Machine Learning for Signal Processing Predicting and Estimation from Time Series: Part 2 i**ng for Signal Processing
and Estimation from
Series: Part 2
Bhiksha Raj**

iksha Raj
11-755/18797
11-755/18797

Preliminaries : P(y|x) for Gaussian

• If $P(x,y)$ is Gaussian:

$$
P(\mathbf{x}, \mathbf{y}) = N \begin{bmatrix} \mu_{\mathbf{x}} \\ \mu_{\mathbf{y}} \end{bmatrix}, \begin{bmatrix} C_{\mathbf{x}\mathbf{x}} & C_{\mathbf{x}\mathbf{y}} \\ C_{\mathbf{y}\mathbf{x}} & C_{\mathbf{y}\mathbf{y}} \end{bmatrix}
$$

• The conditional probability of y given x is also Gaussian

– The slice in the figure is Gaussian

The slice in the figure is Gaussian

\n
$$
P(y \mid x) = N(\mu_y + C_{yx} C_{xx}^{-1} (x - \mu_x), C_{yy} - C_{yx} C_{xx}^{-1} C_{xy})
$$
\nThe mean of this Gaussian is a function of x

\nThe variance of y reduces if x is known

\n– Uncertainty is reduced

\n
$$
11-755/18797
$$

- The mean of this Gaussian is a function of x
- The variance of y reduces if x is known
	- Uncertainty is reduced

Background: Sum of Gaussian RVs

• The conditional probability of O:

• The overall probability of O: $N(AS + \mu_{\varepsilon}, \mathbf{\Theta}_{\varepsilon})$

lity of *O*:

; + $\boldsymbol{\mu}{\varepsilon}$, $A\mathbf{\Theta}_{S}A^{T} + \mathbf{\Theta}_{\varepsilon})$

Background: Joint Prob. of O and S

$$
0 = AS + \varepsilon \qquad Z = \begin{bmatrix} 0 \\ S \end{bmatrix}
$$

Background: Joint Prob. of O and S
 $\mathbf{0} = AS + \varepsilon$ $\mathbf{z} = \begin{bmatrix} 0 \\ S \end{bmatrix}$

• The joint probability of O and S (i.e. P(Z)) is

also Gaussian also Gaussian

$$
P(Z) = P(0, S) = N(\mu_Z, \Theta_Z)
$$

• Where

$$
\mu_Z = \begin{bmatrix} \mu_O \\ \mu_S \end{bmatrix} = \begin{bmatrix} A\mu_S + \mu_E \\ \mu_S \end{bmatrix}
$$

• Where
\n
$$
\mu_Z = \begin{bmatrix} \mu_0 \\ \mu_S \end{bmatrix} = \begin{bmatrix} A\mu_s + \mu_{\varepsilon} \\ \mu_S \end{bmatrix}
$$
\n• $\Theta_Z = \begin{bmatrix} \Theta_0 & \Theta_{OS} \\ \Theta_{SO} & \Theta_S \end{bmatrix} = \begin{bmatrix} A\Theta_S A^T + \Theta_{\varepsilon} & A\Theta_S \\ \Theta_S A^T & \Theta_S \end{bmatrix}$

Preliminaries : Conditional of S given O: P(S|O)

 $P(S|O) = N(\mu_S + \Theta_S A^T (A \Theta_S A^T + \Theta_{\varepsilon})^{-1} (O - A \mu_S - \mu_{\varepsilon}),$ $\Theta_S - \Theta_S A^T (A \Theta_S A^T + \Theta_{\varepsilon})^{-1} A \Theta_S)$

Estimating the state

- The state is estimated from the updated distribution
	- The updated distribution is propagated into time, not the state

Predicting the next observation

- The probability distribution for the observations at the next time is a mixture:
- $P(X_t|X_{0:t-1}) = \sum_{S_t} P(X_t|S_t)P(S_t|X_{0:t-1})$
- The actual observation can be predicted from $P(x_T | x_{0:T-1})$

Predicting the next observation

- Can use any of the various estimators of x_T from $P(x_T|x_{0:T-1})$
- MAP estimate: $-$ argmax_{x_T} $P(x_T|x_{0:T-1})$
- MMSE estimate:
	- $-$ Expectation($x_T|x_{0:T-1})$

Difference from Viterbi decoding

- Estimating only the *current* state at any time
	- Not the state sequence
	- Although we are considering all past observations
- The most likely state at T and T+1 may be such that there is no valid transition between S_T and S_{T+1} e at T and T+1 may be such
d transition between S_T
11-755/18797

The real-valued state model

• A state equation describing the dynamics of the system

$$
S_t = f(S_{t-1}, \mathcal{E}_t)
$$

- s_t is the state of the system at time t
- ε _t is a driving function, which is assumed to be random
- The state of the system at any time depends only on the state at the previous time instant and the driving term at the current time
- An observation equation relating state to observation

$$
o_t = g(s_t, \gamma_t)
$$

 o_t is the observation at time t

- $\gamma_{\rm t}$ is the noise affecting the observation (also random)
- The observation at any time depends only on the current state of the system and the noise ating state to observation
 $\langle, \gamma_t \rangle$

2 t

observation (also random)

depends only on the current state of

11-755/18797

10

Discrete vs. Continuous state systems

Discrete vs. Continuous State Systems 0 1 2 3

$$
S_t = f(S_{t-1}, \mathcal{E}_t)
$$

 $o_t = g(s_t, \gamma_t)$

$\pi = \frac{0.1}{0.1}$
 $\pi = \frac{0.1}{0.1}$

Prediction at time t:
 $t|0_{0:t-1}) = \sum_{S_{t-1}} P(S_{t-1}|0_{0:t-1}) P(S_t|S_{t-1})$

Update after observing O_t:
 $(S_t|O_{0:t}) = C.P(S_t|O_{0:t-1}) P(O_t|S_t)$
 $P(S_t|O_t|S_t)$ $t|U_{0:t-1}| = \int_{-1}^{1} f(S_{t-1}|U_{0:t-1}) F(S_t|S_{t-1})$ $\qquad \qquad \left| P(S_t|U_{0:t-1}) \right| = 0$ S_{t-1}

:

 $P(S_t | O_{0:t}) = C \cdot P(S_t | O_{0:t-1}) P(O_t | S_t)$

$$
P(S_t | O_{0:t-1}) = \int_{-\infty}^{\infty} P(S_{t-1} | O_{0:t-1}) P(S_t | S_{t-1}) dS_{t-1}
$$

$$
P(S_t | O_{0:t}) = C.P(S_t | O_{0:t-1}) P(O_t | S_t)
$$

Discrete vs. Continuous State Systems

Discrete vs. Continuous State S
\n
$$
\begin{array}{ccc}\n & \text{...} & \text{...} & \text{...} \\
\hline\n & \text{...} & \text{...} & \text{...} & \text{...} \\
\hline\n & \text{...} & \text{...} & \text{...} & \text{...} \\
\hline\n & \text{...} & \text{...} & \text{...} & \text{...} \\
\hline\n & \text{...} & \text{...} & \text{...} & \text{...} \\
\hline\n & \text{...} & \text{...} & \text{...} & \text{...} \\
\hline\n & \text{...} & \text{...} & \text{...} & \text{...} \\
\hline\n & \text{...} & \text{...} & \text{...} & \text{...} & \text{...} \\
\hline\n & \text{...} & \text{...} & \text{...} & \text{...} & \text{...} \\
\hline\n & \text{...} & \text{...} & \text{...} & \text{...} & \text{...} \\
\hline\n & \text{...} & \text{...} & \text{...} & \text{...} & \text{...} \\
\hline\n & \text{...} & \text{...} & \text{...} & \text{...} & \text{...} & \text{...} \\
\hline\n & \text{...} & \text{...} & \text{...} & \text{...} & \text{...} & \text{...} \\
\hline\n & \text{...} & \text{...} & \text{...} &
$$

$$
S_t = f(S_{t-1}, \mathcal{E}_t)
$$

 $o_t = g(s_t, \gamma_t)$

$$
P(s)
$$

 $P(s_t | s_{t-1})$

 $P(O|S)$

$$
\mathcal{L}^{\mathcal{L}}(\mathcal{L}
$$

Transition prob $P(s_t = j | s_{t-1} = i)$

Observation prob $P(O|S)$

Special case: Linear Gaussian model

$$
S_t = A_t S_{t-1} + \mathcal{E}_t
$$

 $\mathbf{0} \mathbf{o}_t = B_t \mathbf{s}_t + \gamma_t$

Special case: Linear Gaussian model

\n
$$
s_{t} = A_{t} s_{t-1} + \varepsilon_{t}
$$
\n
$$
P(\varepsilon) = \frac{1}{\sqrt{(2\pi)^{d} |\Theta_{\varepsilon}|}} \exp(-0.5(\varepsilon - \mu_{\varepsilon})^{T} \Theta_{\varepsilon}^{-1} (\varepsilon - \mu_{\varepsilon}))
$$
\n
$$
o_{t} = B_{t} s_{t} + \gamma_{t}
$$
\n
$$
P(\gamma) = \frac{1}{\sqrt{(2\pi)^{d} |\Theta_{\gamma}|}} \exp(-0.5(\gamma - \mu_{\gamma})^{T} \Theta_{\gamma}^{-1} (\gamma - \mu_{\gamma}))
$$
\nlinear state dynamics equations

