Question 1: Why is a Gaussian white noise (or Gaussian random process) characterised by its spectrum only? As the spectrum does not have any phase information, is it necessary to consider the phase of a Gaussian process?

Answer 1: We first refer to the concept of white noise, and then will comment on spectrum and Gaussian processes.

White noise

White noise is a rather abstract—and sometimes intuitively understood—term which depending on the area of study has different implications. Brown [Bro83] defines it as “...a stationary random process having a constant spectral density function.” while Papoulis [Pap91] states that “We shall say that a process \(v(t)\) is white noise if its values \(v(t_i)\) and \(v(t_j)\) are uncorrelated for every \(t_i\) and \(t_j \neq t_i\).”

Perhaps a more physically-intuitive definition of white noise can be stated based on its relationship to the filtering problem and stochastic differential equations in general. Oksendal [Oks10] and Bain [BC09] suggest that white noise can be regarded as a process whose integral is Brownian motion, an almost surely continuous process with independent and normally distributed increments. Observe that although this definition provides physical meaning, it is not thoroughly correct since Brownian motion is nowhere differentiable.

In order to understand the spectral properties of white noise, we first need to agree in what we understand by such term. We therefore, from a signal processing point of view, consider the process \(X_t\) to be white noise if and only if \(X_t\) is wide sense stationary and

\[
\mathbb{E}\{X_t\} = 0 \quad (1)
\]

\[
\mathbb{E}\{X_s^*X_t\} = \sigma^2 \delta(t-s) \quad (2)
\]

Spectral properties of stochastic processes

Spectral analysis of deterministic signals has seen a number of applications from everyday tasks such as tone-touching devices to professional high quality audio and compression. In this context, Fourier analysis provides a robust theoretical framework and intuitive physical interpretation. Strictly speaking, the Fourier transform is a bijective—and therefore invertible—operator between function spaces, meaning that a signal in time is uniquely defined by its Fourier transform (via both its real and imaginary parts, or equivalently via its magnitude and phase).

The extension to the Fourier transform to stochastic processes is not straightforward\(^1\). As \(\{X_t, t > 0\}\) is a stochastic process, that is, a collection of random variables, its Fourier transform

\[
\mathcal{F}\{X_t\}(f) = \int_{-\infty}^{\infty} X_t e^{j2\pi ft} dt, \quad (3)
\]

\(^1\)For an alternative representation of a stochastic process as an infinite linear combination of orthogonal functions, see the Karhunen–Loève theorem.
is a random variable itself.

So, what is the meaning of spectrum and phase of a random signal? Is it the expectation of its Fourier transform? At least for the case of white noise, this does not make much sense due to the linear properties of the integral operator. The expected value of the Fourier transform of a realisation of white noise is given by

$$\mathbb{E}\{\mathcal{F}\{X_t\}(f)\} = \mathcal{F}\{\mathbb{E}\{X_t\}(f)\} = 0,$$

so, if we were to support the spectral analysis of random signals with the expected value of their Fourier transform, we would not achieve much.

Nevertheless, it is possible to use the Fourier transform to analyse the power spectrum of wide-sense stationary stochastic processes. The Wiener–Khinchin theorem [Wie30, Khi34] states that the autocorrelation function of a wide-sense-stationary random process has a spectral decomposition given by the power spectrum of that process. That is,

$$S_X(f) = \int_{-\infty}^{\infty} r_X(\tau)e^{j2\pi f\tau}d\tau,$$

where $r_X(\tau) = \mathbb{E}\{X_t^*X_{t+\tau}\}$.

This is a consequence of the integral forms of the covariance/autocorrelation operator (responsible of the notion of power in the above equation) and the Fourier transform (responsible of the notion of spectrum), which can be interchanged due to their linear properties, therefore, allowing to first compute power and second spectrum.

For the case of white noise, as the correlation function is a Dirac delta, see Eq. (2), its Fourier transform is a constant function in the frequency domain, meaning that its power spectral density is equal for all frequencies just like white light.

What is the phase of white noise then? Well, I have taken a long way to say this and I think it is worth it: the phase of white noise is a random variable, and as such its analysis can only be performed based on its probability density function. We have shown that its first moment (mean) will not give us much information, but perhaps higher order statistics will do.

