x_2^2 >= \sigma^2$$

$$< x_1^2 >=< x_2^2 >= \sigma^2$$

The extra information about the correlation between x_1 and x_2 is contained in $< x_1x_2 >$, i.e. in a_{12} which (again by stationarity) can only be a function of the time separation $\tau = t_1 - t_2$. We can hence write $< E(t)E(t+\tau) >= C(\tau)$ independent of t. $C(\tau)$ is called the autocorrelation function. From (1) above, $C^2(\tau) \leq \sigma^2$. This suggests that the quantity $r(\tau) = C(\tau)/\sigma^2$ is worth defining, as a dimensionless correlation coefficient, normalised so that $r(0) = 1$. The generalisation of all these results for a k variable gaussian is given in the Section 1.8

1.3 The Wiener-Khinchin Theorem

So far, we have only *asserted* that the sum of waves with random phases generates a time-stationary gaussian signal. We now have to check this. It is convenient to start with a signal going from 0 to T, and only later take the limit $T \to \infty$. The usual theory of Fourier series tells us that we can write

$$\begin{aligned} E(t) &\equiv \sum a_n \cos\omega_n t + b_n \sin\omega_n t \\ &\equiv \sum r_n \cos(\omega_n t + \varphi_n) \end{aligned}$$

where,

$$\omega_n = \frac{2\pi}{T},\ r_n = \sqrt{a_n x^2 + b_n^2}, and\ \tan\varphi_n = -b_n/a_n$$

Notice that the frequencies come in multiples of the "fundamental" $2\pi/T$ which is very small since T is large, and hence they form a closely spaced set. We can now compute the autocorrelation

$$C(\tau) = \langle E(t)E(t+\tau)\rangle = \langle \sum_n r_n \cos(\omega_n t + \varphi_n) \sum_m r_m \cos(\omega_m(t+\tau) + \varphi_m)\rangle$$

The averaging on the right hand side has to be carried out by letting each of the phases φ_k vary independently from 0 to 2π. When we do this, only terms with $m = n$ can survive, and we get

$$C(\tau) = \sum \frac{1}{2} r_n^2 \cos\omega_n\tau$$

.

Putting τ equal to zero, we get the variance

$$C(0) = \langle E(t)^2\rangle = \sum \frac{1}{2} r_n^2$$

We note that the autocorrelation is independent of t and hence we have checked time stationarity, at least for this statistical property. We now have to face the limit $T \to \infty$. The number of frequencies in a given range $\Delta\omega$ blows up as

$$\frac{\Delta\omega}{(2\pi/T)} = \frac{T\Delta\omega}{2\pi}.$$

Clearly, the r_n^2 have to scale inversely with T if statistical qualities like $C(\tau)$ are to have a well defined $T \to \infty$ behaviour. Further, since the number of r_n's even in a small interval $\Delta\omega$ blows up, what is important is their combined effect rather than the behaviour of any individual one. All this motivates the definition.

$$\sum_{\omega<\omega_n<\omega+\Delta\omega} \frac{r_n^2}{2} = 2S(\omega)\Delta\omega$$

as $T \to \infty$. Physically, $2S(\omega)\Delta\omega$ is the contribution to the variance $\langle E^2(t)\rangle$ from the interval ω to $\omega + \Delta\omega$. Hence the term "power spectrum" for $S(\omega)$. Our basic result for the autocorrelation now reads

$$C(\tau) = \int_0^\infty 2S(\omega)\cos\omega\tau d\omega = \int_{-\infty}^{+\infty} S(\omega)e^{-i\omega\tau} d\omega$$

if we define $S(-\omega) = S(\omega)$.

This is the "Wiener–Khinchin theorem" stating that the autocorrelation function is the Fourier transform of the power spectrum. It can also be written with the frequency measured in cycles (rather than radians) per second and denoted by ν.

$$C(\tau) = \int_0^{\infty} 2P(\nu)\cos(2\pi\nu\tau)d\nu = \int_{-\infty}^{+\infty} P(\nu)e^{-2\pi i\nu\tau}d\nu$$

and as before, $P(-\nu) = P(\nu)$.

