State Space Models in Stan

Jeffrey B. Arnold

2016-12-01

Contents

Chapter 1

Introduction

This contains documentation for "State Space Models in Stan"

Chapter 2

The Linear State Space Model

[**DurbinKoopman2012**]

The linear Gaussian state space model (SSM)[1] the the n-dimensional observation sequence $\boldsymbol{y}_1, \dots, \boldsymbol{y}_n$,

$$
\begin{aligned}
\boldsymbol{y}_t &= \boldsymbol{d}_t + \boldsymbol{Z}_t \boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t, & \boldsymbol{\varepsilon}_t &\sim N(0, \boldsymbol{H}_t), \\
\boldsymbol{\alpha}_{t+1} &= \boldsymbol{c}_t + \boldsymbol{T}_t \boldsymbol{\alpha}_t + \boldsymbol{R}_t \boldsymbol{\eta}_t, & \boldsymbol{\eta}_t &\sim N(0, \boldsymbol{Q}_t), \\
& & \boldsymbol{\alpha}_1 &\sim N(\boldsymbol{a}_1, \boldsymbol{P}_1).
\end{aligned}
$$

for $t = 1, \dots, n$. The first equation is called the *observation* or *measurement equation*. The second equation is called the *state*, *transition*, or *system equation*. The vector $\boldsymbol{y}_t$ is a $p \times 1$ vector called the *observation vector*. The vector $\alpha\alpha_t$ is a $m \times 1$ vector called the *state vector*. The matrices are vectors, $\boldsymbol{Z}_t$,$\boldsymbol{T}_t$, $\boldsymbol{R}_t$, $\boldsymbol{H}_t$, $\boldsymbol{Q}_t$, c_t, and d_t are called the *system matrices*. The system matrices are considered fixed and known in the filtering and smoothing equations below, but can be parameters themselves. The $p \times m$ matrix $\boldsymbol{Z}_t$ links the observation vector $\boldsymbol{y}_t$ with the state vector $\boldsymbol{\alpha}_t$. The $m \times m$ transition matrix $\boldsymbol{T}_t$ determines the evolution of the state vector, $\boldsymbol{\alpha}_t$. The $q \times 1$ vector $\boldsymbol{\eta}_t$ is called the *state disturbance vector*, and the $p \times 1$ vector $\boldsymbol{\varepsilon}_t$ is called the *observation disturbance vector*. An assumption is that the state and observation disturbance vectors are uncorrelated, $\mathrm{Cov}(\boldsymbol{\varepsilon}_t, \boldsymbol{\eta}_t) = 0$.

In a general state space model, the normality assumptions of the densities of $\boldsymbol{\varepsilon}$ and $\boldsymbol{\eta}$ are dropped.

In many cases $\boldsymbol{R}_t$ is the identity matrix. It is possible to define $\eta^*_t = \boldsymbol{R}_t \boldsymbol{\eta}_t$, and $\boldsymbol{Q}^* = \boldsymbol{R}_t \boldsymbol{Q}'_t \boldsymbol{R}'_t$. However, if $\boldsymbol{R}_t$ is $m \times q$ and $q < m$, and $\boldsymbol{Q}_t$ is nonsingular, then it is useful to work with the nonsingular $\boldsymbol{\eta}_t$ rather than a singular $\boldsymbol{\eta}^*_t$.

The initial state vector $\boldsymbol{\alpha}_1$ is assume to be generated as,

$$
\alpha_1 \sim N(\boldsymbol{a}_1, \boldsymbol{P}_1)
$$

independently of the observation and state disturbances $\boldsymbol{\varepsilon}$ and $\boldsymbol{\eta}$. The values of $\boldsymbol{a}_1$ and $\boldsymbol{P}_1$ can be considered as given and known in most stationary processes. When the process is nonstationary, the elements of $\boldsymbol{a}_1$ need to be treated as unknown and estimated. This is called *initialization*.

