1 GSW… Signal Transforms

The basic idea of all signal transforms is to take a signal of interest, and represent it as a linear sum of a number of other signals. Something like[1]:

$$y(t) \approx \sum_i a_i \, x_i(t) \tag{0.1}$$

That might not sound very useful at first, but if these 'other signals' have some particularly simple mathematical properties, then it's often easier to convert a given signal into a sum of other signals, perform whatever linear mathematical operation[2] is required on the 'other signals', and then add up all the results. That can be easier than trying to do the same operation on the original signal.

1.1 A Very Simple Example

I'll start with just about the simplest example I can think of. Consider a set of four rectangular pulses, each with a width of one-quarter, a height of two, equally spaced between zero and one, as shown below:

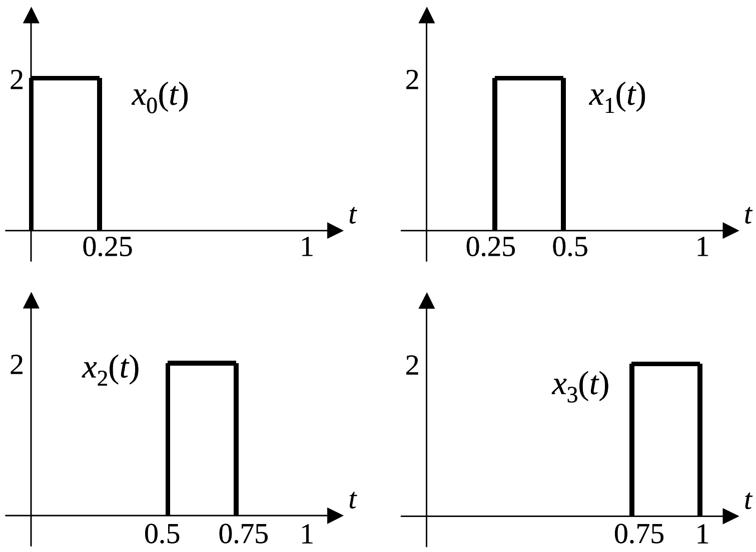

Figure 1 Four Rectangular Basis Functions

[1] I'll use an 'approximately equal' sign here, since it's sometimes not possible to express a given signal $y(t)$ exactly in terms of a linear sum of a chosen set of other signals.

[2] It has to be a linear operation. If you express a signal in terms of a sum of other signals, and then perform a non-linear operation (such as adding a constant, or taking the square) on each of these other signals, and then add the results back together, you don't get the same answer as if you performed the operation on the original signal. Simple example: express a signal $x(t) = 2$ in terms of the sum of two signals $x_1(t) = 1$ and $x_2(t) = 1$. Square both of these signals, and you get $x_1^2(t) = 1$ and $x_2^2(t) = 1$; add these together, and you get $x_1^2(t) + x_2^2(t) = 2$. However, $x^2(t) = 4$. Not the same thing at all. If the operation were linear (multiplying by a constant, integrating, differentiating or delaying by a fixed time), then we'd get the same answer. Fortunately, a lot of the most interesting signal processing algorithms are composed of linear operations, which is just as well. If they weren't, the whole point of doing signal transforms would fall apart.

Call these functions $x_i(t)$, where i ranges from 0 to 3, and we could write:

$$x_i(t) = \begin{cases} 2 & i/4 < t < (i+1)/4 \\ 0 & elsewhere \end{cases} \tag{0.2}$$

The basic idea of signal transforms is to be able to express one function in terms of a linear sum of a set of other functions (known as the *basis signals* or *basis functions*: here, these are the four rectangles). Consider the function $y(t) = t^3$. How do you express this as the sum of a linear set of these rectangles?

