Time Series Analysis

Week 11 – State space models, 2nd part

Peder Bacher
Department of Applied Mathematics and Computer Science
Technical University of Denmark

April 19, 2024

Week 11: Outline of the lecture

State space models, 2nd part:

- Initialization of the Kalman filter
- ML-estimates in state space models, Sec. 10.6
- The Kalman filter when some observations are missing
- ARMA-models on state space form, Sec. 10.4
- Time-varying systems
- Examples

The linear state space model

$$\boldsymbol{X}_t = \boldsymbol{A}\boldsymbol{X}_{t-1} + \boldsymbol{G}\boldsymbol{e}_{1,t}$$
$$\boldsymbol{Y}_t = \boldsymbol{C}\boldsymbol{X}_t + \boldsymbol{e}_{2,t}$$

- $\{\boldsymbol{e}_{1,t}\}$ and $\{\boldsymbol{e}_{2,t}\}$ are mutually uncorrelated normally distributed white noise
- $V(\boldsymbol{e}_{1,t}) = \Sigma_1$ and $V(\boldsymbol{e}_{2,t}) = \Sigma_2$

Kalman Filter – Repetition

- What steps does the Kalman Filter consist of?
- How is the model used and how are the observations used?
- Model is used for prediction and combined with observations for reconstruction.
- How are the predictions/observations weighed in the reconstruction step?

Kalman filter – Initialization

- The Kalman filter has to be initialised - for both analysis and parameter estimation.

Kalman filter – Initialization

- The Kalman filter has to be initialised - for both analysis and parameter estimation.
- If you have no idea: Put $\widehat{\boldsymbol{X}}_{1|1} = \mathbf{0}$ and $\boldsymbol{\Sigma}^{xx}_{1|1} = \alpha \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix and α is a 'large' constant.

Kalman filter – Initialization

- The Kalman filter has to be initialised - for both analysis and parameter estimation.
- If you have no idea: Put $\widehat{\boldsymbol{X}}_{1|1} = \mathbf{0}$ and $\boldsymbol{\Sigma}^{xx}_{1|1} = \alpha \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix and α is a 'large' constant.
- If you know the starting state exactly: Put $\widehat{\boldsymbol{X}}_{1|1} =$ 'Known value' and $\boldsymbol{\Sigma}^{xx}_{1|1} = \mathbf{0}$, whereby the first *observation* has covariance matrix $\boldsymbol{\Sigma}^{yy}_{1|1} = \boldsymbol{\Sigma}_2$

Kalman filter – Initialization

- The Kalman filter has to be initialised - for both analysis and parameter estimation.
- If you have no idea: Put $\widehat{\boldsymbol{X}}_{1|1} = \mathbf{0}$ and $\boldsymbol{\Sigma}^{xx}_{1|1} = \alpha \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix and α is a 'large' constant.
- If you know the starting state exactly: Put $\widehat{\boldsymbol{X}}_{1|1} =$ 'Known value' and $\boldsymbol{\Sigma}^{xx}_{1|1} = \mathbf{0}$, whereby the first *observation* has covariance matrix $\boldsymbol{\Sigma}^{yy}_{1|1} = \boldsymbol{\Sigma}_2$
- If you have a good guess about the starting state: Put $\widehat{\boldsymbol{X}}_{1|1} =$ 'Guess' and $\boldsymbol{\Sigma}^{xx}_{1|1} = \boldsymbol{\Sigma}_{\text{Guess}}$.

Kalman filter – Initialization

- The Kalman filter has to be initialised - for both analysis and parameter estimation.
- If you have no idea: Put $\widehat{\boldsymbol{X}}_{1|1} = \mathbf{0}$ and $\boldsymbol{\Sigma}^{xx}_{1|1} = \alpha \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix and α is a 'large' constant.
- If you know the starting state exactly: Put $\widehat{\boldsymbol{X}}_{1|1} =$ 'Known value' and $\boldsymbol{\Sigma}^{xx}_{1|1} = \mathbf{0}$, whereby the first *observation* has covariance matrix $\boldsymbol{\Sigma}^{yy}_{1|1} = \boldsymbol{\Sigma}_2$
- If you have a good guess about the starting state: Put $\widehat{\boldsymbol{X}}_{1|1} =$ 'Guess' and $\boldsymbol{\Sigma}^{xx}_{1|1} = \boldsymbol{\Sigma}_{\text{Guess}}$.
- The important part is that the (un-)certainty of $\widehat{\boldsymbol{X}}_{1|1}$ is reflected in $\boldsymbol{\Sigma}^{xx}_{1|1}$.

Maximum Likelihood Estimates

- Let $\mathcal{Y}_{N^*}$ contain the available observations and let $\boldsymbol{\theta}$ contain the parameters of the model

Maximum Likelihood Estimates

- Let $\mathcal{Y}_{N^*}$ contain the available observations and let $\boldsymbol{\theta}$ contain the parameters of the model
- The likelihood function is the density of the random vector corresponding to the observations and given the set of parameters:

$$L(\boldsymbol{\theta}; \mathcal{Y}_{N^*}) = f(\mathcal{Y}_{N^*} | \boldsymbol{\theta})$$

Maximum Likelihood Estimates

- Let $\mathcal{Y}_{N^*}$ contain the available observations and let $\boldsymbol{\theta}$ contain the parameters of the model
- The likelihood function is the density of the random vector corresponding to the observations and given the set of parameters:

$$L(\boldsymbol{\theta}; \mathcal{Y}_{N^*}) = f(\mathcal{Y}_{N^*} | \boldsymbol{\theta})$$

