Time Series and Sequence Learning

Lecture 5 – Structural time series and state space models

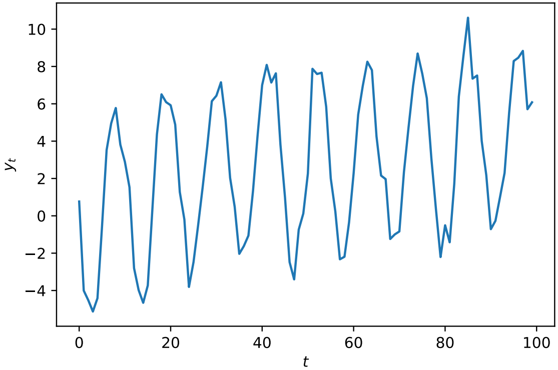
Johan Alenlöv, Linköping University

2021-09-13

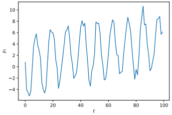
Some updates

- Deadline for computer lab 1 is on **Thursday!**
- Office hours, **Tuesdays 14:30 – 15:30**.
  - Come to my office, 3D:453
  - Or we take it over Zoom
- Exam will be a computerbased exam on campus.

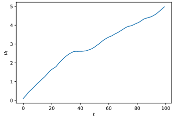
Summary of Lecture 4: Structural time series



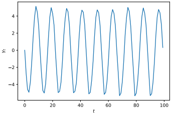
Summary of Lecture 4: Structural time series



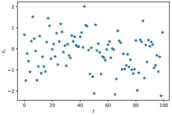
$=$



$+$



$+$



$$y_t \quad = \quad \mu_t \quad + \quad \gamma_t \quad + \quad \varepsilon_t$$

Summary of Lecture 4: Local level model

The simplest structural time series is the **local level model**.

- No seasonal component γ_t.
- The trend component is assumed to follow a random walk.

Def. (Local level model):

$$\mu_t = \mu_{t-1} + \zeta_t, \qquad \zeta_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\zeta^2),$$

$$y_t = \mu_t + \varepsilon_t, \qquad \varepsilon_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$$

with $\mu_1 \sim \mathcal{N}(a_1, P_1)$.

Summary of Lecture 4: Kalman filter

The **Kalman filter** is a recursive algorithm for estimating the unobserved **state** μ_t based on the observed data $y_{1:t}$, for $t = 1, 2, \dots$

Specifically, it allows us to compute the:

- Filtering distribution,

$$p(\mu_t \mid y_{1:t}) = \mathcal{N}(\mu_t \mid \hat{\mu}_{t|t}, P_{t|t})$$

- (1-step) Predictive distribution,

$$p(\mu_{t+1} \mid y_{1:t}) = \mathcal{N}(\mu_{t+1} \mid \hat{\mu}_{t+1|t}, P_{t+1|t})$$

for $t = 1, 2, \dots$.

Structural state space models

Def. (Local linear model): Trend is given by

$$
\begin{aligned}
\alpha_t &= \begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix} \alpha_{t-1} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \zeta_t \\
\mu_t &= \begin{bmatrix} 1 & 0 \end{bmatrix} \alpha_t
\end{aligned}
$$

Cf. For the local level model the trend is given by

$$
\begin{aligned}
\alpha_t &= \alpha_{t-1} + \underline{\zeta_t}, \quad \sim N(\ ,\ ) \\
\mu_t &= \alpha_t
\end{aligned}
$$

Similarly, a $k-1$th order polynomial trend model $\Delta^k \mu_t = \zeta_t$ can be written as

$$\alpha_t = \overset{d\times d \quad T}{\begin{bmatrix} c_1 & c_2 & \cdots & c_{k-1} & c_k \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}} \alpha_{t-1} + \underset{R}{\overset{d\times 1}{\begin{bmatrix} 1 \\ 0 \\ 0 \\ \cdots \\ 0 \end{bmatrix}}} \zeta_t,$$

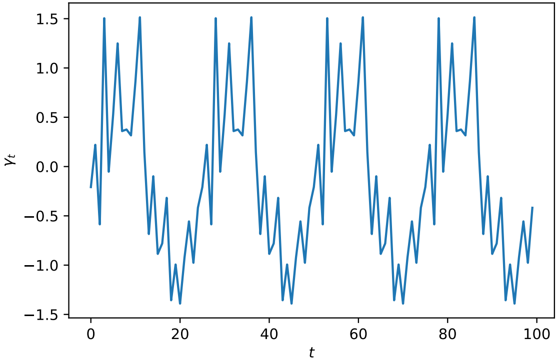
$$y_t = \mu_t = \underset{d\times 1}{\overset{Z}{\begin{bmatrix} 1 & 0 & 0 & \vdots & 0 \end{bmatrix}}} \alpha_t,$$

where the **state vector** is

$$\alpha_t = \begin{bmatrix} \mu_t & \mu_{t-1} & \cdots & \mu_{t-k+1} \end{bmatrix}^{\mathsf{T}}$$

and $c_i = (-1)^{i+1} \binom{k}{i}$.