Well, clearly, you can't. Not exactly, anyway. Any linear combination of this set of rectangles is going to look a bit like a staircase: the value is going to be constant for each time step of 0.25 seconds. However, we can get very close: what we need to know is how to work out the heights of these rectangles that gives the closest possible approximation to the smooth curve $y(t) = t^3$. That means working out the coefficients a_0, a_1, a_2 and a_3 in the expression:

$$y(t) \approx \sum_i a_i\, x_i(t) \tag{0.3}$$

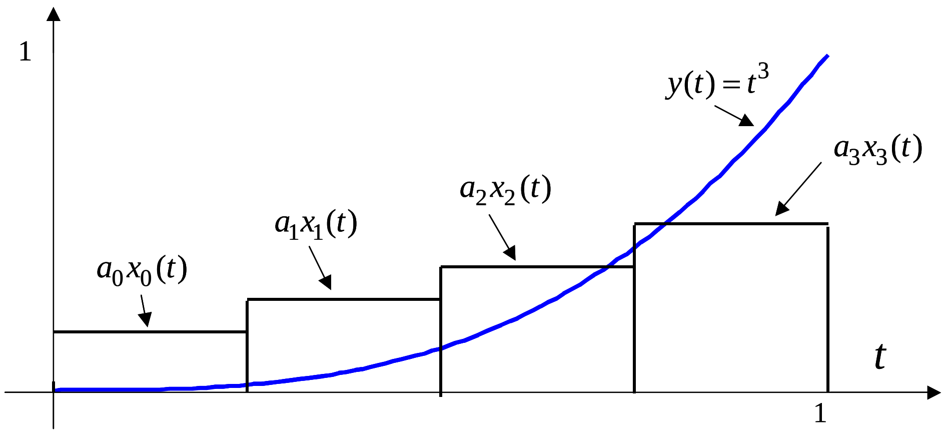

Figure 2 Matching the Basis Functions to the Curve

that give this best fit. As usual in signal processing, we'll define 'best fit' as 'minimising the mean square error'. In the general case this means minimising:

$$\mathrm{E}\left\{e^2(t)\right\} = \frac{1}{(t_2 - t_1)} \int_{t_1}^{t_2} \left(y(t) - \sum_{i=1}^{3} a_i x_i(t) \right)^2 dt \tag{0.4}$$

where the range over which we're trying to match the signal $y(t)$ is from t_1 to t_2. To minimise this mean-square error, we just differentiate the expression for the mean-square error with respect to each value of a_j, and look for a turning point. The coefficients a_j are not functions of time, so we can just do the differentiation inside the integral:

$$\frac{d\,\mathrm{E}\left\{e^2(t)\right\}}{da_j} = \frac{1}{(t_2 - t_1)} \int_{t_1}^{t_2} 2\left(y(t) - \sum_{i=1}^{3} a_i x_i(t) \right)\left(-x_j(t)\right) dt \tag{0.5}$$

and for a turning point (in this case a minimum), we set this to zero, which gives:

$$\int_{t_1}^{t_2} 2y(t)x_j(t)\,dt = \int_{t_1}^{t_2} 2\left(\sum_{i=1}^{3} a_i x_i(t)\right)x_j(t)\,dt \tag{0.6}$$

Expanding the sum into one component with $i = j$ and another component with all the others, gives

$$\int_{t_1}^{t_2} 2y(t)x_j(t)\,dt = 2a_j \int_{t_1}^{t_2} x_j(t)x_j(t)\,dt + 2\sum_{i \neq j} a_i \int_{t_1}^{t_2} x_i(t)x_j(t)\,dt \tag{0.7}$$

This is where two very useful properties of the basis functions make life much easier for us. The functions I chose (the four rectangles) are *orthogonal*, and *orthonormal*.

1.1.1 Orthonormal Basis Functions

The four rectangles don't overlap at all. Therefore, if I multiply any two of them together, I'm just going to get zero at all times. More generally, this means that:

$$\int_{t_1}^{t_2} x_i(t)x_j(t)dt = 0 \tag{0.8}$$

provided $i \neq j$. Sets of functions that have this property are known as *orthogonal functions*.

Now, you might be wondering why I chose these rectangular basis functions to have a height of two. The answer is that I can then write:

$$\int_{t_1}^{t_2} x_i(t)x_i(t)dt = 1 \tag{0.9}$$

The product of one of these rectangle with itself is just the same rectangle with a height of four, and the area under a rectangle of height four and width one-quarter is just one. Sets of orthogonal basis functions that all have this property are known as *orthonormal* functions, and they have the property:

$$\int_{t_1}^{t_2} x_i(t)x_j(t)dt = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \tag{0.10}$$

It's possible to use sets of functions that are not orthonormal, but using orthonormal functions makes the use of these functions very much easier.

