Digital Communication Using GNU Radio

Prof Kumar Appaiah

Department of Electrical Engineering

Indian Institute of Technology Bombay

Week-03

Lecture-16

Welcome back to this lecture on digital communication using GNU radio. In this lecture, we are going to continue our discussion on demodulation, the preliminaries and talk a little more about jointly Gaussian random vectors and then we will briefly discuss a little bit about hypothesis testing as well. Let us now go back to where we left off in the last lecture. We were dealing with collections of random vectors, particularly Gaussian random vectors and we were saying that if you have a random vector $\boldsymbol{X}$, this consists of a column vector consisting of

X_1, X_2, up to X_n

which are random variables, which are jointly Gaussian. One important aspect is that, I have not provided the definition here, a jointly Gaussian random vector has a special definition. It is one where any linear combination of the entries results in a Gaussian random variable.

There is a very classic example of where a pair of random variables, both of them are Gaussian but they are not jointly Gaussian. So for example, let us say you have X is for simplicity $N(0,1)$. So, X is a Gaussian random variable which is normally distributed. Let us make it little neater.

So X is distributed as normal with mean zero variance 1.

α

is independent of X and is +1 or -1 with equal probability. So it is like α is one where you toss a coin and if its heads is +1, the coin is tails is -1 independent of X and you write

$Y=\alpha X$.

Now, what does Y do? Y is essentially just a coin flip multiplied by a Gaussian. So, since X is normally distributed with mean zero and variance 1, you can actually

check the distribution for Y .

If α is 1, it is going to be the same as X which means it is going to be Gaussian. If alpha is -1, it is actually flipped version of X but you know X is normally distributed with mean zero and variance 1. $-X$ is also going to be normally distributed with mean zero and variance 1. So in both the cases, it is very evident that Y is also Gaussian with mean zero and variance 1. But this is a very classic example where X and Y are both individually Gaussian but not jointly Gaussian.

Why is that the case? Because if you look at $X+Y$, $X+Y$ is not a Gaussian random variable. Why? $X+Y$ takes the value zero with probability $1/2$. That is, it takes the value zero with probability $1/2$. That is because whenever

$\alpha=-1$,

$Y=-X$.

So

$X+Y=0$.

So with probability $1/2$, you are getting zero. You are getting a weight on a single number and as you know a continuous random variable like a Gaussian cannot take any single real number with any non-zero probability. So this is a classic case where two individual random variables are Gaussian but they are not jointly Gaussian. Jointly Gaussian means if suppose you have

X_1 , X_2 , up to X_n

and for any real numbers or complex numbers

α_1 , α_2 , α_n ,

is Gaussian with any distribution of course, then these are said to be jointly Gaussian random variables. So this is something we should really keep in mind.

That is the only definition of a jointly Gaussian random variable and only such random variables have this kind of distribution that is

$$\frac{1}{\sqrt{(2\pi)^n |\boldsymbol{C}_x|}} \exp\left(\frac{-(\boldsymbol{x}-\boldsymbol{m}_x)^H \boldsymbol{C}_x^{-1} (\boldsymbol{x}-\boldsymbol{m}_x)}{2} \right)$$

Now as we discussed at the end of the last lecture, independent and uncorrelated are the same for jointly Gaussian random vectors.

This is something which is also very key. If you remember the example we discussed where

$Y = \alpha X$,

Y is basically $+X$ or $-X$ with equal probability, you can actually verify that those two are uncorrelated. But they are definitely not independent because if you know the value of X , then you know that with probability $1/2$ Y is definitely X with other probability $1/2$ Y is definitely $-X$, they are definitely not independent. But they are uncorrelated because if you evaluate

in that case, it will turn out to be 0. So in that sense, independent, equal to uncorrelated only holds for jointly Gaussian random vectors.