Table 2.1: Dimensions of matrices and vectors in the SSM

matrix/vector	dimension	name
$\boldsymbol{y}_t$	$p \times 1$	observation vector
$\boldsymbol{\alpha}_t$	$m \times 1$	(unobserved) state vector
$\boldsymbol{\varepsilon}_t$	$m \times 1$	observation disturbance (error)
$\boldsymbol{\eta}_t$	$q \times 1$	state disturbance (error)

[1] This is also called a dynamic linear model (DLM).

matrix/vector	dimension	name
$\boldsymbol{a}_1$	$m \times 1$	initial state mean
$\boldsymbol{c}_t$	$m \times 1$	state intercept
$\boldsymbol{d}_t$	$p \times 1$	observation intercept
$\boldsymbol{Z}_t$	$p \times m$	design matrix
$\boldsymbol{T}_t$	$m \times m$	transition matrix
$\boldsymbol{H}_t$	$p \times p$	observation covariance matrix
$\boldsymbol{R}_t$	$m \times q$	state covariance selection matrix
$\boldsymbol{Q}_t$	$q \times q$	state covariance matrix
$\boldsymbol{P}_1$	$m \times m$	initial state covariance matrix

Chapter 3

Filtering and Smoothing

3.1 Filtering

From [**DurbinKoopman2012**]

Let $\boldsymbol{a}_t = \mathrm{E}(\boldsymbol{\alpha}_t | y_{1,\dots,t-1})$ be the expected value $\boldsymbol{P}_t = \mathrm{Var}(\boldsymbol{\alpha}_t | y_{1,\dots,t-1})$ be the variance of the state in $t+1$ given data up to time t. To calculate $\boldsymbol{\alpha}_{t+1}$ and $\boldsymbol{P}_{t+1}$ given the arrival of new data at time t,

$$
\begin{aligned}
\boldsymbol{v}_t &= \boldsymbol{y}_t - \boldsymbol{Z}_t \boldsymbol{a}_t - \boldsymbol{d}_t, \\
\boldsymbol{F}_t &= \boldsymbol{Z}_t \boldsymbol{P}_t \boldsymbol{Z}_t' + \boldsymbol{H}_t, \\
\boldsymbol{K}_t &= \boldsymbol{T}_t \boldsymbol{P}_t \boldsymbol{Z}_t' \boldsymbol{F}_t^{-1} \\
\boldsymbol{a}_{t+1} &= \boldsymbol{T}_t \boldsymbol{a}_t + \boldsymbol{K}_t \boldsymbol{v}_t + \boldsymbol{c}_t \\
\boldsymbol{P}_{t+1} &= \boldsymbol{T}_t \boldsymbol{P}_t (\boldsymbol{T}_t - \boldsymbol{K}_t \boldsymbol{Z}_t)' + \boldsymbol{R}_t \boldsymbol{Q}_t \boldsymbol{R}_t'.
\end{aligned}
$$

The vector $\boldsymbol{v}_t$ are the *one-step ahead forecast errors$, and the matrix $\boldsymbol{K}_t$ is called the* Kalman gain*.

The filter can also be written to estimate the *filtered states*, where $\boldsymbol{a}_{t|t} = \mathrm{E}(\boldsymbol{\alpha}_t | y_{1,\dots,t})$ is the expected value and $\boldsymbol{P}_{t|t} = \mathrm{Var}(\boldsymbol{\alpha}_t | y_{1,\dots,t})$ is the variance of the state $\boldsymbol{\alpha}_t$ given information up to *and including* $\boldsymbol{y}_t$. The filter written this way is,