- The ML-estimates are (as always) found by selecting $\boldsymbol{\theta}$ so that the density function is as large as possible at the actual observations

Maximum Likelihood Estimates

- Let $\mathcal{Y}_{N^*}$ contain the available observations and let $\boldsymbol{\theta}$ contain the parameters of the model
- The likelihood function is the density of the random vector corresponding to the observations and given the set of parameters:

$$L(\boldsymbol{\theta}; \mathcal{Y}_{N^*}) = f(\mathcal{Y}_{N^*} | \boldsymbol{\theta})$$

- The ML-estimates are (as always) found by selecting $\boldsymbol{\theta}$ so that the density function is as large as possible at the actual observations
- The random variables $\boldsymbol{Y}_{N^*} | \mathcal{Y}_{N^*-1}$ and $\mathcal{Y}_{N^*-1}$ are independent

Maximum Likelihood Estimates

- Let $\mathcal{Y}_{N^*}$ contain the available observations and let $\boldsymbol{\theta}$ contain the parameters of the model
- The likelihood function is the density of the random vector corresponding to the observations and given the set of parameters:

$$L(\boldsymbol{\theta}; \mathcal{Y}_{N^*}) = f(\mathcal{Y}_{N^*} | \boldsymbol{\theta})$$

- The ML-estimates are (as always) found by selecting $\boldsymbol{\theta}$ so that the density function is as large as possible at the actual observations
- The random variables $\boldsymbol{Y}_{N^*} | \mathcal{Y}_{N^*-1}$ and $\mathcal{Y}_{N^*-1}$ are independent, and so:

$$\begin{aligned} L(\boldsymbol{\theta}; \mathcal{Y}_{N^*}) &= f(\mathcal{Y}_{N^*} | \boldsymbol{\theta}) = f(\boldsymbol{Y}_{N^*} | \mathcal{Y}_{N^*-1}, \boldsymbol{\theta}) f(\mathcal{Y}_{N^*-1} | \boldsymbol{\theta}) \\ &= f(\boldsymbol{Y}_{N^*} | \mathcal{Y}_{N^*-1}, \boldsymbol{\theta}) f(\boldsymbol{Y}_{N^*-1} | \mathcal{Y}_{N^*-2}, \boldsymbol{\theta}) \cdots f(\boldsymbol{Y}_1 | \boldsymbol{\theta}) \end{aligned}$$

Maximum Likelihood Estimates

- Let $\mathcal{Y}_{N^*}$ contain the available observations and let $\boldsymbol{\theta}$ contain the parameters of the model
- The likelihood function is the density of the random vector corresponding to the observations and given the set of parameters:

$$L(\boldsymbol{\theta}; \mathcal{Y}_{N^*}) = f(\mathcal{Y}_{N^*} | \boldsymbol{\theta})$$

- The ML-estimates are (as always) found by selecting $\boldsymbol{\theta}$ so that the density function is as large as possible at the actual observations
- The random variables $\boldsymbol{Y}_{N^*} | \mathcal{Y}_{N^*-1}$ and $\mathcal{Y}_{N^*-1}$ are independent, and so:

$$\begin{aligned} L(\boldsymbol{\theta}; \mathcal{Y}_{N^*}) &= f(\mathcal{Y}_{N^*} | \boldsymbol{\theta}) = f(\boldsymbol{Y}_{N^*} | \mathcal{Y}_{N^*-1}, \boldsymbol{\theta}) f(\mathcal{Y}_{N^*-1} | \boldsymbol{\theta}) \\ &= f(\boldsymbol{Y}_{N^*} | \mathcal{Y}_{N^*-1}, \boldsymbol{\theta}) f(\boldsymbol{Y}_{N^*-1} | \mathcal{Y}_{N^*-2}, \boldsymbol{\theta}) \cdots f(\boldsymbol{Y}_1 | \boldsymbol{\theta}) \end{aligned}$$

- So we need one step predictions including estimates of their variance.

Maximum Likelihood Estimates

- Let $\mathcal{Y}_{N^*}$ contain the available observations and let $\boldsymbol{\theta}$ contain the parameters of the model
- The likelihood function is the density of the random vector corresponding to the observations and given the set of parameters:

$$L(\boldsymbol{\theta}; \mathcal{Y}_{N^*}) = f(\mathcal{Y}_{N^*} | \boldsymbol{\theta})$$

- The ML-estimates are (as always) found by selecting $\boldsymbol{\theta}$ so that the density function is as large as possible at the actual observations
- The random variables $\boldsymbol{Y}_{N^*} | \mathcal{Y}_{N^*-1}$ and $\mathcal{Y}_{N^*-1}$ are independent, and so:

$$\begin{aligned} L(\boldsymbol{\theta}; \mathcal{Y}_{N^*}) = f(\mathcal{Y}_{N^*} | \boldsymbol{\theta}) &= f(\boldsymbol{Y}_{N^*} | \mathcal{Y}_{N^*-1}, \boldsymbol{\theta}) f(\mathcal{Y}_{N^*-1} | \boldsymbol{\theta}) \\ &= f(\boldsymbol{Y}_{N^*} | \mathcal{Y}_{N^*-1}, \boldsymbol{\theta}) f(\boldsymbol{Y}_{N^*-1} | \mathcal{Y}_{N^*-2}, \boldsymbol{\theta}) \cdots f(\boldsymbol{Y}_1 | \boldsymbol{\theta}) \end{aligned}$$

- So we need one step predictions including estimates of their variance. Do we know how to do this?