- A linear state dynamics equation
	- $-$ Probability of state driving term ε is Gaussian
	- Sometimes viewed as a driving term μ_{ε} and additive zero-mean noise
- A *linear* observation equation
	- Probability of observation noise γ is Gaussian
- \bullet A_t , B_t and Gaussian parameters assumed known – May vary with time s a driving term μ_{ε} and additive
equation
ation noise γ is Gaussian
arameters assumed known
11-755/18797

Linear model example The wind and the target

• State: Wind speed at time t depends on speed at time $t-1$

$$
S_t = S_{t-1} + \epsilon
$$

• Observation: Arrow position at time t depends on wind speed at time t -1 + ϵ_t

position at time *t* depends on
 $S_t + \gamma_t$

$$
\boldsymbol{\theta}_t = \boldsymbol{B}\boldsymbol{S}_t + \boldsymbol{\gamma}_t
$$

Model Parameters:
\n**The initial state probability**
\n
$$
P_0(s) = \frac{1}{\sqrt{(2\pi)^d |R|}} \exp(-0.5(s-\overline{s})R^{-1}(s-\overline{s})^T)
$$

 $P_0(s) = Gaussian(s; \overline{s}, R)$

• We also assume the initial state distribution to be Gaussian $\begin{aligned} \text{\emph{a}}\ \text{\emph{b}}\ \text{\emph{b}}\ \text{\emph{b}}\ \text{\emph{b}}\ \text{\emph{b}}\ \text{\emph{b}}\ \text{\emph{b}}\ \text{\emph{b}}\ \text{\emph{c}}\ \text{\emph{c}}\ \text{\emph{c}}\ \text{\emph{c}}\ \text{\emph{d}}\ \text{\emph{c}}\ \text{\emph{e}}\ \text{\emph{e}}\ \text{\emph{b}}\ \text{\emph{c}}\ \text{\emph{e}}\ \text{\emph{b}}\ \text{\emph{c}}\ \text{\emph{b}}\ \text{\emph{c}}$

– Often assumed zero mean

$$
s_t = A_t s_{t-1} + \varepsilon_t
$$

$$
o_t = B_t s_t + \gamma_t
$$

Model Parameters: The observation probability $\mathbf{v}_t = \mathbf{B}_t \mathbf{s}_t + \mathbf{\gamma}_t$ $P(\mathbf{\gamma}) = Gaussian(\mathbf{\gamma}; \boldsymbol{\mu}_{\gamma}, \boldsymbol{\Theta}_{\gamma})$

$$
P(o_t | s_t) = Gaussian(o_t; \mu_{\gamma} + B_t s_t, \Theta_{\gamma})
$$

- The probability of the observation, given the state, is simply the probability of the noise, with the mean shifted
	- Since the only uncertainty is from the noise
- The new mean is the mean of the distribution of the noise + the value of the observation in the absence of noise of the noise, with the mean
nty is from the noise
ean of the distribution of the
eobservation in the absence of
i11-755/18797

Model Parameters: State transition probability

$$
s_{t+1} = A_t s_t + \varepsilon_t \qquad P(\varepsilon) = Gaussian(\varepsilon; \mu_{\varepsilon}, \Theta_{\varepsilon})
$$

$$
P(s_{t+1} | s_t) = Gaussian(s_t; \mu_{\varepsilon} + A_t s_t, \Theta_{\varepsilon})
$$

• The probability of the state at time t, given the state at t-1, is simply the probability of the driving term, with the mean shifted he state at time t, given the
y the probability of the
he mean shifted
11-755/18797

$$
\sum_{s} \sum_{s} s_{t+1} = A_{t}S_{t} + \varepsilon_{t}
$$

Prediction at time 0:

$$
P(S_0) = P_0(S_0)
$$

Update after O_0 : :

$$
P(S_0|O_0) = C.P(S_0)P(O_0|S_0)
$$

Prediction at time 1:

$$
P(S_1|O_0) = \int_{-\infty}^{\infty} P(S_0|O_0) P(S_1|S_0) dS_0
$$

Update after O_1 : :

$$
\sum_{i=1}^{\infty} \sum_{s} s_{t+1} = A_t s_t + \varepsilon_t
$$

Prediction at time 0:

$$
P(S_0) = P_0(S_0)
$$

Update after O_0 : :

$$
P(S_0|O_0) = C.P(S_0)P(O_0|S_0)
$$

Prediction at time 1:

$$
P(S_1|O_0) = \int_{-\infty}^{\infty} P(S_0|O_0) P(S_1|S_0) dS_0
$$

Update after O_1 : :

Model Parameters:
\n**The initial state probability**
\n
$$
P_0(s) = \frac{1}{\sqrt{(2\pi)^d |R_0|}} \exp(-0.5(s-\overline{s}_0)R_0^{-1}(s-\overline{s}_0)^T)
$$
\n
$$
P_0(s) = Gaussian(s; \overline{s}_0, R_0)
$$

- We assume the *initial* state distribution to be Gaussian *ial* state distribution to be
 p mean

11-755/18797

21
	- Often assumed zero mean

$$
\sum_{i=1}^{\infty} \sum_{s} s_{t+1} = A_t s_t + \varepsilon_t
$$

Prediction at time 0:

$$
P(S_0) = P_0(S_0)
$$

a priori probability distribution of state s

$$
= N(\bar{s}_0, R_0)
$$

Update after O_0 : :

$$
P(S_0|O_0) = C.P(S_0)P(O_0|S_0)
$$

Prediction at time 1:

$$
P(S_1|O_0) = \int_{-\infty}^{\infty} P(S_0|O_0) P(S_1|S_0) dS_0
$$

Update after O_1 : :

$$
\begin{array}{c}\n\circ \\
\circ \\
\circ \\
\circ \\
\circ\n\end{array}
$$
\n
$$
\begin{array}{c}\nS_{t+1} = A_t S_t \\
\circ \\
\circ \\
\circ \\
\circ \\
\circ \\
\circ\n\end{array}
$$

$$
S_{t+1} = A_t S_t + \mathcal{E}_t
$$

$$
O_t = B_t S_t + \gamma_t
$$

Prediction at time 0:

$$
P(S_0) = N(\overline{s}_0, R_0)
$$

Update after O_0 : :

$$
P(S_0|O_0) = C.P(S_0)P(O_0|S_0)
$$

Prediction at time 1:

$$
P(S_1|O_0) = \int_{-\infty}^{\infty} P(S_0|O_0) P(S_1|S_0) dS_0
$$

Update after O_1 : :

MLSP Recap: Conditional of S given O: P(S|O) for Gaussian RVs

Recap: Conditional of S given O: MLSP P(S|O) for Gaussian RVs

$$
P(S|O) = N(\mu_S + \Theta_S B^{\mathrm{T}} (B \Theta_S B^{\mathrm{T}} + \Theta_\gamma)^{-1} (O - B\mu_S - \mu_\gamma),
$$

$$
\Theta_S - \Theta_S B^{\mathrm{T}} (B \Theta_S B^{\mathrm{T}} + \Theta_\gamma)^{-1} B\Theta_S)
$$

MLSP Recap: Conditional of S given O: P(S|O) for Gaussian RVs

$$
P(S_0|O_0) = N(\overline{S}_0 + R_0B^T\left(BR_0B^T + \mathcal{O}_{\gamma}\right)^{-1}(O_0 - B\overline{S}_0 - \mu_{\gamma}),
$$

$$
R_0 - R_0B^T\left(BR_0B^T + \mathcal{O}_{\gamma}\right)^{-1}BR_0)
$$

MLS **Recap: Conditional of S given O: P(S|O) for Gaussian RVs** $0 = BS + \gamma$ $K_0 = R_0 B^{\mathrm{T}} (B R_0 B^{\mathrm{T}} + \Theta_{\gamma})^{-1}$ $\hat{s}_0 = \overline{s}_0 + K_0 (O_0 - B \overline{s}_0 - \mu_\gamma)$ $\widehat{R}_0 = (I - K_0) R_0$ $P(S_0|O_0) = N(\hat{s}_0, \hat{R}_0)$

 $P(S_0|O_0) = N(\bar{s_0} + R_0B^T(BR_0B^T + \Theta_\gamma)^{-1}(O_0 - B\bar{s}_0 - \mu_\gamma)),$ $R_0 - R_0 B^{\text{T}} (B R_0 B^{\text{T}} + \mathcal{O}_{\gamma})^{-1} B R_0$

$$
\int_{\alpha}^{\infty} \cos \theta \, d\theta \, d\theta
$$
\n
$$
S_{t+1} = A_t S_t + \varepsilon_t
$$
\n
$$
O_t = B_t S_t + \gamma_t
$$

$$
S_{t+1} = A_t S_t + \mathcal{E}_t
$$

$$
O_t = B_t S_t + \gamma_t
$$

Prediction at time 0:

$$
P(S_0) = N(\overline{s}_0, R_0)
$$

Update after O_0 : :

