Observation Theory

Script V4PB –Non-linear least-squares

Now that you have seen some examples of non-linear estimation problems and the associated observation equations, in this video we will introduce the principle of non-linear least-squares estimation

So the question is how to estimate the unknown parameters x if the model is non-linear.

The system of observation equations will now look like this if we have m observations.

You can see that each observable is a function of the n by 1 parameter vector x.

The principle of non-linear estimation will first be introduced for the most simple situation where we have only 1 observation which is a non-linear function of 1 unknown parameter x.

Hence, y is observed , and x is unknown and a() is the non-linear function.

This is an inverse problem, and if function a is not invertible, it cannot be directly solved.

The trick we will apply here is to approximate a() with a linear function, assuming that this is a good approximation in the neighbourhood of x.

Let's see how that works.

First we will make an initial guess of x.

Here our guess is called x0.

This guess x0 can be inserted in the function a() since this is the so-called forward model.

In this way we obtain the function value at x0.

It is at this point that we approximate the function a() by a linear function as follows.

We take the derivative of a() to find the gradient at x0, which is the slope of the tangent line at x0 – the blue line shown here.

So, the idea is that the tangent line is an approximation of function a.

Remember that the observation we had was given by y.

Let's briefly look at the tangent line only .

By definition the gradient is equal to Delta-y over Delta-x as visualised here.

The equation is shown in the top left.

In the figure you can see that Delta y is equal to y-a(x0), which is the difference of the two function values we have.

If we use this result in the top equation we get this result.

We now have that Delta y is a LINEAR function of Delta x, so we can easily calculate Delta x.

And what does that give us finally?

It is a new approximation for x, called x1 being equal to our initial guess x0 plus Delta x.

And this x1 could be our "estimate" of x… here you see the result with the actual function a() shown as well.

Note that this approximation with a linear function is in fact the so-called first-order Taylor series approximation.

If you are interested, a short note on Taylor's theorem is included in the reading materials.

Having the estimate for x, we may ask ourselves though: how good is this estimate?

What we can do is to solve the forward model again to get the function value y1 at x1.

We can then see how y1 deviates from the observed y.

What we could do now is to repeat the whole procedure, but now with x1 as our new guess.

Again we approximate function a() by a linear function, in this case the tangent line at x1.

This will then give us x2 which is much closer to the true x.

This process can be continued.

Hence the procedure is as follows.

We start with our initial guess x_0, based on the linear approximation we find Delta x1, which provides us with a new approximation of our unknown parameter.

The new x_1 is then used to get a better approximation x_2 Again the same equations, but now x0 is replaced by x1, and from that we get x2.

And we continue and continue…

It is thus an iterative procedure.

But after how many iterations do we stop?

We will do that once Delta x is very small our final estimate of the unknown x will thus be equal to xi.

Note that in this procedure it is straightforward to calculate the Delta x in each step, since we are only dealing with a function of 1 variable.

Delta-x at step I can thus be calculated with this equation, in which we will denote the numerator with delta y

Let x now be a vector with n parameters.

The first order linear approximation is then given by this equation Note that Delta x is here an n by 1 vector.

The vector highlighted here is the gradient vector with partial derivatives to the individual elements of vector x.

This gradient vector is also referred to as the Jacobian.

In this way again we expressed Delta y as a linear function of Delta x.

We call this the LINEARIZED observation equation.

This means that we can now apply linear least-squares estimation to estimate Delta x from Delta y.

The only problem in this case is that we have one observable and n unknowns.

Therefore it is time to get back to our original problem where we have m functions of n unknown parameters.

Similarly as on the previous slide, you can set up the linearized observation equation for each observable.

This gives you then again a linear system of m observation equations in n unknowns, which can be solved with linear least-squares estimation.

The iterative procedure as we presented for the scalar case is still applicable and will be presented in the reading materials.

In the next video we will look at some properties of non-linear least-squares estimation.

But let's first work on some examples in the exercises.