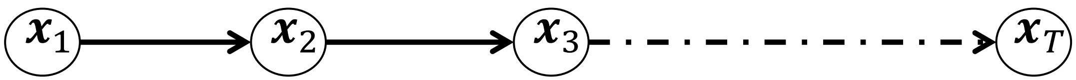
Linear Dynamical Systems (Kalman filter)

(a) Overview of HMMs

(b) From HMMs to Linear Dynamical Systems (LDS)

Markov Chains with Discrete Random Variables



Let's assume we have discrete random variables (e.g., taking 3 discrete values $\boldsymbol{x}_t = \{\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}\}$)

Markov Property: $p(\boldsymbol{x}_t|\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{t-1}) = p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$

$$\text{e.g. } p(\boldsymbol{x}_t = \begin{bmatrix}1\\0\\0\end{bmatrix} |\boldsymbol{x}_{t-1} = \begin{bmatrix}0\\1\\0\end{bmatrix})$$

Stationary, Homogeneous or Time-Invariant if the distribution $p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$ does not depend on t

Markov Chains with Discrete Random Variables

$$p(\boldsymbol{x}_1, .., \boldsymbol{x}_T) = p(\boldsymbol{x}_1) \prod_{t=2}^{\mathrm{T}} p(\boldsymbol{x}_t | \boldsymbol{x}_{t-1})$$

What do we need in order to describe the whole procedure?

(1) A probability for the first frame/timestamp etc $p(\boldsymbol{x}_1)$. In order to define the probability we need to define the vector $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_{\mathrm{K}})$

$$p(\boldsymbol{x}_1 | \boldsymbol{\pi}) = \prod_{c=1}^{K} {\pi_c}^{x_{1c}}$$

(2) A transition probability $p(\boldsymbol{x}_t | \boldsymbol{x}_{t-1})$. In order to define it we need a KxK transition matrix $\boldsymbol{A} = [a_{ij}]$

$$p(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}, \boldsymbol{A}) = \prod_{j=1}^{K} \prod_{k=1}^{K} {a_{jk}}^{x_{t-1j} x_{tk}}$$

Markov Chains with Discrete Random Variables

(1) Using the transition matrix we can compute various probabilities regarding future

$$p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_{t-1},\boldsymbol{A}) = \boldsymbol{A}^2$$

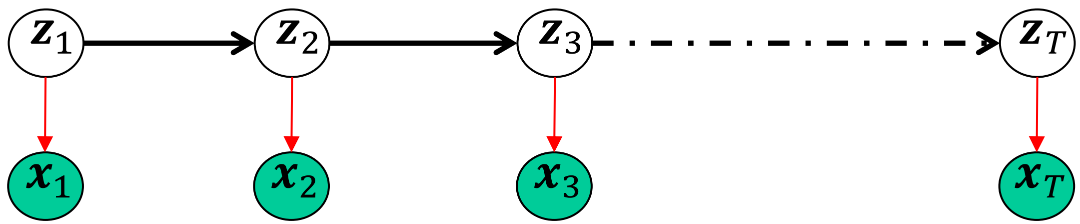
$$p(\boldsymbol{x}_{t+n}|\boldsymbol{x}_{t-1},\boldsymbol{A}) = \boldsymbol{A}^n$$

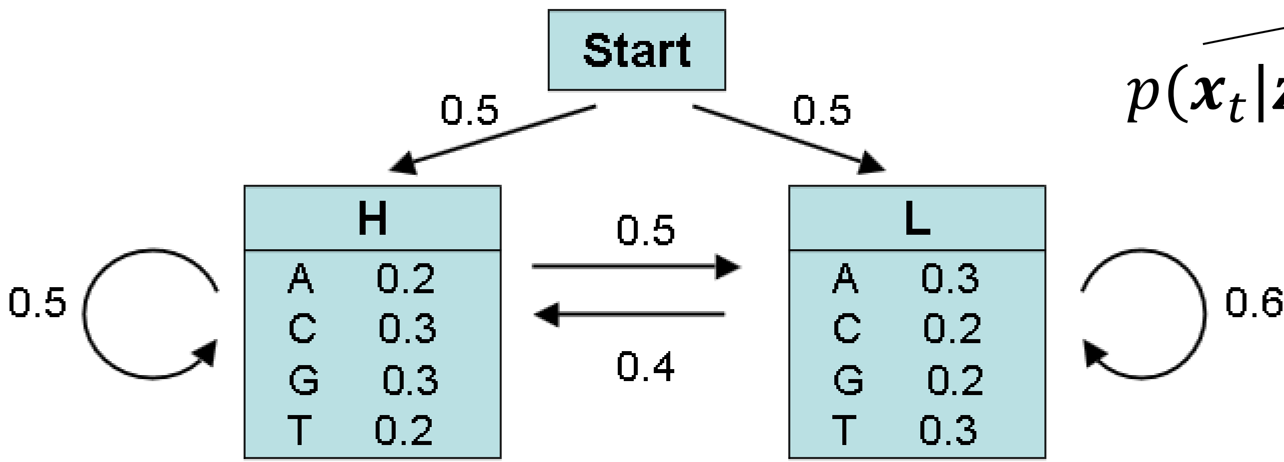
$$p(\boldsymbol{x}_{t+2}|\boldsymbol{x}_{t-1},\boldsymbol{A}) = \boldsymbol{A}^3$$