 $P(S_0|O_0) = C.P(S_0)P(O_0|S_0)$

 $P(S_0|O_0) = N(\hat{s}_0, \hat{R}_0)$

Prediction at time 1: $f(1|U_0) = \int F(0|U_0)F(0|1|0)dU_0$ ∞ $-\infty$

Update after O_1 : :

$$
\sum_{s} \sum_{s} S_{t+1} = A_{t}S_{t} + \varepsilon_{t}
$$

Prediction at time 0:

$$
P(S_0) = N(\overline{s}_0, R_0)
$$

Update after O_0 : : Prediction at time 1: \sim $\frac{\partial |U_0|}{\partial I} = N(S_0, K_0)$ $\hat{s}_0 = \bar{s}_0 + K_0 (0_0 - B\bar{s_0} - \mu_\gamma)$ $\hat{R}_0 = (I - K_0) R_0$ ∞ $K_0 = R_0 B^T (B R_0 B^T + \theta_\gamma)^{-1}$
 $\hat{s}_0 = \bar{s}_0 + K_0 (O_0 - B \bar{s_0} - \mu_\gamma)$ $\boxed{\hat{R}_0 = (I - K_0) R_0}$ $\boldsymbol{K_0} = \boldsymbol{R}_0 \boldsymbol{B}^\mathrm{T} \big(\boldsymbol{B} \boldsymbol{R}_0 \boldsymbol{B}^\mathrm{T} + \boldsymbol{\varTheta}_{\boldsymbol{\gamma}} \big)^{-1} \hspace{2mm} \Bigg[$

$$
P(S_1|O_0) = \int_{-\infty} P(S_0|O_0) P(S_1|S_0) dS_0
$$

Update after O_1 : :

$$
\int_{\alpha}^{\infty} \sqrt{\frac{S_{t+1} - A_t S_t + \varepsilon_t}{O_t = B_t S_t + \gamma_t}}
$$

$$
o_t = B_t s_t + \gamma_t
$$

Prediction at time 0:

$$
\frac{P(S_0) = N(\bar{s}_0, R_0)}{\text{Update after O0:}\n\left| \frac{P(S_0) = N(\bar{s}_0, R_0)}{P(S_0 | O_0) = C \cdot P(S_0) P(O_0 | S_0)} \right|}\n= N(\bar{s}_0 + R_0 B^{\text{T}} (B R_0 B^{\text{T}} + \Theta_\gamma)^{-1} (O_0 - B \bar{s}_0 - \mu_\gamma),
$$
\n
$$
R_0 - R_0 B^{\text{T}} (B R_0 B^{\text{T}} + \Theta_\gamma)^{-1} B R_0
$$

Prediction at time 1:

$$
P(S_1|O_0) = \int_{-\infty}^{\infty} P(S_0|O_0) P(S_1|S_0) dS_0
$$

Update after O_1 :

Introducting shorthand notation

$$
P(S_0|O_0) = N(\bar{s}_0 + R_0 B^{\text{T}} (BR_0 B^{\text{T}} + \Theta_\gamma))^{-1} (O_0 - B\bar{s}_0 - \mu_\gamma),
$$

$$
R_0 - R_0 B^{\text{T}} (BR_0 B^{\text{T}} + \Theta_\gamma))^{-1} BR_0
$$

$$
\hat{S}_0 = \overline{S}_0 + R_0 B^{\text{T}} \left(B R_0 B^{\text{T}} + \Theta_\gamma \right)^{-1} \left(O - B \overline{S}_0 - \mu_\gamma \right)
$$

$$
\hat{R}_0 = R_0 - R_0 B^{\text{T}} \left(B R_0 B^{\text{T}} + \Theta_\gamma \right)^{-1} B R_0
$$

$$
P(S_0|O_0) = N(\hat{s}_0, \hat{R}_0)
$$

Introducting shorthand notation

$$
P(S_0|O_0) = N(\bar{s}_0 + R_0 B^{\text{T}} (BR_0 B^{\text{T}} + \Theta_\gamma))^{-1} (O_0 - B\bar{s}_0 - \mu_\gamma),
$$

$$
R_0 - R_0 B^{\text{T}} (BR_0 B^{\text{T}} + \Theta_\gamma))^{-1} BR_0
$$

$$
K_0 = R_0 B^T (BR_0 B^T + \Theta_\gamma)^{-1}
$$

$$
\hat{S}_0 = \overline{S}_0 + K_0 (0 - B\overline{S}_0 - \mu_\gamma)
$$

$$
\hat{R}_0 = (I - K_0 B)R_0
$$

$$
P(S_0|O_0) = N(\hat{s}_0, \hat{R}_0)
$$

$$
\sum_{s} \sum_{s} S_{t+1} = A_{t}S_{t} + \varepsilon_{t}
$$

Prediction at time 0:

$$
P(S_0) = N(\overline{s}_0, R_0)
$$

Update after O_0 : : Prediction at time 1: \sim $\frac{\partial |U_0|}{\partial I} = N(S_0, K_0)$ $\hat{s}_0 = \bar{s}_0 + K_0 (0_0 - B\bar{s_0} - \mu_\gamma)$ $\hat{R}_0 = (I - K_0) R_0$ ∞ $K_0 = R_0 B^T (B R_0 B^T + \theta_\gamma)^{-1}$
 $\hat{s}_0 = \bar{s}_0 + K_0 (O_0 - B \bar{s_0} - \mu_\gamma)$ $\boxed{\hat{R}_0 = (I - K_0) R_0}$ $\boldsymbol{K_0} = \boldsymbol{R}_0 \boldsymbol{B}^\mathrm{T} \big(\boldsymbol{B} \boldsymbol{R}_0 \boldsymbol{B}^\mathrm{T} + \boldsymbol{\varTheta}_{\boldsymbol{\gamma}} \big)^{-1} \hspace{2mm} \Bigg[$

$$
P(S_1|O_0) = \int_{-\infty} P(S_0|O_0) P(S_1|S_0) dS_0
$$

Update after O_1 : :

$$
\sum_{i=1}^{\infty} \sum_{s} s_{t+1} = A_t s_t + \varepsilon_t
$$

Prediction at time 0:

$$
P(S_0) = N(\overline{s}_0, R_0)
$$

Update after O_0 : :

$$
S_{t+1} = A_t S_t + \mathcal{E}_t
$$

\n
$$
O_t = B_t S_t + \gamma_t
$$

\n
$$
P(S_0) = N(\bar{S}_0, R_0)
$$

\n
$$
K_0 = R_0 B^T (B R_0 B^T + \theta_\gamma)^{-1}
$$

\n
$$
P(S_0 | O_0) = N(\hat{S}_0, \hat{R}_0)
$$

\n
$$
\hat{s}_0 = \bar{s}_0 + K_0 (O_0 - B \bar{s}_0 - \mu_\gamma)
$$

\n
$$
\hat{R}_0 = (I - K_0) R_0
$$

Prediction at time 1:

$$
P(S_1|O_0) = \int_{-\infty}^{\infty} P(S_0|O_0) P(S_1|S_0) dS_0
$$

Update after O_1 : :

The prediction equation

$$
P(S_1|O_0) = \int_{-\infty}^{\infty} P(S_0|O_0)P(S_1|S_0)dS_0
$$

\n
$$
P(S_0|O_0) = N(\hat{s}_0, \hat{R}_0)
$$

\n
$$
P(S_1|S_0) = N(AS_0 + \mu_{\varepsilon}, \Theta_{\varepsilon})
$$

\n
$$
S_{t+1} = A_t S_t + \varepsilon_t
$$

• The integral of the product of two Gaussians

$$
P(S_1|O_0) = \int_{-\infty}^{\infty} Gaussian(S_0; \hat{s}_0, \hat{R}_0) Gaussian(S_1; AS_0, \Theta_{\varepsilon})dS_0
$$

The Prediction Equation

• The integral of the product of two Gaussians is Gaussian!

$$
P(S_1|O_0) = \int_{-\infty}^{\infty} Gaussian(S_0; \hat{s}_0, \hat{R}_0) Gaussian(S_1; AS_0 + \mu_{\varepsilon}, \Theta_{\varepsilon})dS_0
$$

$$
= \int_{-\infty}^{\infty} C_1 \exp(-0.5(S_0 - \hat{s}_0) \hat{R}_0^{-1} (S_0 - \hat{s}_0)^T). C_2 \exp(-0.5(S_1 - AS_0 - \mu_{\varepsilon}) \Theta_{\varepsilon}^{-1} (S_1 - AS_0 - \mu_{\varepsilon})^T) dS_0
$$

$$
= Gaussian(S_1; A\hat{s}_0 + \mu_{\varepsilon}, \Theta_{\varepsilon} + A\hat{R}_0 A^T)
$$

$$
S_{t+1} = A_t S_t + \varepsilon_t
$$

$$
P(S_1|O_0) = N(A\hat{s}_0 + \mu_{\varepsilon}, \Theta_{\varepsilon} + A\hat{R}_0A^T)
$$
$$
\sum_{s}^{\infty} \sum_{s} s_{t+1} = A_{t}S_{t} + \varepsilon_{t}
$$

Prediction at time 0:

$$
P(S_0) = N(\overline{s}_0, R_0)
$$

Update after O_0 :

$$
P(S_0|O_0) = N(\hat{s}_0, \hat{R}_0) \quad \frac{P_0}{\hat{s}_0 = \bar{s}_0 + K_0(O_0 - B\bar{s}_0 - \mu_\gamma)} \quad \hat{R}_0 = (I - K_0) R_0
$$

 $K_0 = R_0 B^{\text{T}} (B R_0 B^{\text{T}} + \theta_0)^{-1}$

Prediction at time 1:

$$
P(S_1|O_0) = \int_{-\infty}^{\infty} P(S_0|O_0) P(S_1|S_0) dS_0 \qquad \boxed{= N(A\hat{s}_0 + \mu_{\varepsilon}, \Theta_{\varepsilon} + A\hat{R}_0 A^T)}
$$

Update after O_1 :

