Visual Tracking
Part I: Foundation

CSE P576 Autumn 2021

Vitaly Ablavsky

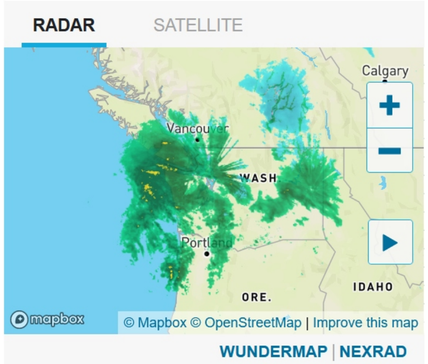
What Is Tracking?

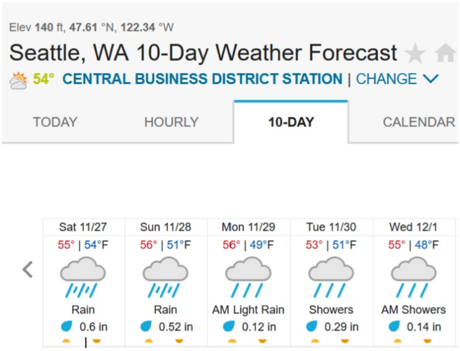

$x_{t+1} = x_t + v_t + \eta_t$
x_t
GO! GO!
GOOOOOAL!
FELINEBETS.NET
ARG vs BRA

Tracking Implies Prediction

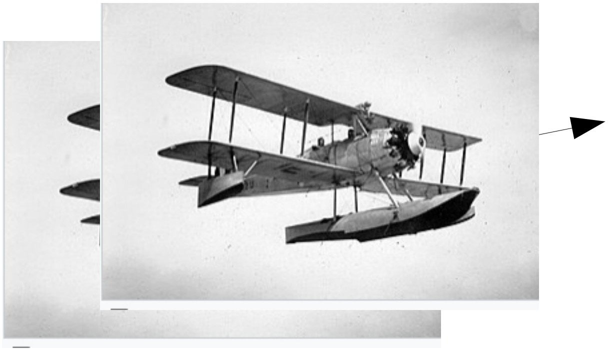
- Tracking a weather system: not just telling you it’s raining outside your house, but whether it will rain tomorrow (easy to guess if you’re in Seattle)

- Tracking an airplane: an estimate where it is currently and where it will be delta-T seconds from now





https://www.wunderground.com/forecast/us/wa/seattle/KWASEATT2713



https://en.wikipedia.org/wiki/Floatplane

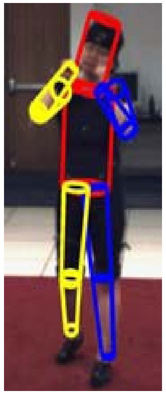
Tracking Applications: Parameters

- Number of objects
- Types of objects
- Number of sensors
- Types of sensors
- Distance to sensor

Tracking Applications: Examples

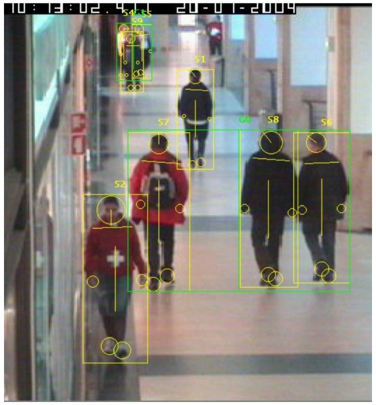
1 object, articulated, 300 pixels tall

body
motion
analysis



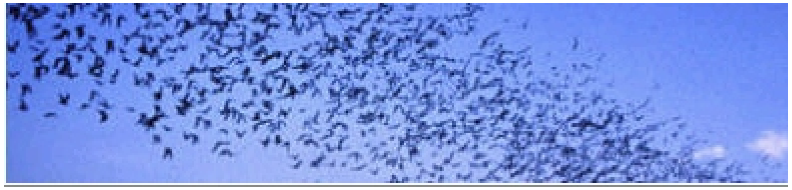
10 objects, articulated, 150 pixels tall

foot
traffic
analysis



100 objects, 15 – 30 pixels tall

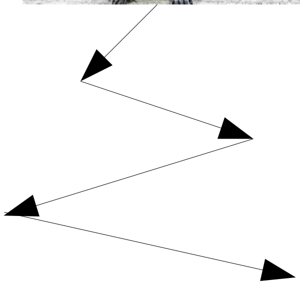
bat census



Factors that Make Tracking Hard

- (unknown) target dynamics: how fast does the state change?

vs.