Seasonal component



Seasonal component

Seasonal component model:

$$\gamma_t = -\sum_{j=1}^{s-1} \gamma_{t-j} + \omega_t, \qquad \omega_t \sim \mathcal{N}(0, \sigma_\omega^2).$$

Matrix form: By defining the **state vector**

$$\alpha_t = \begin{bmatrix} \gamma_t & \gamma_{t-1} & \cdots & \gamma_{t-s+2} \end{bmatrix}^\mathsf{T}$$

we can write this as

$$\alpha_t = \begin{bmatrix} -1 & -1 & \cdots & -1 & -1 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \alpha_{t-1} + \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \omega_t$$

$$\gamma_t = \begin{bmatrix} 1 & 0 & 0 & \vdots & 0 \end{bmatrix} \alpha_t,$$

Trend component ($k = 2$):

$$\alpha_t = \begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix} \alpha_{t-1} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \zeta_t$$

$$\mu_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \alpha_t$$

Seasonal component ($s = 3$):

$$\alpha_t = \begin{bmatrix} -1 & -1 \\ 1 & 0 \end{bmatrix} \alpha_{t-1} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \omega_t$$

$$\gamma_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \alpha_t$$

Structural time series: $y_t = \mu_t + \gamma_t + \varepsilon_t$

Putting it together: Let $\alpha_t = \begin{bmatrix} \mu_t & \mu_{t-1} & \gamma_t & \gamma_{t-1} \end{bmatrix}^\mathsf{T}$, then

$$\alpha_t = \begin{bmatrix} 2 & -1 & & \\ 1 & 0 & & \\ & & -1 & -1 \\ & & 1 & 0 \end{bmatrix} \alpha_{t-1} + \begin{bmatrix} 1 & \\ 0 & \\ & 1 \\ & 0 \end{bmatrix} \begin{bmatrix} \zeta_t \\ \omega_t \end{bmatrix}$$

4×2

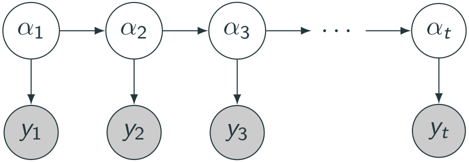
$$y_t = \begin{bmatrix} 1 & 0 & 1 & 0 \end{bmatrix} \alpha_t + \varepsilon_t$$

A general state space model

Def. A **Linear Gaussian State-Space (LGSS)** model is given by:

$$\alpha_t = T\alpha_{t-1} + R\eta_t, \qquad \eta_t \sim \mathcal{N}(0, Q),$$
$$y_t = Z\alpha_t + \varepsilon_t \qquad \varepsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2),$$

and initial distribution $\alpha_1 \sim \mathcal{N}(a_1, P_1)$.



Structural time series – block matrix model

A general structural time series model

$$y_t = \mu_t + \gamma_t + \varepsilon_t$$

can be written in state space form using **block matrices**.

State vector:

$$\alpha_t = \begin{bmatrix} \mu_t & \mu_{t-1} & \cdots & \mu_{t-k+1} & \gamma_t & \gamma_{t-1} & \cdots & \gamma_{t-s+2} \end{bmatrix}^\mathsf{T}$$

State space model:

$$\begin{aligned} \alpha_t &= \begin{bmatrix} T_{[\mu]} & \\ & T_{[\gamma]} \end{bmatrix} \alpha_{t-1} + \begin{bmatrix} R_{[\mu]} & \\ & R_{[\gamma]} \end{bmatrix} \eta_t, & \eta_t &\sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\zeta^2 & 0 \\ 0 & \sigma_\omega^2 \end{bmatrix} \right), \\ y_t &= \begin{bmatrix} Z_{[\mu]} & Z_{[\gamma]} \end{bmatrix} \alpha_t + \varepsilon_t, & \varepsilon_t &\sim \mathcal{N}(0, \sigma_\varepsilon^2). \end{aligned}$$

The Kalman filter

Filtering, Smoothing, and Predicting

Given a time-series $y_{1:n} = (y_1, y_2, \ldots, y_n)$ we wish to calculate the distribution of α_t conditioned on the observed time-series $y_{1:n}$.

