

COMP6247 - Kalman Filter

Luke McClure
lam3g17@soton.ac.uk
29573904

1 BACKGROUND

Kalman Filter is a method to estimate the parameters of a time series model in an online fashion. This is completed by estimating the parameters θ and covariance P and updating these as new information is revealed.

$$\theta(n|n-1) = \theta(n-1|n-1) \quad (1)$$

$$P(n|n-1) = P(n-1|n-1) + Q \quad (2)$$

$$\theta(n|n) = \theta(n|n-1) + k(n)e(n) \quad (3)$$

$$P(n|n) = (I - k(n)x(n)^T)P(n|n-1) \quad (4)$$

$$k(n) = \frac{P(n|n-1)x(n)}{R + x(n)^T P(n|n-1)x(n)} \quad (5)$$

2 CONVERGENCE OVER DIFFERENT INITIAL CONDITIONS

By standardising A , R , and Q , we can investigate how different random initial values of θ converge when using Kalman filtering. This allows us to build a general picture of how well Kalman filtering performs on a general series using these hyperparameters.

Over all samples shown in Figure 1, Kalman Filter was able to converge θ within 50 samples, or was already close to A by chance. This is fairly independent of how far the initial value of θ was from A suggesting that the gradient of convergence tends to get steeper with a larger difference between the two values. This makes sense as a larger difference between the values of A and θ results in larger updates being applied (due to a larger prediction error).

In samples a and b there is less of a convergence pattern as the error stays focused between 0.5 and 1.5. The starting parameters of θ in these instances were very close to A , so it acted as if the filter had already converged. Samples c, d and e show the classic convergence pattern with error starting high (5-7) and dropping very quickly within 50 samples to between 0 and 2.

It is interesting to note that in all the plots of $\theta[0]$ and $\theta[1]$ you can see similar patterns as the filters independently react to the same key features within the original AR series and are "surprised" by them.

Even when the error takes only about 50 samples to converge, it takes much longer for θ . In samples c, d and e the components of θ typically take at least 100 iterations to reach close to their respective component of A . With components of θ that start closer to their target of A , the plot of their change is a lot more noisy. Smaller variations show clearer within these plots and θ reacts seemingly more strongly to the key features of S that are being encountered by the filters on that iteration.

3 EFFECTS OF HYPERPARAMETERS ON CONVERGENCE

In Figure 2 the left column shows the effects of varying the signal noise variance R , and the right shows the effects of varying the process noise covariance Q . Both variations take the initial parameters from Figure 1 and vary the order of magnitude ± 1 . In both cases the other variable is left at the same value from Figure 1 and both A and θ are left constant. The same AR series S has been used over all experiments within this section.

3.1 Varying signal noise variance R

Signal noise covariance is the amount of noise from the measurement of the signal. Convergence speed is inversely proportional to the magnitude of R from this data, with experiment a taking around 100 samples to converge whilst experiment e converging almost instantly. There seems to be a slight stability correlation with the magnitude of R as well, with experiment a seeming to settle much more than experiment e when it has finally converged.