1.1.2 Finding the Best Linear Combination

Using an orthonormal set of basis functions means that equation (0.7) can be simplified to:

$$\int_{t_1}^{t_2} 2y(t)x_j(t)\,dt = 2a_j \tag{0.11}$$

and therefore:

$$a_j = \int_{t_1}^{t_2} y(t) x_j(t)\, dt \tag{0.12}$$

It's a very simple way to work out the optimum linear combination of the basis vectors to use (optimum in the 'minimise mean square error' sense, as usual.)

For the example here, we get:

$$a_0 = \int_0^{0.25} 2t^3\, dt = \left[\frac{t^4}{2}\right]_0^{0.25} = \frac{1}{512} \tag{0.13}$$

$$a_1 = \int_{0.25}^{0.5} 2t^3\, dt = \left[\frac{t^4}{2}\right]_{0.25}^{0.5} = \frac{1}{32} - \frac{1}{512} = \frac{15}{512} \tag{0.14}$$

$$a_2 = \int_{0.5}^{0.75} 2t^3\, dt = \left[\frac{t^4}{2}\right]_{0.5}^{0.75} = \frac{81}{512} - \frac{1}{32} = \frac{65}{512} \tag{0.15}$$

$$a_3 = \int_{0.75}^{1} 2t^3\, dt = \left[\frac{t^4}{2}\right]_{0.75}^{1} = \frac{1}{2} - \frac{81}{512} = \frac{175}{512} \tag{0.16}$$

so the best possible linear combination of our four basis functions is:

$$y_e(t) = \frac{1}{512} x_0(t) + \frac{15}{512} x_1(t) + \frac{65}{512} x_2(t) + \frac{175}{512} x_3(t) \tag{0.17}$$

Plot these on a graph (along with the exact function $y(t) = t^3$) and it looks like this:

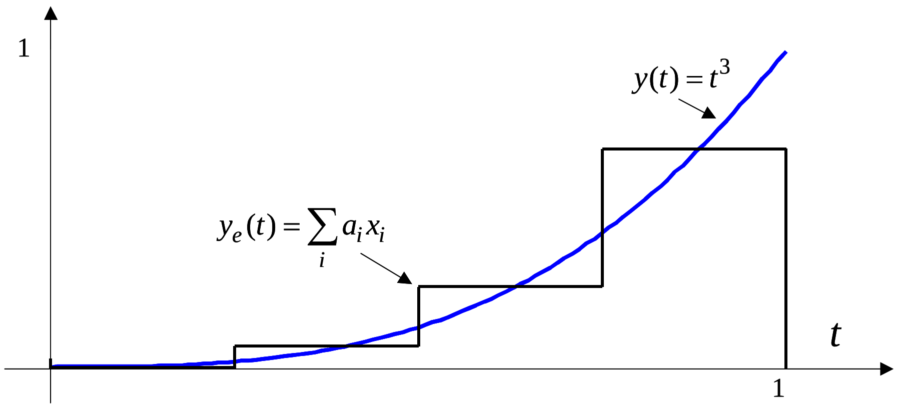

Figure 3 Exact and Approximate Forms of $y(t) = t^3$

Note that the approximate form, $y_e(t)$ is accurate over a restricted range only: in this case from zero to one. Trying to use this approximate expression outside this range can lead to very large errors (in this case, the approximate expression has a value of zero for all values of t greater than one, which is clearly a very bad approximation to t^3).

1.1.3 The Expectation of the Error in the Approximation

With any approximation technique, it's very useful to have some measure of how good an approximation it is, and yet again the usual method is to quantify this in terms of the mean square error (after all, that's the error we were trying to minimise in the first place). In this case, this gives:

$$\mathrm{E}\left\{e^2(t)\right\} = \frac{1}{(t_2 - t_1)} \int_{t_1}^{t_2} \left(y(t) - \sum_{i=1}^{3} a_i x_i(t) \right)^2 dt \tag{0.18}$$

Multiply out the bracket, and we get:

$$\begin{aligned} \mathrm{E}\left\{e^2(t)\right\} &= \frac{1}{(t_2 - t_1)} \int_{t_1}^{t_2} \left(y(t) - \sum_i a_i x_i(t) \right)^2 dt \\ &= \frac{1}{(t_2 - t_1)} \left(\int_{t_1}^{t_2} y^2(t)\,dt - 2\sum_i a_i \int_{t_1}^{t_2} y(t)\,x_i(t)\,dt + \sum_k \sum_i a_i a_k \int_{t_1}^{t_2} x_i(t)\,x_k(t)\,dt \right) \end{aligned} \tag{0.19}$$