This problem changes depending on the relationship of n and t:

$n < t$: This is known as the **forecasting** problem.

$n = t$: This is known as the **filtering** problem.

$n > t$: This is known as the **smoothing** problem.

The Kalman filter

For any s, t, denote by $\hat{\alpha}_{t|s} = \mathbb{E}[\alpha_t \mid y_{1:s}]$ and $P_{t|s} = \mathrm{Cov}(\alpha_t \mid y_{1:s})$.

Thm. For an LGSS model, $p(\alpha_t \mid y_{1:s}) = \mathcal{N}(\alpha_t \mid \hat{\alpha}_{t|s}, P_{t|s})$.

Of particular interest are:

- Filtering distribution,

$$p(\alpha_t \mid y_{1:t}) = \mathcal{N}(\alpha_t \mid \hat{\alpha}_{t|t}, P_{t|t}).$$

- (1-step) Predictive distributions,

$$p(\alpha_t \mid y_{1:t-1}) = \mathcal{N}(\alpha_t \mid \hat{\alpha}_{t|t-1}, P_{t|t-1}),$$
$$p(y_t \mid y_{1:t-1}) = \mathcal{N}(y_t \mid \hat{y}_{t|t-1}, F_{t|t-1}).$$

Kalman filter

Kalman filter: For $t = 1, 2, \ldots$

- **Predict:**

  - Predict α_t:
$$\begin{cases} \hat{\alpha}_{t|t-1} = T\hat{\alpha}_{t-1|t-1}, \\ P_{t|t-1} = TP_{t-1|t-1}T^{\mathsf{T}} + RQR^{\mathsf{T}} \end{cases}$$

  - Predict y_t:
$$\begin{cases} \hat{y}_{t|t-1} = Z\hat{\alpha}_{t|t-1}, \\ F_{t|t-1} = ZP_{t|t-1}Z^{\mathsf{T}} + \sigma_\varepsilon^2 \end{cases} \qquad (\star)$$

- **Update:**

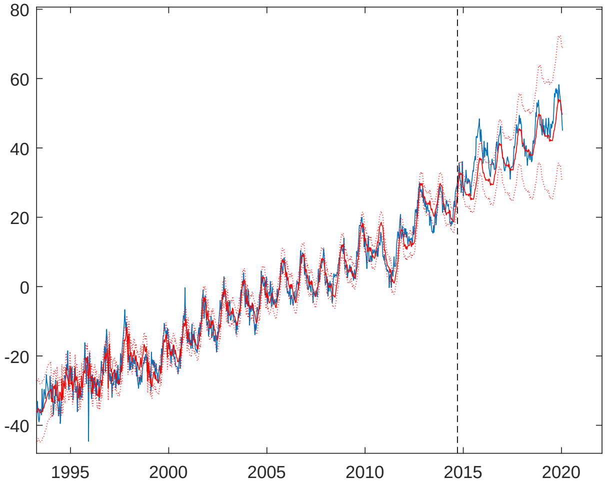
  - Kalman gain: $K_t = P_{t|t-1}Z^{\mathsf{T}}F_{t|t-1}^{-1}$

  - Update filter:
$$\begin{cases} \hat{\alpha}_{t|t} = \hat{\alpha}_{t|t-1} + K_t(y_t - \hat{y}_{t|t-1}), \\ P_{t|t} = (I - K_tZ)P_{t|t-1} \end{cases} \qquad (\star\star)$$

($\star$) At time $t = 1$ we initialize $\hat{\alpha}_{1|0} = a_1$ and $P_{1|0} = P_1$.

($\star\star$) If y_t is missing we skip the update and set $\hat{\alpha}_{t|t} = \hat{\alpha}_{t|t-1}$ and $P_{t|t} = P_{t|t-1}$.

ex) GMSL data



AR and ARMA models

Auto-regressive models in state space form

Recall, seasonal component model:

$$\gamma_t = -\sum_{j=1}^{s-1} \gamma_{t-j} + \omega_t,$$

On matrix form:

$$\alpha_t = \begin{bmatrix} -1 & -1 & \cdots & -1 & -1 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \alpha_{t-1} + \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \omega_t,$$

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

Auto-regressive model in state space form

State space formulation of AR model: The AR(p) model,

$$y_t = \sum_{j=1}^{p} a_j y_{t-j} + \eta_t,$$

can equivalently be expressed in **state space form** as

$$\alpha_t = \begin{bmatrix} a_1 & a_2 & \cdots & a_{p-1} & a_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \alpha_{t-1} + \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \eta_t,$$

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

A scalar AR(p) model can be written as a vector-valued AR(1) model!