$$
\begin{aligned}
\boldsymbol{v}_t &= \boldsymbol{y}_t - \boldsymbol{Z}_t \boldsymbol{a}_t - \boldsymbol{d}_t, \\
\boldsymbol{F}_t &= \boldsymbol{Z}_t \boldsymbol{P}_t \boldsymbol{Z}_t' + \boldsymbol{H}_t, \\
\boldsymbol{a}_{t|t} &= \boldsymbol{a}_t + \boldsymbol{P}_t \boldsymbol{Z}_t' \boldsymbol{F}_t^{-1} v_t, \\
\boldsymbol{P}_{t|t} &= \boldsymbol{P}_t - \boldsymbol{P}_t \boldsymbol{Z}_t' \boldsymbol{F}_t^{-1} \boldsymbol{Z}_t \boldsymbol{P}_t, \\
\boldsymbol{a}_{t+1} &= \boldsymbol{T}_t \boldsymbol{a}_{t|t} + \boldsymbol{c}_t, \\
\boldsymbol{P}_{t+1} &= \boldsymbol{T}_t \boldsymbol{P}_{t|t} \boldsymbol{T}_t' + \boldsymbol{R}_t \boldsymbol{Q}_t \boldsymbol{R}_t'.
\end{aligned}
$$

Table 3.1: Dimensions of matrices and vectors in the SSM

matrix/vector	dimension
$\boldsymbol{v}_t$	$p \times 1$
$\boldsymbol{a}_t$	$m \times 1$
$\boldsymbol{a}_{t\|t}$	$m \times 1$
$\boldsymbol{F}_t$	$p \times p$
$\boldsymbol{K}_t$	$m \times p$
$\boldsymbol{P}_t$	$m \times m$
$\boldsymbol{P}_{t\|T}$	$m \times m$

matrix/vector	dimension
$\boldsymbol{x}_t$	$m \times 1$
$\boldsymbol{L}_t$	$m \times m$

See [**DurbinKoopman2012**]: For a time-invariant state space model, the Kalman recursion for $\boldsymbol{P}_{t+1}$ converges to a constant matrix $\bar{\boldsymbol{P}}$,

$$\bar{\boldsymbol{P}} = \boldsymbol{T}\bar{\boldsymbol{P}}\boldsymbol{T}' - \boldsymbol{T}\bar{\boldsymbol{P}}\boldsymbol{Z}'\bar{\boldsymbol{F}}^{-1}\boldsymbol{Z}\bar{\boldsymbol{P}}\boldsymbol{T}' + \boldsymbol{R}\boldsymbol{Q}\boldsymbol{R}',$$

where $\bar{\boldsymbol{F}} = \boldsymbol{Z}\bar{\boldsymbol{P}}\boldsymbol{Z}' + \boldsymbol{H}$.

See [**DurbinKoopman2012**]: The *state estimation error* is,

$$\boldsymbol{x}_t = \boldsymbol{\alpha}_t - \boldsymbol{a}_t,$$

where $\mathrm{Var}(\boldsymbol{x}_t) = \boldsymbol{P}_t$. The v_t are sometimes called *innovations*, since they are the part of $\boldsymbol{y}_t$ not predicted from the past. The innovation analog of the state space model is

$$\begin{aligned}
\boldsymbol{v}_t &= \boldsymbol{Z}_t\boldsymbol{x}_t + \boldsymbol{\varepsilon}_t, \\
\boldsymbol{x}_{t+1} &= \boldsymbol{L}\boldsymbol{x}_t + \boldsymbol{R}_t\boldsymbol{\eta}_t - \boldsymbol{K}_t\boldsymbol{\varepsilon}_t, \\
\boldsymbol{K}_t &= \boldsymbol{T}_t\boldsymbol{P}_t\boldsymbol{Z}_t'\boldsymbol{F}_t^{-1}, \\
\boldsymbol{L}_t &= \boldsymbol{T}_t - \boldsymbol{K}_t\boldsymbol{Z}_t, \boldsymbol{P}_{t+1} && = \boldsymbol{T}_t\boldsymbol{P}_t\boldsymbol{L}_t' + \boldsymbol{R}_t\boldsymbol{Q}_t\boldsymbol{R}_T'.
\end{aligned}$$

These recursions allow for a simpler derivation of $\boldsymbol{P}_{t+1}$, and are useful for the smoothing recursions. Moreover, the one-step ahead forecast errors are indendendent, which allows for a simple derivation of the log-likelihood.