Maximum Likelihood Estimates

- Let $\mathcal{Y}_{N^*}$ contain the available observations and let $\boldsymbol{\theta}$ contain the parameters of the model
- The likelihood function is the density of the random vector corresponding to the observations and given the set of parameters:

$$L(\boldsymbol{\theta}; \mathcal{Y}_{N^*}) = f(\mathcal{Y}_{N^*} | \boldsymbol{\theta})$$

- The ML-estimates are (as always) found by selecting $\boldsymbol{\theta}$ so that the density function is as large as possible at the actual observations
- The random variables $\boldsymbol{Y}_{N^*} | \mathcal{Y}_{N^*-1}$ and $\mathcal{Y}_{N^*-1}$ are independent, and so:

$$\begin{aligned} L(\boldsymbol{\theta}; \mathcal{Y}_{N^*}) = f(\mathcal{Y}_{N^*} | \boldsymbol{\theta}) &= f(\boldsymbol{Y}_{N^*} | \mathcal{Y}_{N^*-1}, \boldsymbol{\theta}) f(\mathcal{Y}_{N^*-1} | \boldsymbol{\theta}) \\ &= f(\boldsymbol{Y}_{N^*} | \mathcal{Y}_{N^*-1}, \boldsymbol{\theta}) f(\boldsymbol{Y}_{N^*-1} | \mathcal{Y}_{N^*-2}, \boldsymbol{\theta}) \cdots f(\boldsymbol{Y}_1 | \boldsymbol{\theta}) \end{aligned}$$

- So we need one step predictions including estimates of their variance. Do we know how to do this?
- The Kalman filter!

MLE / KF – Prediction

- Assume that at time t we have:

$$\widehat{\boldsymbol{X}}_{t|t} = E\left[\boldsymbol{X}_t|\mathcal{Y}_t\right] \quad \text{and} \quad \boldsymbol{\Sigma}^{xx}_{t|t} = V\left[\boldsymbol{X}_t|\mathcal{Y}_t\right]$$

MLE / KF – Prediction

- Assume that at time t we have:

$$\widehat{\boldsymbol{X}}_{t|t} = E\left[\boldsymbol{X}_t|\mathcal{Y}_t\right] \quad \text{and} \quad \boldsymbol{\Sigma}^{xx}_{t|t} = V\left[\boldsymbol{X}_t|\mathcal{Y}_t\right]$$

- Using the model we obtain predictions for time $t+1$:

$$\begin{aligned}
\widehat{\boldsymbol{X}}_{t+1|t} &= \boldsymbol{A}\widehat{\boldsymbol{X}}_{t|t} \\
\boldsymbol{\Sigma}^{xx}_{t+1|t} &= \boldsymbol{A}\boldsymbol{\Sigma}^{xx}_{t|t}\boldsymbol{A}^T + \boldsymbol{G}\boldsymbol{\Sigma}_1\boldsymbol{G}^T \\
\widehat{\boldsymbol{Y}}_{t+1|t} &= \boldsymbol{C}\widehat{\boldsymbol{X}}_{t+1|t} \\
\boldsymbol{\Sigma}^{yy}_{t+1|t} &= \boldsymbol{C}\boldsymbol{\Sigma}^{xx}_{t+1|t}\boldsymbol{C}^T + \boldsymbol{\Sigma}_2
\end{aligned}$$

MLE / KF – Prediction

- Assume that at time t we have:

$$\widehat{\boldsymbol{X}}_{t|t} = E\left[\boldsymbol{X}_t|\mathcal{Y}_t\right] \quad \text{and} \quad \boldsymbol{\Sigma}^{xx}_{t|t} = V\left[\boldsymbol{X}_t|\mathcal{Y}_t\right]$$

- Using the model we obtain predictions for time $t+1$:

$$\begin{aligned}
\widehat{\boldsymbol{X}}_{t+1|t} &= \boldsymbol{A}\widehat{\boldsymbol{X}}_{t|t} \\
\boldsymbol{\Sigma}^{xx}_{t+1|t} &= \boldsymbol{A}\boldsymbol{\Sigma}^{xx}_{t|t}\boldsymbol{A}^T + \boldsymbol{G}\boldsymbol{\Sigma}_1\boldsymbol{G}^T \\
\widehat{\boldsymbol{Y}}_{t+1|t} &= \boldsymbol{C}\widehat{\boldsymbol{X}}_{t+1|t} \\
\boldsymbol{\Sigma}^{yy}_{t+1|t} &= \boldsymbol{C}\boldsymbol{\Sigma}^{xx}_{t+1|t}\boldsymbol{C}^T + \boldsymbol{\Sigma}_2
\end{aligned}$$

- Due to the gaussian white noise process, $f(\boldsymbol{Y}_{t+1}|\mathcal{Y}_t, \boldsymbol{\theta})$ is the (multivariate) normal density (see Chapter 2) with mean $\widehat{\boldsymbol{Y}}_{t+1|t}$ and variance-covariance $\boldsymbol{\Sigma}^{yy}_{t+1|t}$

MLE / KF – Prediction

- Assume that at time t we have:

$$\widehat{\boldsymbol{X}}_{t|t} = E\left[\boldsymbol{X}_t|\mathcal{Y}_t\right] \quad \text{and} \quad \boldsymbol{\Sigma}^{xx}_{t|t} = V\left[\boldsymbol{X}_t|\mathcal{Y}_t\right]$$