(1) The stationary probability of a Markov Chain is very important (it's an indication of how probable ending in one of states in random move) (Google Page Rank).

$$\boldsymbol{\pi}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{\pi}^{T}$$

Latent Variables in Markov Chain

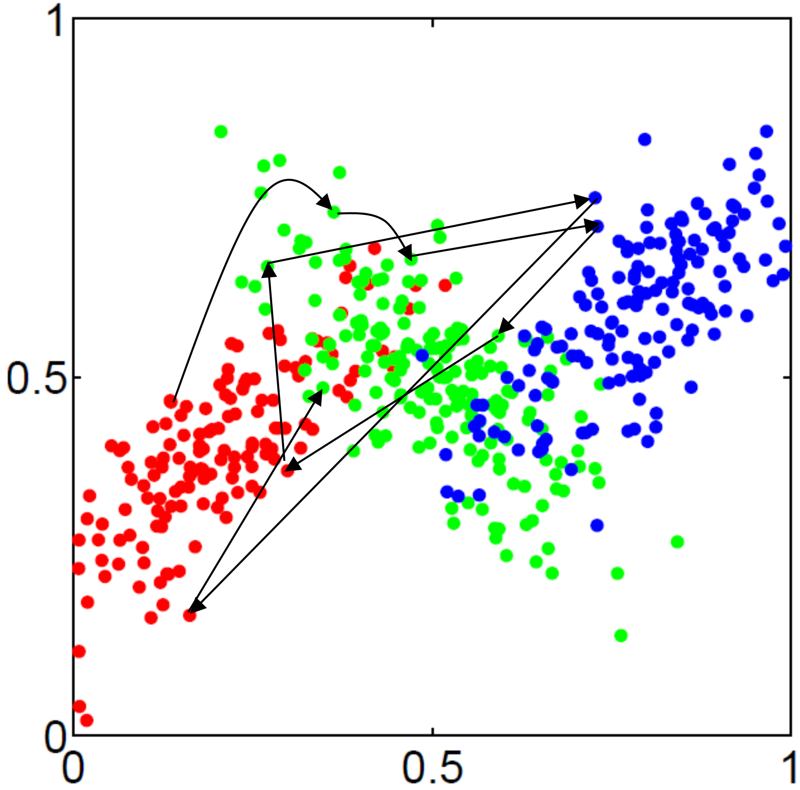


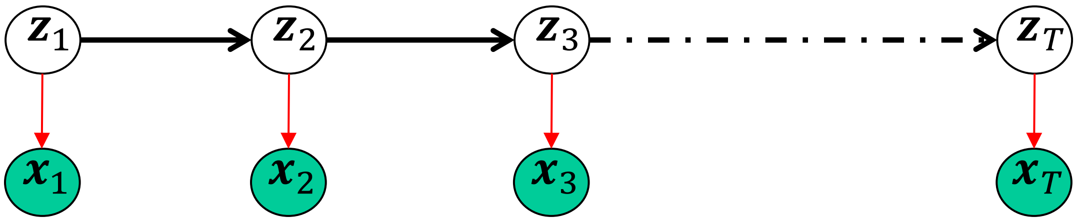


$p(\boldsymbol{x}_t|\boldsymbol{z}_t)$

LxK emission probability matrix **B**

Latent Variables in a Markov Chain





$$p(\boldsymbol{x}_t|\boldsymbol{z}_t) = \prod_{k=1}^{K} N(\boldsymbol{x}_t|\boldsymbol{\mu}_\kappa, \boldsymbol{\Sigma}_\kappa)^{z_k}$$

K Gaussian distributions

Factorization of an HMM

$$p(\boldsymbol{z}_1,..,\boldsymbol{z}_T) = p(\boldsymbol{z}_1)\prod_{t=2}^{\mathrm{T}} p(\boldsymbol{z}_t|\boldsymbol{z}_{t-1})$$

$$\begin{aligned} &p(\boldsymbol{X},\boldsymbol{Z}|\theta) \\ &= p\left(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_T,\boldsymbol{z}_{1,}\boldsymbol{z}_2,\cdots,\boldsymbol{z}_T|\theta\right) \\ &= \prod_{t=1}^{T} p(\boldsymbol{x}_t|\boldsymbol{z}_t)\, p(\boldsymbol{z}_1)\prod_{t=2}^{\mathrm{T}} p(\boldsymbol{z}_t|\boldsymbol{z}_{t-1}) \end{aligned}$$

What can we do with an HMM ?