Alternative methods **TODO**

- square-root filtering
- precision filters
- sequential filtering

3.2 Smoothing

While filtering calculates the conditional densities the states and disturbances given data prior to or up to the current time, smoothing calculates the conditional densities states and disturbances given the entire series of observations, $\boldsymbol{y}_{1:n}$.

State smoothing calculates the conditional mean, $\hat{\boldsymbol{\alpha}}_t = \mathrm{E}(\boldsymbol{\alpha}_t|\boldsymbol{y}_{1:n})$, and variance, $\boldsymbol{V}_t = \mathrm{Var}(\boldsymbol{\alpha}_t|\boldsymbol{y}_{1:n})$, of the states. Observation disturbance smoothing calculates the conditional mean, $\hat{\boldsymbol{\varepsilon}}_t = \mathrm{E}(\boldsymbol{\varepsilon}_t|\boldsymbol{y}_{1:n})$, and variance, $\mathrm{Var}(\boldsymbol{\varepsilon}_t|\boldsymbol{y}_{1:n})$, of the state disturbances. Likewise, state disturbance smoothing calculates the conditional mean, $\hat{\boldsymbol{\eta}}_t = \mathrm{E}(\boldsymbol{\eta}_t|\boldsymbol{y}_{1:n})$, and variance, $\mathrm{Var}(\boldsymbol{\eta}_t|\boldsymbol{y}_{1:n})$, of the state disturbances.

Table 3.2: Dimensions of vectors and matrices used in smoothing recursions

Vector/Matrix	Dimension
$\boldsymbol{r}_t$	$m \times 1$
$\boldsymbol{\alpha}_t$	$m \times 1$
$\boldsymbol{u}_t$	$p \times 1$
$\hat{\boldsymbol{\varepsilon}}_t$	$p \times 1$
$\hat{\boldsymbol{\eta}}_t$	$r \times 1$
$\boldsymbol{N}_t$	$m \times m$
$\boldsymbol{V}_t$	$m \times m$
$\boldsymbol{D}_t$	$p \times p$

3.2.1 State Smoothing

Smoothing calculates conditional density of the states given all observations, $p(\boldsymbol{\alpha}|\boldsymbol{y}_{1:n})$. Let $\hat{\boldsymbol{\alpha}} = \mathrm{E}(\boldsymbol{\alpha}_t|\boldsymbol{y}_{1:n})$ be the mean and $\boldsymbol{V}_t = \mathrm{Var}(\boldsymbol{\alpha}|\boldsymbol{y}_{1:n})$ be the variance of this density. The following recursions can be used to calculate these densities [**DurbinKoopman2012**],

$$
\begin{aligned}
\boldsymbol{r}_{t-1} &= \boldsymbol{Z}_t' \boldsymbol{F}_t^{-1} \boldsymbol{v}_t + \boldsymbol{L}_t' \boldsymbol{r}_t, & \boldsymbol{N}_{t-1} &= \boldsymbol{Z}_t' \boldsymbol{F}_t^{-1} \boldsymbol{Z}_t + \boldsymbol{L}_t' \boldsymbol{N}_t \boldsymbol{L}_t, \\
\hat{\boldsymbol{\alpha}}_t &= \boldsymbol{a}_t + \boldsymbol{P}_t \boldsymbol{r}_{t-1}, & \boldsymbol{V}_t &= \boldsymbol{P}_t - \boldsymbol{P}_t \boldsymbol{N}_{t-1} \boldsymbol{P}_t,
\end{aligned}
$$

for $t = n, \dots, 1$, with $\boldsymbol{r}_n = \boldsymbol{0}$, and $\boldsymbol{N}_n = \boldsymbol{0}$.

