FORECASTING
STAT0010

Alexandros Beskos

a.beskos@ucl.ac.uk

'Lecture 10' Outline: Structural time series models

1. Local level model
2. Linear trend model
3. State space models
4. Kalman filter

A structural time series describes a process by components such as 'trend', 'seasonality', and 'noise':

$$\text{observation} = \text{trend} + \text{seasonality} + \text{noise}\,.$$

rather than in terms of, e.g., *AR* and *MA* components. Flexibility can be introduced by allowing each term to change (stochastically) over time.

Definition 1 (Local level model)

Consider

$$\underbrace{Y_t}_{\textit{observation}} = \underbrace{\mu_t}_{\textit{trend}} + \underbrace{e_t}_{\textit{noise}} \qquad e_t \sim \mathcal{WN}(0, \sigma_e^2)\,, \qquad (1)$$

where the trend is modelled by a random walk:

$$\mu_t = \mu_{t-1} + h_t\,, \qquad h_t \sim \mathcal{WN}(0, \sigma_h^2)\,, \qquad (2)$$

and where $\mathrm{cov}(e_s, h_t) = 0 \ \forall\, s, t$*.*

Then equations (1) and (2) describe the local level model. (Y_t is known as the observation process of the local level model). Eqn. (1) is a.k.a observation equation; Eqn. (2) is a.k.a transition (or state) equation.

Lemma 2

Local-level model is $ARIMA(0,1,1)$.

Proof First note that Y_t is non-stationary (contains random walk trend). Differencing gives:

$$\nabla Y_t = \nabla \mu_t + \nabla e_t\,.$$

Now, note, from Eqn. (2): $\mu_t = \mu_{t-1} + h_t$, we have that $\nabla \mu_t = h_t$. Hence

$$Z_t := \nabla Y_t = h_t + \nabla e_t = h_t + e_t - e_{t-1}\,.$$

We have $\mathbb{E}(Z_t) = 0$, and noting that $e_t \sim \mathcal{WN}(0, \sigma_e^2)$, $h_t \sim \mathcal{WN}(0, \sigma_h^2)$, $\operatorname{cov}(e_s, h_t) = 0\ \forall\, s, t$:

$$\begin{aligned}
\gamma_Z(0) &= \mathbb{E}(Z_t^2) = \mathbb{E}\big((h_t + e_t - e_{t-1})(h_t + e_t - e_{t-1})\big) = \sigma_h^2 + 2\sigma_e^2 \\
\gamma_Z(1) &= \mathbb{E}(Z_t Z_{t-1}) = \mathbb{E}\big((h_t + e_t - e_{t-1})(h_{t-1} + e_{t-1} - e_{t-2})\big) = -\sigma_e^2\,, \\
\gamma_Z(k) &= \mathbb{E}(Z_t Z_{t-k}) = 0\,, k \geq 2\,.
\end{aligned}$$

Hence,

$$\rho_Z(k) = \gamma_Z(k)/\gamma_Z(0) = \begin{cases} \dfrac{-\sigma_e^2}{\sigma_h^2 + 2\sigma_e^2}\,, & k = 1 \\ 0\,, & k \geq 2 \end{cases}$$

ACF cuts off at lag 1 $\Rightarrow \nabla Y_t$ is $MA(1) \Rightarrow Y_t$ is $ARIMA(0,1,1)$ ■

Definition 3

The ARIMA$(0,1,1)$ form of the local level model, namely $\nabla Y_t = h_t + e_t - e_{t-1}$, is called the reduced form of the model.

Alternative way to derive reduced form, which generalises better for more complicated models is to use backshift operator, as follows. Rewrite (transition) Eqn. (2): $\mu_t = \mu_{t-1} + h_t$ in terms of backshift:

$$\mu_t = (1-B)^{-1} h_t\,,$$

and substitute into (observation) Eqn. (1): $Y_t = \mu_t + e_t$:

$$Y_t = (1-B)^{-1} h_t + e_t$$

Multiplying both sides by $(1-B)$ gives:

$$(1-B)Y_t = h_t + (1-B)e_t\,,$$

which is the reduced form.

The local level model can be extended to a local linear trend model by introducing a slope in the trend term. If the trend term were, say, deterministic then we would write a linear trend as

$$\mu_t = \alpha + \beta t\,, \qquad \alpha, \beta \in \mathbb{R}\,,$$

i.e.

$$\nabla \mu_t = \beta \Rightarrow \mu_t = \mu_{t-1} + \beta\,.$$

Allowing the slope to change stochastically over time, we have...

Definition 4 (Local linear trend)

Consider

$$\begin{aligned} Y_t &= \mu_t + e_t && (3)\\ \mu_t &= \mu_{t-1} + \beta_{t-1} + h_t && (4)\\ \beta_t &= \beta_{t-1} + z_t && (5) \end{aligned}$$

where e_t, h_t, z_t are mutually and serially uncorrelated white noise terms with variances $\sigma_e^2, \sigma_h^2, \sigma_z^2$. Then equations (3), (4), and (5) describe the local linear trend model. Eqn. (3) is the observation equation and eqns. (4) and (5) are the transition (or state) equations.

Lemma 5

Local linear trend model is $ARIMA(0,2,2)$.

Proof Recall:

$$
\begin{aligned}
Y_t &= \mu_t + e_t \\
\mu_t &= \mu_{t-1} + \beta_{t-1} + h_t \\
\beta_t &= \beta_{t-1} + z_t \, .
\end{aligned}
$$

Then

$$
\begin{aligned}
\nabla\beta_t &= z_t \\
\beta_t &= (1-B)^{-1} z_t \\
\mu_t &= (1-B)^{-1}(\beta_{t-1} + h_t) \\
&= (1-B)^{-1}((1-B)^{-1} z_{t-1} + h_t) \\
&= (1-B)^{-2} z_{t-1} + (1-B)^{-1} h_t
\end{aligned}
$$