 $P(S_1|O_{0:1}) = C \cdot P(S_1|O_0)P(O_1|S_1)$

More shorthand notation

$$
P(S_1|O_0) = N(A\hat{s}_0 + \mu_{\varepsilon}, \Theta_{\varepsilon} + A\hat{R}_0A^T)
$$

$$
\bar{s}_1 = A\hat{s}_0 + \mu_{\varepsilon}
$$

$$
R_1 = \Theta_{\varepsilon} + A\widehat{R}_0 A^T
$$

$$
P(S_1|O_0) = N(\overline{s}_1, R_1)
$$

$$
\sum_{s}^{\infty} \sum_{s} s_{t+1} = A_{t}S_{t} + \varepsilon_{t}
$$

Prediction at time 0:

$$
P(S_0) = N(\overline{s}_0, R_0)
$$

Update after O_0 :

$$
P(S_0|O_0) = N(\hat{s}_0, \hat{R}_0) \qquad \hat{s}_0 = \bar{s}_0 + K_0(O_0 - B\bar{s}_0 - \mu_Y) \qquad \hat{R}_0 = (I - K_0) R_0
$$

 \vec{a} - \vec{a} + \vec{a}

 $K_0 = R_0 B^{\mathrm{T}} (B R_0 B^{\mathrm{T}} + \mathcal{O}_\nu)^{-1}$

Prediction at time 1:

$$
P(S_1|O_0) = N(\overline{s}_1, R_1) \qquad R_1 = \theta_{\varepsilon} + A\widehat{R}_0 A^T
$$

Update after O_1 :

 $P(S_1|O_{0:1}) = C \cdot P(S_1|O_0)P(O_1|S_1)$

$$
\sum_{i=1}^{\infty} \sum_{s} s_{t+1} = A_t s_t + \varepsilon_t
$$

Prediction at time 0:

$$
P(S_0) = N(\overline{s}_0, R_0)
$$

Update after O_0 : : Prediction at time 1: $\hat{\mathbf{S}}_0 = N(S_0, K_0)$ $\hat{\mathbf{S}}_0 = \bar{\mathbf{S}}_0 + K_0 (\mathbf{0}_0 - \mathbf{B} \bar{\mathbf{S}}_0 - \mathbf{\mu}_{\gamma})$ $\hat{\mathbf{R}}_0 = (I - K_0) R_0$ $R_1 = \theta_{\varepsilon} + A\hat{R}_0 A^T$ $\bar{s}_1 = A\hat{s}_0 + \mu_{\varepsilon}$ \boldsymbol{T} $K_0 = R_0 B^T (B R_0 B^T + \theta_\gamma)^{-1}$
 $\hat{s}_0 = \bar{s}_0 + K_0 (O_0 - B \bar{s_0} - \mu_\gamma)$ $\hat{R}_0 = (I - K_0) R_0$
 $\bar{s}_1 = A \hat{s}_0 + \mu_\varepsilon$ $\boldsymbol{K_0} = \boldsymbol{R}_0 \boldsymbol{B}^\mathrm{T} \big(\boldsymbol{B} \boldsymbol{R}_0 \boldsymbol{B}^\mathrm{T} + \boldsymbol{\varTheta}_{\boldsymbol{\gamma}} \big)^{-1}$

Update after O_1 : :

 $P(S_1|O_{0:1}) = C \cdot P(S_1|O_0)P(O_1|S_1)$

$$
\sum_{i=1}^{\infty} \sum_{s} s_{t+1} = A_t s_t + \varepsilon_t
$$

Prediction at time 0:

$$
P(S_0) = N(\overline{s}_0, R_0)
$$

Update after O_1 : : Update after O_0 : : Prediction at time 1: $P(S_1|O_{0:1}) = C.P(S_1|O_0)P(O_1|S_1) = N(\hat{s}_1, \hat{R}_1)$ $\hat{\mathbf{S}}_0 = N(S_0, K_0)$ $\hat{\mathbf{S}}_0 = \bar{\mathbf{S}}_0 + K_0 (\mathbf{0}_0 - \mathbf{B} \bar{\mathbf{S}}_0 - \mathbf{\mu}_{\gamma})$ $\hat{\mathbf{R}}_0 = (I - K_0 \mathbf{R}) \mathbf{R}_0$ $R_1 = \theta_{\varepsilon} + A\hat{R}_0 A^T$ $\bar{s}_1 = A\hat{s}_0 + \mu_{\varepsilon}$ \boldsymbol{T} 1, \mathbf{K}_1) \mathbf{K}_2 $\left(\mathbf{\Theta}_{\gamma}\right)^{-1}$
 $\mathbf{\widehat{R}}_{0} = (I - K_{0}B) R_{0}$ $\hat{R}_1 = (I - K_1 B) R_1$ $\boldsymbol{K_0} = \boldsymbol{R}_0 \boldsymbol{B}^\mathrm{T} \big(\boldsymbol{B} \boldsymbol{R}_0 \boldsymbol{B}^\mathrm{T} + \boldsymbol{\varTheta}_{\boldsymbol{\gamma}} \big)^{-1}$ $\hat{s}_1 = \bar{s}_1 + K_1 (\bm{\theta}_1 - B \bar{s}_1 - \mu_\gamma)$ $K_1 = R_1 B^T (BR_1 B^T + \mathcal{O}_\gamma)^{-1}$

$$
\sum_{i=1}^{\infty} \sum_{s} s_{t+1} = A_t s_t + \varepsilon_t
$$

Prediction at time 0:

$$
P(S_0) = N(\overline{s}_0, R_0)
$$

Update after O_1 : : Update after O_0 : : Prediction at time 1: $P(S_0|O_0) = N(\hat{s}_0, \hat{R}_0)$ $\hat{s}_0 = \bar{s}_0 + K_0(O_0 - B\bar{s}_0 - \mu_Y)$ $\hat{R}_0 = (I - K_0B)R_0$ $R_1 = \theta_{\varepsilon} + A\hat{R}_0 A^T$ $\overline{s}_1 = A\hat{s}_0 + \mu_{\varepsilon}$ \boldsymbol{T} $P(S_1 | O_{0.1}) = N(\hat{s}_1, \hat{R}_1)$ $K_0 = R_0 B^{T} (BR_0 B^{T} + \theta_{\gamma})^{-1}$
 $\hat{s}_0 = \bar{s}_0 + K_0 (O_0 - B \bar{s_0} - \mu_{\gamma})$ $\hat{R}_0 = (I - K_0 B) R_0$
 $\bar{s}_1 = A \hat{s}_0 + \mu_{\varepsilon}$ $\boldsymbol{K_0} = \boldsymbol{R}_0 \boldsymbol{B}^\mathrm{T} \big(\boldsymbol{B} \boldsymbol{R}_0 \boldsymbol{B}^\mathrm{T} + \boldsymbol{\varTheta}_{\boldsymbol{\gamma}} \big)^{-1}$ $\hat{s}_1 = \bar{s}_1 + K_1 (\theta_1 - B \bar{s}_1 - \mu_\gamma)$ $\hat{R}_1 = (I - K_1 B) R_1$ $\pmb{K}_\mathbf{1} = \pmb{R}_1 \pmb{B}^\mathrm{T} \big(\pmb{B} \pmb{R}_1 \pmb{B}^\mathrm{T} + \pmb{\varTheta}_{\pmb{\gamma}} \big)^{-\mathbf{1}}$

Gaussian Continuous State Linear Systems Gaussian Continuou

Linear System
 $s_{t+1} = A_t s_t + \varepsilon_t$
 $o_t = B_t s_t + \gamma_t$

Prediction at time t:
 $P(S_t | O_{0:t-1}) = \int_0^\infty P(S_{t-1} | O_{0:t-1}) P_s$ \overrightarrow{s} $o_t = B_t s_t + \gamma_t$ $S_{t+1} = A_t S_t + \varepsilon_t$

 $P_{\rm o}(\rm s)$

$$
P(S_t|O_{0:t-1}) = \int_{-\infty}^{\infty} P(S_{t-1}|O_{0:t-1}) P(S_t|S_{t-1}) dS_{t-1}
$$

Update after observing O_t : :

$$
P(S_t | O_{0:t}) = C.P(S_t | O_{0:t-1})P(O_t | S_t)
$$

Gaussian Continuous State Linear Systems $S_{t+1} = A_t S_t + \varepsilon_t$ $o_t = B_t s_t + \gamma_t$ \overline{s}

Prediction at time t:

 $P_o(s)$

$$
P(S_t|O_{0:t-1}) = N(\bar{s}_t, R_t)
$$

$$
\bar{s}_t = A\hat{s}_{t-1} + \mu_{\varepsilon}
$$

$$
R_t = \Theta_{\varepsilon} + A\hat{R}_{t-1}A^T
$$

Update after observing O.:

 $P(S_t|O_{0:t}) = N(\hat{S}_t, \hat{R}_t)$

$$
K_t = R_1 B^T (BR_1 B^T + \Theta_\gamma)^{-1}
$$

$$
\hat{s}_t = \bar{s}_t + Kt (Ot - B\bar{s}_t - \mu_\gamma)
$$

$$
\hat{R}_t = (I - KtB) R_t
$$

Gaussian Continuous State Linear Systems

$$
S_{t+1} = A_t S_t + \mathcal{E}_t
$$

$$
O_t = B_t S_t + \gamma_t
$$

Prediction at time t:

$$
P(S_t|O_{0:t-1}) = N(\bar{s}_t, R_t)
$$

Update after observing O.:

 $P(S_t|O_{0:t}) = N(\hat{s}_t, \hat{R}_t)$

KALMAN FILTER

$$
\bar{s}_t = A\hat{s}_{t-1} + \mu_{\varepsilon}
$$

$$
R_t = \Theta_{\varepsilon} + A\hat{R}_{t-1}A^T
$$

$$
K_t = R_1 B^T (BR_1 B^T + \Theta_\gamma)^{-1}
$$

$$
\hat{s}_t = \bar{s}_t + Kt (Ot - B\bar{s}_t - \mu_\gamma)
$$

$$
\hat{R}_t = (I - KtB) R_t
$$

The Kalman filter
(based on state equation)

• Prediction (based on state equation)

$$
\overline{s}_t = A_t \hat{s}_{t-1} + \mu_{\varepsilon} \qquad \qquad s_t = A_t s_{t-1} + \varepsilon_t
$$

$$
R_t = \Theta_{\varepsilon} + A_t \hat{R}_{t-1} A_t^T
$$

• Update (using observation and observation equation) -1 $\overline{\Omega} = R \overline{S} + \gamma$

$$
K_{t} = R_{t}B_{t}^{T}\left(B_{t}R_{t}B_{t}^{T} + \Theta_{\gamma}\right)^{-1} \quad \ \ \sigma_{t} = B_{t}s_{t} + \gamma_{t}
$$
\n
$$
\hat{s}_{t} = \overline{s}_{t} + K_{t}\left(\sigma_{t} - B_{t}\overline{s}_{t} - \mu_{\gamma}\right)
$$
\n
$$
\hat{R}_{t} = \left(I - K_{t}B_{t}\right)R_{t}
$$
\n
$$
11-755/18797
$$

$$
\hat{R}_t = (I - K_t B_t) R_t
$$

Explaining the Kalman Filter
ediction
ediction

• Prediction

$$
\overline{s}_t = A_t \hat{s}_{t-1} + \mu_{\varepsilon}
$$

$$
S_t = A_t S_{t-1} + \mathcal{E}_t
$$

$$
o_t = B_t s_t + \gamma_t
$$

$$
R_t = \Theta_{\varepsilon} + A_t \hat{R}_{t-1} A_t^T
$$

• The Kalman filter can be explained but working through
 $K_t(o_t - B_t\overline{s}_t - \mu_\gamma)$
 $I - K_tB_t$) R_t <u>UT WOPKING TNP(</u> $\overline{}$ uthout work<mark>ı</mark>r Explaining the Nation
 $s_t = A_ts_{t-1} + \varepsilon_t$
 $\overline{s}_t = A_t\hat{s}_{t-1} + \mu_{\varepsilon}$
 $\overline{s}_t = \Theta_{\varepsilon} + A_t\hat{R}_{t-1}A_t^T$

The Kalman filter can be explained

intuitively without working through intuitively without working through the math

$$
\hat{S}_t = \overline{S}_t + K_t (o_t - B_t \overline{S}_t - \mu_\gamma)
$$

$$
\hat{R}_t = (I - K_t B_t) R_t
$$

Explaining the Kalman Filter
ediction
ediction

• Prediction

$$
\overline{s}_t = A_t \hat{s}_{t-1} + \mu_{\varepsilon}
$$

$$
S_t = A_t S_{t-1} + \mathcal{E}_t
$$

$$
o_t = B_t s_t + \gamma_t
$$

$$
R_t = \Theta_{\varepsilon} + A_t \hat{R}_{t-1} A_t^T
$$

• The Kalman fil <u>UT WOPKING TNP(</u> $\overline{}$ uthout work<mark>ı</mark>r intuitively without working through \vert **Explaining the Nation

Frediction**
 $\overline{s}_t = A_t \hat{s}_{t-1} + \mu_{\varepsilon}$
 $\overline{s}_t = \Theta_{\varepsilon} + A_t \hat{R}_{t-1} A_t^T$

The Kalman filter can be explained

intuitively without working through the math

 $\hat{s} = \bar{s} + K I \hat{a} - R \bar{s} - H I$ $\overline{\mathbf{c}}$ but working through
 $K \log_B R \bar{s} = \mu$
 \pm think of the filter

the state, and (b) the

ne estimate a (a) the state, a ˆ as estimating (a) the state, and (b) the \vert To do so, we must think of the filter uncertainty of the estimate

$$
s_t = A_t s_{t-1} + \varepsilon_t
$$

$$
o_t = B_t s_t + \gamma_t
$$

- If our best guess for the state at time $t-1$ is \hat{S}_{t-1} , what is our best prediction for S_t ?
- If the guess \hat{s}_{t-1} as uncertainty (variance)
 \hat{R}_{t-1} , what is the uncertainty of the prediction r best guess for the state at time $t-1$ is
what is our best prediction for s_t ?
e guess \hat{s}_{t-1} as uncertainty (variance)
, what is the uncertainty of the prediction
e state at t ? of the state at t ? External prediction for s_t ?

uncertainty (variance)

ncertainty of the prediction

11-755/18797

 $\begin{array}{ccc} \n\text{unphy by } \text{up output} \ \n\text{...} \quad \text{...} \$ state dynamics equation
 $(B_t R_t B_t^+ + \Theta_\gamma)$
 $K_t (o_t - B_t \overline{s}_t - \mu_\gamma)$
 $(I - K_t B_t) R_t$ The predicted state at time t is obtained $= K_t B_t \left(B_t K_t B_t^+ + \Theta_{\gamma} \right)$ at t-1 through the state dynamics equation \blacksquare $K_t = K_t B_t \left(B_t K_t B_t \right)$ simply by propagating the estimated state

$$
\hat{S}_t = \overline{S}_t + K_t \big(O_t - B_t \overline{S}_t - \mu_\gamma \big)
$$

$$
\hat{R}_t = (I - K_t B_t) R_t
$$

This is the uncert my in the prediction.

predictor =

ance of As_{t-1}

because ε_t is not $\frac{1}{1-t}$ predictor = 1 the predictor The variance of the predictor = This is the uncertainty in the prediction. variance of ε_t + variance of As_{t-1}

 $\frac{1}{\epsilon}$ ˆ <u>I</u> The two simply add because $\varepsilon_{\rm t}$ is not $correlated with s_t$

52

We can also predi using the observ 1 tate using the the predicted state using the observation \vert We can also predict the observation from | equation

$$
S_t = S_t + K_t (O_t - B_t S_t - \mu_\gamma)
$$

$$
\hat{R}_t = (I - K_t B_t) R_t
$$

$$
s_t = A_t s_{t-1} + \varepsilon_t
$$

$$
o_t = B_t s_t + \gamma_t
$$

• If our best prediction for the state at time t is \bar{s}_t , what is our best prediction for o_t ? 11-755/18797 53

$$
s_t = A_t s_{t-1} + \varepsilon_t
$$

$$
o_t = B_t s_t + \gamma_t
$$

- If our best prediction for the state at time t is \bar{s}_t , what is our best prediction for o_t ?
	- If \bar{s}_t has uncertainty (variance) R_t , what is the uncertainty of the prediction of the observation at t ? 11- π ance) R_t , what is the uncertainty of the stion at t ?
htion at t ?
htion at t ?

$$
s_t = A_t s_{t-1} + \varepsilon_t
$$

$$
o_t = B_t s_t + \gamma_t
$$

- If our best prediction for the state at time t is \bar{s}_t , what is our best prediction for o_t ?
	- If \bar{s}_t has uncertainty (variance) R_t , what is the uncertainty of the prediction of the observation at t ?
- Will the predicted $\hat{\sigma}_t$ be the same as the actual observation of o_t ? The same as the uncertainty of the same as the actual observation
the same as the actual observation
 $\frac{1}{11-755/18797}$

$$
S_t = A_t S_{t-1} + \varepsilon_t
$$

$$
o_t = B_t s_t + \gamma_t
$$

- If our best prediction for the state at time t is \bar{s}_t , what is our best prediction for o_t ?
	- If \bar{s}_t has uncertainty (variance) R_t , what is the uncertainty of the prediction of the observation at t ?
- Will the predicted $\hat{\sigma}_t$ be the same as the actual observation of O_t ? ¹
ance) R_t , what is the uncertainty of the
tion at t ?
the same as the actual observation
ir guess \bar{s}_t to account for this
11-755/18797
	- How should we adjust our guess \bar{s}_t to account for this difference?