During the filtering pass $\boldsymbol{v}_t$, $\boldsymbol{F}_t$, $\boldsymbol{K}_t$, and $\boldsymbol{P}_t$ for $t = 1, \dots, n$ need to be stored. Alternatively, $\boldsymbol{a}_t$ and $\boldsymbol{P}_t$ only can be stored, and $\boldsymbol{v}_t$, $\boldsymbol{F}_t$, $\boldsymbol{K}_t$ recalculated on the fly. However, since the dimensions of $\boldsymbol{v}_t$, $\boldsymbol{F}_t$, $\boldsymbol{K}_t$ are usually small relative to $\boldsymbol{a}_t$ and $\boldsymbol{P}_t$ is is usually worth storing them.

3.2.2 Disturbance smoothing

Disturbance smoothing calculates the density of the state and observation disturbances ($\boldsymbol{\eta}_t$ and $\boldsymbol{\varepsilon}_t$) given the full series of observations $\boldsymbol{y}_{1:n}$. Let $\hat{\varepsilon}_t = \mathrm{E}(\boldsymbol{\varepsilon}|\boldsymbol{y}_{1:n})$ be the mean and $\mathrm{Var}(\boldsymbol{\varepsilon}_t|\boldsymbol{y}_{1:n})$ be the variance of the smoothed density of the observation disturbances at time t, $p(\boldsymbol{\varepsilon}_t|\boldsymbol{y}_{1:n})$. Likewise, let $\hat{\boldsymbol{\eta}} = \mathrm{E}(\boldsymbol{\eta}_t|\boldsymbol{y}_{1:n})$ be the mean and $\mathrm{Var}(\boldsymbol{\eta}_t|\boldsymbol{y}_{1:n})$ be the variance of the smoothed density of the state disturbances at time t, $p(\boldsymbol{\eta}_t|\boldsymbol{y}_{1:n})$. The following recursions can be used to calculate these values [**DurbinKoopman2012**]:

$$
\begin{aligned}
\hat{\varepsilon}_t &= \boldsymbol{H}_t(\boldsymbol{F}^{-1}\boldsymbol{v}_t - \boldsymbol{K}_t'\boldsymbol{r}_t), & \mathrm{Var}(\boldsymbol{\varepsilon}_t|\boldsymbol{Y}_n) &= \boldsymbol{H}_t - \boldsymbol{H}_t(\boldsymbol{F}_t^{-1} + \boldsymbol{K}_t'\boldsymbol{N}_t\boldsymbol{K}_t)\boldsymbol{H}_t, \\
\hat{\boldsymbol{\eta}}_t &= \boldsymbol{Q}_t\boldsymbol{R}_t'\boldsymbol{r}_t, & \mathrm{Var}(\boldsymbol{\eta}_t|\boldsymbol{Y}_n) &= \boldsymbol{Q}_t - \boldsymbol{Q}_t\boldsymbol{R}_t'\boldsymbol{N}_t\boldsymbol{R}_t\boldsymbol{Q}_t, \\
\boldsymbol{r}_{t-1} &= \boldsymbol{Z}_t'\boldsymbol{F}_t^{-1}\boldsymbol{v}_t + \boldsymbol{L}_t'\boldsymbol{r}_t, & \boldsymbol{N}_{t-1} &= \boldsymbol{Z}_t'\boldsymbol{F}_t^{-1}\boldsymbol{Z}_t + \boldsymbol{L}_t'\boldsymbol{N}_t\boldsymbol{L}_t
\end{aligned}
$$

Alternatively, these equations can be rewritten as [**DurbinKoopman2012**]:

$$
\begin{aligned}
\hat{\varepsilon}_t &= \boldsymbol{H}_t\boldsymbol{u}_t, & \mathrm{Var}(\boldsymbol{\varepsilon}_t|\boldsymbol{Y}_n) &= \boldsymbol{H}_t - \boldsymbol{H}_t\boldsymbol{D}_t\boldsymbol{H}_t, \\
\hat{\boldsymbol{\eta}}_t &= \boldsymbol{Q}_t\boldsymbol{R}_t'\boldsymbol{r}_t, & \mathrm{Var}(\boldsymbol{\eta}_t|\boldsymbol{Y}_n) &= \boldsymbol{Q}_t - \boldsymbol{Q}_t\boldsymbol{R}_t'\boldsymbol{N}_t\boldsymbol{R}_t\boldsymbol{Q}_t, \\
\boldsymbol{u}_t &= \boldsymbol{F}^{-1}\boldsymbol{v}_t - \boldsymbol{K}_t'\boldsymbol{r}_t, & \boldsymbol{D}_t &= \boldsymbol{F}_t^{-1} + \boldsymbol{K}_t'\boldsymbol{N}_t\boldsymbol{K}_t, \\
\boldsymbol{r}_{t-1} &= \boldsymbol{Z}_t'\boldsymbol{u}_t + \boldsymbol{T}_t'\boldsymbol{r}_t, & \boldsymbol{N}_{t-1} &= \boldsymbol{Z}_t'\boldsymbol{D}_t\boldsymbol{Z}_t + \boldsymbol{T}_t'\boldsymbol{N}_t\boldsymbol{T}_t - \boldsymbol{Z}_t'\boldsymbol{K}_t'\boldsymbol{N}_t\boldsymbol{T}_t - \boldsymbol{T}_t'\boldsymbol{N}_t\boldsymbol{K}_t\boldsymbol{Z}_t.
\end{aligned}
$$

This reformulation can be computationally useful since it relies on the system matrices $\boldsymbol{Z}_t$ and $\boldsymbol{T}_t$ which are often sparse. The disturbance smoothing recursions require only $\boldsymbol{v}_t$, $\boldsymbol{f}_t$, and $\boldsymbol{K}_t$ which are calculated with a forward pass of the Kalman filter. Unlike the state smoother, the disturbance smoothers do not require either the mean ($\boldsymbol{a}_t$) or variance ($\boldsymbol{P}_t$) of the predicted state density.

3.2.3 Fast state smoothing

If the variances of the states do not need to be calculated, then a faster smoothing algorithm can be used (Koopman 1993). The fast state smoother is defined as [**DurbinKoopman2012**],

$$
\begin{aligned}
\hat{\boldsymbol{\alpha}}_t &= \boldsymbol{T}_t\hat{\boldsymbol{\alpha}}_t + \boldsymbol{R}_t\boldsymbol{Q}_t\boldsymbol{R}_t'\boldsymbol{r}_t, \quad t = 2, \dots, n \\
\hat{\boldsymbol{\alpha}}_1 &= \boldsymbol{a}_1 + \boldsymbol{P}_1\boldsymbol{r}_0.
\end{aligned}
$$

The values of $\boldsymbol{r}_t$ come from the recursions in the disturbance smoother.

3.3 Simulation smoothers

Simulation smoothing draws samples of the states, $p(\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_n|\boldsymbol{y}_{1:n})$, or disturbances, $p(\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_n|\boldsymbol{y}_{1:n})$ and $p(\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_n|\boldsymbol{y}_{1:n})$.[^simsmo]

3.3.1 Mean correction simulation smoother

The mean-correction simulation smoother was introduced in **DurbinKoopman2002** . See **DurbinKoopman2012** (Sec 4.9) for an exposition of it. It requires only the previously described filters and smoothers, and generating samples from multivariate distributions.

3.3.1.1 Disturbances

1. Run a filter and disturbance smoother to calculate $\hat{\boldsymbol{\varepsilon}}_{1:n}$ and $\hat{\boldsymbol{\eta}}_{1:(n-1)}$
2. Draw samples from the unconditional distribution of the disturbances,

$$
\begin{aligned}
\boldsymbol{\eta}_t^+ &\sim N(0, \boldsymbol{H}_t) \quad t = 1, \dots, n-1 \\
\boldsymbol{\varepsilon}_t^+ &\sim N(0, \boldsymbol{Q}_t) \quad t = 1, \dots, n
\end{aligned}
$$