Recall

$$\mu_t = (1-B)^{-2} z_{t-1} + (1-B)^{-1} h_t$$

Then,

$$\begin{aligned} Y_t &= \mu_t + e_t \\ &= (1-B)^{-2} z_{t-1} + (1-B)^{-1} h_t + e_t \\ (1-B)^2 Y_t &= z_{t-1} + (1-B) h_t + (1-B)^2 e_t \\ \nabla^2 Y_t &= z_{t-1} + \nabla h_t + \nabla^2 e_t \,. \end{aligned}$$

I.e. $\nabla^2 Y_t$ is stationary (is sum of white noise terms), and $\mathbb{E}(\nabla^2 Y_t) = 0$. The autocovariance of $\nabla^2 Y_t$ is

$$\begin{aligned} \gamma_{\nabla^2 Y}(0) &= 2\sigma_h^2 + \sigma_z^2 + 6\sigma_e^2 \\ \gamma_{\nabla^2 Y}(1) &= -\sigma_h^2 - 4\sigma_e^2 \\ \gamma_{\nabla^2 Y}(2) &= \sigma_e^2 \\ \gamma_{\nabla^2 Y}(k) &= 0\,, \quad \text{for } k \geq 3\,. \end{aligned}$$

Hence $\nabla^2 Y_t$ is $MA(2) \Rightarrow Y_t$ is $ARIMA(0,2,2)$.

To perform estimation (of trend, slope, etc) and prediction on a structural model, it is convenient to express the model in state-space form. Recall local linear trend model:

$$\begin{aligned} Y_t &= \mu_t + e_t \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + h_t \\ \beta_t &= \beta_{t-1} + z_t \,. \end{aligned}$$

This can be written in the form

$$Y_t = \begin{bmatrix} 1 \\ 0 \end{bmatrix}^T \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} + \epsilon_t \qquad \text{[observation or measurement equation]}$$

with $\epsilon_t \sim \mathcal{WN}(0, \sigma^2)$, and

$$\begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_{t-1} \\ \beta_{t-1} \end{bmatrix} + \begin{bmatrix} h_t \\ z_t \end{bmatrix} \qquad \text{[state or transition equation(s)]}$$

I.e.

$$Y_t = \mathbf{B}^T \mathbf{S}_t + \epsilon_t \qquad \text{[measurement]}$$

$$\mathbf{S}_t = \mathbf{C}\mathbf{S}_{t-1} + \mathbf{H}_t \qquad \text{[transition]}$$

$$Y_t = \mathbf{B}^T \mathbf{S}_t + \epsilon_t \qquad \text{[measurement]}$$

$$\mathbf{S}_t = \mathbf{C}\mathbf{S}_{t-1} + \mathbf{H}_t \qquad \text{[transition]}$$

with

$$\begin{array}{rcllll}
\mathbf{B} & = & [1, 0]^T & \in \mathbb{R}^2 & & \text{[observation model/matrix]} \\
\mathbf{S}_t & = & [\mu_t, \beta_t]^T & \in \mathbb{R}^2, \forall t & & \text{[state or state vector]} \\
\mathbf{C} & = & \left[\begin{array}{cc} 1 & 1 \\ 0 & 1 \end{array}\right] & \in \mathbb{R}^{2\times 2} & & \text{[state transition model/matrix]} \\
\mathbf{H}_t & = & [h_t, z_t]^T & \in \mathbb{R}^2, \forall t & & \text{[process noise]}
\end{array}$$

where $\mathbf{H}_t \sim \mathcal{WN}(\mathbf{0}, \mathbf{V})$, with variance (-covariance) matrix

$$\mathbf{V} = \text{var}(\mathbf{H}_t) = \mathbb{E}(\mathbf{H}_t \mathbf{H}_t^T) = \left[\begin{array}{cc} \mathbb{E}(h_t^2) & \mathbb{E}(h_t z_t) \\ \mathbb{E}(z_t h_t) & \mathbb{E}(h_t^2) \end{array}\right] = \left[\begin{array}{cc} \sigma_h^2 & 0 \\ 0 & \sigma_z^2 \end{array}\right]$$

which is assumed constant w.r.t. time. Also $\epsilon_t \sim \mathcal{WN}(0, \sigma^2)$.

'Generally' (for our purposes)

$$Y_t = \mathbf{B}^T \mathbf{S}_t + \epsilon_t \qquad \text{[measurement]}$$

$$\mathbf{S}_t = \mathbf{C}\mathbf{S}_{t-1} + \mathbf{H}_t \qquad \text{[transition]}$$

with

observable at time t	Y_t	$\in \mathbb{R}, \forall t$	[observation]
known	$\mathbf{B}$	$\in \mathbb{R}^n$	[observation model/matrix]
unknown	$\mathbf{S}_t$	$\in \mathbb{R}^n, \forall t$	[state or state vector]
known	$\mathbf{C}$	$\in \mathbb{R}^{n \times n}$	[state transition model/matrix]
mean & variance known	$\mathbf{H}_t$	$\in \mathbb{R}^n, \forall t$	[process noise]

where $\mathbf{H}_t \sim \mathcal{WN}(\mathbf{0}, \mathbf{V})$ and $\epsilon_t \sim \mathcal{WN}(0, \sigma^2)$ are uncorrelated with all other terms and where $\mathbf{V}$ and σ^2 are known.

State space models are very general and flexible. *SARIMA* type models merely constitute one specific subclass.

$$Y_t = \mathbf{B}^T\mathbf{S}_t + \epsilon_t \qquad \text{[measurement]}$$

$$\mathbf{S}_t = \mathbf{C}\mathbf{S}_{t-1} + \mathbf{H}_t \qquad \text{[transition]}$$

- Kalman filter uses predictor/corrector recursive approach to calculate optimal estimator (predictor) of state $\mathbf{S}_t$.
- A key idea in the Kalman filter development is to compute posterior estimate $\hat{\mathbf{S}}_{t|t}$ as an 'optimal' linear combination of prior estimate $\hat{\mathbf{S}}_{t|t-1}$ and a weighted difference between actual measurement Y_t and measurement prediction $\mathbf{B}^T\hat{\mathbf{S}}_{t|t-1}$.
- Unlike the forecasting approach discussed in previous lectures, there is no need to wait for T-many observations before we start making predictions.
- It is a very general, flexible, method with myriad extensions.
- Has been used in: autopilot and navigation systems (e.g. **Apollo programme**), economics, radar tracking, speech enhancement, weather forecasting, plus too many more to mention!