Once again, the choice of orthonormal basis functions lets us simplify this expression considerably, since all the integrals in the final (double) summation in which $i \neq k$ are zero, and for $i = k$ they are one. Furthermore, we've got the result that:

$$\int_0^1 y(t) x_i(t)\,dt = a_i \tag{0.20}$$

since that's how we worked out a_i in the first place. Substituting these results in equation (0.19) gives:

$$\begin{aligned} \mathrm{E}\left\{e^2(t)\right\} &= \frac{1}{(t_2 - t_1)} \left(\int_{t_1}^{t_2} y^2(t)\,dt - 2\sum_i a_i{}^2 + \sum_i a_i{}^2 \right) \\ &= \frac{1}{(t_2 - t_1)} \left(\int_{t_1}^{t_2} y^2(t)\,dt - \sum_i a_i{}^2 \right) \end{aligned} \tag{0.21}$$

For the example here, that gives:

$$\begin{aligned} \mathrm{E}\left\{e^2(t)\right\} &= \int_0^1 t^6\,dt - \left(\frac{1}{512^2} + \frac{15^2}{512^2} + \frac{65^2}{512^2} + \frac{175^2}{512^2} \right) \\ &= \left[\frac{t^7}{7} \right]_0^1 - \frac{35076}{262144} = \frac{1}{7} - \frac{8769}{65536} = \frac{4153}{458752} = 0.0091 \end{aligned} \tag{0.22}$$

That's quite a small mean-square error. You can produce a reasonable good approximation to the $y(t) = t^3$ graph between zero and one by adding together suitable amounts of our four orthonormal basis functions, and you can work out the best possible linear combination of them to use very easily.

1.1.4 Warning: The Expectation of Error in the Answers

The whole point of doing signal transforms is that some mathematical operations are easier to perform on the basis functions than on the original function, and can provide good approximations to the right answer.

For example, suppose you wanted to integrate the function $y = t^3$ between zero and one. You could do this analytically:

$$\int_0^1 t^3\,dt = \left[\frac{t^4}{4}\right]_0^1 = \frac{1}{4} \tag{0.23}$$

or by just adding up the areas of the four rectangular basis functions:

$$\int_0^1 t^3\,dt \approx \sum_{i=0}^{3} a_i \frac{1}{2} = \frac{1}{2}\left(\frac{1+15+65+175}{512}\right) = \frac{1}{4} \tag{0.24}$$

in this case, the answer is perfectly correct, and avoids having to do any integration at all (after you've worked out the coefficients a_i, that is).

Of course, just because $y_e(t)$ is a good approximation to the original signal $y(t)$ doesn't mean that the results of any linear operation on the transformed signal will be a good approximation to the results of the linear operation on the original signal. For example, what if you wanted to differentiate the function $y = t^2$ rather than integrate it? The differential of $y = t^3$ is $y = 3t^2$, a smooth function. Differentiate the series of rectangles, and the result is zero everywhere except at five times (0, 0.25, 0.5, 0.75 and 1), where the gradient is infinite.

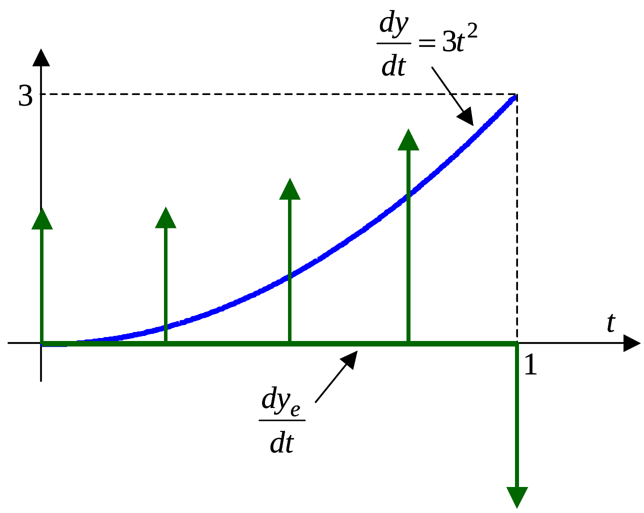

Figure 4 Errors in Differentiating a Transform

Nowhere near right: you have to be a little careful about what basis functions you use, and what you do with them.