- Using the model we obtain predictions for time $t+1$:

$$\begin{aligned}
\widehat{\boldsymbol{X}}_{t+1|t} &= \boldsymbol{A}\widehat{\boldsymbol{X}}_{t|t} \\
\boldsymbol{\Sigma}^{xx}_{t+1|t} &= \boldsymbol{A}\boldsymbol{\Sigma}^{xx}_{t|t}\boldsymbol{A}^T + \boldsymbol{G}\boldsymbol{\Sigma}_1\boldsymbol{G}^T \\
\widehat{\boldsymbol{Y}}_{t+1|t} &= \boldsymbol{C}\widehat{\boldsymbol{X}}_{t+1|t} \\
\boldsymbol{\Sigma}^{yy}_{t+1|t} &= \boldsymbol{C}\boldsymbol{\Sigma}^{xx}_{t+1|t}\boldsymbol{C}^T + \boldsymbol{\Sigma}_2
\end{aligned}$$

- Due to the gaussian white noise process, $f(\boldsymbol{Y}_{t+1}|\mathcal{Y}_t, \boldsymbol{\theta})$ is the (multivariate) normal density (see Chapter 2) with mean $\widehat{\boldsymbol{Y}}_{t+1|t}$ and variance-covariance $\boldsymbol{\Sigma}^{yy}_{t+1|t}$
- Explain to each other how the likelihood function is found constructed and used for estimation.

MLE / KF – The likelihood function

- Using the prediction errors and variances

$$\widetilde{\boldsymbol{Y}}_i = \boldsymbol{Y}_i - \widehat{\boldsymbol{Y}}_{i|i-1}$$

MLE / KF – The likelihood function

- Using the prediction errors and variances

$$\widetilde{\boldsymbol{Y}}_i = \boldsymbol{Y}_i - \widehat{\boldsymbol{Y}}_{i|i-1}$$

- The likelihood function can be expressed as

$$L\left(\boldsymbol{\theta}; \mathcal{Y}_{N^*}\right) = \prod_{i=1}^{N^*} \left[(2\pi)^m \det \boldsymbol{\Sigma}_{t|t-1}^{yy} \right]^{-\frac{1}{2}} \exp\left[-\frac{1}{2} \widetilde{\boldsymbol{Y}}_i^T (\boldsymbol{\Sigma}_{t+1|t}^{yy})^{-1} \widetilde{\boldsymbol{Y}}_i \right]$$

MLE / KF – The likelihood function

- Using the prediction errors and variances

$$\widetilde{\boldsymbol{Y}}_i = \boldsymbol{Y}_i - \widehat{\boldsymbol{Y}}_{i|i-1}$$

- The likelihood function can be expressed as

$$L\left(\boldsymbol{\theta}; \mathcal{Y}_{N^*}\right) = \prod_{i=1}^{N^*} \left[(2\pi)^m \operatorname{det} \boldsymbol{\Sigma}_{t|t-1}^{yy}\right]^{-\frac{1}{2}} \exp\left[-\frac{1}{2} \widetilde{\boldsymbol{Y}}_i^T (\boldsymbol{\Sigma}_{t+1|t}^{yy})^{-1} \widetilde{\boldsymbol{Y}}_i\right]$$

- Yielding the log-likelihood function:

$$\log L\left(\boldsymbol{\theta}; \mathcal{Y}_{N^*}\right) = -\frac{1}{2} \sum_{i=1}^{N} \left(\log \operatorname{det} \boldsymbol{\Sigma}_{t|t-1}^{yy} + \widetilde{\boldsymbol{Y}}_i^T (\boldsymbol{\Sigma}_{t+1|t}^{yy})_i^{-1} \widetilde{\boldsymbol{Y}}_i\right) + c$$

- The variance of the estimates can be approximated by the 2nd order derivatives of the log-likelihood.

MLE / KF – Reconstruction – Missing data

At time $t+1$ there are two possibilities for the reconstruction part:

The observation $\boldsymbol{Y}_{t+1}$ is available:

We update the state estimate using the reconstruction step of the Kalman Filter:

$$\boldsymbol{K}_{t+1} = \boldsymbol{\Sigma}^{xx}_{t+1|t}\boldsymbol{C}^T\left(\boldsymbol{\Sigma}^{yy}_{t+1|t}\right)^{-1}$$

$$\widehat{\boldsymbol{X}}_{t+1|t+1} = \widehat{\boldsymbol{X}}_{t+1|t} + \boldsymbol{K}_{t+1}\left(\boldsymbol{Y}_{t+1} - \widehat{\boldsymbol{Y}}_{t+1|t}\right)$$

$$\boldsymbol{\Sigma}^{xx}_{t+1|t+1} = \boldsymbol{\Sigma}^{xx}_{t+1|t} - \boldsymbol{K}_{t+1}\boldsymbol{\Sigma}^{yy}_{t+1|t}\boldsymbol{K}^T_{t+1}$$

MLE / KF – Reconstruction – Missing data

At time $t+1$ there are two possibilities for the reconstruction part:

The observation $\boldsymbol{Y}_{t+1}$ is available:

We update the state estimate using the reconstruction step of the Kalman Filter:

$$\boldsymbol{K}_{t+1} = \boldsymbol{\Sigma}^{xx}_{t+1|t} \boldsymbol{C}^T \left(\boldsymbol{\Sigma}^{yy}_{t+1|t} \right)^{-1}$$
$$\widehat{\boldsymbol{X}}_{t+1|t+1} = \widehat{\boldsymbol{X}}_{t+1|t} + \boldsymbol{K}_{t+1} \left(\boldsymbol{Y}_{t+1} - \widehat{\boldsymbol{Y}}_{t+1|t} \right)$$
$$\boldsymbol{\Sigma}^{xx}_{t+1|t+1} = \boldsymbol{\Sigma}^{xx}_{t+1|t} - \boldsymbol{K}_{t+1} \boldsymbol{\Sigma}^{yy}_{t+1|t} \boldsymbol{K}^T_{t+1}$$