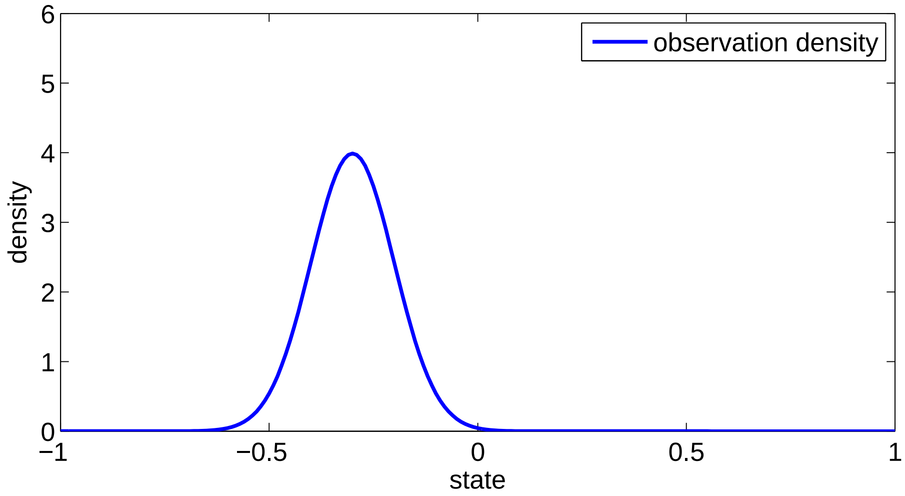


Figure: Make prediction (prior) with uncertainty; observe data drawn from some distribution; update prediction (posterior) as ‘optimal’ linear combination of prediction and observation. Note that the updated variance is smaller than observed or prior. In this example (local-level model), the mean of the posterior is used as the mean of the prior estimate in the next step...

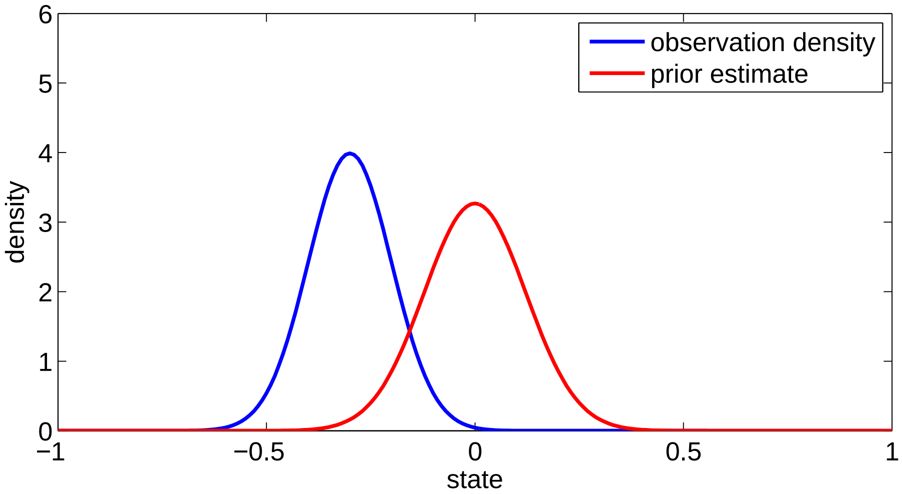


Figure: Make prediction (prior) with uncertainty; observe data drawn from some distribution; update prediction (posterior) as 'optimal' linear combination of prediction and observation. Note that the updated variance is smaller than observed or prior. In this example (local-level model), the mean of the posterior is used as the mean of the prior estimate in the next step...

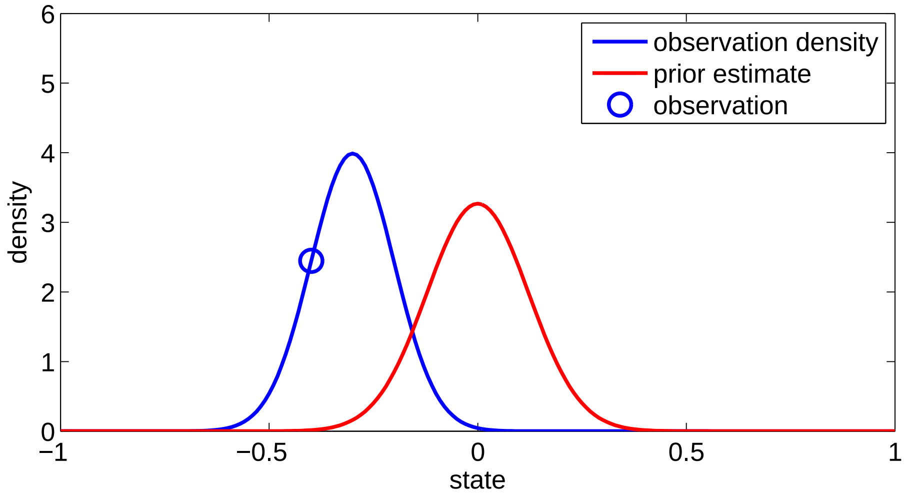


Figure: Make prediction (prior) with uncertainty; observe data drawn from some distribution; update prediction (posterior) as ‘optimal’ linear combination of prediction and observation. Note that the updated variance is smaller than observed or prior. In this example (local-level model), the mean of the posterior is used as the mean of the prior estimate in the next step...