How? So in the case of uncorrelated random variables, you have a

$\boldsymbol{C}_x$

which has only diagonal entries and 0s here, 0s here. Why? Because what are the off diagonal entries of the covariance matrix? The off diagonal entries of the covariance matrix are the correlations multiplied by root variances of the individual variables, which means if you say that those random variables are uncorrelated, their covariance is 0, which means your $\boldsymbol{C}_x$ is now a diagonal matrix. If your $\boldsymbol{C}_x$ is a diagonal matrix, you can easily verify $\boldsymbol{C}_x$ inverse is also a diagonal matrix and if you then look at this

$(\boldsymbol{x}-\boldsymbol{m}_x)^H \boldsymbol{C}_x^{-1} (\boldsymbol{x}-\boldsymbol{m}_x)$,

of course there is a hermitian here, you can easily verify that what will happen is that this will essentially have a form, this will result in something of the form, maybe I will just erase this, some constant times

x_1^2

or some constant times

$(x_1 - m_{x_1})^2$

plus some constant times

x_2^2

or

$(x_2 - m_{x_2})^2$

and so on. And these, so the way it will appear is that you will have

$e^{- \sim x_1^2 - \sim x_2^2 - \sim ..}$

and this can actually split into

$e^{-x_1^2} \sim e^{-x_2^2} \sim e^{-..}$

and so on. Now you know that whenever you have a joint distribution which splits into the marginals and products, you can verify that that is going to be a, that means that from all the random variables in the joint distribution are independent because you can show that the uncorrelatedness which reflects in C_x will result in the PDFs being appearing as the product of the marginals.

So that is something special only for Gaussian, maybe there may be some other variables for Gaussian definitely uncorrelated implies independence and this is definitely not the case for most other random variables you can check. Finally, of course joint Gaussianness is preserved under affine transformations. You can see that this is an affine transformation over here, over here and that preserves the joint Gaussianness that is if you take a jointly Gaussian random vector, apply an affine transformation to it, then the resulting random variable is also going to be jointly Gaussian. Now again these things are going to come in handy whenever you deal with vectors like jointly detecting multiple symbols and so on. So just keep these things in mind, we will refer back to these as and when the situation arises.

The next thing that we have to consider is Gaussian random processes. Now one thing you must remember is that whenever we deal with these practical communication systems you have aspects like noise and you know channel and all those things. Typically what happens is that the random process essentially is like a varying waveform that is you have, there are two pictures of a random process that is one picture is, there is one particular random process, one particular realization, there is another realization and then there is a third realization. These are all sample path wise realizations. For example, let us say that you know this is like one particular path which you take, another

particular path which you take, another particular path which you take.

There are multiple paths, each of these can be chosen randomly. This is the picture when you want to basically fix the realization and view as a function of time. The other picture is if you fix time, okay I will go back to the blue pen one second, if you fix time then it can be this red value, this blue value, this green value and so on. That is

$x(t)$

is a Gaussian random variable. So there are two pictures in the case of a Gaussian random, I mean any random process or a Gaussian random process.

If you want to look at it as a function of time, it is a realization. If you fix t and want to look at it, then it becomes a random variable. So typically this picture is you know used. We say

$x(t)$

is a random variable for every $t \in \mathbb{R}$. While we are silent about the exact interpretation of t , you can assume that t is like time.

So for every time t you have a random variable. Now in the random process, what is the significance of our understanding and why do we have to look at these things? See whenever you deal with random processes, suppose that you make an observation now, one question that we may ask is, is this observation useful for the next time instant? For example, if it is sunny today, then if it is likely to be sunny tomorrow as well, you can say that over time, you know the level of let's say the sun is definitely correlated across time. That is something which you can say. But suppose that you are in an environment where you know today it may be sunny, tomorrow all bets are off, it may be sunny, cloudy, rainy or whatever, then you can say that for this particular random process, there is no real correlation among the weather on successive days. So you have these kinds of characterizations of how closely related the random variables are.

So this is something which we will try to exploit even when designing it for our communication system. So for real numbers

t_1 , t_2 , up to t_n

and complex numbers

a_1 , a_2 up to a_n ,

now just hold on and we will tell you what

t_1 , t_2 , up to t_n

are. Let us say that we observe the random process at time t_1 , time t_2 and time t_n and then we take a linear combination of those random variables you observed at time t_1 , time t_2 and time t_n that is

$a_1 x(t_1)+a_2 x(t_2)+..$.

Then this is a Gaussian random variable. This is the definition of a Gaussian random process.

