Time Series Analysis

Henrik Madsen

hm@imm.dtu.dk

Informatics and Mathematical Modelling
Technical University of Denmark
DK-2800 Kgs. Lyngby

Outline of the lecture

State space models, 2nd part:

- The Kalman filter when some observations are missing
- ARMA-models on state space form, Sec. 10.4 (not 10.4.1)
- ML-estimates of state space models, Sec. 10.6

Cursory material:

- Signal extraction, Sec. 10.4.1
- Time series with missing observations, Sec. 10.5

The linear stochastic state space model

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

- $\boldsymbol{X}$: State vector
- $\boldsymbol{Y}$: Observation vector
- $\boldsymbol{u}$: Input vector
- $\boldsymbol{e}_1$: System noise
- $\boldsymbol{e}_2$: Observation noise

- $\dim(X_t) = m$ is called the order of the system
- $\{\boldsymbol{e}_{1,t}\}$ and $\{\boldsymbol{e}_{2,t}\}$ mutually independent white noise
- $V[\boldsymbol{e}_1] = \boldsymbol{\Sigma}_1$, $V[\boldsymbol{e}_2] = \boldsymbol{\Sigma}_2$
- $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{\Sigma}_1$, and $\boldsymbol{\Sigma}_2$ are **known** matrices

The Kalman filter

Initialization: $\widehat{\boldsymbol{X}}_{1|0} = \boldsymbol{\mu}_0$, $\boldsymbol{\Sigma}^{xx}_{1|0} = \boldsymbol{V}_0 \Rightarrow \boldsymbol{\Sigma}^{yy}_{1|0} = \boldsymbol{C}\boldsymbol{\Sigma}^{xx}_{1|0}\boldsymbol{C}^T + \boldsymbol{\Sigma}_2$

For: $t = 1, 2, 3, \ldots$

Reconstruction:

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

Prediction:

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

The Kalman filter

Initialization: $\widehat{\boldsymbol{X}}_{1|0} = \boldsymbol{\mu}_0,\ \boldsymbol{\Sigma}^{xx}_{1|0} = \boldsymbol{V}_0 \ \Rightarrow\ \boldsymbol{\Sigma}^{yy}_{1|0} = \boldsymbol{C}\boldsymbol{\Sigma}^{xx}_{1|0}\boldsymbol{C}^T + \boldsymbol{\Sigma}_2$

For: $t = 1, 2, 3, \ldots$

Reconstruction:

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

Prediction:

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

What happens if the observation Y_t is missing for some t?

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:

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

- 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.

$ARMA(p, q)$-models on state space form

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

State space form:

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

$$\boldsymbol{X}_t = (X_{1,t}, X_{2,t}, \ldots, X_{d,t})^T, \quad d = \max(p, q+1)$$

$$\boldsymbol{A} = \begin{bmatrix} -\phi_1 & 1 & 0 & \cdots & 0 \\ -\phi_2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ -\phi_{d-1} & 0 & 0 & & 1 \\ -\phi_d & 0 & 0 & & 0 \end{bmatrix} \qquad \boldsymbol{e}_{1,t} = \boldsymbol{G}\boldsymbol{\varepsilon}_t = \begin{bmatrix} 1 \\ \theta_1 \\ \vdots \\ \theta_{d-1} \end{bmatrix} \varepsilon_t$$

$$C = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}$$

ML-estimates in state space models

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

- $\{\boldsymbol{e}_{1,t}\}$ and $\{\boldsymbol{e}_{2,t}\}$ are mutually uncorrelated normally distributed white noise
- $V(\boldsymbol{e}_{1,t}) = \Sigma_1$ and $V(\boldsymbol{e}_{2,t}) = \Sigma_2$
- For $ARMA(p,q)$-models we have $\boldsymbol{A}$, $\boldsymbol{C}$, and $\boldsymbol{G}$ as stated on the previous slide. Furthermore, $\boldsymbol{e}_{1,t} = \varepsilon_t$, $\boldsymbol{\Sigma}_1 = \sigma_\varepsilon^2$, and $\boldsymbol{\Sigma}_2 = 0$

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 is 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:

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

- The conditional densities can be found using the Kalman filter

MLE / KF

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

DTU

MLE / KF (cont'nd)

At time $t+1$ there is two possibilities:

The observation $\boldsymbol{Y}_{t+1}$ is available: We update the state estimate using the reconstruction step of the Kalman Filter:

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

The observation $\boldsymbol{Y}_{t+1}$ is missing: We got no new information and we use:

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

And then we predict for time $t+2$

MLE / KF (cont'nd)

- Using the prediction errors and variances

$$\begin{aligned}\widetilde{\boldsymbol{Y}}_i &= \boldsymbol{Y}_i - \widehat{\boldsymbol{Y}}_{i|i-1} \\ \boldsymbol{R}_i &= \boldsymbol{\Sigma}^{yy}_{i|i-1}\end{aligned}$$

- The likelihood function can be expressed as

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

- In practice optimization is based on $\log L(\boldsymbol{\theta}; \mathcal{Y}_{N^*})$ and the variance of the estimates can be approximated by the 2'nd order derivatives of log-likelihood.

MLE / KF (cont'nd)

- The only outstanding issue is "prediction" of $\boldsymbol{Y}_1$, i.e. calculation of $\widehat{\boldsymbol{Y}}_{1|0}$
- This can be done by setting $\widehat{\boldsymbol{X}}_{0|0} = \mathbf{0}$ and $\boldsymbol{\Sigma}^{xx}_{0|0} = \alpha \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix and α is a 'large' constant (we don't know what it is)
- Alternatively, we can estimate the initial state $\widehat{\boldsymbol{X}}_{0|0}$ and set $\boldsymbol{\Sigma}^{xx}_{0|0} = \mathbf{0}$, whereby $\boldsymbol{\Sigma}^{xx}_{1|0} = \boldsymbol{G}\boldsymbol{\Sigma}_1\boldsymbol{G}^T$