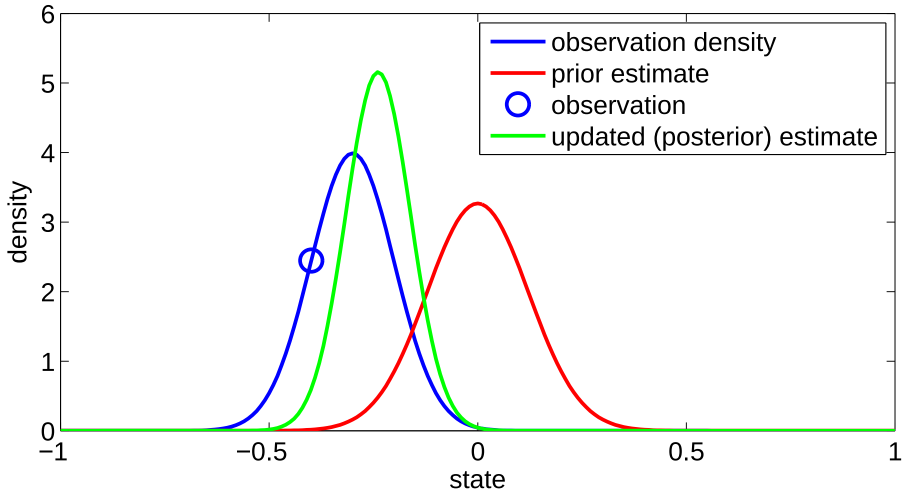


Figure: Make prediction (prior) with uncertainty; observe data drawn from some distribution; update prediction (posterior) as 'optimal' linear combination of prediction and observation. Note that the updated variance is smaller than observed or prior. In this example (local-level model), the mean of the posterior is used as the mean of the prior estimate in the next step...

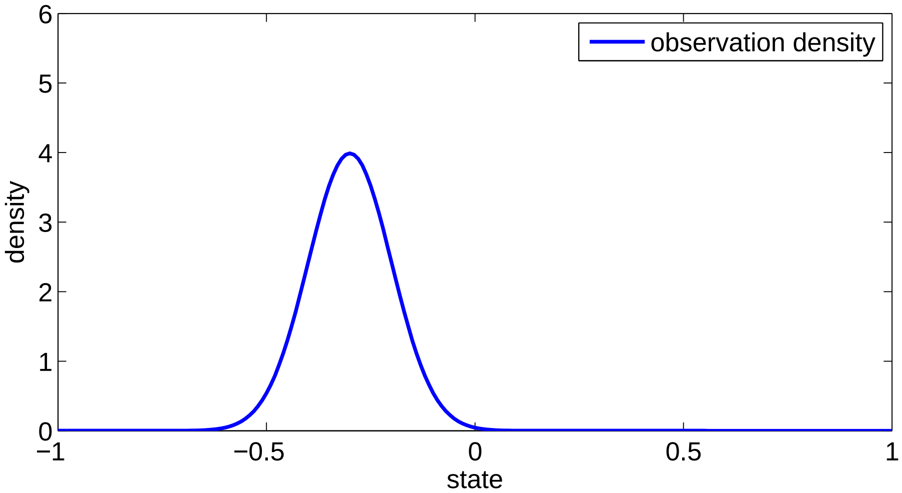


Figure: Prior based on transition model and posterior density from previous step (add on uncertainty); observe data (in this example, the observation distribution is fixed); update posterior as before. Note that the updated variance is smaller than observed or prior and it is smaller than the posterior estimate in the previous step.

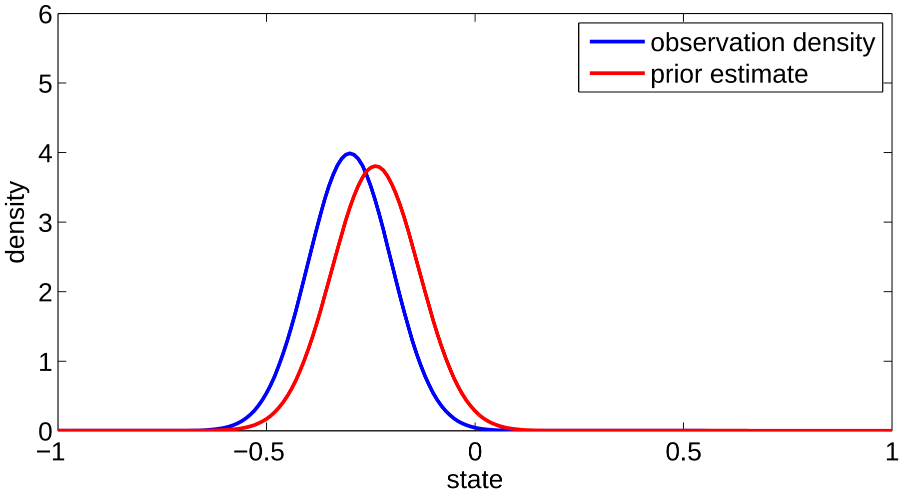


Figure: Prior based on transition model and posterior density from previous step (add on uncertainty); observe data (in this example, the observation distribution is fixed); update posterior as before. Note that the updated variance is smaller than observed or prior and it is smaller than the posterior estimate in the previous step.

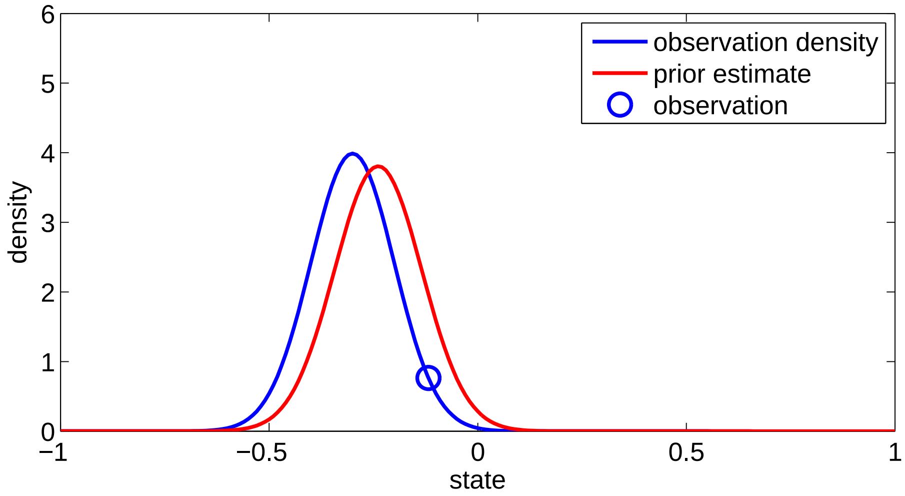


Figure: Prior based on transition model and posterior density from previous step (add on uncertainty); observe data (in this example, the observation distribution is fixed); update posterior as before. Note that the updated variance is smaller than observed or prior and it is smaller than the posterior estimate in the previous step.