I think you can recall that I just mentioned about jointly Gaussian random vector. If you take

t_1 , t_2 , up to t_n

and measure the random process at that time,

$x(t_1)$, $x(t_2)$, $x(t_n)$

forms a random vector and I am claiming that it is a jointly Gaussian random vector because

$a_1 x(t_1)+a_2 x(t_2)+..$

is Gaussian and therefore the Gaussian random process is defined in this manner. If you basically take the random variables that constitute this random process for various times and take any linear combination, if that results in a Gaussian random variable then this process is said to be a Gaussian random process. Now the Gaussian random process is completely characterized by its mean and its autocorrelation function. Like if you recall we mentioned that jointly Gaussian random vector, you just need to know its mean, you just need to know its covariance matrix.

In a similar way, the covariance matrix is transformed to autocorrelation function. So the covariance matrix was useful in the random vector case because you had finite set of random vectors. In this case, the autocorrelation function tells you the relationship between pairs of random variables for every pairs of times. Now one aspect is that you have these aspects related to

$x(t_1)$ and $x(t_2)$

have some correlation. What about

and $x(t_2+\delta t)$,

do they have the same correlation property? That is if you look elsewhere, do they have the same correlation properties? That particular aspect is called stationarity.

So if you recall a wide sense stationary process is one where the mean does not vary with time and the autocorrelation depends only on the time gap. That is if you want to measure the correlation between

$x(t_1)$ and $x(t_2)$,

let us say covariance, it depends only on

$t_2 - t_1$.

That is if this holds true, then you say that such a process is called a wide sense stationary process. So therefore, another way of looking at it is that the covariance function or the autocorrelation function, in this case we may write autocorrelation because we typically say that it is 0 mean or we subtract out the mean. So autocorrelation, autocovariance, it does not matter.

So the autocorrelation process depends only on the gap $t_2 - t_1$. We can always characterize the autocorrelation by calling this gap $t_2 - t_1$ as τ and such a process is called a wide sense stationary random process. It is very typical to assume that random processes are wide sense stationary when analyzing several communication systems. In the Gaussian case, there is a special bonus. Wide sense stationarity implies strict sense stationarity because it is almost like once you have characterized the covariance, it is like you know the distribution like in the case of random vector $\boldsymbol{X}$ is known.

In a similar way, the $\boldsymbol{C}_x$ or the autocorrelation which is basically getting the covariance for all pairwise random variables is fixed. Therefore, the distribution itself can be fully characterized. This is something which you can look at from references. But most important is Gaussian random processes have a very nice characterization. Any linear combination of

$x(t_1)$, $x(t_2)$, $x(t_n)$

is Gaussian.

Now, now that we have our Gaussian random processes in our toolkit, let us look at

$n(t)$.

We define n of t in this particular situation as the noise process. It has zero mean and we define its power spectral density $N_0/2$ as σ^2 . Now power spectral density, I am

sure you would have seen it in the context of you know random processes, but if you want to be refreshed about the definition, the power spectral density is the autocorrelation of the, sorry, is the Fourier transform of the autocorrelation. As we discussed, we are going to restrict our consideration to wide sense stationary processes.

So, if you take the autocorrelation which only depends on the lag $t_2 - t_1$ and find its autocorrelation, that is the power spectral density. So, the power spectral density function is $N_0/2$ which is same as σ^2 . That is if you look at the power spectral density, I will say

$S_n(f)$,

the power spectral density is flat and its value here is $N_0/2$. Now as you very well know, whenever you have a random variable which is flat like this, then let me just fix this a little bit.

So, this is $S_n(f)$. Whenever you have a random variable with a flat or rather not random, whenever you have a Fourier transform that is flat, this is $N_0/2$, you know that its corresponding time domain function is $\delta(t)$. You know, basically I am using the fact that $\delta(t)$ has Fourier transform 1. So, since $N_0/2$ is the Fourier transform, we have

.

This is the autocorrelation or autocovariance of the noise function is $N_0/2$ times $\delta(\tau)$. We can also write it as

$\sigma^2 \delta(\tau)$.

Now one question which you may ask is why this choice $N_0/2$? This $N_0/2$ choice is made for various reasons. Of course, you can go back and look at the Boltzmann constant based derivation and so on. But more importantly, we want $N_0/2$ because we will eventually move to complex baseband signaling. In the context of complex baseband signaling, if we choose $N_0/2$ as the noise along the real axis and $N_0/2$ as the noise along the imaginary axis and treat those as independent, then we will eventually get N_0 as the complex noise.