In this particular case of the autocorrelation, we did not use independence of the φ 's. Thus the theorem is valid even for a non-gaussian random process. (for which different φ 's are not independent). Notice also that we could have averaged over t instead of over all the φ's and we would have obtained the same result, viz. that contributions are nonzero only when we multiply a given frequency with itself. One could even argue that the operation of integrating over the φ's is summing over a fictitious collection (i.e "ensemble") of signals, while integrating over t and dividing by T is closer to what we do in practice. The idea that the ensemble average can be realised by the more practical time average is called "ergodicity" and like everything else here, needs better proof than we have given it. A rigorous treatment would in fact start by worrying about existence of a well-defined $T \to \infty$ limit for all statistical quantities, not just the autocorrelation. This is called "proving the existence of the random process".

The autocorrelation $C(\tau)$ and the power spectrum $S(\omega)$ could in principle be measured in two different kinds of experiments. In the time domain, one could record samples of the voltage and calculate averages of lagged products to get C. In the frequency domain one would pass the signal through a filter admitting a narrow band of frequencies around ω, and measure the average power that gets through.

A simple but instructive application of the Wiener Khinchin theorem is to a power spectrum which is constant ("flat band") between $\nu_0 - B/2$ and $\nu_0 + B/2$. A simple calculation shows that

$$C(\tau) \;=\; 2KB\left(\cos(2\pi\nu_0\tau)\right)\left(\frac{\sin(\pi B\tau)}{\pi B\tau}\right)$$

The first factor $2KB$ is the value at $\tau = 0$, hence the total power/variance to radio astronomers/statisticians. The second factor is an oscillation at the centre frequency. This is easily understood. If the bandwidth B is very small compared to ν_0, the third factor would be close to unity for values of τ extending over say $1/4B$, which is still many cycles of the centre frequency. This approaches the limiting case of a single sinusoidal wave, whose autocorrelation is sinusoidal. The third sinc function factor describes "bandwidth decorrelation[1]", which occurs when τ becomes comparable to or larger than $1/B$.

Another important case, in some ways opposite to the preceding one, occurs when $\nu_0 = B/2$, so that the band extends from 0 to B. This is a so-called "baseband". In this case, the autocorrelation is proportional to a sinc function of $2\pi B\tau$. Now, the correlation between a pair of voltages measured at an interval of $1/2B$ or any multiple (except zero!) thereof is zero, a special property of our flat band. In this case, we see very clearly that a set of samples measured at this interval of $1/2B$, the so-called "Nyquist sampling interval", would actually be statistically independent since correlations between any pair vanish (this would be clearer after going through Section 1.8). Clearly, this is the minimum number of measurements which would have to be made to reproduce the signal, since if we missed one of them the others would give us no clue about it. As we will now see, it is also the maximum number for this bandwidth!

1.4 The Sampling Theorem

This more general property of a band-limited signal (one with zero power outside a bandwidth B) goes by the name of the "Shannon Sampling Theorem". It states that a set of

[1] also called "fringe washing" in Chapter 4

samples separated by $1/2B$ is sufficient to reconstruct the signal. One can obtain a preliminary feel for the theorem by counting Fourier coefficients. The number of parameters defining our signal is twice the number of frequencies, (since we have an a and a b, or an r and a φ, for each ω_n). Hence the number of real values needed to specify our signal for a time T is

$$2 \times \frac{\Delta\omega T}{2\pi} = 2\left(\frac{\Delta\omega}{2\pi}\right) T = 2BT$$

This rate at which new real numbers need to be measured to keep pace with the signal is $2B$. The so called "Nyquist sampling interval" is therefore $(2B)^{-1}$. A real proof (sketched in Section 1.8) would give a reconstruction of the signal from these samples!

In words, the Shannon criterion is two samples per cycle of the maximum frequency *difference* present. The usual intuition is that the centre frequency ν_0 does not play a role in these considerations. It just acts a kind of rapid modulation which is completely known and one does not have to sample variations at this frequency. This intuition is consistent with radio engineers/astronomers fundamental right to move the centre frequency around by heterodyning[2] with local (or even imported[3]) oscillators, but a more careful examination shows that the centre frequency should satisfy $\nu_0 = (n + \frac{1}{2})B$ for the sampling at a rate $2B$ to work.