Given a string of observations and parameters:

(1) We want to find for a timestamp t the probabilities of $\boldsymbol{z}_t$ given the observations **that far**.

This process is called Filtering: $p(\boldsymbol{z}_t | \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_t)$

(2) We want to find for a timestamp t the probabilities of $\boldsymbol{z}_t$ given **the whole string**.

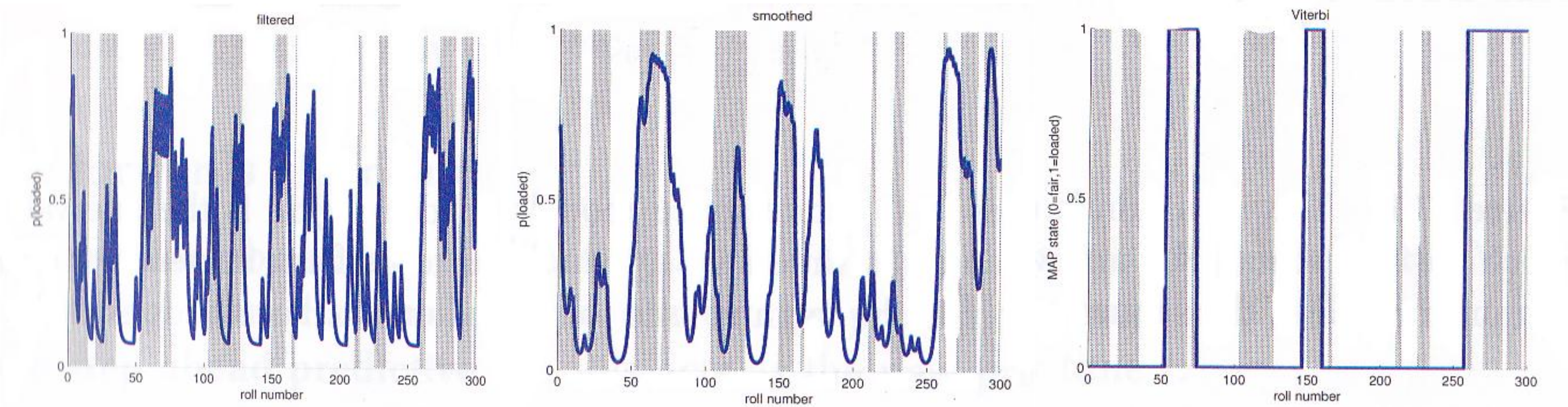
This process is called Smoothing: $p(\boldsymbol{z}_t | \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_T)$

(3) Given the observation string find the string of hidden variables that maximize the posterior.

This process is called Decoding (Viterbi).

$$\arg\max_{\boldsymbol{z}_1 \ldots \boldsymbol{z}_t} p(\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_t | \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_t)$$

Hidden Markov Models



Filtering Smoothing Decoding

Taken from **Machine Learning: A Probabilistic Perspective by K. Murphy**

Hidden Markov Models

(4) Find the probability of the model.

This process is called Evaluation

$$p(\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_T)$$

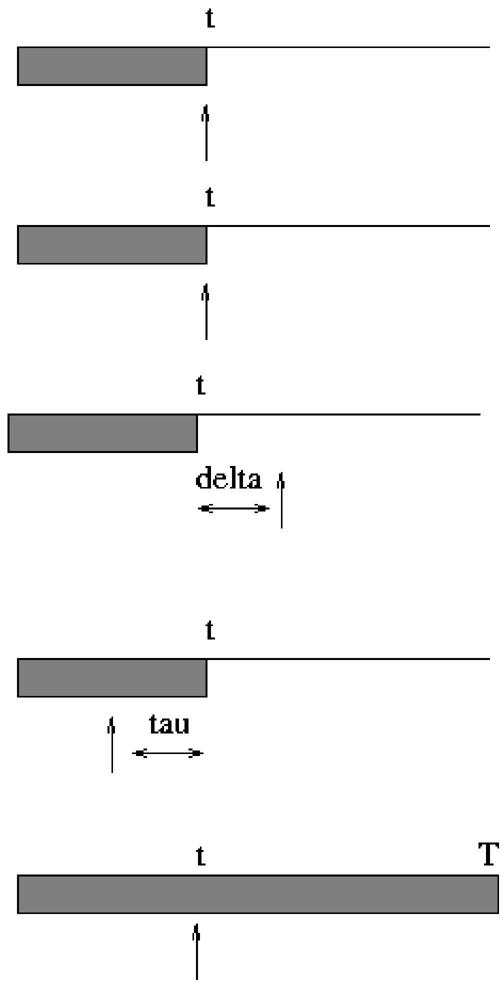
(5) Prediction

$$p(\boldsymbol{z}_{t+\delta} | \boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_{\boldsymbol{t}}) \qquad p(\boldsymbol{x}_{t+\delta} | \boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_{\boldsymbol{t}})$$