So, that is where we are getting at. For now, we are just going to define the noise along this particular real axis as $N_0/2$. $N_0/2$ is the real noise process with zero mean and power spectral density $N_0/2$. So, that is $n(t)$. So,

$N_0 = k T_0$,

k is the Boltzmann constant and T_0 is the operating temperature.

You can choose it as 300 Kelvin or something. It has to be in Kelvin, remember and the Boltzmann constant is as defined in your list of physical constants. Now, there is a problem with this definition of noise. That is, it is like noise is flat and has a very wide spectrum and this particular noise is actually, this is the frequency axis because of the Fourier transform. So, across all frequencies it keeps going. This means that your noise essentially has an infinite amount of energy in it.

But that does not make sense because you can actually start extracting the energy from the noise and build something called a perpetual motion machine, which is something you may have heard about in physics. You can actually try to get free energy from the noise. So, the answer to, you know, if you have a question as to whether that is possible, the answer is no. This is a noise model and what happens is that the noise essentially affects the frequency range within which you employ the communication system because wherever you send the signal that is where the noise is. So, to that extent within the band of interest, within the frequency range of interest, the noise exists and can be treated as flat.

That is basically the assumption we are making. So, let us not get into pathological questions of whether the noise actually has infinite energy, infinite power. No. We will just assume that if this is the range of bandwidths you are using, remember your pass band signaling and all those, so the noise here, this is the amount that affects your signaling. So, that is the summary. Now we will take a very brief look at hypothesis testing.

So, hypothesis testing can be looked at as a framework for deciding which of M possible hypotheses best explains an observation Y . This is based on a statistical model. So, we assume that, let us say that there is an observation Y , but there is something called i that actually causes this. If you want to have a very simple example, let us say that you send a message i . So, this can be 1, 2, 3, 4, you just send a message i and you receive Y .

The question is from Y , how do we decide which message was sent? So, that is why we have these hypotheses. Hypothesis H_i said message i is sent. Now we have to decide how these hypotheses are tested and how we decide which of these hypotheses is likely to be true. That is we have H_1 , H_2 up to H_M , let us say we send messages from 1 to M , which of these messages are sent is the question. So,

probability of H_i are the Bayesian priors that are known.

That is in this particular case, the simple way to understand is probability of H_i is the probability that message i is sent. For example, if you live in a city where let us say most of the days it is sunny and every day you send a message regarding the weather, then the probability that H_i, let us say that probability that you send the message that it is sunny is much more likely than the probability that you send the message that it is rainy or snowy or cold or something like that. So, in that sense $\mathrm{P}[H_i]$ are the Bayesian priors. But in the case of messages when you send in digital communication, let us say 0s and 1s. Typically, these 0s and 1s or 1, 2, 3, 4 all those are equally likely and given that they are equally likely, these $\mathrm{P}[H_i]$ are going to be $\frac{1}{M}$, which means all of them have the same prior probability.

So, this is something again that you have to keep in mind. A Bayesian prior indicates what the probability that that particular event of that particular message being sent is. If all messages are equally likely, then that does not help you make a decision. Intuitively, if you have a situation where it is more sunny most of the times, then you may say even if I get a rainy answer, I have to be really convinced that it is actually really rainy because most of the times he sends a sunny message. What if suddenly he or she sends a rainy message? These kinds of questions are what this essentially is meant to handle.

So, let us go to the hypothesis testing case with a very simple example. Before we go into that particular example, I will give you another simple test. Let us say that you have a car windshield and let us say that you count the number of drops and let us say that the number of drops is what you measure. Let us say it is something like 10,000, 100,000, 5 or something like that. So, your y is the number of drops on the car windshield and you have to decide whether it is raining or it is not raining.

For a minute, let us assume that you are only counting the drops and not checking the clouds. So, the question which we may ask is, there are two hypotheses. It is raining, it is not raining. Then based on the observation which is the number of drops which we call y, how will you conclude whether it is raining or not? In other words, we have to come up with a way. For example, we can make a statement like if the number of drops is greater than 1,25,000, then it is raining.

If it is less than or equal to 1,25,000, it is not raining. Someone actually just sprayed those to clean the windshield or something like that. In this manner, we have to come up with a means or a metric to break our tie or to find out which hypothesis we are going to decide on. Hypothesis testing can have errors. It could be that you are under a tree, so

you do not get enough number of drops, yet it is raining. So, this can lead to an error or someone essentially came with a bucket and poured all the water.