1.5 The Central Limit and Pairing Theorems

We now come to the statistics of $E(t)$. For example, we already know that $\langle E^2(t)\rangle = \sum r_n^2/2$. How about $\langle E^3(t)\rangle$? Quite easy to check that it is zero because

$$\langle r_l r_m r_n \cos(\omega_m t + \varphi_m)\cos(\omega_n t + \varphi_n)\cos(\omega_l t + \varphi_l)\rangle = 0$$

when we let the φ's each vary independently over the full circle 0 to 2π. This is true whether l, m, n are distinct or not. But coming to even powers like $\langle E^4(t)\rangle$, something interesting happens. When we integrate a product like $r_l r_m r_n r_p \cos(\omega_m t + \varphi_m)\cos(\omega_n t + \varphi_n)\cos(\omega_\ell t + \varphi_l)\cos(\omega_p t + \varphi_p)$ over all the four φ's we can get non-zero answers, provided the φ's occur in pairs, i.e., if $l = m$ and $n = p$, then we encounter $\cos^2\varphi_l \times \cos^2\varphi_n$ which has a non-zero average. (We saw a particular case of this when we calculated $\langle E(t)E(t+\tau)\rangle$ and only r_m^2 type terms survived).

Because of the random and independent phases of the large number of different frequencies, we can now state the "pairing theorem".

$$\langle E(t_1)E(T_2)\ldots E(t_{2k})\rangle = \sum_{\text{pairs}} \langle E(t_1)E(t_2)\rangle \ldots \langle E(t_{2k-1})E(t_{2k})\rangle$$

As discussed in Section 1.8, this pairing theorem proves that the statistics is gaussian. (A careful treatment shows that only the $r_m^2 r_n^2$ terms are equal on the two sides- we have not quite got the r_m^4 terms right, but there are many more (of the order of N times more) of the former type and they dominate as $T \to \infty$ and the numbers of sines and cosines we are adding is very large). This result — that the sum of a large number of small, finite variance, independent terms has a gaussian distribution — is a particular case of the "central limit theorem". We only need the particular case where these terms are cosines with random phases.

[2]see Chapter 3

[3]aaaaagggh! beware of weak puns. (eds.)

1.6 Quasimonochromatic and Complex Signals

For a strictly monochromatic signal, electrical engineers have known for a long time that it is very convenient to use a *complex* voltage $V(t) = E_0 \exp(i(\omega t + \varphi))$ whose real part gives the actual signal $E_r(t) = E_0 \cos(\omega t + \varphi)$. One need not think of the imaginary part as a pure fiction since it can be obtained from the given signal by a phase shift of $\pi/2$, viz. as $E_i(t) = E_0 \cos(\omega t + \phi - \pi/2)$. In practice, since one invariably deals with signals at an intermediate frequency derived by beating with a local oscillator, both the real and imaginary parts are available by using two such oscillators $\pi/2$ out of phase. Squaring and adding the real and imaginary parts give $E_r^2(t) + E_i^2(t) = V(t)^* V(t) = E_0^2$ which is the power averaged over a cycle. This is actually closer to what is practically measured than the instantaneous power, which fluctuates at a frequency 2ω.

These ideas go through even when we have a range of frequencies present, by simply imagining the complex voltages corresponding to each of the monochromatic components to be added. In mathematical terms, this operation of deriving $E_i(t)$ from $E_r(t)$ goes by the name of the "Hilbert Transform", and the time domain equivalent is described in Section 1.8 But the physical interpretation is easiest when the different components occupy a range $\Delta\omega$ - the so called "bandwidth" - which is small compared to the "centre frequency" ω_0. Such a signal is called "quasimonochromatic", and can be represented as below

$$E_q(t) = Re\ \exp(i\omega_0 t) \sum_{-\Delta\omega/2<\omega_1<\Delta\omega/2} E(\omega_1)\ \exp(i\omega_1 t + i\varphi(\omega_1))$$

In this expression, ω_1 is a frequency offset from the chosen centre ω_0, so that $E(\omega_1)$ actually represents the amplitude at a frequency $\omega_0 + \omega_1$, and $\varphi(\omega_1)$ the phase. We can now think of our quasimonochromatic signal as a rapidly varying phasor at the centre frequency ω_0, modulated by a complex voltage

$$V_m(t) = \sum_{-\Delta\omega/2<\omega_1<\Delta\omega/2} E(\omega_1)\ \exp(i\omega_1 t + i\varphi\omega_1)$$

This latter phasor varies much more slowly than $\exp(-i\omega_0 t)$. In fact, it takes a time $\Delta\omega^{-1}$ for $V_m(t)$ to vary significantly since the highest frequencies present are of order $\Delta\omega$. This time scale is much longer than the timescale ω^{-1} associated with the centre frequency. Writing $V_m(t)$ in the polar form as $R(t)\exp(i\alpha(t))$, our original real signal reads

$$E_q(t) = R(t)\cos(\omega_0 t + \alpha(t))$$

.