(6) EM Parameter estimation (Baum-Welch algorithm)

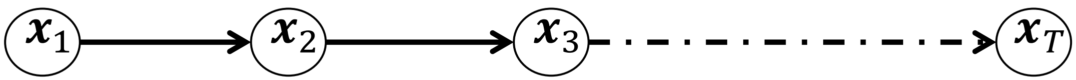
$$\boldsymbol{A}, \boldsymbol{\pi}, \theta$$

Hidden Markov Models



filtering	$p(\boldsymbol{z}_t \mid \boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_t)$
Viterbi	$\arg\max_{z_1 \dots z_t} p(\boldsymbol{z}_1, \dots, \boldsymbol{z}_t \mid \boldsymbol{x}_1, \dots, \boldsymbol{x}_t)$
prediction	$p(\boldsymbol{z}_{t+\delta} \mid \boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_t)$ $p(\boldsymbol{x}_{t+\delta} \mid \boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_t)$
fixed−lag smoothing	$p(\boldsymbol{z}_{t-\tau} \mid \boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_t)$
fixed interval smoothing (offline)	$p(\boldsymbol{z}_t \mid \boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_T)$

Linear Dynamical Systems (LDS)



- Up until now we had Markov Chains with discrete variables
- How can define a transition relationship with continuous valued variables?

$$\boldsymbol{x}_1 = \boldsymbol{\mu}_0 + \boldsymbol{u} \qquad\qquad \boldsymbol{x}_t = \boldsymbol{A}\boldsymbol{x}_{t-1} + \boldsymbol{v}_t$$

$$\boldsymbol{u} \sim N(\boldsymbol{u}|\boldsymbol{0}, \boldsymbol{P}_0) \qquad\qquad \boldsymbol{v} \sim N(\boldsymbol{v}|\boldsymbol{0}, \boldsymbol{\Gamma})$$

or

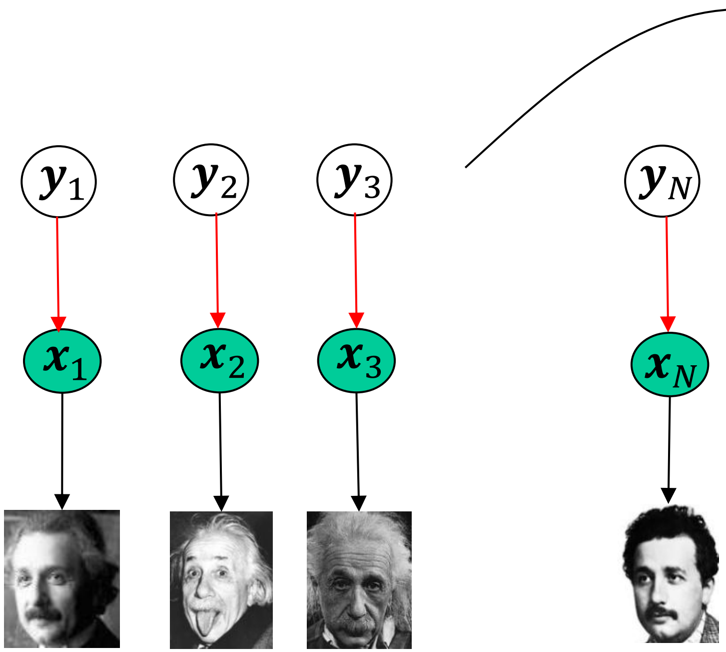
$$\boldsymbol{x}_1 \sim N(\boldsymbol{x}_1|\boldsymbol{\mu}_0, \boldsymbol{P}_0) \qquad\qquad \boldsymbol{x}_t \sim N(\boldsymbol{x}_t|\boldsymbol{A}\boldsymbol{x}_{t-1}, \boldsymbol{\Gamma})$$

or

$$p(\boldsymbol{x}_1) = N(\boldsymbol{x}_1|\boldsymbol{\mu}_0, \boldsymbol{P}_0) \qquad\qquad \boldsymbol{p}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) = N(\boldsymbol{A}\boldsymbol{x}_{t-1}|\boldsymbol{0}, \boldsymbol{\Gamma})$$