Notes on Gaussianity

We have assumed only three properties to define white noise, i.e. wide-sense stationary, zero-mean, uncorrelated. In this regard, all the stochastic properties of such signal are uniquely defined by unique pdf, or strictly speaking, a collection of identical pdfs which joint distribution has a covariance matrix that is a factor of the identity matrix.

Furthermore, the definition of Gaussian processes is very clear: \{X_t, t > 0\} is a Gaussian process if and only if, for any finite set of indices \{t_1, t_2, \ldots, t_n\} the collection of random variables \{X_{t_1}, X_{t_2}, \ldots, X_{t_n}\} denotes a multivariate Gaussian random variable and has a joint Gaussian distribution of the form

$$p(X_{t_1}, X_{t_2}, \ldots, X_{t_n}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(X - \mu)\Sigma^{-1}(X - \mu)\right),$$

where $X = [X_{t_1}, X_{t_2}, \ldots, X_{t_n}]^T$, and $\mu$ and $\Sigma$ are parameters representing the mean and covariance of $X$ respectively.

Based on these two premises, white noise does not need to be Gaussian, but only zero mean, uncorrelated and stationary. Furthermore, for the particular case of Gaussian white noise, the correlation matrix $\Sigma$ is a factor of the identity matrix due to Eq. (2).
**Question 2a:** Does a signal have to be stationary to apply adaptive filters?

**Answer 2a:** No. There are three different concepts here: (i) the nature of the filter, (ii) the nature of the data, and (iii) the implementation of the filter.

An adaptive filter is a particular type of filter whose structure (e.g. parameters, order, transfer function, etc.) is updated in an online fashion based on some error measure. Depending on the design of the adaptive capability of the filter, that is, the minimisation function considered or the update criteria, adaptive filters may have different convergence properties, and are optimal only for a specific type of data, however, this does not imply that such a filter cannot be applied to data of a different nature.

I will borrow an example from non-adaptive filtering to illustrate this. An optimal solution for the general nonlinear filtering problem is given by the Kushner-Stratonovich equation [Kus64, Str60], which is a differential equation for the posterior probability density function of the state of a system [BC09]. A closed form solution to this equation can only be found in two particular cases: the linear and Gaussian case that yields the Kalman filter [Kal60], and the nonlinear case satisfying the Benes condition [Ben81]. By saying that the Kalman filter is the optimal least squares filter for linear and Gaussian systems, we mean that there is no other filter that performs better for this type of systems.

In real-world applications we are yet to find systems satisfying the linear, Gaussian or Benes conditions, however, we try and implement optimal filters designed for ideal systems meeting these theoretical conditions. Usually, they provide satisfactory results, such as the case of the Kalman filter applied to guidance, navigation and control of vehicles, econometric analysis and general signal processing applications.

It is worth emphasising that the properties of convergence and optimality of both adaptive and fixed filters, only reflect their design criterion and by no means restrict the type of data they can be applied to but only represents the type of data for which the filter is optimal. An optimal filter can indeed provide acceptable, but not optimal, results for data that do not meet the theoretical requirements for optimality.

**Question 2b:** can I say that a well designed adaptive filter can be applied to estimate a nonlinear and nonstationary signal without explicitly specifying the nonlinearity and nonstationarity of the signal, i.e., the changing properties of the data over time?

**Answer 2b:** In line with my previous answer, you certainly can apply the filter but the performance will be far from optimal.

It is a somewhat unclear what well designed filter means in this case. To design a filter you need to know some properties of the data or the underlying generating system, otherwise there is no design. You cannot design an optimal filter for a black-box time-varying system, although I would recommend some particle-filtering-based algorithms [DdFG01].

**Question 3:** If I only calculate coefficients of a filter when estimation errors are lower than a certain threshold, other than feeding the errors back to the calculation, whenever a new sample is available, is this filter an "adaptive filter"?

**Answer 3:** It is indeed. This type of adaptive filter would be, for instance, one with a non-convex cost function which is insensitive to large errors (those larger that your threshold.)
References