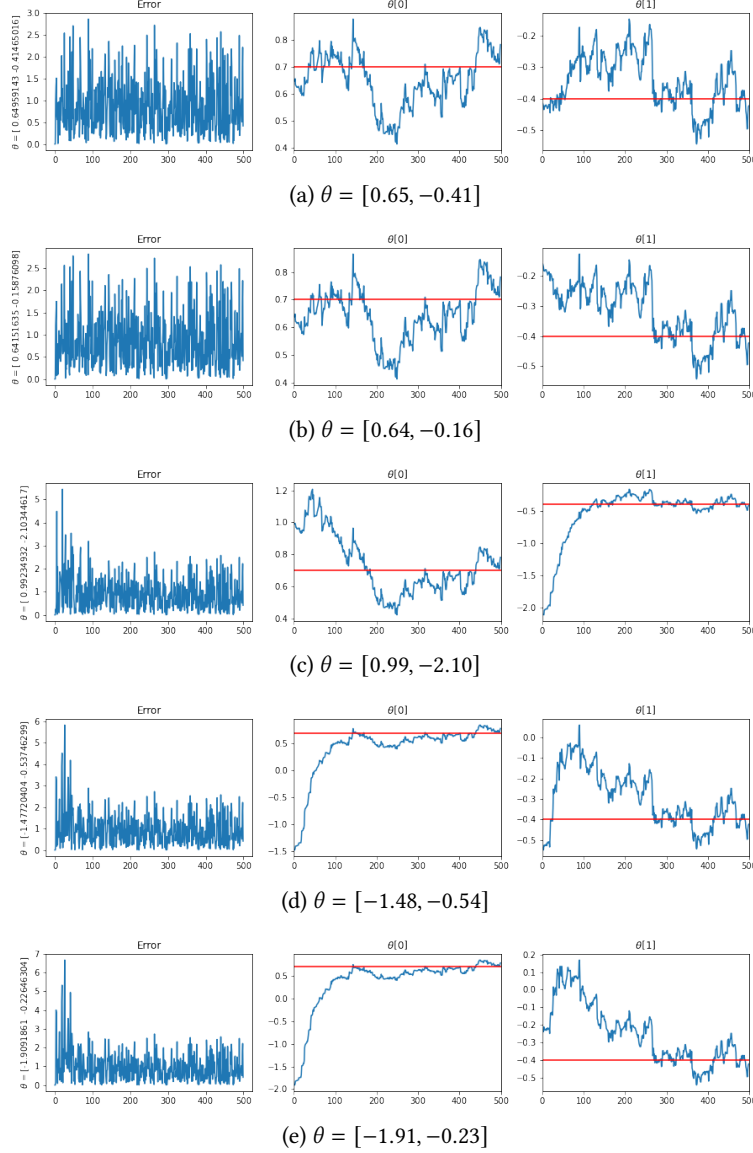
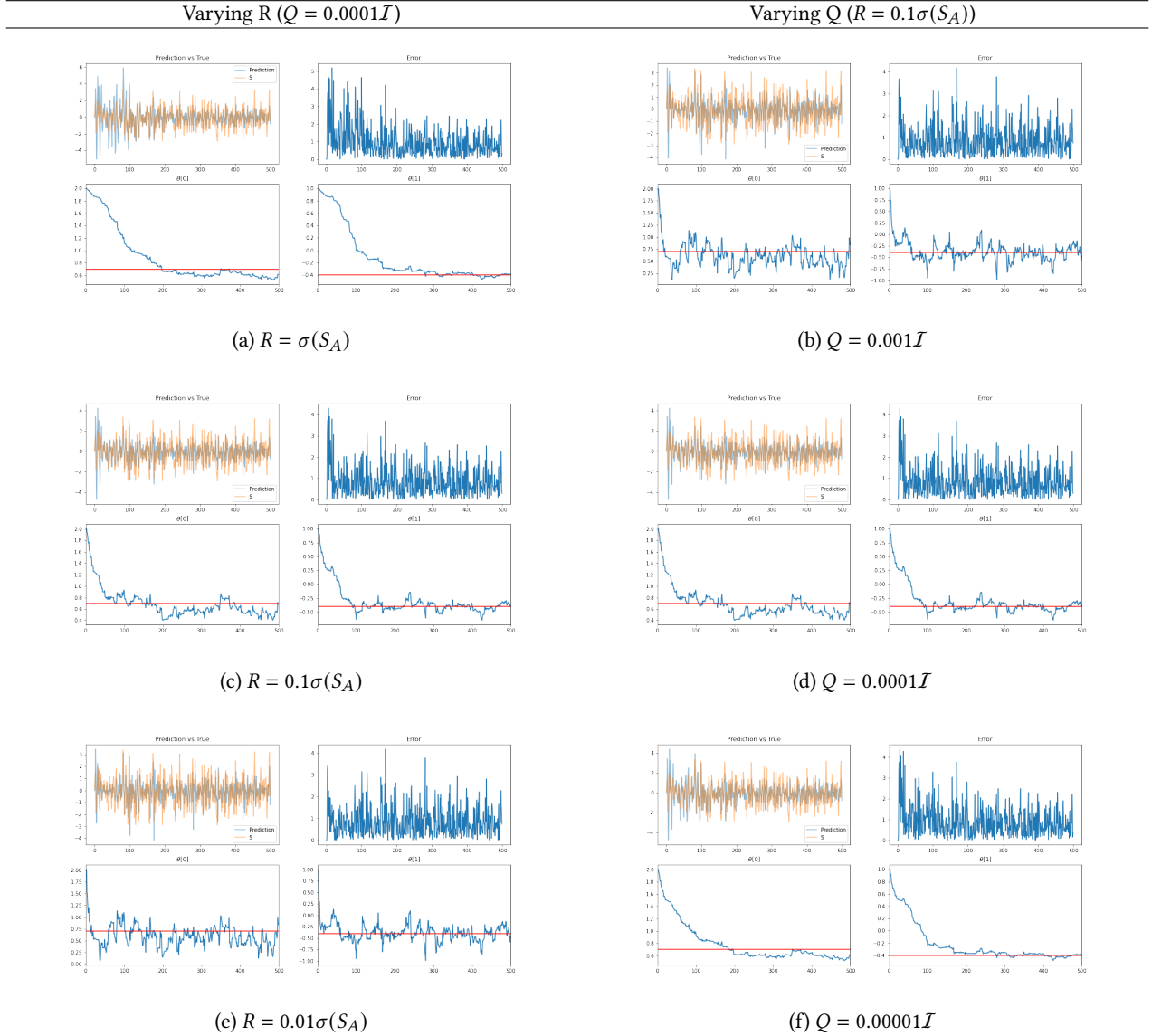