Latent Variable Models (Dynamic, Continuous)

Linear Dynamical Systems (LDS)



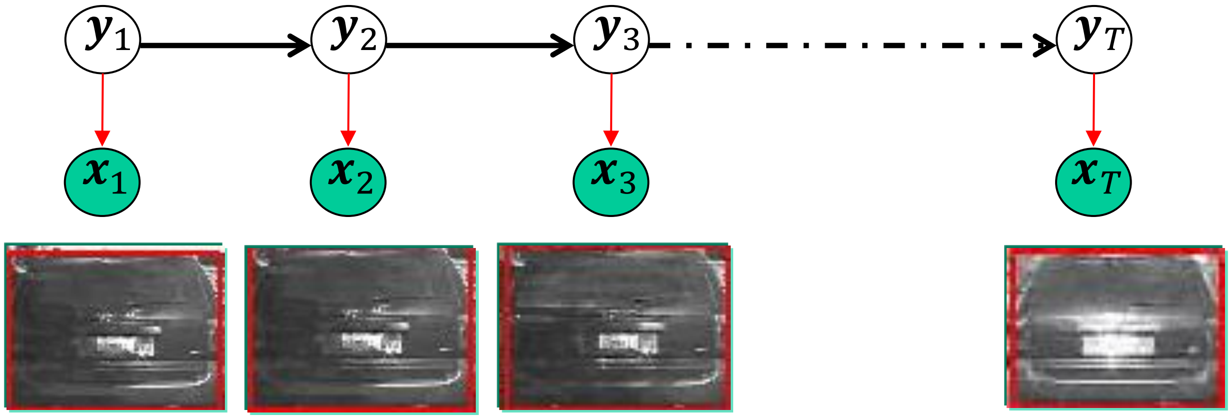
Share a common linear structure

$$\boldsymbol{x} = \boldsymbol{W}\boldsymbol{y} + \boldsymbol{\mu} + \boldsymbol{e}$$
$$\boldsymbol{e} \sim N(\boldsymbol{e}|\boldsymbol{0}, \sigma^2\boldsymbol{I})$$
$$\boldsymbol{y} \sim N(\boldsymbol{y}|\boldsymbol{0}, \boldsymbol{I})$$

We want to find the parameters:

$$\theta = \{\boldsymbol{W}, \boldsymbol{\mu}, \sigma^2\}$$

Linear Dynamical Systems (LDS)



$$\boldsymbol{x}_t = \mathbf{W}\boldsymbol{y}_t + \boldsymbol{e}_t \qquad \mathbf{e} \sim N(\boldsymbol{e}|\mathbf{0}, \mathbf{\Sigma})$$

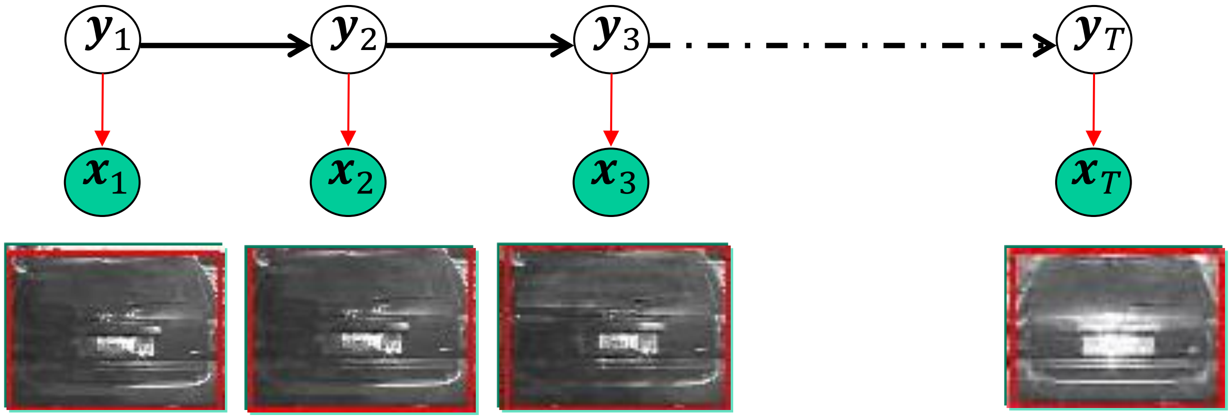
Transition model

$$\boldsymbol{y}_1 = \boldsymbol{\mu}_0 + \boldsymbol{u} \qquad \boldsymbol{u} \sim N(\boldsymbol{u}|\mathbf{0}, \boldsymbol{P}_0)$$

$$\boldsymbol{y}_t = \boldsymbol{A}\boldsymbol{y}_{t-1} + \boldsymbol{v}_t \qquad \boldsymbol{v} \sim N(\boldsymbol{v}|\mathbf{0}, \mathbf{\Gamma})$$