1.2 A Slightly Different Example

Consider the following three functions of t, in the range between $t = 0$ and $t = 1$:

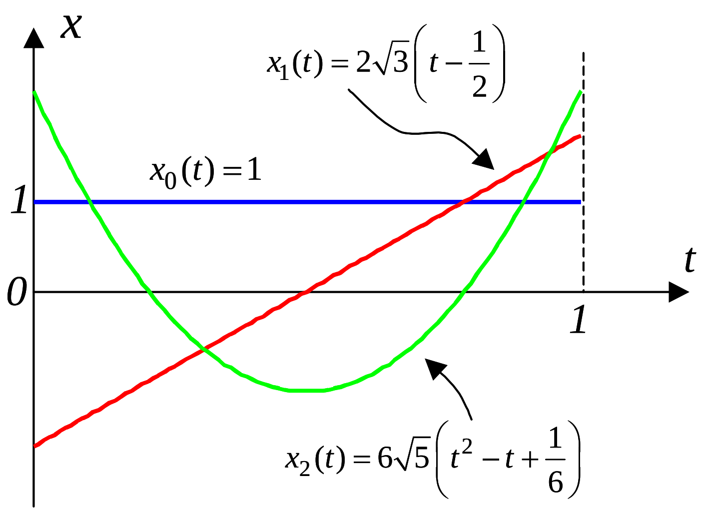

Figure 5 Three Orthonormal Functions in the Range 0 to 1

These three functions are orthonormal, in other words they obey the equation:

$$\int_0^1 x_i(t)x_j(t)dt = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \tag{0.25}$$

Now, suppose we had some other function of time that could be expressed in the form:

$$y(t) = A + Bt + Ct^2 \tag{0.26}$$

where A, B and C are constants. We could follow a similar procedure as before to find the optimum amounts of the three functions to add together to approximate $y(t)$, but in this case there's a short-cut. The function $y(t)$ is the sum of a constant term, a term proportional to t, and a term proportional to t^2. So are the three basis functions.

That suggests that by equating terms in the co-efficients of t, we should be able to express the function $y(t)$ in terms of the three basis functions exactly. All we need to do is solve the equations:

$$A = a_0 - \sqrt{3}a_1 + \sqrt{5}a_2 \tag{0.27}$$

$$B = 2\sqrt{3}a_1 - 6\sqrt{5}a_2 \tag{0.28}$$

$$C = 6\sqrt{5}a_2 \tag{0.29}$$

and these are simple to solve, giving:

$$a_2 = \frac{C}{6\sqrt{5}} \tag{0.30}$$

$$a_1 = \frac{B+C}{2\sqrt{3}} \tag{0.31}$$

$$a_0 = A + \frac{B}{2} - \frac{C}{3} \tag{0.32}$$

In this case, with this particular function $y(t)$, and this choice of orthonormal basis functions, there is no error in the signal transform. The 'best fit' is perfect.

1.2.1 Complete Sets of Orthogonal Functions

It is sometimes possible to find a set of orthogonal basis functions that ensures that the 'best fit' is perfect for a wide range of different input functions y(t), preferably including all of the input functions of interest. A set of orthogonal functions with this property is called a *complete set*.

One such complete set is an infinite number of infinitely-thin rectangles, equally spaced in time. This is just the limiting case of the situation illustrated above where the time period from zero to one second was split into four rectangles, although we'd now have an infinite number of rectangles. You can express any function in terms of an infinite number of infinitely thin rectangles in this way: you just have to set the height of the small rectangle equal to the value of the function at that time.

The limiting case of a rectangle with a width of Δt and height of $1 / \Delta t$ (and therefore an area of one) as Δt tends to zero is the delta function, such a rectangle at time $t = \tau$ is written as $\delta(t - \tau)$.