You counted the number of drops. There are many, many drops. So, you have greater than 1,25,000, but it is not raining. So, you made an error. So, under various situations, errors are possible. Your aim is to take into account all of these situations and minimize the error. So, this is like a very basic kind of example to motivate this hypothesis testing, but in a more practical scenario, let us restrict our consideration to the Gaussian case.

Let us say that you have a very basic Gaussian system. There are two symbols 0 or 1, but what you see is Y and Y manifests as, if 0 is sent, Y manifests as a Gaussian with mean 0 and variance σ^2. If 1 is sent, I am sorry, this should be 1, if 1 is sent, Y manifests as a Gaussian with mean m and variance σ^2. That is, if you send 0, then you get a Y which looks like this. This is essentially the distribution. Of course, it should not go below 0, sorry, please don't, you know, let me just redo this.

So, if you send, you get this. If you send 1, you get this, where this is m. So this is what is happening. So, if you, now let us look at this intuitively. Intuitively, it is very evident that if you send, if you receive some value which is here, then most likely what was sent, you know, it is like most likely that message 1 was sent and this Gaussian took it far to the right. It is more likely that a value close to m is taken far to the right than a value close to 0.

But suppose that you get something which is, you know, over here, let us say, over here, then it is more likely that a 0 was carried to this place rather than m, assuming m is a positive number. So, what is the decision rule? Now, we have to decide, just like my windshield example, we have to decide where we can draw a line and to the left of that line we will say message 0 was sent, to the right of that line we will say message 1 was sent. An intuitive rule is, this is 0, this is m. Of course, by symmetry, because the variance is the same, we can decide that $m/2$ is one place where, you know, we can draw the line and to the left we will decide that H_0 was sent, to the right we will decide H_1 was sent.

This looks like a possible way to make this decision. But so, I mean, when we decide this decision rule, we want to partition the observation space. We want to basically say anything here is H_0, anything here is H_1 and that kind of thing is what you want to say. But what is the correct statement or what is the mathematically correct statement which optimizes something is something that we are yet to see. So, let us briefly look at the conditional error probabilities. Errors can happen in hypothesis testing, that is, even though you know, you know, you, what message 1 was sent, it could be that the message

was carried to the left side because of the noise.

So, in this particular case, you know, instead of m you got something over here because of the behavior of the Gaussian. So, you can have this kind of errors. So, what is the probability that H_j is decided? You conclude that H_j is decided for some j not equal to i, but H_i is true. That is, what is the probability that you will decide this message 1 was sent even though message 0 was actually sent? What is the probability that you decide that message 0 was sent even though actually message 1 was sent? So, that is basically the

,

that is, let us say that you have multiple symbols, you decide one of the wrong symbols given that hi was sent, okay, and now this is same as

$$1-\mathrm{P}[\,Y\in\Gamma_i|H_i]\ ,$$

that is, this is basically the same as 1 minus the probability that y belongs to the correct region.

This is a conditional, you know, error probability just so you know. With priors, actually there is a little bit, it gets a little more tricky. With priors what happens is that you have to account for the prior, that is, what is the probability that, you know, message i is sent, what is the probability that message j is sent, this is the case where, you know, all messages are not equally likely. Right now we will just skip this for a minute. We will focus on this particular aspect. So, if we look at conditional error probabilities and maximum likelihood, that is what we call as ML, maximum likelihood.

The maximum likelihood essentially maximizes the, some likelihood function, that is, it tries to give you something which has the most probable way of deciding. Let us look at our example. What is the probability of error given 0 is sent? If you look back at our plot, once you send 0, you will make an error if you cross into $m/2$ or to the right. Similarly, if you send 1, you will make an error if you cross from $m/2$ to the, if you cross $m/2$ to the left. So now, coming back to that, what is the probability that you will make an error? Let us look at this particular case.

So, you will make an error if this particular Gaussian crosses $m/2$, that is, this is

$$\int_{m/2}^{\infty}\frac{1}{\sigma\sqrt{2\pi}}e^{-x^2/2\sigma^2}dx\ .$$

Now, again if you just do some quick normalizations and you know, do the transformation to z and stuff, you can easily find out that this is

$Q(m/2\sigma)$.