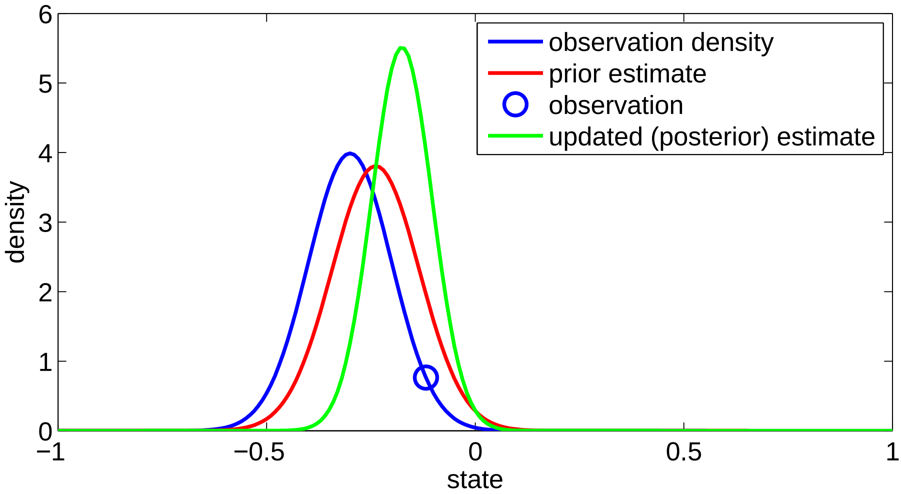
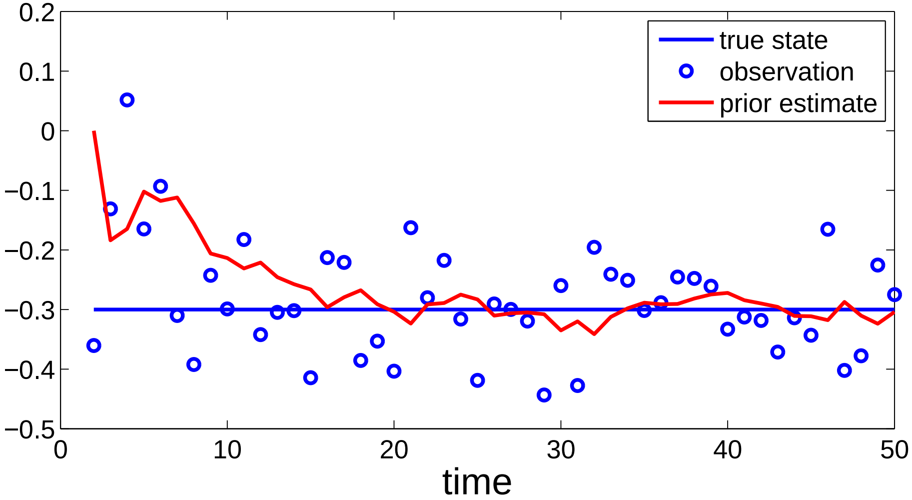


Figure: Prior based on transition model and posterior density from previous step (add on uncertainty); observe data (in this example, the observation distribution is fixed); update posterior as before. Note that the updated variance is smaller than observed or prior and it is smaller than the posterior estimate in the previous step.



true state
observation
prior estimate
0.2
0.1
0
−0.1
−0.2
−0.3
−0.4
−0.5
0
10
20
30
40
50
time

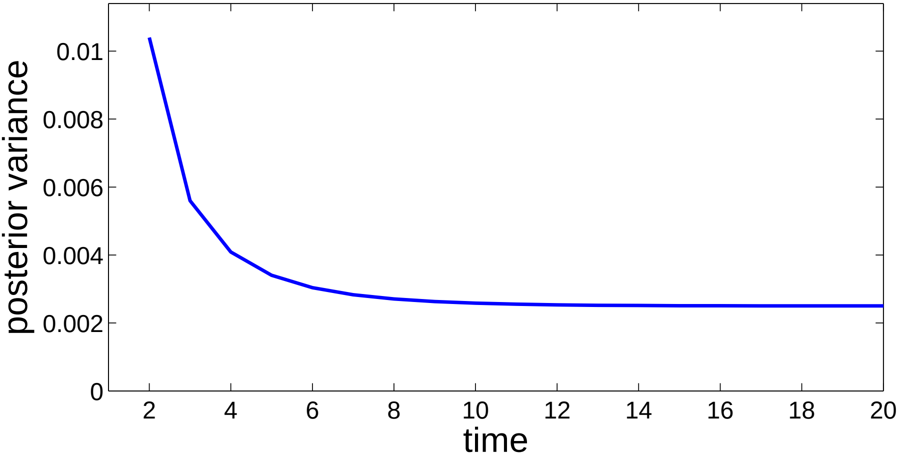


Figure: Posterior error variance decreases to equilibrium over time

$$Y_t = \mathbf{B}^T\mathbf{S}_t + \epsilon_t \qquad \text{[measurement]}$$

$$\mathbf{S}_t = \mathbf{C}\mathbf{S}_{t-1} + \mathbf{H}_t \qquad \text{[transition]}$$

Kalman filter uses predictor/corrector recursive approach to calculate optimal estimator (predictor) of state $\mathbf{S}_t$.

Definition 6 (prior state estimate)

Define $\hat{\mathbf{S}}_{t|t-1}$ *as (prior) estimate of (state)* $\mathbf{S}_t$*, given observations* $Y_{1:t-1}$.

Prior state estimate can be used to predict Y_t using $\hat{Y}_{t|t-1} = \mathbf{B}^T\hat{\mathbf{S}}_{t|t-1}$.

Definition 7 (posterior state estimate)

Define $\hat{\mathbf{S}}_{t|t}$ *as (posterior) estimate of (state)* $\mathbf{S}_t$*, given observations* $Y_{1:t}$.