The observation $\boldsymbol{Y}_{t+1}$ is missing:

We get no new information and we use:

$$\widehat{\boldsymbol{X}}_{t+1|t+1} = \widehat{\boldsymbol{X}}_{t+1|t}$$
$$\boldsymbol{\Sigma}^{xx}_{t+1|t+1} = \boldsymbol{\Sigma}^{xx}_{t+1|t}$$

MLE / KF – Reconstruction – Missing data

At time $t+1$ there are two possibilities for the reconstruction part:

The observation $\boldsymbol{Y}_{t+1}$ is available:

We update the state estimate using the reconstruction step of the Kalman Filter:

$$\boldsymbol{K}_{t+1} = \boldsymbol{\Sigma}^{xx}_{t+1|t} \boldsymbol{C}^T \left(\boldsymbol{\Sigma}^{yy}_{t+1|t} \right)^{-1}$$

$$\widehat{\boldsymbol{X}}_{t+1|t+1} = \widehat{\boldsymbol{X}}_{t+1|t} + \boldsymbol{K}_{t+1} \left(\boldsymbol{Y}_{t+1} - \widehat{\boldsymbol{Y}}_{t+1|t} \right)$$

$$\boldsymbol{\Sigma}^{xx}_{t+1|t+1} = \boldsymbol{\Sigma}^{xx}_{t+1|t} - \boldsymbol{K}_{t+1} \boldsymbol{\Sigma}^{yy}_{t+1|t} \boldsymbol{K}^T_{t+1}$$

The observation $\boldsymbol{Y}_{t+1}$ is missing:

We get no new information and we use:

$$\widehat{\boldsymbol{X}}_{t+1|t+1} = \widehat{\boldsymbol{X}}_{t+1|t}$$

$$\boldsymbol{\Sigma}^{xx}_{t+1|t+1} = \boldsymbol{\Sigma}^{xx}_{t+1|t}$$

Note: The same technique is used for multi-step predictions.

The ARMA(p, q) model as a state space model

$$Y_t + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

State space form:

$$\boldsymbol{X}_t = \boldsymbol{A}\boldsymbol{X}_{t-1} + \boldsymbol{G}\boldsymbol{\varepsilon}_{1,t}$$
$$\boldsymbol{Y}_t = \boldsymbol{C}\boldsymbol{X}_t + \boldsymbol{\varepsilon}_{2,t}$$

The ARMA(p, q) model as a state space model

$$Y_t + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

State space form:

$$\begin{aligned} \boldsymbol{X}_t &= \boldsymbol{A}\boldsymbol{X}_{t-1} + \boldsymbol{G}\boldsymbol{\varepsilon}_{1,t} \\ \boldsymbol{Y}_t &= \boldsymbol{C}\boldsymbol{X}_t \end{aligned}$$

Consider the following state space model, where row i is given by how Y_t influences Y_{t+i}:

$$\begin{aligned} \boldsymbol{X}_t &= \begin{bmatrix} -\phi_1 & 1 & 0 & \cdots & 0 \\ -\phi_2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\phi_{d-1} & 0 & 0 & 0 & 1 \\ -\phi_d & 0 & 0 & \cdots & 0 \end{bmatrix} \boldsymbol{X}_{t-1} + \begin{pmatrix} 1 \\ \theta_1 \\ \vdots \\ \theta_{d-1} \end{pmatrix} \boldsymbol{\varepsilon}_t \\ \boldsymbol{Y}_t &= \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix} \boldsymbol{X}_t \end{aligned}$$

where $d = max(p, q+1)$ and any extra parameter is fixed to zero.

The ARMA(p, q) model as a state space model

$$Y_t + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

State space form:

$$\boldsymbol{X}_t = \boldsymbol{A}\boldsymbol{X}_{t-1} + \boldsymbol{G}\boldsymbol{\varepsilon}_{1,t}$$
$$\boldsymbol{Y}_t = \boldsymbol{C}\boldsymbol{X}_t$$

Consider the following state space model, where row i is given by how Y_t influences Y_{t+i}:

$$\boldsymbol{X}_t = \begin{bmatrix} -\phi_1 & 1 & 0 & \cdots & 0 \\ -\phi_2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\phi_{d-1} & 0 & 0 & 0 & 1 \\ -\phi_d & 0 & 0 & \cdots & 0 \end{bmatrix} \boldsymbol{X}_{t-1} + \begin{pmatrix} 1 \\ \theta_1 \\ \vdots \\ \theta_{d-1} \end{pmatrix} \boldsymbol{\varepsilon}_t$$
$$\boldsymbol{Y}_t = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix} \boldsymbol{X}_t$$

where $d = max(p, q+1)$ and any extra parameter is fixed to zero.
What is the advantage of writing ARMA-processes on state-space form?