Similarly, if you do the same analysis for error given 1, you will make an error if what you observe is to the left of $m/2$ even though 1 was sent and that you can also verify, it will turn out to be

$Q(m/2\sigma)$.

This is because why is

$N(0,\sigma^2)$

under hypothesis H_0 , why is

$N(m,\sigma^2)$

under hypothesis H_1 ? So, it is very evident that for both of these, you can essentially evaluate these probabilities.

It turns out that both of these are the same. Now, that means that since, you know, probability of sending 0 and probability of sending 1 for both equiprobable

$Q(m/2\sigma)$

is indeed the probability of making an error. So, the maximum likelihood decision rule actually is going to say that I want you to find me

$\arg\max_{i\in 0,1,\ldots M-1} \mathrm{p}(y|i) = \arg\max_{i\in 0,1,\ldots M-1} \log \mathrm{p}(y|i)$.

So, that is you write the pdf

$\mathrm{p}(y|i)$

and find out that particular i which is the maximum. In the case of Gaussian, this is just going to turn out to be the minimum distance.

That is something which we will see. The other is the minimum probability of error rule which is the one with Bayesian priors. So now, if you want to minimize the probability of error, then this delta minimum probability of error

$\delta_{MPE}(y) = \arg\max_{i\in 0,1,\ldots M-1} \pi(i)\mathrm{p}(y|i)$.

The only difference is we do $\pi(i)$ times $\mathrm{p}(y|i)$ which includes the prior probability. That is in this case, let us say that there is a probability that you will send 1 more often than 0, you can account for that in the form of $\pi(0)$.

That is what essentially this is. Other than that the rule is similar. If all the $\pi(i)$'s are the same, let us say in our case it is half and half, then this particular thing can be removed because for all of these func $p(y|i)$, the same multiplier exists. So, you can get rid of this. So, maximum likelihood is the same of minimum probability of error also known as MAP, maximum a posteriori under the condition of equal priors that is when all symbols are equally likely. Now, the minimum probability of error is also the same as the maximum a posteriori rule which is the MAP rule because it maximizes the probability that H_i occurs given that y is observed.

So, that is something which you can verify. The final thing which we want to just briefly dwell upon is the aspect of irrelevant statistics. Sometimes when we have Y and it is complicated, you know it is a complicated thing to process because it has been much of unnecessary or extraneous information. The question which we have is can we decompose the statistic into Y_1 , Y_2 where only Y_1 is relevant and Y_2 is not relevant. Let us take a simple example.

Let us say that our symbols are like a PAM system, everything is real. But let us say that the noise is going to take it to some place like this. There is complex noise. So, there is real noise and imaginary noise. So, can we not decompose Y into two parts? So, in my example we will just take the real part and imaginary part. The imaginary part of noise is irrelevant because it does not affect your decision on what is actually sent on the real axis.

In a technically speaking for M -merry hypothesis testing where you want to check which one of M hypotheses were sent, if you decompose your Y into Y_1 , Y_2 , you can decompose it in various ways like break into a vector or you know subtract or add or something like that. If you find some part that is irrelevant to the conditional distribution of Y_2 given Y_1 , then H_i is independent of i . That is if it is not going to aid you in deciding which hypothesis is sent, then

$$\mathrm{p}(y_2|y_1, i)=\mathrm{p}(y_2|y_1)$$

for all i . That is intuitively speaking if you have some extraneous information that is not relevant, then you can ignore it. We will look at this more technically when we deal with projecting noise onto the signal space and we will show that that part that is not projected is irrelevant.

That is something which we will see in the next few lectures. So, to summarize what we have learnt over the past few lectures, practical communication systems have some kind of random effects such as noise and these necessitate optimal detection techniques. Inference, inferences from meaning concluding what was sent from the received signals may be complicated and it may be easier to take them to a different space or convert them. In this situation, we will use the relevant space with sufficient statistics and that is something we will see. Hypothesis testing to obtain the maximum likelihood or minimum error decision on transmitted symbols is something that is a tool that we will use in order to recover our data and also find out how much error there is while recovering the data. These are some aspects that we will see in the next lecture followed by implementation of these on GNU radio as well. Thank you. Thank you.