Definition 8 (prior state prediction)

From model equations

$$\hat{\mathbf{S}}_{t|t-1} = \mathbf{C}\hat{\mathbf{S}}_{t-1|t-1}$$

$$Y_t = \mathbf{B}^T \mathbf{S}_t + \epsilon_t \qquad \text{[measurement]}$$

$$\mathbf{S}_t = \mathbf{C}\mathbf{S}_{t-1} + \mathbf{H}_t \qquad \text{[transition]}$$

Definition 9 (variance of prior state estimate error)

$$\mathbf{P}_{t|t-1} := \text{var}\left(\mathbf{S}_t - \hat{\mathbf{S}}_{t|t-1}\right) = \mathbb{E}\left((\mathbf{S}_{t|t-1} - \hat{\mathbf{S}}_{t|t-1})(\mathbf{S}_{t|t-1} - \hat{\mathbf{S}}_{t|t-1})^T\right)$$

Definition 10 (variance of posterior state estimate error)

$$\mathbf{P}_{t|t} := \text{var}\left(\mathbf{S}_t - \hat{\mathbf{S}}_{t|t}\right) = \mathbb{E}\left((\mathbf{S}_{t|t} - \hat{\mathbf{S}}_{t|t})(\mathbf{S}_{t|t} - \hat{\mathbf{S}}_{t|t})^T\right)$$

Definition 11 (measurement residual)

$$\widetilde{Y}_t = Y_t - \hat{Y}_{t|t-1} = Y_t - \mathbf{B}^T \hat{\mathbf{S}}_{t|t-1}$$

In the interests of the recursive approach explained fully later, the prior state prediction error variance, for time t, can be computed from the posterior error variance at time $t-1$.

$$
\begin{aligned}
\mathbf{P}_{t|t-1} &= \text{var}\left(\mathbf{S}_t - \hat{\mathbf{S}}_{t|t-1}\right) \\
&= \text{var}\left(\mathbf{C}\mathbf{S}_{t-1} + \mathbf{H}_t - \hat{\mathbf{S}}_{t|t-1}\right) \quad \text{[from transition eqn.]} \\
&= \text{var}\left(\mathbf{C}\mathbf{S}_{t-1} + \mathbf{H}_t - \mathbf{C}\hat{\mathbf{S}}_{t-1|t-1}\right) \quad \text{[from prior state prediction, defn. 8]} \\
&= \text{var}\left(\mathbf{C}(\mathbf{S}_{t-1} - \hat{\mathbf{S}}_{t-1|t-1}) + \mathbf{H}_t\right) \\
&= \text{var}\left(\mathbf{C}(\mathbf{S}_{t-1} - \hat{\mathbf{S}}_{t-1|t-1})\right) + \text{var}\left(\mathbf{H}_t\right) \quad \text{[process noise uncorrelated with other terms]} \\
&= \mathbf{C}\,\text{var}\left(\mathbf{S}_{t-1} - \hat{\mathbf{S}}_{t-1|t-1}\right)\mathbf{C}^T + \mathbf{V} \quad \text{[revision section in course notes]} \\
&= \mathbf{C}\mathbf{P}_{t-1|t-1}\mathbf{C}^T + \mathbf{V}
\end{aligned}
$$

Definition 12 (prior state prediction error variance)

$$\mathbf{P}_{t|t-1} = \mathbf{C}\mathbf{P}_{t-1|t-1}\mathbf{C}^T + \mathbf{V}$$

A key idea in the Kalman filter development is to compute posterior estimate $\hat{\mathbf{S}}_{t|t}$ as a linear combination of prior estimate $\hat{\mathbf{S}}_{t|t-1}$ and a weighted difference between actual measurement Y_t and measurement prediction $\mathbf{B}^T\hat{\mathbf{S}}_{t|t-1}$:

Definition 13 (updated posterior state estimate)

$$\hat{\mathbf{S}}_{t|t} = \hat{\mathbf{S}}_{t|t-1} + \mathbf{K}_t \left(Y_t - \mathbf{B}^T\hat{\mathbf{S}}_{t|t-1} \right)$$

Remark 14

Note: from defn. of measurement residual (Defn. 11) the updated state estimate can be written as $\hat{\mathbf{S}}_{t|t} = \hat{\mathbf{S}}_{t|t-1} + \mathbf{K}_t \widetilde{Y}_t$.

Remark 15

The matrix $\mathbf{K}_t$ *is chosen to minimise the posterior state estimate error variance*

$$\mathbf{P}_{t|t} := \mathrm{var}\left(\mathbf{S}_t - \hat{\mathbf{S}}_{t|t} \right) .$$

Now want to find an expression for posterior error variance $\mathbf{P}_{t|t}$ in terms of prior error variance $\mathbf{P}_{t|t-1}$.

$$
\begin{aligned}
\mathbf{P}_{t|t} &= \text{var}\left(\mathbf{S}_t - \hat{\mathbf{S}}_{t|t}\right) \quad \text{[defn. 10]} \\
&= \text{var}\left(\mathbf{S}_t - (\hat{\mathbf{S}}_{t|t-1} + \mathbf{K}_t(Y_t - \mathbf{B}^T\hat{\mathbf{S}}_{t|t-1}))\right) \quad \text{[defn. 13]} \\
&= \text{var}\left(\mathbf{S}_t - (\hat{\mathbf{S}}_{t|t-1} + \mathbf{K}_t(\mathbf{B}^T\mathbf{S}_t + \epsilon_t - \mathbf{B}^T\hat{\mathbf{S}}_{t|t-1}))\right) \quad \text{[measurement equation]} \\
&= \text{var}\left(\mathbf{S}_t - \hat{\mathbf{S}}_{t|t-1} - \mathbf{K}_t\mathbf{B}^T\mathbf{S}_t - \mathbf{K}_t\epsilon_t + \mathbf{K}_t\mathbf{B}^T\hat{\mathbf{S}}_{t|t-1}\right) \\
&= \text{var}\left(\mathbf{S}_t - \hat{\mathbf{S}}_{t|t-1} - \mathbf{K}_t\mathbf{B}^T(\mathbf{S}_t - \hat{\mathbf{S}}_{t|t-1}) - \mathbf{K}_t\epsilon_t\right) \\
&= \text{var}\left((\mathbf{I} - \mathbf{K}_t\mathbf{B}^T)(\mathbf{S}_t - \hat{\mathbf{S}}_{t|t-1}) - \mathbf{K}_t\epsilon_t\right) \\
&= \text{var}\left((\mathbf{I} - \mathbf{K}_t\mathbf{B}^T)(\mathbf{S}_t - \hat{\mathbf{S}}_{t|t-1})\right) + \text{var}\left(\mathbf{K}_t\epsilon_t\right) \quad \text{[measurement error uncorrelated with other terms]} \\
&= (\mathbf{I} - \mathbf{K}_t\mathbf{B}^T)\,\text{var}\left(\mathbf{S}_t - \hat{\mathbf{S}}_{t|t-1}\right)(\mathbf{I} - \mathbf{K}_t\mathbf{B}^T)^T + \mathbf{K}_t\,\text{var}\left(\epsilon_t\right)\mathbf{K}_t^T \\
&= (\mathbf{I} - \mathbf{K}_t\mathbf{B}^T)\mathbf{P}_{t|t-1}(\mathbf{I} - \mathbf{K}_t\mathbf{B}^T)^T + \sigma_\epsilon^2\mathbf{K}_t\mathbf{K}_t^T \quad \text{[defn. 9]}
\end{aligned}
$$