Auto-regressive moving average model in state space form

State space formulation of ARMA: Consider the ARMA(p, q) model,

$$y_t = \sum_{j=1}^{p} a_j y_{t-j} + \sum_{j=1}^{q} b_j \eta_{t-j} + \eta_t.$$

Let $d = \max(p, q+1)$ and define $a_j = 0$ for $j > p$ and $b_j = 0$ for $j > q$. Then, an equivalent **state space form** is given by

$$\alpha_t = \begin{bmatrix} a_1 & a_2 & \cdots & a_{d-1} & a_d \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \alpha_{t-1} + \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \eta_t,$$

$$y_t = \begin{bmatrix} 1 & b_1 & \cdots & b_{d-2} & b_{d-1} \end{bmatrix} \alpha_t$$

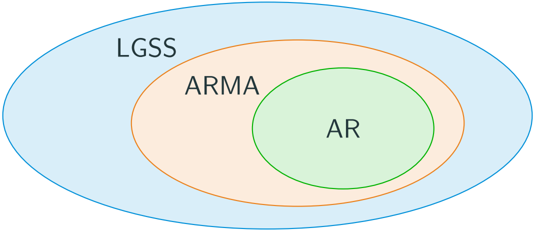
A general state space model

Both AR and ARMA models are special cases of the general linear Gaussian state space model,

Recall, the LGSS model is defined by

$$\alpha_t = T\alpha_{t-1} + R\eta_t, \qquad \eta_t \sim \mathcal{N}(0, Q),$$
$$y_t = Z\alpha_t + \varepsilon_t \qquad \varepsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2),$$

and initial distribution $\alpha_1 \sim \mathcal{N}(a_1, P_1)$.



State space model

The **state vector** at time t is a summary of **all the relevant information** needed to predict future states or observations.

More specifically:

If we know the **state vector** α_t, then there is no further information available in past states α_s, $s < t$, or observations y_s, $s \leq t$, regarding the future states and observations.

Even more specifically:

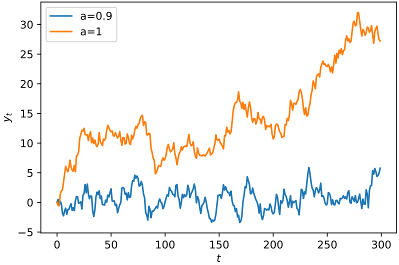
Conditionally on α_t, any future state or observation variable is **conditionally independent** of past states and observations,

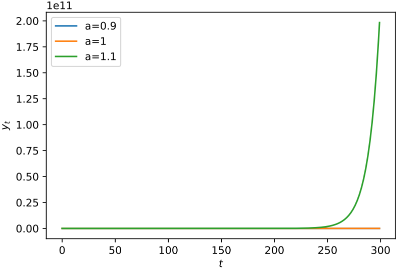
$$p(\alpha_\tau \mid \alpha_{1:t}, y_{1:t}) = p(\alpha_\tau \mid \alpha_t) \qquad \text{for any } \tau > t.$$

Stability of state space models

ex) Simulation of AR(1)

Simulation of $y_t = a y_{t-1} + \varepsilon_t$





The AR(1) model is:

- Stable if $|a| < 1 \Rightarrow$ converges to stationary
- Marginally stable if $|a| = 1 \Rightarrow$ linear drift
- Unstable if $|a| > 1 \Rightarrow$ exponential explosion

Stability of state space model

Can this be generalized to an LGSS model?

AR(1): $y_t = a y_{t-1} + \varepsilon_t$

LGSS: $\alpha_t = T\alpha_{t-1} + R\eta_t$

Intuitively, the state process is unstable if "size(T)" > 1.