MAP Recap (for Gaussians)

• If $P(x,y)$ is Gaussian:

$$
\hat{y} = \mu_y + C_{yx} C_{xx}^{-1} (x - \mu_x)
$$

MAP Recap: For Gaussians

• If $P(x,y)$ is Gaussian:

The Kalman filter $\frac{s_t = A_t s_{t-1} + \varepsilon_t}{o_t = B_t s_t + \gamma_t}$

$$
o_t = B_t s_t + \gamma_t
$$

 $\hat{o}_t = B_t \overline{s}_t + \mu_\gamma$

 O_t

 $s_t = A_t s_{t-1} + \varepsilon_t$

• Prediction

$$
R_t = \Theta_{\varepsilon} + A_t \hat{R}_{t-1} A_t^T
$$

 $\overline{s}_t = A_t \hat{s}_{t-1} + \mu_{\varepsilon}$

• Update

$$
K_t = R_t B_t^T \left(B_t R_t B_t^T + \Theta_{\gamma} \right)^{-1}
$$

 $\frac{1}{T}\left(B_{t}R_{t}B_{t}^{T}+\Theta_{\gamma}\right)^{-1}$

the MAP estimator

(a) = C_{00}

he Kalman Gain $RBT+ Θ) = $C_{oo}$$ \mathbf{z} $RB^{T} = C_{so}$, $(BRB^{T} + \Theta) = C_{oo}$ This is the slope of the MAP estimator that predicts s from o

This is also called the Kalman Gain

I IS THE UITTERENCE $\begin{array}{c} \begin{array}{c} \text{111} \\ \text{11} \end{array} \end{array}$ the actual observation and the predicted observation, scaled by the Kalman Gain

I IS THE UITTERENCE \overline{a} $\begin{array}{c} \begin{array}{c} \text{111} \\ \text{11} \end{array} \end{array}$ The correction is the difference between the actual observation and the predicted observation, scaled by the Kalman Gain

The Kalman filter
 $s_t = A_t s_{t-1} + \varepsilon_t$

• Prediction

$$
\left| \overline{S}_t = A_t \hat{S}_{t-1} + \mu_{\varepsilon} \right|
$$

$$
S_t = A_t S_{t-1} + \varepsilon_t
$$

$$
o_t = B_t S_t + \gamma_t
$$

$$
R_t = \Theta_{\varepsilon} + A_t \hat{R}_{t-1} A_t^T
$$

• Update:

Ke a correction h observe the data and make a correction The uncertainty in state decreases if we

hrinkage" based on l ^t ^t ^t ^t ^t ^t s muitiplicative The reduction is a multiplicative • Update:

• Update:

The uncertainty in state decreases if we

observe the data and make a correction

The reduction is a multiplicative

"shrinkage" based on Kalman gain and B
 $\hat{R}_t = (I - K_t B_t)R_t$

$$
\left| \hat{R}_t = \left(I - K_t B_t \right) R_t \right|
$$

The Kalman filter
 $s_t = A_t s_{t-1} + \varepsilon_t$

• Prediction

$$
\left| \overline{S}_t = A_t \hat{S}_{t-1} + \mu_{\varepsilon} \right|
$$

$$
R_t = \Theta_{\varepsilon} + A_t \hat{R}_{t-1} A_t^T
$$

• Update:

• Update

$$
K_{t} = R_{t}B_{t}^{T}\left(B_{t}R_{t}B_{t}^{T} + \Theta_{\gamma}\right)^{-1}
$$

$$
\hat{s}_{t} = \overline{s}_{t} + K_{t}\left(\sigma_{t} - B_{t}\overline{s}_{t} - \mu_{\gamma}\right)
$$

$$
\hat{R}_{t} = \left(I - K_{t}B_{t}\right)R_{t}
$$

$$
\hat{S}_t = \overline{S}_t + K_t \left(O_t - B_t \overline{S}_t - \mu_\gamma \right)
$$

$$
\hat{R}_t = (I - K_t B_t) R_t
$$

$$
S_t = A_t S_{t-1} + \varepsilon_t
$$

$$
O_t = B_t S_t + \gamma_t
$$

Kalman filter

- **Predict state** \bullet
- **Predict measurement** \bullet
- Compute measurement error \bullet
- Update state \bullet

Kalman filter

- **Predict state** \bullet
- **Predict measurement** \bullet
- Compute measurement error \bullet
- **Update state** \bullet
- Note: Progress of Kalman gain is not actually dependent on observations or estimated state... \bullet 66

The Kalman Filter

- Very popular for tracking the state of processes
	- Control systems
	- Robotic tracking
		- Simultaneous localization and mapping
	- Radars
	- Even the stock market..
- What are the parameters of the process? eding and mapping

11-755/18797

11-755/18797

11-755/18797

Kalman filter contd.
 $s_t = A_t s_{t-1} + \varepsilon_t$

$$
S_t = A_t S_{t-1} + \mathcal{E}_t
$$

$$
O_t = B_t S_t + \gamma_t
$$

- Model parameters A and B must be known – Often the state equation includes an additional driving term: $s_t = A_t s_{t-1} + G_t u_t + \varepsilon_t$
	- The parameters of the driving term must be known $\mu_t s_{t-1} + G_t u_t + \varepsilon_t$

	the driving term must be

	ribution must be known

	ili-755/18797

	68
- The initial state distribution must be known

Defining the parameters

- State state must be carefully defined
- **Defining the parameters**
State state must be carefully defined
- E.g. for a robotic vehicle, the state is an extended
vector that includes the current velocity and
acceleration vector that includes the current velocity and acceleration
	- $S = [X, dX, d^2X]$
- State equation: Must incorporate appropriate constraints In the appropriate

Interation and velocity, velocity at

delocity + acc. * time step

0 1]

11-755/18797

11-755/18797
	- If state includes acceleration and velocity, velocity at next time = current velocity + acc. $*$ time step

$$
-St = AS_{t-1} + e
$$

• $A = [1 t 0.5t^2; 0 1 t; 0 0 1]$

Parameters

- Observation equation:
	- Critical to have accurate observation equation
	- Must provide a valid relationship between state and observations
- Observations typically high-dimensional – May have higher or lower dimensionality than $11y$ high-dimensional
lower dimensionality than
11-755/18797
	- state

Problems

$$
S_t = f(S_{t-1}, \mathcal{E}_t)
$$

$$
O_t = g(S_t, \gamma_t)
$$

- f() and/or $g($) may not be nice linear functions – Conventional Kalman update rules are no longer valid
- ϵ and/or γ may not be Gaussian – Gaussian based update rules no longer valid the gaussian
be Gaussian
late rules no longer valid

All distributions remain Gaussian

Problems

$$
S_t = f(S_{t-1}, \mathcal{E}_t)
$$

$$
O_t = g(S_t, \gamma_t)
$$

- Nonlinear f() and/or g() : The Gaussian assumption breaks down $\begin{aligned} s_t &= f(s_{t-1}, \varepsilon_t) \ \frac{o_t = g(s_t, \gamma_t)}{\end{aligned}$
Nonlinear f() and/or g() : The Gaussian
assumption breaks down
— Conventional Kalman update rules are no longer
valid
	- valid $\begin{align*} \mathsf{in} & \mathsf{un} \ \mathsf{update} & \mathsf{rule} \ \mathsf{are} & \mathsf{no} \ \mathsf{longer} \end{align*}$

The problem with non-linear functions

$$
P(s_t | o_{0:t-1}) = \int_{-\infty}^{\infty} P(s_{t-1} | o_{0:t-1}) P(s_t | s_{t-1}) ds_{t-1}
$$

$$
P(s_t | o_{0:t}) = CP(s_t | o_{0:t-1}) P(o_t | s_t)
$$

- Estimation requires knowledge of $P(o|s)$
	- $-$ Difficult to estimate for nonlinear $g()$
	- Even if it can be estimated, may not be tractable with update loop ge of $P(o|s)$

	Ilinear g()

	may not be tractable with update loop

	wledge of $P(s_t|s_{t-1})$

	osed form integration

	11-755/18797

	74
- Estimation also requires knowledge of $P(s_t|s_{t-1})$
	- $-$ Difficult for nonlinear f()

 $o_t = g(s_t, \gamma_t)$

 $S_t = f(S_{t-1}, \mathcal{E}_t)$

– May not be amenable to closed form integration

The problem with nonlinearity

$$
o_t = g(s_t, \gamma_t)
$$

• The PDF may not have a closed form

$$
P(o_t | s_t) = \sum_{\gamma: g(s_t, \gamma) = o_t} \frac{P_{\gamma}(\gamma)}{|J_{g(s_t, \gamma)}(o_t)|}
$$

$$
|J_{g(s_t, \gamma)}(o_t)| = \begin{vmatrix} \frac{\partial o_t(1)}{\partial \gamma(1)} & \cdots & \frac{\partial o_t(1)}{\partial \gamma(n)} \\ \vdots & \ddots & \vdots \\ \frac{\partial o_t(n)}{\partial \gamma(1)} & \cdots & \frac{\partial o_t(n)}{\partial \gamma(n)} \end{vmatrix}
$$

• Even if a closed form exists initially, it will typically become intractable very quickly $\begin{array}{l} \left| \frac{\partial o_r(1)}{\partial \gamma(1)} \right| \cdots \left| \frac{\partial o_r(1)}{\partial \gamma(n)} \right| \ \vdots \ \left| \frac{\partial o_r(n)}{\partial \gamma(1)} \right| \cdots \left| \frac{\partial o_r(n)}{\partial \gamma(n)} \right| \ \end{array}$
cists initially, it will typically
y quickly

Example: a simple nonlinearity

$$
o = \gamma + \log(1 + \exp(s))
$$

• $P(O|S) = ?$

 $-$ Assume γ is Gaussian

- Assume γ is Gaussian
\n
$$
- P(\gamma) = Gaussian(\gamma; \mu_{\gamma}, \Theta_{\gamma})
$$
\n
$$
11-755/18797
$$
\n
$$
11-755/18797
$$

Example: a simple nonlinearity

 $\frac{1}{2}$
bservation (o)

 $^{0}_{8}$

 -6

$$
o = \gamma + \log(1 + \exp(s))
$$

• $P(O|S) = ?$

$$
P(\gamma) = Gaussian(\gamma; \mu_{\gamma}, \Theta_{\gamma})
$$

 $\begin{split} &\frac{sign(\gamma;\mu_\gamma,\Theta_\gamma)}{11000}(1+exp(s)),\Theta_\gamma) \ &\frac{11000}{2} \end{split}$ $P(o | s) = Gaussian(o; \mu_{\gamma} + log(1 + exp(s)), \Theta_{\gamma})$

 $\sqrt[4]{x} = 0$; $\sin^{\circ} s$

 \overline{a}

 $\overline{4}$

 6

Example: At T=0.

$$
o = \gamma + \log(1 + \exp(s))
$$

Assume initial probability P(s) is Gaussian

$$
P(s_0) = P_0(s) = Gaussian(s; \overline{s}, R)
$$

 $\frac{1}{2}$ s=0

• Update $P(s_0 | o_0) = CP(o_0 | s_0) P(s_0)$

 $S(11, 12)$ is Gaussian

= $Gaussian(s; \overline{s}, R)$

= $CP(o_0 | s_0)P(s_0)$
 $log(1+exp(s_0)), \Theta_\gamma)Gaussian(s_0; \overline{s}, R)$ $P(s_0 | o_0) = CGaussian(o; \mu_\gamma + log(1 + exp(s_0)), \Theta_\gamma)Gaussian(s_0; \overline{s}, R)$

UPDATE: At T=0.