$$\begin{aligned}
\mathbf{P}_{t|t} &= (\mathbf{I}-\mathbf{K}_t\mathbf{B}^T)\mathbf{P}_{t|t-1}(\mathbf{I}-\mathbf{K}_t\mathbf{B}^T)^T + \sigma_\epsilon^2\mathbf{K}_t\mathbf{K}_t^T \\
&= \mathbf{P}_{t|t-1} - \mathbf{K}_t\mathbf{B}^T\mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{B}\mathbf{K}_t^T + \mathbf{K}_t\mathbf{B}^T\mathbf{P}_{t|t-1}\mathbf{B}\mathbf{K}_t^T + \sigma_\epsilon^2\mathbf{K}_t\mathbf{K}_t^T \\
&= \mathbf{P}_{t|t-1} - \mathbf{K}_t\mathbf{B}^T\mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{B}\mathbf{K}_t^T + \mathbf{K}_t(\mathbf{B}^T\mathbf{P}_{t|t-1}\mathbf{B} + \sigma_\epsilon^2)\mathbf{K}_t^T \quad (6)
\end{aligned}$$

The trace (sum of diagonals) of $\mathbf{P}_{t|t}$ can be miminimised by finding $\mathbf{K}_t$ s.t.

$$\frac{\partial\,\mathrm{tr}\left(\mathbf{P}_{t|t}\right)}{\partial\mathbf{K}_t} = 0\,. \qquad \text{[c.f. least squares minimisation]}$$

It can be shown that the solution is...

Definition 16 (Optimal Kalman gain)

$$\boxed{\mathbf{K}_t = \frac{\mathbf{P}_{t|t-1}\mathbf{B}}{\mathbf{B}^T\mathbf{P}_{t|t-1}\mathbf{B} + \sigma_\epsilon^2}}$$

Note denominator (for our purposes) is a number $\in \mathbb{R}$ (it is what the lecture notes calls f_t)

Recall Eqn. (6)

$$\mathbf{P}_{t|t-1} - \mathbf{K}_t\mathbf{B}^T\mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{B}\mathbf{K}_t^T + \mathbf{K}_t(\mathbf{B}^T\mathbf{P}_{t|t-1}\mathbf{B} + \sigma_\epsilon^2)\mathbf{K}_t^T$$

Note, from Kalman gain (defn. 16):

$$\mathbf{K}_t = \frac{\mathbf{P}_{t|t-1}\mathbf{B}}{\mathbf{B}^T\mathbf{P}_{t|t-1}\mathbf{B} + \sigma_\epsilon^2}$$

we have (multiply both sides by denominator and post multiply by $\mathbf{K}_t^T$)

$$\mathbf{K}_t(\mathbf{B}^T\mathbf{P}_{t|t-1}\mathbf{B} + \sigma_\epsilon^2)\mathbf{K}_t^T = \mathbf{P}_{t|t-1}\mathbf{B}\mathbf{K}_t^T\,.$$

Substituting this into Eqn. (6) gives:

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{K}_t\mathbf{B}^T\mathbf{P}_{t|t-1} - \cancel{\mathbf{P}_{t|t-1}\mathbf{B}\mathbf{K}_t^T} + \cancel{\mathbf{P}_{t|t-1}\mathbf{B}\mathbf{K}_t^T}$$

Definition 17 (updated posterior estimate covariance)

$$\boxed{\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{K}_t\mathbf{B}^T\mathbf{P}_{t|t-1}}$$

Remark 18

Recall Kalman gain:

$$\mathbf{K}_t = \frac{\mathbf{P}_{t|t-1}\mathbf{B}}{\mathbf{B}^T\mathbf{P}_{t|t-1}\mathbf{B} + \sigma_\epsilon^2},$$

and state update equation

$$\hat{\mathbf{S}}_{t|t} = \hat{\mathbf{S}}_{t|t-1} + \mathbf{K}_t\left(Y_t - \mathbf{B}^T\hat{\mathbf{S}}_{t|t-1}\right) = \hat{\mathbf{S}}_{t|t-1} + \mathbf{K}_t\left(Y_t - \hat{Y}_{t|t-1}\right).$$

- *If we have confidence in the measurement then the measurement noise is small (i.e. if σ_ϵ^2 is small) and the Kalman gain puts heavier weight onto the residuals.*
- *As we become more confident about the predictions, then the prediction variance $\mathbf{P}_{t|t-1}$ becomes small and the Kalman gain weights the residuals less (i.e. the prediction $\hat{\mathbf{S}}_{t|t-1}$ is weighted relatively more heavily).*

Definition 19 (initial estimates)

Either prior information provides initial state estimate and variance. Otherwise, $\hat{\mathbf{S}}_{0|0} = \mathbf{0}$ *and* $\mathbf{P}_{0|0} = \mathbf{V}$*, or* $\mathbf{P}_{0|0} = \lambda\mathbf{I}$*, (where* λ *sufficiently large) are often used.*