Estimation in ARMA(p, q)-models using the KF

- Using the Kalman filter we can get the mean and variance of the one-step predictions of the observations:

$$\widehat{\boldsymbol{Y}}_{t+1|t} = \boldsymbol{C}\widehat{\boldsymbol{X}}_{t+1|t}$$
$$\boldsymbol{\Sigma}^{yy}_{t+1|t} = \boldsymbol{C}\boldsymbol{\Sigma}^{xx}_{t+1|t}\boldsymbol{C}^T + \boldsymbol{\Sigma}_2$$

- The Kalman filter can handle missing observations

Estimation in ARMA(p, q)-models using the KF

- Using the Kalman filter we can get the mean and variance of the one-step predictions of the observations:

$$\widehat{\boldsymbol{Y}}_{t+1|t} = \boldsymbol{C}\widehat{\boldsymbol{X}}_{t+1|t}$$
$$\boldsymbol{\Sigma}^{yy}_{t+1|t} = \boldsymbol{C}\boldsymbol{\Sigma}^{xx}_{t+1|t}\boldsymbol{C}^T + \boldsymbol{\Sigma}_2$$

- The Kalman filter can handle missing observations
- An ARMA(p, q)-model can be written as a state space model

Estimation in ARMA(p, q)-models using the KF

- Using the Kalman filter we can get the mean and variance of the one-step predictions of the observations:

$$\widehat{\boldsymbol{Y}}_{t+1|t} = \boldsymbol{C}\widehat{\boldsymbol{X}}_{t+1|t}$$
$$\boldsymbol{\Sigma}^{yy}_{t+1|t} = \boldsymbol{C}\boldsymbol{\Sigma}^{xx}_{t+1|t}\boldsymbol{C}^T + \boldsymbol{\Sigma}_2$$

- The Kalman filter can handle missing observations
- An ARMA(p, q)-model can be written as a state space model
- This gives us a way of calculating ML-estimates in the ARMA(p, q)-model even when some observations are missing.

Falling body - Revisited

Remember the discretised state-space model of a falling body

$$\boldsymbol{X}_t = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \boldsymbol{X}_{t-1} - \begin{bmatrix} 0.5 \\ 1 \end{bmatrix} g + \epsilon_t$$

$$Y_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \boldsymbol{X}_t + e_t$$

Falling body - Revisited

Remember the discretised state-space model of a falling body

$$
\begin{aligned}
\boldsymbol{X}_t &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \boldsymbol{X}_{t-1} - \begin{bmatrix} 0.5 \\ 1 \end{bmatrix} g + \epsilon_t \\
Y_t &= \begin{bmatrix} 1 & 0 \end{bmatrix} \boldsymbol{X}_t + e_t
\end{aligned}
$$

Imagine that we believe g to be changing over time, but we don't know how or why. How can we incorporate this?

Falling body - Revisited

Remember the discretised state-space model of a falling body

$$\boldsymbol{X}_t = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \boldsymbol{X}_{t-1} - \begin{bmatrix} 0.5 \\ 1 \end{bmatrix} g + \epsilon_t$$
$$Y_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \boldsymbol{X}_t + e_t$$

Imagine that we believe g to be changing over time, but we don't know how or why. How can we incorporate this? We can rewrite the model with g as a state!

Falling body - Revisited

Remember the discretised state-space model of a falling body

$$\boldsymbol{X}_t = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \boldsymbol{X}_{t-1} - \begin{bmatrix} 0.5 \\ 1 \end{bmatrix} g + \epsilon_t$$
$$Y_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \boldsymbol{X}_t + e_t$$

Imagine that we believe g to be changing over time, but we don't know how or why. How can we incorporate this? We can rewrite the model with g as a state!

$$\boldsymbol{X}_t = \begin{bmatrix} 1 & 1 & -0.5 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \boldsymbol{X}_{t-1} + \epsilon_t$$
$$Y_t = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \boldsymbol{X}_t + e_t$$

Falling body - Revisited

Remember the discretised state-space model of a falling body

$$
\begin{aligned}
\boldsymbol{X}_t &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \boldsymbol{X}_{t-1} - \begin{bmatrix} 0.5 \\ 1 \end{bmatrix} g + \epsilon_t \\
Y_t &= \begin{bmatrix} 1 & 0 \end{bmatrix} \boldsymbol{X}_t + e_t
\end{aligned}
$$

Imagine that we believe g to be changing over time, but we don't know how or why. How can we incorporate this? We can rewrite the model with g as a state!

$$
\begin{aligned}
\boldsymbol{X}_t &= \begin{bmatrix} 1 & 1 & -0.5 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \boldsymbol{X}_{t-1} + \epsilon_t \\
Y_t &= \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \boldsymbol{X}_t + e_t
\end{aligned}
$$

Under what conditions should the process noise for g be non-zero?

Falling body - Revisited

Remember the discretised state-space model of a falling body

$$\boldsymbol{X}_t = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \boldsymbol{X}_{t-1} - \begin{bmatrix} 0.5 \\ 1 \end{bmatrix} g + \epsilon_t$$
$$Y_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \boldsymbol{X}_t + e_t$$

Imagine that we believe g to be changing over time, but we don't know how or why. How can we incorporate this? We can rewrite the model with g as a state!

$$\boldsymbol{X}_t = \begin{bmatrix} 1 & 1 & -0.5 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \boldsymbol{X}_{t-1} + \epsilon_t$$
$$Y_t = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \boldsymbol{X}_t + e_t$$

Under what conditions should the process noise for g be non-zero?
When does this trick work?