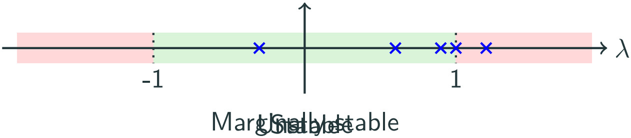
Stability of state space model

Thm. A state space model of dimension $d = \dim(\alpha_t)$ is:

- Stable iff $|\lambda_j| < 1, \quad j = 1, \ldots, d,$
- Marginally stable iff $|\lambda_j| \leq 1, \quad j = 1, \ldots, d,$
- Unstable iff $|\lambda_j| > 1$ for any $j = 1, \ldots, d,$

where λ_j, $j = 1, \ldots, d$ are the **eigenvalues of T.**

If all eigenvalues are real numbers,

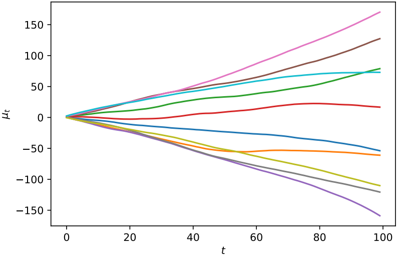


ex) Eigenvalues of linear trend model

Linear trend model:

$$\alpha_t = \overbrace{\begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix}}^{F} \alpha_{t-1} + \overbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}^{R} \zeta_t$$

$$\mu_t = \underbrace{\begin{bmatrix} 1 & 0 \end{bmatrix}}_{Z} \alpha_t$$



Check for stability by computing the eigenvalues

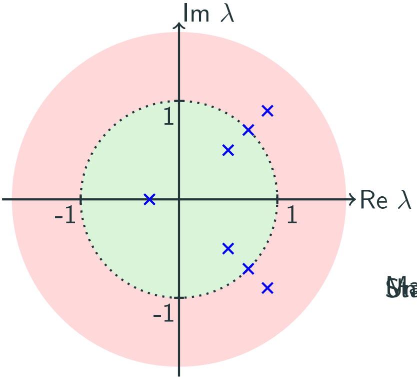
$$\mathrm{eig}(F) \Rightarrow \lambda_1 = \lambda_2 = 1.$$

The trend model is **marginally stable!**

Complex eigenvalues

Note. The eigenvalues can be complex numbers in general. Thus, the stability condition reads

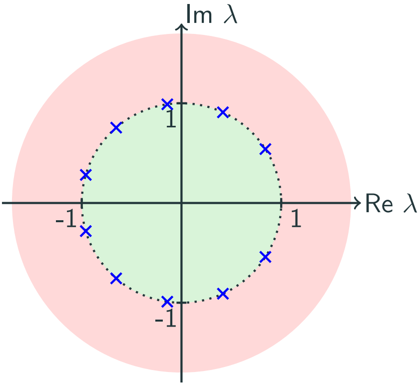
all eigenvalues of T are within the unit circle in the complex plane.



Marginally stable
Stable
Unstable

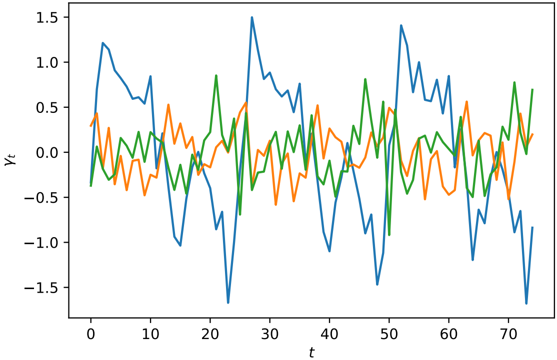
Seasonal model:

$$T = \begin{bmatrix} -1 & -1 & \cdots & -1 & -1 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$



The seasonal model is **marginally stable!**

ex) Sample trajectories for seasonal model



Stability of structural time series

The structural time series models that we have proposed are

designed to be marginally stable!

Marginal stability results in desirable properties:

- Real eigenvalues $\lambda_j = 1 \Rightarrow$ polynomial drift/trend.
- Complex eigenvalues with $|\lambda_j| = 1 \Rightarrow$ periodicity/seasonality.

Summary

A few concepts to summarize lecture 5:

- **State space model:** The observed time series $\{y_t\}$ is modeled using an unobserved state process $\{\alpha_t\}$
- **State vector:** The state vector α_t contains all relevant information about the process at time t.
- **Structural time series:** Block-matrix construction of a state space model, where the different "blocks" correspond to time series components such as trend, seasonality, etc.
- **Kalman filter:** A recursive algorithm for estimating the unobserved state vector based on the observed data in a (linear and Gaussian) state space model.
- **Stability:** Results in a stationary state process (possibly after an initial transient). Corresponds to all eigenvalues of T being strictly within the unit circle.
- **Marginal stability:** Results in a state process that grows polynomially and/or shows a non-diminishing periodic pattern. Corresponds to some eigenvalues of T being *on* the unit circle.