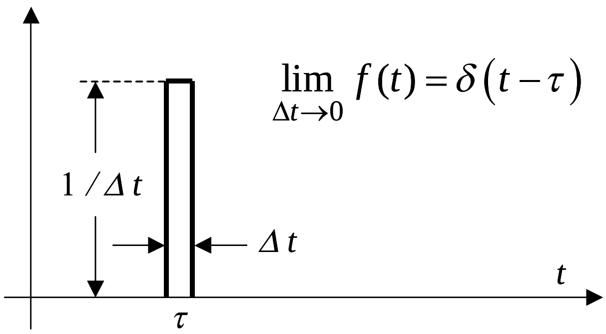

Figure 6 The Delta Function

Any function $y(t)$ can be expressed in terms of these basis functions by multiplying each delta function by $y(t)$

$$y(t) = \lim_{\delta t \to 0} \sum y(\tau)\delta(t-\tau)d\tau = \int_{-\infty}^{\infty} y(\tau)\delta(t-\tau)d\tau \tag{0.33}$$

1.3 Orthogonal, but not Orthonormal

There is one very important signal transform[3] in which the basis functions are not orthonormal, although they are still orthogonal. In this case we can write:

$$\int_{t_1}^{t_2} x_i(t)x_j(t)dt = \begin{cases} E_i & i = j \\ 0 & i \neq j \end{cases} \tag{0.34}$$

where:

$$E_i = \int_{t_1}^{t_2} x_i(t)x_i(t)dt \tag{0.35}$$

and E_i can be thought of as the energy of the signal $x_i(t)$ in the time period between t_1 and t_2. (That's why I chose the letter E to represent it.)

Using an orthogonal (but not orthonormal) set of basis functions means that equation (0.7) can be simplified to:

$$\int_{t_1}^{t_2} 2y(t)x_j(t)\,dt = 2E_j a_j \tag{0.36}$$

and therefore:

$$a_j = \frac{1}{E_j}\int_{t_1}^{t_2} y(t)x_j(t)\,dt \tag{0.37}$$

Working out the minimum mean-square error in this case, we get:

$$\begin{aligned} \mathrm{E}\left\{e^2(t)\right\} &= \frac{1}{(t_2 - t_1)}\int_{t_1}^{t_2}\left(y(t) - \sum_i a_i x_i(t)\right)^2 dt \\ &= \frac{1}{(t_2 - t_1)}\left(\int_{t_1}^{t_2} y^2(t)\,dt - \sum_i E_i a_i^2 \right) \end{aligned} \tag{0.38}$$

1.4 Working with Complex Basis Functions

Not all signals are real. If we have a complex signal, then we can still use complex basis signals, we just have to change the derivations above slightly. The problem is that we're no

[3] The Fourier transform: perhaps the most important of them all.

longer just looking for the minimum mean square error, we're looking for the minimum mean square of the absolute value of the error[4]. For any complex number z, this is given by:

$$|z|^2 = z\,z^* \tag{0.39}$$

so here, we're trying to minimise:

$$\mathrm{E}\left\{e^2(t)\right\} = \frac{1}{(t_2 - t_1)} \int_{t_1}^{t_2} \left(y(t) - \sum_i a_i x_i(t) \right) \left(y(t) - \sum_i a_i x_i(t) \right)^* dt \tag{0.40}$$

Multiply this out, and we get:

$$\mathrm{E}\left\{e^2(t)\right\} = \frac{1}{(t_2 - t_1)} \begin{pmatrix} \int_{t_1}^{t_2} y(t) y^*(t)\,dt - \sum_i a_i^* \int_{t_1}^{t_2} y(t) x_i^*(t)\,dt \\ -\sum_i a_i \int_{t_1}^{t_2} y^*(t) x_i(t)\,dt + \sum_i \sum_j a_i a_j^* \int_{t_1}^{t_2} x_i(t) x_j^*(t) dt \end{pmatrix} \tag{0.41}$$

For this to work out easily, we'll have to slightly modify our definition of *orthogonal*, to:

$$\int_{t_1}^{t_2} x_i(t) x_j^*(t) dt = 0 \qquad i \neq j \tag{0.42}$$

and if we can find a set of basis functions with this property, then we can simplify the last term to:

$$\sum_i \sum_j a_i a_j^* \int_{t_1}^{t_2} x_i(t) x_j^*(t) dt = \sum_i a_i a_i^* E_i \tag{0.43}$$

where E_i is the energy in the signal $x_i(t)$ between t_1 and t_2, defined by:

$$E_i = \int_{t_1}^{t_2} x_i(t) x_i^*(t)\,dt \tag{0.44}$$

and note that by definition, E_i is real.