Parameter estimation as state-estimation

For the linear state-space model

$$\begin{aligned} \text{System equation:} \quad & \boldsymbol{X}_t = \boldsymbol{A}\boldsymbol{X}_{t-1} + \boldsymbol{B}\boldsymbol{u}_{t-1} + \boldsymbol{e}_{1,t} \\ \text{Observation equation:} \quad & \boldsymbol{Y}_t = \boldsymbol{C}\boldsymbol{X}_t + \boldsymbol{e}_{2,t}, \end{aligned}$$

when $\boldsymbol{u}_t$ unknown, (with observations or not), it can be estimated as a state by

$$\begin{aligned} \text{System equation:} \quad \boldsymbol{X}_t &= \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{t-1} \\ \boldsymbol{u}_{t-1} \end{bmatrix} + \boldsymbol{e}_{1,t} \\ \text{Observation equation:} \quad \boldsymbol{Y}_t &= \begin{bmatrix} \boldsymbol{C} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_t \\ \boldsymbol{u}_t \end{bmatrix} + \boldsymbol{e}_{2,t}, \end{aligned}$$

Parameter estimation as state-estimation

For the linear state-space model

$$\begin{aligned} \text{System equation:} \quad & \boldsymbol{X}_t = \boldsymbol{A}\boldsymbol{X}_{t-1} + \boldsymbol{B}\boldsymbol{u}_{t-1} + \boldsymbol{e}_{1,t} \\ \text{Observation equation:} \quad & \boldsymbol{Y}_t = \boldsymbol{C}\boldsymbol{X}_t + \boldsymbol{e}_{2,t}, \end{aligned}$$

when $\boldsymbol{u}_t$ unknown, (with observations or not), it can be estimated as a state by

$$\begin{aligned} \text{System equation:} \quad \boldsymbol{X}_t &= \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{t-1} \\ \boldsymbol{u}_{t-1} \end{bmatrix} + \boldsymbol{e}_{1,t} \\ \text{Observation equation:} \quad \boldsymbol{Y}_t &= \begin{bmatrix} \boldsymbol{C} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_t \\ \boldsymbol{u}_t \end{bmatrix} + \boldsymbol{e}_{2,t}, \text{ or} \end{aligned}$$

Parameter estimation as state-estimation

For the linear state-space model

$$\text{System equation:} \quad \boldsymbol{X}_t = \boldsymbol{A}\boldsymbol{X}_{t-1} + \boldsymbol{B}\boldsymbol{u}_{t-1} + \boldsymbol{e}_{1,t}$$
$$\text{Observation equation:} \quad \boldsymbol{Y}_t = \boldsymbol{C}\boldsymbol{X}_t + \boldsymbol{e}_{2,t},$$

when $\boldsymbol{u}_t$ unknown, (with observations or not), it can be estimated as a state by

$$\text{System equation:} \quad \boldsymbol{X}_t = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{t-1} \\ \boldsymbol{u}_{t-1} \end{bmatrix} + \boldsymbol{e}_{1,t}$$
$$\text{Observation equation:} \quad \boldsymbol{Y}_t = \begin{bmatrix} \boldsymbol{C} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_t \\ \boldsymbol{u}_t \end{bmatrix} + \boldsymbol{e}_{2,t}, \quad \text{or} \quad \boldsymbol{Y}_t = \begin{bmatrix} \boldsymbol{C} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_t \\ \boldsymbol{u}_t \end{bmatrix} + \boldsymbol{e}_{2,t}$$

Parameter estimation as state-estimation

For the linear state-space model

$$\begin{aligned} \text{System equation:} \quad \boldsymbol{X}_t &= \boldsymbol{A}\boldsymbol{X}_{t-1} + \boldsymbol{B}\boldsymbol{u}_{t-1} + \boldsymbol{e}_{1,t} \\ \text{Observation equation:} \quad \boldsymbol{Y}_t &= \boldsymbol{C}\boldsymbol{X}_t + \boldsymbol{e}_{2,t}, \end{aligned}$$

when $\boldsymbol{u}_t$ unknown, (with observations or not), it can be estimated as a state by

$$\text{System equation:} \quad \boldsymbol{X}_t = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{t-1} \\ \boldsymbol{u}_{t-1} \end{bmatrix} + \boldsymbol{e}_{1,t}$$

$$\text{Observation equation:} \quad \boldsymbol{Y}_t = \begin{bmatrix} \boldsymbol{C} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_t \\ \boldsymbol{u}_t \end{bmatrix} + \boldsymbol{e}_{2,t}, \text{ or } \boldsymbol{Y}_t = \begin{bmatrix} \boldsymbol{C} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_t \\ \boldsymbol{u}_t \end{bmatrix} + \boldsymbol{e}_{2,t}$$

When would you do this?

Parameter estimation as state-estimation

For the linear state-space model

$$\text{System equation:} \quad \boldsymbol{X}_t = \boldsymbol{A}\boldsymbol{X}_{t-1} + \boldsymbol{B}\boldsymbol{u}_{t-1} + \boldsymbol{e}_{1,t}$$
$$\text{Observation equation:} \quad \boldsymbol{Y}_t = \boldsymbol{C}\boldsymbol{X}_t + \boldsymbol{e}_{2,t},$$

when $\boldsymbol{u}_t$ unknown, (with observations or not), it can be estimated as a state by

$$\text{System equation:} \quad \boldsymbol{X}_t = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{t-1} \\ \boldsymbol{u}_{t-1} \end{bmatrix} + \boldsymbol{e}_{1,t}$$
$$\text{Observation equation:} \quad \boldsymbol{Y}_t = \begin{bmatrix} \boldsymbol{C} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_t \\ \boldsymbol{u}_t \end{bmatrix} + \boldsymbol{e}_{2,t}, \text{ or } \boldsymbol{Y}_t = \begin{bmatrix} \boldsymbol{C} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_t \\ \boldsymbol{u}_t \end{bmatrix} + \boldsymbol{e}_{2,t}$$

When would you do this? Unknown or uncertain inputs.