Figure 1: Varying initial starting conditions θ with $A = [0.7, -0.4]$, $R = 0.2\sigma(S_A)$, $Q = 0.0001I$

Looking at θ it is easy to confirm all of the hypotheses from the trend of error across experiments a, c and e. It is clear how θ takes much longer to converge in experiment a than in e, showing that R is inversely proportional to convergence speed. These plots also show how stability of θ is proportional to R , experiment e shows a very volatile search in the neighbourhood of A while in both experiments c and a when θ eventually converges on A it stays relatively close.

This makes sense when thinking about the role R has to play within Kalman Filtering. An increase of R indicates that the system should not trust the data as much, and so must take longer to eventually converge. This can be seen clearly in Equation 5 where R only appears in the denominator. The inverse relationship with stability is also a result of the inverse nature of R and $k(n)$, with a larger R resulting in smaller updates to θ . Settling on the parameters within θ over a longer time results in every sample not having such a pronounced impact, leading to more stability.

Figure 2: Varying hyper parameters with $A = [0.7, -0.4]$, $\theta = [2.0, 1.0]$

3.2 Varying process noise covariance Q

Process noise covariance is a measure of the noise within the signal itself, as Q gets smaller it takes longer for KF to converge on A but this convergence is more stable.

Varying process noise covariance has a similar but opposite effect from signal noise variance, there is a proportional relationship between Q and convergence speed. In this case the convergence is less pronounced in the error function compared to signal noise covariance, the only clear definition of lack of convergence in the error plots are the thicker portions at the immediate start of the plot. All of these non-converged sections end by 50 iterations, but they appear longer with a smaller Q .

Speed of θ convergence is proportional to Q . The relationship between convergence speed vs stability has the same opposing nature in both R and Q , where as the convergence speed decreases the eventual convergence is more stable or vice versa.

Hyperparameter	Error	Stability
R	Inversely proportional	Proportional
Q	Proportional	Inversely proportional

Figure 3: Summary of hyperparameters proportionality with convergence speed

4 CONSTANT VS TIME VARYING AR PROCESSES

Kalman filtering works not only with stationary AR's, but also time varying AR's where the parameters A shift with time according to defined functions.