3. Simulate observations from the system using the simulated disturbances,

$$
\begin{aligned}
\boldsymbol{y}_t^+ &= \boldsymbol{d}_t + \boldsymbol{Z}_t \boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t^+, \\
\boldsymbol{\alpha}_{t+1} &= \boldsymbol{c}_t + \boldsymbol{T}_t \boldsymbol{\alpha}_t + \boldsymbol{R}_t \boldsymbol{\eta}_t^+,
\end{aligned}
$$

   where $\boldsymbol{\alpha}_1 \sim N(\boldsymbol{a}_1, \boldsymbol{P}_1)$.
4. Run a filter and disturbance smoother on the simulated observations $\boldsymbol{y}^+$ to calculate $\hat{\boldsymbol{\varepsilon}}_t^+ = \mathrm{E}(\boldsymbol{\varepsilon}_t | \boldsymbol{y}_{1:n}^+)$ and $\hat{\boldsymbol{\eta}}_t^+ = \mathrm{E}(\boldsymbol{\eta}_t | \boldsymbol{y}_{1:n}^+)$.
5. A sample from $p(\hat{\boldsymbol{\eta}}_{1:(n-1)}, \hat{\boldsymbol{\varepsilon}}_{1:n} | \boldsymbol{y}_{1:n})$ is

$$
\begin{aligned}
\tilde{\boldsymbol{\eta}}_t &= \boldsymbol{\eta}_t^+ - \hat{\boldsymbol{\eta}}_t^+ + \hat{\boldsymbol{\eta}}_t, \\
\tilde{\boldsymbol{\varepsilon}}_t &= \boldsymbol{\varepsilon}_t^+ - \hat{\boldsymbol{\varepsilon}}_t^+ + \hat{\boldsymbol{\varepsilon}}_t.
\end{aligned}
$$

3.3.1.2 States

1. Run a filter and disturbance smoother to calculate the mean of the states conditional on the full series of observations, $\hat{\boldsymbol{\alpha}}_{1:n} = \mathrm{E}(\boldsymbol{\alpha}_{1:n} | \boldsymbol{y}_{1:n})$.
2. Draw samples from the unconditional distribution of the disturbances,

$$
\begin{aligned}
\boldsymbol{\eta}_t^+ &\sim N(0, \boldsymbol{H}_t) \quad t = 1, \dots, n-1 \\
\boldsymbol{\varepsilon}_t^+ &\sim N(0, \boldsymbol{Q}_t) \quad t = 1, \dots, n
\end{aligned}
$$

3. Simulate states and observations from the system using the simulated disturbances,

$$
\begin{aligned}
\boldsymbol{y}_t^+ &= \boldsymbol{d}_t + \boldsymbol{Z}_t \boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t^+, \\
\boldsymbol{\alpha}_{t+1}^+ &= \boldsymbol{c}_t + \boldsymbol{T}_t \boldsymbol{\alpha}_t + \boldsymbol{R}_t \boldsymbol{\eta}_t^+,
\end{aligned}
$$

   where $\boldsymbol{\alpha}_1^+ \sim N(\boldsymbol{a}_1, \boldsymbol{P}_1)$.
4. Run a filter and smoother on the simulated observations $\boldsymbol{y}^+$ to calculate $\hat{\boldsymbol{\alpha}}_t^+ = \mathrm{E}(\boldsymbol{\alpha}_t | \boldsymbol{y}_{1:n}^+)$.
5. A sample from $p(\hat{\boldsymbol{\alpha}}_{1:n} | \boldsymbol{y}_{1:n})$ is

$$
\tilde{\boldsymbol{\alpha}}_t = \boldsymbol{\alpha}_t^+ - \hat{\boldsymbol{\alpha}}_t^+ + \hat{\boldsymbol{\alpha}}_t.
$$

One convenient feature of this method is that since only the conditional means of the states are required, the fast state smoother can be used, since the variances of the states are not required.