Definition 20 (prediction equations)

$$\hat{\mathbf{S}}_{t|t-1} = \mathbf{C}\hat{\mathbf{S}}_{t-1|t-1}$$

From Defn. 8: prior state prediction

$$\mathbf{P}_{t|t-1} = \mathbf{C}\mathbf{P}_{t-1|t-1}\mathbf{C}^T + \mathbf{V}$$

From Defn. 12: prior state prediction error variance

Definition 21 (update equations)

$$\mathbf{K}_t = \frac{\mathbf{P}_{t|t-1}\mathbf{B}}{\mathbf{B}^T\mathbf{P}_{t|t-1}\mathbf{B} + \sigma_\epsilon^2}$$

From Defn. 16: Optimal Kalman gain

$$\hat{\mathbf{S}}_{t|t} = \hat{\mathbf{S}}_{t|t-1} + \mathbf{K}_t\left(Y_t - \mathbf{B}^T\hat{\mathbf{S}}_{t|t-1}\right)$$

From Defn. 13: updated posterior state estimate

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{K}_t\mathbf{B}^T\mathbf{P}_{t|t-1}$$

From Defn. 17: updated posterior estimate covariance

Example 22

Consider the following state space model

$$\begin{aligned} Y_t &= \mathbf{B}^T \mathbf{S}_t \\ \mathbf{S}_t &= \mathbf{C}\mathbf{S}_{t-1} + \mathbf{H}_t \,, \end{aligned}$$

with

$$\mathbf{S}_t = \left[\mu_{1,t}\,,\ \mu_{2,t}\right]^T \,, \qquad \mathbf{B} = [0, 1]^T$$

$$\mathbf{C} = \begin{bmatrix} \zeta_1 & 1 \\ \zeta_2 & 0 \end{bmatrix} , \qquad \mathbf{H}_t = \begin{bmatrix} h_t \\ 0 \end{bmatrix} , \quad h_t \sim \mathcal{WN}(0, \sigma_h^2)$$

Show Y_t is $AR(2)$.

The model equations say that

$$\begin{aligned} Y_t = \mu_{1,t} &= \zeta_1\,\mu_{1,t-1} + \mu_{2,t-1} + h_t \\ \mu_{2,t} &= \zeta_2\,\mu_{1,t-1} \end{aligned}$$

Substituting the latter into the former furnishes:

$$Y_t = \mu_{1,t} = \zeta_1\,\mu_{1,t-1} + \zeta_2\,\mu_{1,t-2} + h_t \,. \qquad \blacksquare$$

$$Y_t = \mu_{1,t} = \zeta_1 \, \mu_{1,t-1} + \zeta_2 \, \mu_{1,t-2} + h_t$$

But, this is equivalent to a zero mean $AR(2)$ model:

$$Y_t = \phi_1 \, Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t \,,$$

with

$$\zeta_1 = \phi_1 \,, \quad \zeta_2 = \phi_2 \,, \quad h_t = \epsilon_t \,.$$

Example 23

Consider the state space model in Example 22. Let $\sigma_h = 1$, $\phi_1 = 1/2$, $\phi_2 = -1/4$. *Use the initial conditions:*

$$\hat{\mathbf{S}}_{0|0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} , \qquad \mathbf{P}_{0,0} = \mathbf{V},$$

to predict the first observation and prior state error variance.

$$\hat{\mathbf{S}}_{0|0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \qquad \mathbf{P}_{0|0} = \mathbf{V} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Using Kalman filter predictor equations, the prior state estimate is therefore $\hat{\mathbf{S}}_{1|0} = \mathbf{C}\hat{\mathbf{S}}_{0|0} = [0, 0]^T$. Prior state error variance is:

$$\begin{aligned} \mathbf{P}_{1|0} &= \mathbf{C}\mathbf{P}_{0|0}\mathbf{C}^T + \mathbf{V} \\ &= \begin{bmatrix} 1/2 & 1 \\ -1/4 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1/2 & -1/4 \\ 1 & 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 5/4 & -1/8 \\ -1/8 & 1/16 \end{bmatrix} \qquad \blacksquare \end{aligned}$$

Example 24

Consider state space model in Example 22. At time $t = 1$, we observe that $Y_t = 1/2$. Compute Kalman gain, posterior state update, and posterior state error variance.

From Kalman filter equations, the Kalman gain is

$$\mathbf{K}_1 = \frac{\mathbf{P}_{1|0}\mathbf{B}}{\mathbf{B}^T\mathbf{P}_{1|0}\mathbf{B} + \sigma_\epsilon^2}, \qquad \text{[note: in following line that } \sigma = 0]$$

$$= \begin{bmatrix} 5/4 & -1/8 \\ -1/8 & 1/16 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}^T \begin{bmatrix} 5/4 & -1/8 \\ -1/8 & 1/16 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right)^{-1}$$

$$= \begin{bmatrix} 1 \\ -1/10 \end{bmatrix}$$

Posterior state estimate is

$$\begin{aligned}\hat{\mathbf{S}}_{1|1} &= \hat{\mathbf{S}}_{1|0} + \mathbf{K}_1(Y_1 - \mathbf{B}^T\hat{\mathbf{S}}_{1|0}) \\ &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ -1/10 \end{bmatrix} (1/2 - 0) \\ &= \begin{bmatrix} 1/2 \\ -1/20 \end{bmatrix}\end{aligned}$$

Posterior state error variance

$$\begin{aligned}\mathbf{P}_{1|1} &= \mathbf{P}_{1|0} - \mathbf{K}_1\mathbf{B}^T\mathbf{P}_{1|0} \\ &= \begin{bmatrix} 5/4 & -1/8 \\ -1/8 & 1/16 \end{bmatrix} - \begin{bmatrix} 1 \\ -1/10 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}^T \begin{bmatrix} 5/4 & -1/8 \\ -1/8 & 1/16 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 \\ 0 & 1/20 \end{bmatrix}. \qquad \blacksquare\end{aligned}$$