 $P(s_0 | o_0) = CGaussian(o; \mu_\nu + \log(1 + \exp(s_0)), \Theta_\nu)$ Gaussian $(s_0; \overline{s}, R)$

$$
P(s_0 | o_0) = CGaussian(o; \mu_{\gamma} + log(1 + exp(s_0)), \Theta_{\gamma})Gaussian(s_0; \bar{s}, R)
$$

\n
$$
P(s_0 | o_0) = C exp\left(\frac{-0.5(\mu_{\gamma} + log(1 + exp(s_0)) - o)^T \Theta_{\gamma}^{-1}(\mu_{\gamma} + log(1 + exp(s_0)) - o)}{-0.5(s_0 - \bar{s})^T R^{-1} (s_0 - \bar{s})}\right)
$$

\n• = Not Gaussian

• = Not Gaussian

Prediction for T = 1

$$
S_t = S_{t-1} + \mathcal{E}
$$
 $P(\mathcal{E}) = Gaussian(\mathcal{E}; 0, \Theta_{\mathcal{E}})$

 $P(s_1 | o_0) = \int P(s_0 | o_0) P(s_1 | s_0) ds_0$ ∞ $-\infty$ Prediction $P(s_1 | o_0) =$ ■ Trivial, linear state transition equation $P(s_t | s_{t-1}) = Gaussian(s_t; s_{t-1}, \Theta_{\varepsilon})$

11-755/18797 80 ¹ ⁰ ⁰ 1 1 0 0 1 0 0 1 0 ¹ ⁰ exp 0.5() () 0.5(log(1 exp())) (log(1 exp())) (| o) exp s s s s ds s s R s s s o s o P s C T T T ^e ^g ^g ^g

= intractable

Update at T=1 and later

• Update at T=1 $P(s_t | o_{0:t}) = CP(s_t | o_{0:t-1}) P(o_t | s_t)$

– Intractable

• Prediction for T=2

table

\nion for T=2

\n
$$
P(s_t \mid o_{0:t-1}) = \int_{-\infty}^{\infty} P(s_{t-1} \mid o_{0:t-1}) P(s_t \mid s_{t-1}) ds_{t-1}
$$
\nstable

\n
$$
11-755/18797
$$

– Intractable

The State prediction Equation

$$
S_t = f(S_{t-1}, \mathcal{E}_t)
$$

- Similar problems arise for the state prediction equation
- $P(s_t|s_{t-1})$ may not have a closed form
- Even if it does, it may become intractable within the prediction and update equations e a closed form

v become intractable within

pdate equations

ction equation, which includes an

1

11-755/18797

82
	- Particularly the prediction equation, which includes an integration operation

The tangent at any point is a good local approximation if the function is sufficiently smooth

The tangent at any point is a good local approximation if the function is sufficiently smooth

The tangent at any point is a good local approximation if the function is sufficiently smooth

• The tangent at any point is a good local approximation if the function is sufficiently smooth

Linearizing the observation function

$$
P(s_t | o_{0:t-1}) = Gaussian(\overline{s}_t, R_t)
$$

$$
o = \gamma + g(s) \qquad o \approx \gamma + g(\overline{s}_t) + J_g(\overline{s}_t)(s - \overline{s}_t)
$$

- Simple first-order Taylor series expansion – J() is the Jacobian matrix
	- Simply a determinant for scalar state
- Expansion around current predicted a priori (or predicted) mean of the state natrix
nt for scalar state
urrent predicted *a priori*
n of the state
n changes with time
11-755/18797
	- Linear approximation changes with time

- $P(s_t)$ is small where approximation error is large
	- Most of the probability mass of s is in low-error regions

Linearizing the observation function

$$
P(s_t | o_{0:t-1}) = Gaussian(\overline{s}_t, R_t)
$$

$$
o = \gamma + g(s) \qquad o \approx \gamma + g(\overline{s}_t) + J_g(\overline{s}_t)(s - \overline{s}_t)
$$

- With the linearized approximation the system becomes "linear"
- The observation PDF becomes Gaussian

 $P(\gamma) = Gaussian(\gamma; 0, \Theta_{\gamma})$

becomes "linear"
\nThe observation PDF becomes Gaussian
\n
$$
P(\gamma) = Gaussian(\gamma; 0, \Theta_{\gamma})
$$
\n
$$
P(o \mid s) = Gaussian(o; g(\bar{s}) + J_g(\bar{s})(s - \bar{s}), \Theta_{\gamma})
$$

The state equation?

$$
S_t = f(S_{t-1}) + \varepsilon \qquad P(\varepsilon) = Gaussian(\varepsilon; 0, \Theta_{\varepsilon})
$$

- Again, direct use of f() can be disastrous
- Solution: Linearize

$$
P(s_{t-1} | o_{0:t-1}) = Gaussian(s_{t-1}; \hat{s}_{t-1}, \hat{R}_{t-1})
$$

an(s_{t-1}; \hat{s}_{t-1} , \hat{R}_{t-1})
 $\varepsilon + f(\hat{s}_{t-1}) + J_f(\hat{s}_{t-1})(s_{t-1} - \hat{s}_{t-1})$

mean of the updated

- 1

p a linear one

11-755/18797 $s_t = f(s_{t-1}) + \varepsilon$ $\sum_{t=1}^{t} s_t \approx \varepsilon + f(\hat{s}_{t-1}) + J_f(\hat{s}_{t-1})(s_{t-1} - \hat{s}_{t-1})$

- Solution: Linearize
 $P(S_{t-1} | o_{0:t-1}) = Gaussian(s_{t-1}; \hat{s}_{t-1}, \hat{R}_{t-1})$
 $= f(s_{t-1}) + \varepsilon$ $\qquad s_t \approx \varepsilon + f(\hat{s}_{t-1}) + J_f(\hat{s}_{t-1})(s_{t-1} \hat{s}_{t-1})$

Linearize around the mean of the updated distribution of s at $t 1$ distribution of s at $t-1$
	- Converts the system to a linear one

Linearized System

$$
o = \gamma + g(s)
$$

\n
$$
s_{t} = f(s_{t-1}) + \varepsilon
$$

\n
$$
o \approx \gamma + g(\overline{s}_{t}) + J_{g}(\overline{s}_{t})(s - \overline{s}_{t})
$$

\n
$$
s_{t} \approx \varepsilon + f(\hat{s}_{t-1}) + J_{f}(\hat{s}_{t-1})(s_{t-1} - \hat{s}_{t-1})
$$

- Now we have a simple time-varying linear system $o \approx \gamma + g(\bar{s}_t) + J_g(\bar{s}_t)(s - \bar{s}_t)$
 $s_t \approx \varepsilon + f(\hat{s}_{t-1}) + J_f(\hat{s}_{t-1})(s_{t-1} - \hat{s}_{t-1})$

• Now we have a simple time-varying linear

system

• Kalman filter equations directly apply
 $\frac{41.755/18797}{41.755/18797}$ $J_{-1})+J_{f}(\hat{s}_{t-1})(s_{t-1}-\hat{s}_{t-1})$
ple time-varying linear
ions directly apply
-

• Prediction

$$
\overline{s}_t = f(\hat{s}_{t-1})
$$

$$
R_t = \Theta_{\varepsilon} + A_t \hat{R}_{t-1} A_t^T
$$

$$
s_t = f(s_{t-1}) + \varepsilon
$$

 $o_t = g(s_t) + \gamma$

$$
A_t = J_f(\hat{S}_{t-1})
$$

$$
B_t = J_g(\bar{S}_t)
$$

Jacobians used in Linearization

 $\frac{1}{T}\left(B_{t}R_{t}B_{t}^{T}+\Theta_{\gamma}\right)^{-1}$
 $+K_{t}\left(\sigma_{t}-g(\overline{s}_{t})\right)$
 $=K_{t}B_{t}B_{t}$
 $=K_{$ Assuming ε and γ are 0 mean for simplicity