Parameter estimation as state-estimation

For the linear state-space model

$$\text{System equation:} \quad \boldsymbol{X}_t = \boldsymbol{A}\boldsymbol{X}_{t-1} + \boldsymbol{B}\boldsymbol{u}_{t-1} + \boldsymbol{e}_{1,t}$$
$$\text{Observation equation:} \quad \boldsymbol{Y}_t = \boldsymbol{C}\boldsymbol{X}_t + \boldsymbol{e}_{2,t},$$

when $\boldsymbol{u}_t$ unknown, (with observations or not), it can be estimated as a state by

$$\text{System equation:} \quad \boldsymbol{X}_t = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{t-1} \\ \boldsymbol{u}_{t-1} \end{bmatrix} + \boldsymbol{e}_{1,t}$$
$$\text{Observation equation:} \quad \boldsymbol{Y}_t = \begin{bmatrix} \boldsymbol{C} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_t \\ \boldsymbol{u}_t \end{bmatrix} + \boldsymbol{e}_{2,t}, \text{ or } \boldsymbol{Y}_t = \begin{bmatrix} \boldsymbol{C} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_t \\ \boldsymbol{u}_t \end{bmatrix} + \boldsymbol{e}_{2,t}$$

When would you do this? Unknown or uncertain inputs. Update parameter estimate as more information becomes available.

Parameter estimation as state-estimation

For the linear state-space model

$$\text{System equation:} \quad \boldsymbol{X}_t = \boldsymbol{A}\boldsymbol{X}_{t-1} + \boldsymbol{B}\boldsymbol{u}_{t-1} + \boldsymbol{e}_{1,t}$$
$$\text{Observation equation:} \quad \boldsymbol{Y}_t = \boldsymbol{C}\boldsymbol{X}_t + \boldsymbol{e}_{2,t},$$

when $\boldsymbol{u}_t$ unknown, (with observations or not), it can be estimated as a state by

$$\text{System equation:} \quad \boldsymbol{X}_t = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{t-1} \\ \boldsymbol{u}_{t-1} \end{bmatrix} + \boldsymbol{e}_{1,t}$$
$$\text{Observation equation:} \quad \boldsymbol{Y}_t = \begin{bmatrix} \boldsymbol{C} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_t \\ \boldsymbol{u}_t \end{bmatrix} + \boldsymbol{e}_{2,t}, \text{ or } \boldsymbol{Y}_t = \begin{bmatrix} \boldsymbol{C} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_t \\ \boldsymbol{u}_t \end{bmatrix} + \boldsymbol{e}_{2,t}$$

When would you do this? Unknown or uncertain inputs. Update parameter estimate as more information becomes available. Assumption of varying parameters.

Summary of state-space models

- Two kinds of noise.

Summary of state-space models

- Two kinds of noise. Why is this useful?

Summary of state-space models

- Two kinds of noise. Why is this useful? Instant versus sustained effect.

Summary of state-space models

- Two kinds of noise. Why is this useful? Instant versus sustained effect.
- Model identification.

Summary of state-space models

- Two kinds of noise. Why is this useful? Instant versus sustained effect.
- Model identification. What is the general idea?

Summary of state-space models

- Two kinds of noise. Why is this useful? Instant versus sustained effect.
- Model identification. What is the general idea? Formulate physical equations or at least sensible equations.

Summary of state-space models

- Two kinds of noise. Why is this useful? Instant versus sustained effect.
- Model identification. What is the general idea? Formulate physical equations or at least sensible equations.
- Kalman filter.

Summary of state-space models

- Two kinds of noise. Why is this useful? Instant versus sustained effect.
- Model identification. What is the general idea? Formulate physical equations or at least sensible equations.
- Kalman filter. What are the two steps that it consists of?

Summary of state-space models

- Two kinds of noise. Why is this useful? Instant versus sustained effect.
- Model identification. What is the general idea? Formulate physical equations or at least sensible equations.
- Kalman filter. What are the two steps that it consists of? Reconstruction and prediction.

Summary of state-space models

- Two kinds of noise. Why is this useful? Instant versus sustained effect.
- Model identification. What is the general idea? Formulate physical equations or at least sensible equations.
- Kalman filter. What are the two steps that it consists of? Reconstruction and prediction.
- Handling missing values.

Summary of state-space models

- Two kinds of noise. Why is this useful? Instant versus sustained effect.
- Model identification. What is the general idea? Formulate physical equations or at least sensible equations.
- Kalman filter. What are the two steps that it consists of? Reconstruction and prediction.
- Handling missing values. Why is this so easy?

Summary of state-space models

- Two kinds of noise. Why is this useful? Instant versus sustained effect.
- Model identification. What is the general idea? Formulate physical equations or at least sensible equations.
- Kalman filter. What are the two steps that it consists of? Reconstruction and prediction.
- Handling missing values. Why is this so easy? Just don't update during reconstruction step.

Summary of state-space models

- Two kinds of noise. Why is this useful? Instant versus sustained effect.
- Model identification. What is the general idea? Formulate physical equations or at least sensible equations.
- Kalman filter. What are the two steps that it consists of? Reconstruction and prediction.
- Handling missing values. Why is this so easy? Just don't update during reconstruction step.
- Adaptive parameter estimates.

Summary of state-space models

- Two kinds of noise. Why is this useful? Instant versus sustained effect.
- Model identification. What is the general idea? Formulate physical equations or at least sensible equations.
- Kalman filter. What are the two steps that it consists of? Reconstruction and prediction.
- Handling missing values. Why is this so easy? Just don't update during reconstruction step.
- Adaptive parameter estimates. Include parameters as states.