3.3.2 de Jong-Shephard method

These recursions were developed in **DeJongShephard1995** . Although the the mean-correction simulation smoother will work in most cases, there are a few in which it will not work.

TODO

3.3.3 Forward-filter backwards-smoother (FFBS)

This was the simulation method developed in **CarterKohn1994** and **Fruehwirth-Schnatter1994** .

TODO

3.4 Missing observations

When all observations at time t are missing, the filtering recursions become [**DurbinKoopman2012**],

$$\begin{aligned}
\boldsymbol{a}_{t|t} &= \boldsymbol{a}_t, \\
\boldsymbol{P}_{t|t} &= \boldsymbol{P}_t, \\
\boldsymbol{a}_{t+1} &= \boldsymbol{T}_t \boldsymbol{a}_t + \boldsymbol{c}_t \\
\boldsymbol{P}_{t+1} &= \boldsymbol{T}_t \boldsymbol{P}_t \boldsymbol{T}'_t + \boldsymbol{R}_t \boldsymbol{Q}_t \boldsymbol{R}'_t
\end{aligned}$$

This is equivalent to setting $\boldsymbol{Z}_t = \boldsymbol{0}$ (implying also that $\boldsymbol{K}_t = \boldsymbol{0}$) in the filtering equations. For smoothing, also replace $\boldsymbol{Z}_t = \boldsymbol{0}$,

$$\begin{aligned}
\boldsymbol{r}_{t-1} &= \boldsymbol{T}'_t \boldsymbol{r}_t, \\
\boldsymbol{N}_{t-1} &= \boldsymbol{T}'_t \boldsymbol{N}_t \boldsymbol{T}_t,
\end{aligned}$$

When some, but not all observations are missing, replace the observation equation by,

$$\boldsymbol{y}^*_t = \boldsymbol{Z}^*_t \boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}^*_t, \quad \boldsymbol{\varepsilon}^*_t \sim N(\boldsymbol{0}, \boldsymbol{H}^*_t),$$

where,

$$\begin{aligned}
\boldsymbol{y}^*_t &= \boldsymbol{W}_t \boldsymbol{y}_t, \\
\boldsymbol{Z}^* &= \boldsymbol{W}_t \boldsymbol{Z}_t, \\
\boldsymbol{\varepsilon}_t &= \boldsymbol{W}_t \boldsymbol{\varepsilon}_t, \\
\boldsymbol{H}^*_t &= \boldsymbol{W}_t \boldsymbol{H}_t \boldsymbol{W}'_t,
\end{aligned}$$

and $\boldsymbol{W}_t$ is a selection matrix to select non-missing values. In smoothing the missing elements are estimated by the appropriate elements of $\boldsymbol{Z}_t \hat{\boldsymbol{alpha}}_t$, where $\hat{\boldsymbol{\alpha}}_t$ is the smoothed state.

Note If $y_{t,j}$ is missing, then setting the relevant entries in the forecast precision matrix, $F^{-1}_{t,j,.} = \boldsymbol{0}$ and $F^{-1}_{t,.,j} = \boldsymbol{0}$, and Kalman gain matrix, $K_{t,.,j} = \boldsymbol{0}$, will handle missing values in the smoothers without having to pass that information to the smoother. However, it may be computationally more efficient if the values of the locations of the missing observations are known.

Note For the disturbance and state simulation smoothers, I think the missing observations need to be indicated and used when doing the simulations on the state smoother.

3.5 Forecasting matrices

Forecasting future observations are the same as treating the future observations as missing [**DurbinKoopman2012**],

$$\begin{aligned}
\bar{\boldsymbol{y}}_{n+j} &= \boldsymbol{Z}_{n+j} \bar{\boldsymbol{a}}_{n+j} \\
\bar{\boldsymbol{F}}_{n+j} &= \boldsymbol{Z}_{n+j} \bar{\boldsymbol{P}}_{n+j} \boldsymbol{Z}'_{n+j} + \boldsymbol{H}_{n+j}.
\end{aligned}$$