• Update

$$
K_t = R_t B_t^T \left(B_t R_t B_t^T + \Theta_{\gamma} \right)^{-1}
$$

$$
\hat{S}_t = \overline{S}_t + K_t (o_t - g(\overline{S}_t))
$$

$$
\hat{R}_t = (I - K_t B_t) R_t
$$

• Prediction

$$
\overline{s}_t = f(\hat{s}_{t-1})
$$

$$
S_t = f(S_{t-1}) + \varepsilon
$$

$$
o_t = g(s_t) + \gamma
$$

 $\begin{array}{ccc} \n\text{unphy by } \text{up output} \ \n\text{...} \quad \text{...} \$ state dynamics equation
 $\frac{(\mathcal{B}_t K_t \mathcal{B}_t + \mathfrak{S}_y)}{(\mathcal{B}_t K_t (\mathit{o}_t - g(\bar{s}_t)))}$
 $I - K_t B_t R_t$ state at time t is obtained $= K_t B_t \left(B_t K_t B_t^{\dagger} + \mathbf{\Theta}_{\gamma} \right)$ at t-1 through the state dynamics equation \blacksquare $K_t = K_t B_t \left(B_t K_t B_t \right)$ $\frac{1}{4}$ $\left(\hat{z} \right)$ 1 I $I(f)$ The predicted state at time t is obtained simply by propagating the estimated state

$$
\hat{S}_t = \overline{S}_t + K_t (o_t - g(\overline{S}_t))
$$

$$
\hat{R}_t = (I - K_t B_t) R_t
$$

• Prediction

$$
\overline{s}_t = f(\hat{s}_{t-1})
$$

$$
R_t = \Theta_{\varepsilon} + A_t \hat{R}_{t-1} A_t^T
$$

$$
s_t = f(s_{t-1}) + \varepsilon
$$

$$
o_t = g(s_t) + \varepsilon
$$

$$
A_t = J_f(\hat{S}_{t-1})
$$

$$
B_t = J_g(\overline{s}_t)
$$

Incertainty of nre predictor =

ance of As_{t-1}
 $\frac{2\arizing f()}{\sqrt{f'(\frac{f(t)-f(t))}{f(t-1)+f(t-1)+f(t-1)}}}$ $\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$ 1 the predictor The variance of the predictor $=$ Uncertainty of prediction. variance of ε_t + variance of As_{t-1}

 $\mathbf{r}_t - \mathbf{r}_t - \mathbf{r}_t \mathbf{r}_t + \mathbf{r}_t$ ˆ $\frac{1}{\sqrt{1}}$ A is obtained by linearizing f()

The Extended Kalman filter

• Prediction

$$
\overline{s}_t = f(\hat{s}_{t-1})
$$

$$
s_t = f(s_{t-1}) + \varepsilon
$$

$$
o_t = g(s_t) + \varepsilon
$$

$$
R_t = \Theta_{\varepsilon} + A_t \hat{R}_{t-1} A_t^T
$$

• Update

$$
B_t = J_g(\overline{s}_t)
$$

$$
K_t = R_t B_t^T \left(B_t R_t B_t^T + \Theta_{\gamma} \right)^{-1}
$$

 $\frac{d^T(B_t R_t B_t^T + \Theta_\gamma)^{-1}}{dt$
the slope of the MAP
dicts s from o
 Θ) = C_{∞}
earizing g() $RB^{T}+\Theta$) = C_{00} ˆ $RBT = C_{so}$, $(BRB^{T}+ \Theta) = C_{oo}$ The Kalman gain is the slope of the MAP $($ estimator that predicts s from o B is obtained by linearizing g()

The Extended Kalman filter

• Prediction

$$
s_t = f(s_{t-1}) + \varepsilon
$$

$$
\overline{S}_t = f(\hat{S}_{t-1}) \qquad \qquad O_t = g(S_t) + \varepsilon
$$

$$
R_t = \Theta_{\varepsilon} + A_t \hat{R}_{t-1} A_t^T
$$

ve can aiso preak $\begin{aligned} &\text{using the observation}\ &\frac{+K_t(o_t-g(\bar{s}_t))}{\sqrt{1-K_t B_t} R_t} \ &\text{if } \bar{o}_t=g(\bar{s}_t) \ &\text{if } \bar{o}_t=g(\bar{s}_t) \end{aligned}$ ad.i.g 0200. ル a state doing the We can also predict the observation from | the predicted state using the observation equation

$$
\hat{S}_t = \overline{S}_t + K_t (o_t - g(\overline{S}_t))
$$

$$
\hat{R}_t = (I - K_t B_t) R_t
$$

 $\overline{o}_t = g(\overline{s}_t)$

• Prediction

$$
\overline{S}_t = f(\hat{S}_{t-1})
$$

$$
s_t = f(s_{t-1}) + \varepsilon
$$

$$
o_t = g(s_t) + \varepsilon
$$

$$
R_t = \Theta_{\varepsilon} + A_t \hat{R}_{t-1} A_t^T
$$

We must correct the predicted value of Airig an observail 12 **TIER MUKING AN OD** t t t the state after making an observation

$$
\hat{S}_t = \overline{S}_t + K_t \big(o_t - g(\overline{S}_t) \big) \qquad \qquad \overline{O}_t = g(\overline{S}_t)
$$

aking an observation
 $+ K_t (o_t - g(\bar{s}_t))$ $\overline{o_t} = g(\bar{s}_t)$

he difference between

tion and the predicted

by the Kalman Gain I IS THE UITTERENCE \overline{a} $\begin{array}{c} \begin{array}{c} \text{111} \\ \text{11} \end{array} \end{array}$ The correction is the difference between the actual observation and the predicted observation, scaled by the Kalman Gain

• Prediction

$$
\overline{s}_t = f(\hat{s}_{t-1})
$$

$$
R_t = \Theta_{\varepsilon} + A_t \hat{R}_{t-1} A_t^T
$$

$$
o_t = g(s_t) + \varepsilon
$$

 $s_t = f(s_{t-1}) + \varepsilon$

$$
B_t = J_g(\bar{s}_t)
$$

The uncertainty in make a correction $\overline{\mathbf{c}}$ observe the data and make a correction The uncertainty in state decreases if we

I make a correction

ultiplicative "shrinkage"

and B
 $I - K_t B_t R_t$

11-755/18797 s a multiplicative shrinkd The reduction is a multiplicative "shrinkage" based on Kalman gain and B

$$
\hat{R}_t = (I - K_t B_t) R_t
$$

• Prediction

$$
\overline{s}_t = f(\hat{s}_{t-1})
$$

$$
R_t = \Theta_{\varepsilon} + A_t \hat{R}_{t-1} A_t^T
$$

$$
s_t = f(s_{t-1}) + \varepsilon
$$

$$
o_t = g(s_t) + \varepsilon
$$

$$
A_t = J_f(\hat{S}_{t-1})
$$

$$
B_t = J_g(\bar{S}_t)
$$

• Update

$$
K_{t} = R_{t}B_{t}^{T}\left(B_{t}R_{t}B_{t}^{T} + \Theta_{\gamma}\right)^{-1}
$$

$$
\hat{S}_{t} = \overline{S}_{t} + K_{t}\left(\sigma_{t} - g(\overline{S}_{t})\right)
$$

$$
\hat{R}_{t} = \left(I - K_{t}B_{t}\right)R_{t}
$$

$$
\hat{S}_t = \overline{S}_t + K_t (o_t - g(\overline{S}_t))
$$

$$
\hat{R}_t = (I - K_t B_t) R_t
$$

Extended Kalman filter

- Predict state
- Predict measurement
- Compute measurement error
- Update state

Kalman filter

- Predict state
- Predict measurement
- Compute measurement error
- Update state
- Note: Progress of Kalman gain is dependent on estimated state through the Jacobian... 101

EKFs

- **EKFS**
• EKFs are probably the most commonly used algorithm
for tracking and prediction
– Most systems are non-linear for tracking and prediction
	- Most systems are non-linear
	- Specifically, the relationship between state and observation is usually nonlinear
	- The approach can be extended to include non-linear functions of noise as well
- Most systems are non-linear
• Specifically, the relationship between state and
• Specifically, the relationship between state and
• The approach can be extended to include non-linear
• The term "Kalman filter" often simp - Most systems are non-linear
- Specifically, the relationship between state and
observation is usually nonlinear
- The approach can be extended to include non-lin
functions of noise as well
The term "Kalman filter" often Exterided to include non-imear

r" often simply refers to an

in most contexts.

11-755/18797

11-755/18797
- But..

EKFs have limitations

- If the non-linearity changes too quickly with s, the linear approximation is invalid First the non-linearity changes too quickly with s, the linear
pproximation is invalid
— Unstable
The estimate is often biased
— The true function lies entirely on one side of the approximation
Various extensions have been
	- Unstable
- The estimate is often biased
- The true function lies entirely on one side of the approximation ely on one side of the approximation
n proposed:
filters (IEKF)
JKF)
^{11-755/18797}
- Various extensions have been proposed:
	-
	-

Conclusions

- HMMs are predictive models
- Continuous-state models are simple extensions of HMMs
	- Same math applies
- Prediction of linear, Gaussian systems can be HMMs are predictive models

Continuous-state models are simple

extensions of HMMs

– Same math applies

Prediction of linear, Gaussian systems can be

performed by Kalman filtering

Prediction of non-linear, Gaussian syst
- Prediction of non-linear, Gaussian systems can extensions of HMMs

— Same math applies
Prediction of linear, Gaussian systems can be
performed by Kalman filtering
Prediction of non-linear, Gaussian systems can
be performed by Extended Kalman filtering
 **Gaussian systems can be
an filtering
near, Gaussian systems can
tended Kalman filtering
11-755/18797**