Parameters: $\theta = \{\boldsymbol{W}, \boldsymbol{A}, \boldsymbol{\mu}_0, \mathbf{\Sigma}, \mathbf{\Gamma}, \boldsymbol{P}_0\}$

Linear Dynamical Systems (LDS)



First timestamp: $p(\boldsymbol{y}_1) = N(\boldsymbol{y}_1 | \boldsymbol{\mu}_0, \boldsymbol{P}_0)$

Transition Probability : $p(\boldsymbol{y}_t | \boldsymbol{y}_{t-1}) = N(\boldsymbol{y}_t | \boldsymbol{A}\boldsymbol{y}_{t-1}, \boldsymbol{\Gamma})$

Emission: $p(\boldsymbol{x}_t | \boldsymbol{y}_t) = N(\boldsymbol{x}_t | \boldsymbol{W}\boldsymbol{y}_t, \boldsymbol{\Sigma})$

HMM vs LDS

HMM

Markov Chain

with discrete latent variables

$p(\boldsymbol{y_1})$ $\boldsymbol{\pi}$ $Kx1$

$p(\boldsymbol{y_t}|\boldsymbol{y_{t-1}})$ $\boldsymbol{A}$ KxK

$p(\boldsymbol{x_t}|\boldsymbol{y_t})$ $\boldsymbol{B}$ LxK

or

$p(\boldsymbol{x_t}|\boldsymbol{y_t})$ K distributions

LDS

Markov Chain

with continuous latent variables

$$p(\boldsymbol{y}_1) = N(\boldsymbol{y}_1|\boldsymbol{\mu}_0, \boldsymbol{P}_0)$$

$$p(\boldsymbol{y}_t|\boldsymbol{y}_{t-1}) = N(\boldsymbol{A}\boldsymbol{y}_{t-1}|\boldsymbol{0}, \boldsymbol{\Gamma})$$

$$p(\boldsymbol{x}_t|\boldsymbol{y}_t) = N(\boldsymbol{x}_t|\boldsymbol{W}\boldsymbol{y}_t, \boldsymbol{\Sigma})$$

What can we do with LDS?

Global Positioning System (GPS)