We can think of R and α as time dependent, slowly varying, amplitude and phase modulation of an otherwise (hence "quasi") monochromatic signal.

While the mathematics did not assume smallness of $\Delta\omega$, the physical interpretation does. If $R(t)$ changes significantly during a cycle, some of its values may not be attained as maxima and hence its square cannot be regarded as measuring average power. This is as it should be. No amount of algebra can uniquely extract *two* real functions $R(t)$ and $\alpha(t)$ from a single real signal without further conditions (and the condition imposed is explained in section 1.8).

But returning to the quasimonochromatic case, we can now think of $V_m(t)^* V_m(t)$ as the (slowly) time varying power in the signal. Likewise we can think of $\langle V_m^*(t) V_m(t+\tau)\rangle$ as the autocorrelation. (A little algebra checks that this is the same as the autocorrelation of the original real signal). One advantage in working with the complex signal is that the centre frequency cancels in any such product containing one voltage and one complex conjugate voltage. We can therefore think of such products as referring to properties of the fluctuations of the signal amplitude and phase, and measure them even after heterodyning has changed the centre frequency.

1.7 Cross Correlations

We have so far thought of the signal as a function of time, after it enters the antenna. Let us now liberate ourselves from one dimension (time) and think of the electric field as existing in space and time, before it is collected by the antenna. In this view, one can obtain a delayed version of the signal by moving along the longitudinal direction (direction of the source). Thus, the frequency content is obtained by Fourier transforming a *longitudinal* spatial correlation. As explained in Chapter 2, the spatial correlations *transverse* to the direction of propagation carry information on the angular power spectrum of the signal, i.e. the energy as a function of direction in the sky. With hindsight, this can be viewed as a generalisation of the Wiener- Khinchin theorem to spatial correlations of a complex electric field which is the sum of waves propagating in many different directions. Historically, it arose quite independently (and about at the same time!) in the context of optical interference. This is the van Cittert-Zernike theorem of Chapter 2. Since one is now multiplying and averaging signals coming from different antennas, this is called a "cross correlation function". To get a non-vanishing average, one needs to multiply $E_1(x,t)$ by $E_2^*(y,t)$. The complex conjugate sign in one of the terms ensures that this kind of product looks at the phase *difference*. Writing out each signal as a sum with random phases, the terms which leave a non-zero average are the ones in which an $e^{i\varphi_n}$ in an E cancels a $e^{-i\varphi_n}$ in an E^*. An (ill-starred?) product of two complex E's with zero (or two!) complex conjugate signs would average to zero.

1.8 Mathematical details

This section gives some more mathematical details of topics mentioned in the main text of the chapter.

We first give the generalisation of the two variable gaussian to the joint distribution of k variables. Defining the covariance matrix $C_{ij} = \langle x_i x_j \rangle$, and $A = C^{-1}$, then we have

$$P(x_1 \ldots x_k) = (2\pi)^{-k/2}(det\ A)^{1/2} \exp\left(-\frac{1}{2}x^T Ax\right)$$

The quadratic function Q in the exponent has been written in matrix notation with T for transpose. In full, it is $Q = \sum_{ij} x_i a_{ij} x_j$. Notice that the only information we need for the statistics of the amplitudes at k different times is the autocorrelation function $C(\tau)$, evaluated at all time differences $t_i - t_j$. Formally this is stated as "the gaussian process is defined by its second order statistics".

What would be practically useful is an explicit formula for the average value of an arbitrary product $x_i x_j x_l \ldots$ in terms of the second order statistics $\langle x_1 x_2 \rangle \langle x_3 x_7 \rangle \ldots$ etc. The first step is to see that a product of an **odd** number of x's averages to **zero**. (The contributions from $x_1 \ldots x_k$ & $-x_1 \ldots - x_k$ cancel).