To investigate the KF on both these, I used the same values for A to generate a stationary AR, and as a in a time varying AR with Equation 7. The only difference between each output from the same category is the randomly generated excitation.

$$A[0] = a[0] + \cos(2 * \pi * n) \quad (6)$$

$$A[1] = a[1] + 1.4 * \sin(6 * \pi * n) \quad (7)$$

Looking at Figure 4, KF is able to adapt and converge to the time varying AR as fast as the stationary AR. Within experiment b the Kalman Filter quickly picks up of the underlying function in $\theta[1]$ but struggles to fully match the function in $\theta[0]$, although there appears to be a hint of the cosine wave within $\theta[0]$ albeit slightly lagged behind the $A[0]$. The θ convergences seem to be slightly faster in the experiment b too but this is up for interpretation in $\theta[0]$, $\theta[0]$ is matched much better in experiment a.

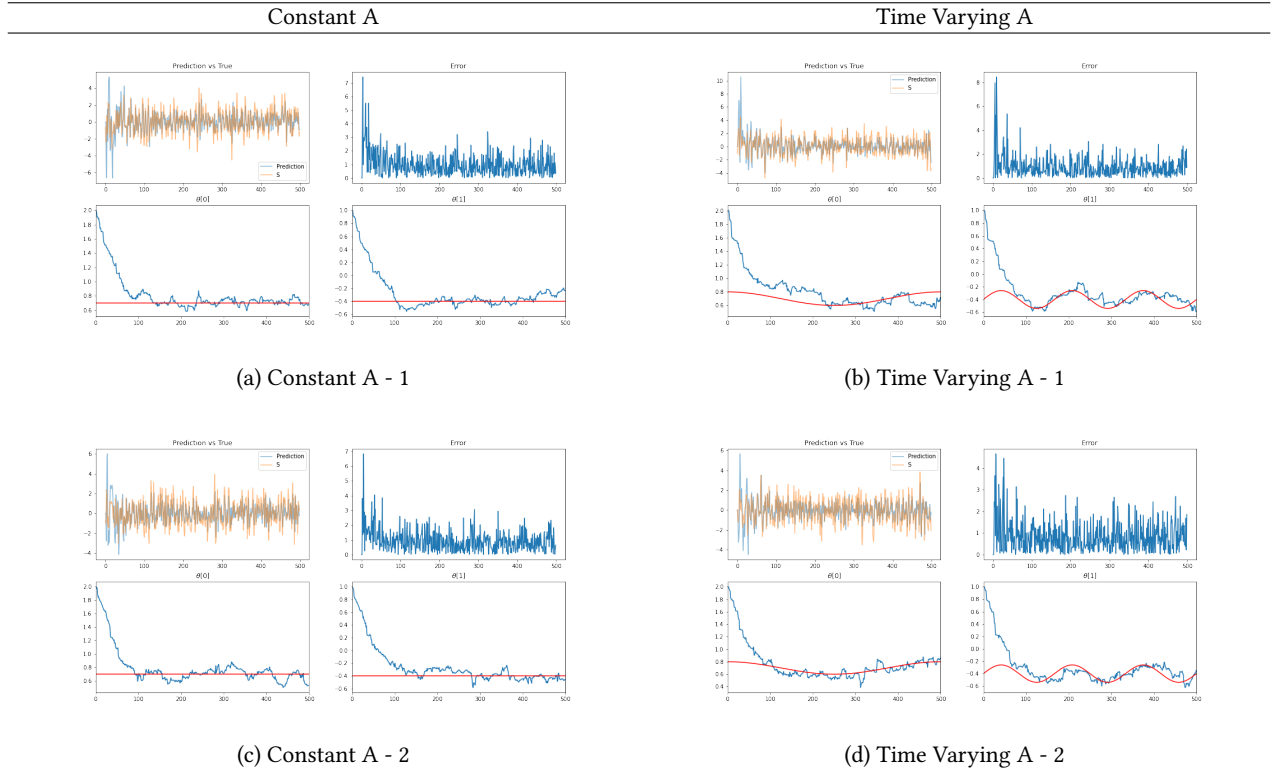


Figure 4: Varying Constant and Time Varying A with $a = [0.7, -0.4]$, $\theta = [2.0, 1.0]$, $R = 0.2\sigma(S_A)$, $Q = 0.0001I$

Experiment d shows a lot more accurate matching of the time varying components by θ , with the periodic functions reflected between iterations 100-200 and by iteration 300 both components of θ have almost matched their counterparts in A .