Concepts of “Least Squares”

Hello everyone! Welcome to the third week of the course: “Observation theory: estimating the unknown”. Week three is devoted to the principle of Least Squares Estimation.

In the first video-lecture, we are discussing the main principle behind the Least Squares Estimation.

To get started, let’s look back to our video of the sea level rise.

As we saw in that video-clip, in order to find, or to “estimate”, the rate of the sea level rise from the tide gauge observations, the problem is to find a line that is the “best” fit to the measurements.

Before solving the problem for this real-data example, let’s look at a similar problem in a more simplified form.

Assume that instead of large number of observations, we have only five observations in time as visualised in this plot.

The question is:

Which line can better explain or model this data?

This line? This one? Or this? How do we find the closest line to this set of observations?

For this specific example of line fitting, as we saw in the previous week lecture, the functional model can be formulated as this system of observation equations, with observations in the vector y, the two unknown parameters in the vector x, and the design matrix, denoted by A. The design matrix explains the functional relation between observations and unknowns.

I want to remind you that due to the errors or uncertainties in the measurements, there is no “solution” for this system of five equations in two unknowns. I repeat: no “solution”.

In other words, we can not find even a single line that goes through all these points or observations simultaneously. Remember, if there is no error, and assuming the linear model is the correct model, then there is a single solution for this problem.

In this noise-free scenario, the system of equations is consistent, and in order to solve it, is enough to take just two of the equations, for example, these two, which are associated with these two points, and simply solve this system of two observations in two unknowns.

And since the system is consistent, any set of two equations, for example, these two, or these ones, provides exactly the same solution.

However, in practice, the system is inconsistent and we are dealing with this unsolvable system of equations. In fact any selected system of two equations in two unknowns provides a different solution than another combination, for example these two observations give this solution and these two provide yet another solution. We say that this system of equations is “inconsistent”.

As we saw, the source of this inconsistency is the errors in the measurements or the deviations in observations with respect to the assumed model, in this case: this straight line.

If we express the error in every observation by variable e, then we can add the errors to this plot and also to the system of equations.

Now, in this new system of equations, we have (m+n) unknown parameters: (n) original unknowns, and (m) number of new, and unknown, errors.

The number of unknowns (m+n) is larger than number of observations (m). So how many different solutions does this new system have? Let’s think about it.

An infinite number of solutions. Why? In fact, any line in this plane with a set of its associated errors is actually a plausible solution for this new system.

This one … this one … and even this one!

If I insert this vector of very large errors together with the estimated line parameters in the functional model equation, it gives us exactly, exactly the original observations in the vector y.

In summary, by adding the errors to the system of equations, we transformed the inconsistent system of equations with no solution into a consistent system, but now with an infinite number of solutions

But which solution is the best one? Which line is the ‘best’ line? In fact we find ourselves asking the same question that we posed at the beginning of this lecture. Which criteria do we use to select the best line?

Most of us intuitively choose the best line, as the closest line to the observations. Something like this, which goes through the middle of the noisy observations. This intuitive idea is in fact the main principle behind the Least Squares Estimation!