For the case of an even number of gaussian variables to be multiplied and averaged, there is a standard trick to evaluate an integral like $\int P(x_1 \ldots x_k) x_3 x_7 \ldots dx_1 \ldots$. Define the Fourier transform of P,

$$G(k_1 \ldots k_k) = \int\int P(x_1 \ldots x_k) e^{-ik_1 x_1 \ldots ik_k x_k} dx_1 \ldots dx_k$$

It is a standard result, derived by the usual device of completing the square, that this Fourier transform is itself a gaussian function of the k's, given by

$$G(k_1, \ldots, k_k) = \exp\left(-\frac{1}{2}\sum_{ij} C_{ij}k_ik_j\right) \equiv \exp\left(-\frac{1}{2}k^TCk\right).$$

Differentiating with respect to k_1 and then k_2, and putting all k's equal to zero, pulls down a factor $-x_1x_2$ into the integral and gives the desired average of x_1x_2. This trick now gives the average of the product of a string of x's in the form of the "pairing theorem". This is easier to state by an example.

$$\begin{aligned}\langle x_1x_2x_3x_4\rangle &= \langle x_1x_2\rangle\langle x_3x_4\rangle + \langle x_1x_3\rangle\langle x_2x_4\rangle + \langle x_1x_4\rangle\langle x_2x_3\rangle \\ &\equiv C_{12}C_{34} + C_{13}C_{24} + C_{14}C_{23}\end{aligned}$$

A sincere attempt to differentiate G with respect to $k_1k_2k_3$ and k_4 and then put all k's to zero will show that the C's get pulled down in precisely this combination. Deeper thought shows that the pairing rule works even when the x's are not all identical, i.e.,

$$\langle x^4\rangle = \langle x^2\rangle\langle x^2\rangle + \langle x^2\rangle\langle x^2\rangle + \langle x^2\rangle\langle x^2\rangle = 3\langle x^2\rangle^2 = 3\sigma^4$$

or even $\langle x^{2n}\rangle = 1, 3, 5 \ldots (2n-1)\sigma^{2n}$.

The last property is easily checked from the single variable gaussian

$$(2\pi\sigma^2)^{-1/2}\exp(-x^2/2\sigma^2)$$

Since the pairing theorem allows one to calculate **all** averages, it could even be taken to define a gaussian signal, and that is what we do in the main text.

We now sketch a proof of the sampling theorem. Start with a band limited (i.e containing only frequencies less than B) signal sampled at the Nyquist rate, $E_r(n/2B)$. The following expression gives a way of constructing a continuous signal $E_c(t)$ from our samples.

$$E_c(t) = \sum_n E_r(n/2B)\ \mathrm{sinc}(2\pi B(t - \frac{n}{2B}))$$

It is also known as Whitaker's interpolation formula. Each sinc function is diabolically chosen to give unity at one sample point and zero at all the others, so $E_c(t)$ is guaranteed to agree with our samples of $E_r(t)$. It is also band limited (Fourier transform of a flat function extending from $-B$ to $+B$). All that is left to check is that it has the same Fourier coefficients as $E_r(t)$ (it does). And hence, we have reconstructed a band limited function from its Nyquist samples, as promised.

We add a few comments on the notion of Hilbert transform mentioned in the context of associating a complex signal with a real one. It looks rather innocent in the frequency domain, just subtract $\pi/2$ from the phase of each cosine in the Fourier series of $E_r(t)$ and reassemble to get $E_i(t)$. In terms of complex Fourier coefficients, it is a multiplication of the positive frequency component by $-i$ and of the corresponding negative frequency component by $+i$, Apart from the i, this is just multiplication by a step function of the symmetric type, jumping from minus 1 to plus 1 at zero frequency. Hence, in the time domain, it is a convolution of $E_r(t)$ by a kernel which is the Fourier transform of this step function, viz $1/t$ (the value t=0 being excluded by the usual principal value rule). Explicitly, we have

$$E_i(t) = \int E_r(s) P \left[1/(t-s) \right] \; ds/\pi$$

There is a similar formula relating E_r to E_i which only differs by a minus sign. This is sufficient to show that one needs values from the infinite past, and more disturbingly, future, of t to compute $E_i(t)$. This is beyond the reach of ordinary mortals, even those equipped with the best filters and phase shifters. Practical schemes to derive the complex signal in real time thus have to make approximations as a concession to causality.

As remarked in the main text, there are many complex signals whose real parts would give our measured $E_r(t)$. The choice made above seemed natural because it was motivated by the quasimonochromatic case. It also has the mathematical property of creating a function which is very well behaved in the upper half plane of t regarded as a complex variable, (should one ever want to go there). The reason is that $V(t)$ is constructed to have terms like $e^{i\omega t}$ with only positive values of ω. Hence the pedantic name of "analytic signal" for this descendant of the humble phasor. It was the more general problem of continuing something given on the real axis to be well behaved in the upper half plane which attracted someone of Hilbert's IQ to this transform.