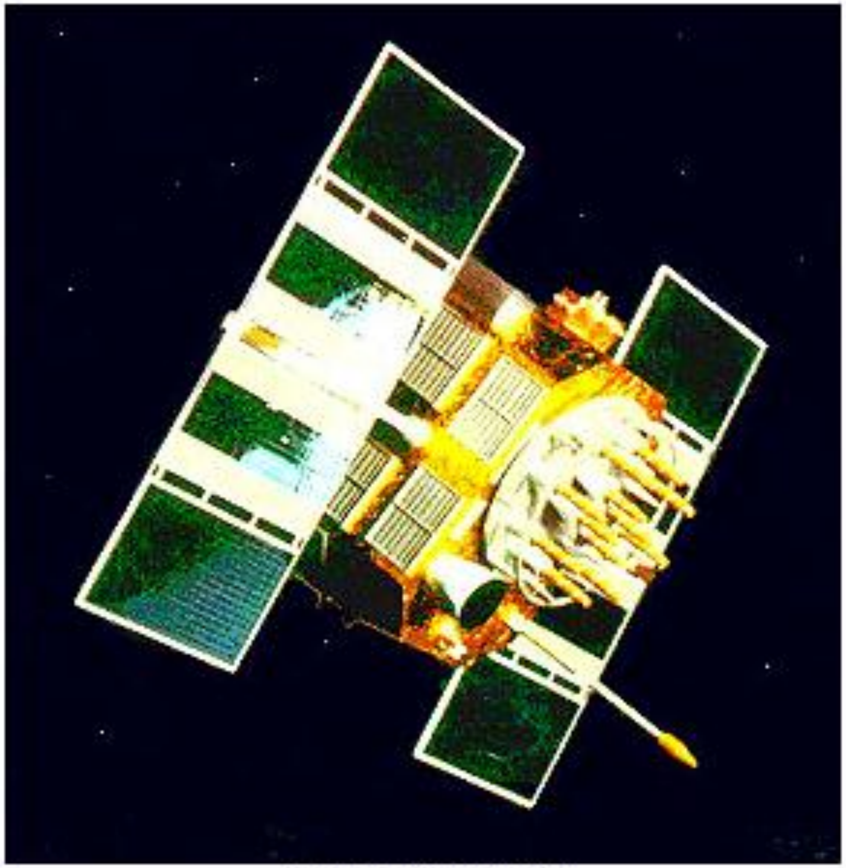


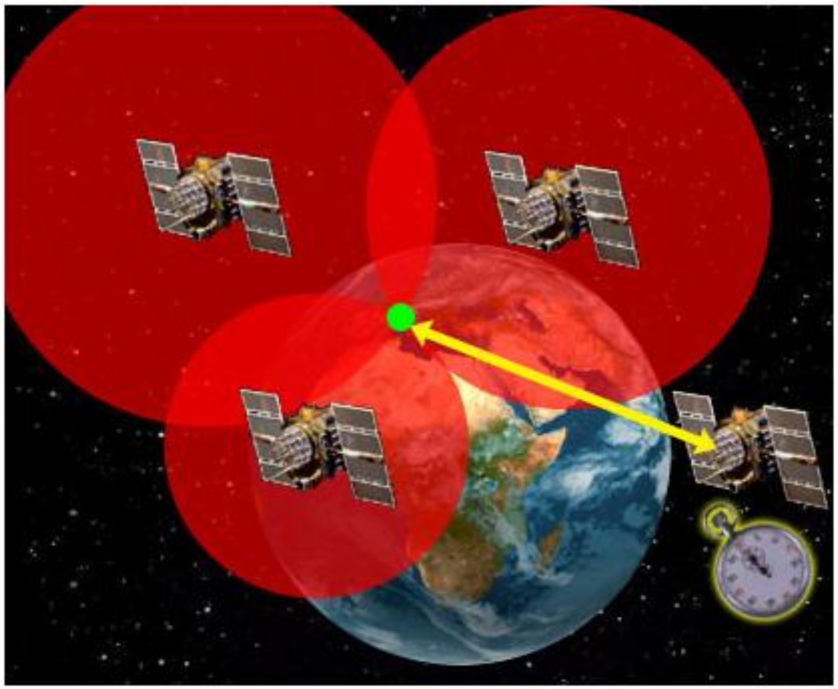
Photo courtesy NASA

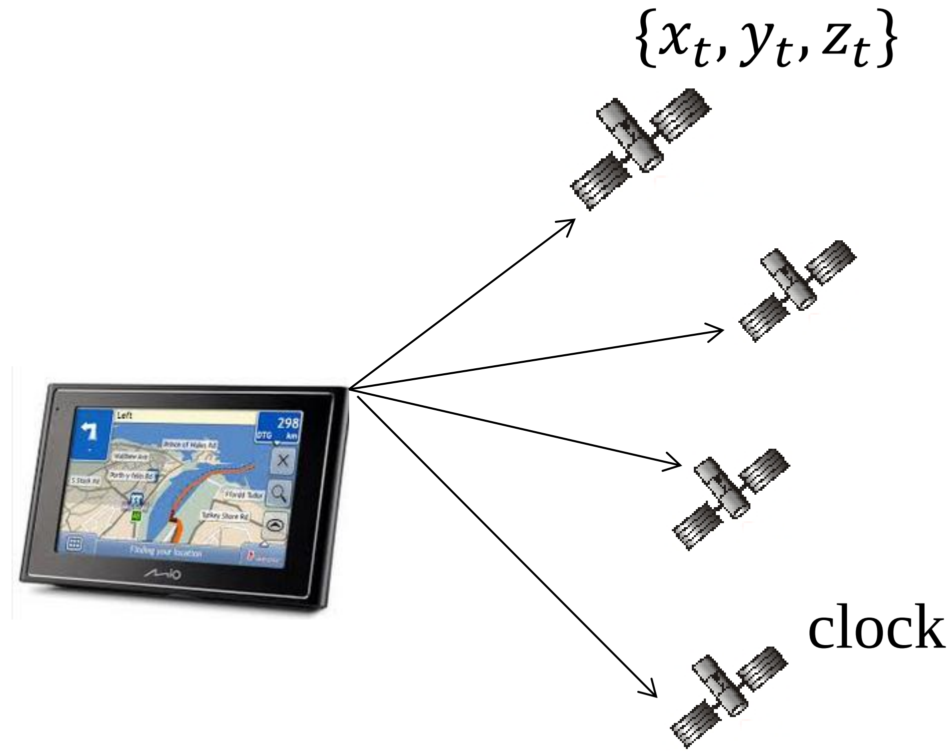
NAVSTAR GPS satellite

Photo courtesy U.S. Department of Defense

Artist's concept of the GPS satellite constellation

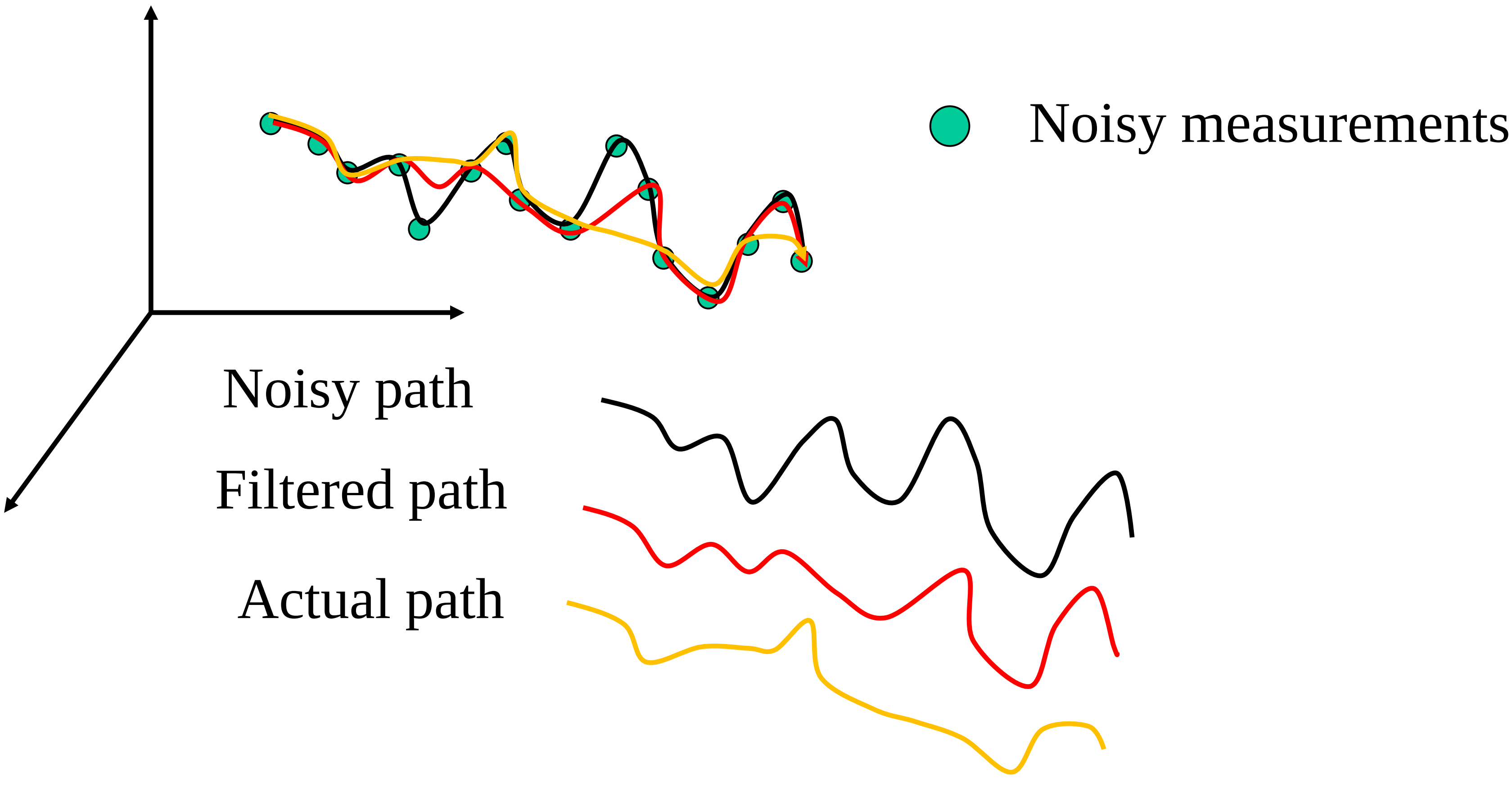
What can we do with LDS?





What can we do with LDS?

We still have noisy measurements



What can we do with LDS?

Given a string of observations and parameters:

(1) We want to find for a timestamp t the probabilities of $\boldsymbol{z}_t$ given the observations **that far**.

This process is called Filtering: $p(\boldsymbol{y}_t|\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_t)$

(2) We want to find for a timestamp t the probabilities of $\boldsymbol{z}_t$ given **the whole time series**.

This process is called Smoothing: $p(\boldsymbol{y}_t|\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_T)$

All above probabilities are Gaussians. Means and covariance matrices are computed recursively!!!!

What can we do with LDS?

(3) Find the probability of the model.

This process is called Evaluation: $p(\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_T)$

(5) Prediction

$$p(\boldsymbol{y}_{t+\delta} | \boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_{\boldsymbol{t}}) \qquad p(\boldsymbol{x}_{t+\delta} | \boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_{\boldsymbol{t}})$$

(6) EM Parameter estimation (Baum-Welch algorithm)

$$\theta = \{\boldsymbol{W}, \boldsymbol{A}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}, \boldsymbol{\Gamma}, \boldsymbol{P}_0\}$$