Since the first term is not a function of the coefficients a_i, we don't need to consider differentiating this term when finding the optimum coefficients. What we do need to do is consider the real and imaginary parts of the coefficients separately. If we write:

[4] When dealing with complex numbers, it's not obvious how to minimise a square, since the square of a complex number is complex. For example, is 6 + 8*j* bigger than 9*j* or not? The usual way to avoid any confusion is to try and minimise the absolute value of the square of the error, corresponding to minimising the distance on the complex plane between the two points. The distance between two points is a well-defined real scalar quantity, and it's obvious when that's getting smaller.

$$a_k = u_k + jv_k \tag{0.45}$$

where u_k and v_k are both real scalar numbers, and the real and imaginary parts of the coefficients respectively, then for a minimum mean square absolute error, we need to differentiate the error with respect to both u_k and jv_k.

First, differentiating equation (0.41) with respect to u_k gives:

$$\frac{\partial \mathrm{E}\left\{e^2(t)\right\}}{\partial u_k} = \frac{1}{(t_2 - t_1)}\left(-\int_{t_1}^{t_2} y(t)x_k^*(t)\,dt - \int_{t_1}^{t_2} y^*(t)\,x_k(t)\,dt + 2u_k E_k \right) \tag{0.46}$$

and then with respect to jv_k gives:

$$\frac{\partial \mathrm{E}\left\{e^2(t)\right\}}{\partial jv_k} = \frac{1}{(t_2 - t_1)}\left(\int_{t_1}^{t_2} y(t)x_k^*(t)\,dt - \int_{t_1}^{t_2} y^*(t)\,x_k(t)\,dt - 2\,jv_k E_k \right) \tag{0.47}$$

and setting these both to zero gives the two equations:

$$\begin{aligned} u_k &= \frac{1}{2E_k}\left(\int_{t_1}^{t_2} y(t)x_k^*(t)\,dt + \int_{t_1}^{t_2} y^*(t)\,x_k(t)\,dt \right) \\ jv_k &= \frac{1}{2E_k}\left(\int_{t_1}^{t_2} y(t)x_k^*(t)\,dt - \int_{t_1}^{t_2} y^*(t)\,x_k(t)\,dt \right) \end{aligned} \tag{0.48}$$

and adding these together gives the result we want:

$$a_k = u_k + jv_k = \frac{1}{E_k}\int_{t_1}^{t_2} y(t)x_k^*(t)\,dt \tag{0.49}$$

Similarly, the mean value of the error is now given by:

$$\begin{aligned} \mathrm{E}\left\{e^2(t)\right\} &= \frac{1}{(t_2 - t_1)}\left(\begin{aligned} &\int_{t_1}^{t_2} y(t)y^*(t)\,dt - \sum_{i=1} a_i^* \int_{t_1}^{t_2} y(t)a_i^* x_i^*(t)\,dt \\ &\qquad - \sum_i a_i \int_{t_1}^{t_2} y^*(t)\,a_i x_i(t)\,dt + \sum_i a_i a_i^* E_i \end{aligned} \right) \\ &= \frac{1}{(t_2 - t_1)}\left(\int_{t_1}^{t_2} y(t)y^*(t)\,dt - \sum_{i=1} a_i^* a_i E_i - \sum_i a_i a_i^* E_i^* + \sum_i a_i a_i^* E_i \right) \end{aligned} \tag{0.50}$$

However, E_i is a real quantity by definition, so $E_i = E_i^*$, and that gives the simple expression for the ‘mean square’ error:

$$\mathrm{E}\left\{e^2(t)\right\} = \frac{1}{(t_2 - t_1)}\left(\int_{t_1}^{t_2} y(t)\,y^*(t)\,dt - \sum_{i=1} a_i a_i^* E_i\right)$$
$$= \frac{1}{(t_2 - t_1)}\left(\int_{t_1}^{t_2} y(t)\,y^*(t)\,dt - \sum_{i=1} |a_i|^2\, E_i\right) \quad (0.51)$$

1.5 Problems

1) Try expressing $y(t) = t^3$ in terms of the second series of orthonormal functions (those used in section 1.2). What is the mean-square error in this case?

2) Consider the three functions $x_1(t) = 1$, $x_2(t) = \exp(jt)$ and $x_3(t) = \exp(2jt)$. Show that they are orthogonal in the range from zero to 2π, and find the best way to express the function $y = x^2$ in terms of a linear sum of these functions.

What is the expectation value of the mean square absolute error in this case?