- target observations/ measurements: how noisy and (in)frequent?

Tracking Research circa 1975

Automatica, Vol. 11, pp. 451–460. Pergamon Press, 1975. Printed in Great Britain

Tracking in a Cluttered Environment With Probabilistic Data Association*

Dépistage dans une Ambiance Encombrée, avec Association Probabilistique des Données

Verfolgung in einer örtlich gestörten Umgebung unter Verwendung von Wahrscheinlichkeitsdaten

YAAKOV BAR-SHALOM† and EDISON TSE†

Tracking a target with uncertainty in the origin of the measurements is accomplished with an algorithm, suitable for real-time implementation, which utilizes the a posteriori *probabilities of the measurements having originated from the target.*

Probabilistic Formulation

Probability Space

triplet $(\Omega, \mathcal{F}, P)$

sample space,
all possible
outcomes
e.g., {1, 2, 3, 4, 5, 6}

event space, e.g.,
set of all subsets of Ω
including
“die lands even” {2, 4, 6}

probability function
$P(A) \geq 0, A \in \mathcal{F}$
$P(\Omega) = 1, P(\emptyset) = 0$

credit: https://en.wikipedia.org/wiki/Probability_space

Random Variables

Random variable X: a function

X: $\Omega \rightarrow \mathbb{R}$

For every Borel subset B of the real line $X^{-1}(B)$ in $\mathcal{F}$

Discrete

Continuous

Stochastic Processes

Stochastic process is an indexed collection of random variables

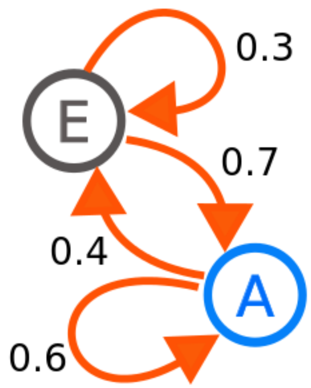
discrete time

$X = \{X_n,\ n = 0, 1, 2, \ldots\}$

continuous time

$X = \{X_t,\ 0 \le t < \infty\}, X : T \times \Omega \to \mathbb{R}$

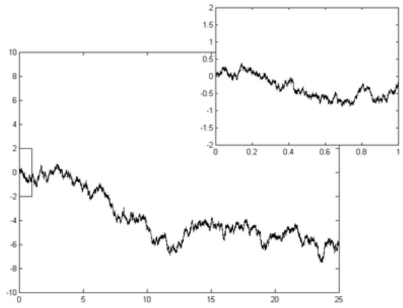
discrete time Markov process



$t \to X(t, \omega)$ called *sample path*

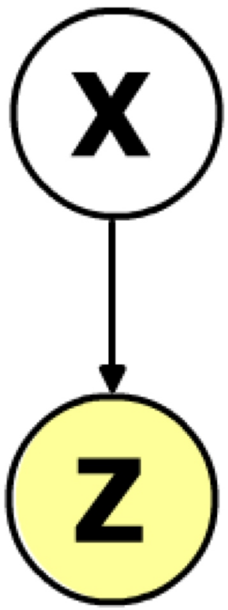
ex. 1: $X_t = Z * t^2$, $Z \sim N(0,1)$

ex. 2: Wiener process
a Gaussian process;
limit of *random walk*



https://en.wikipedia.org/wiki/Markov_chain https://en.wikipedia.org/wiki/Wiener_process

Probabilistic Formulation

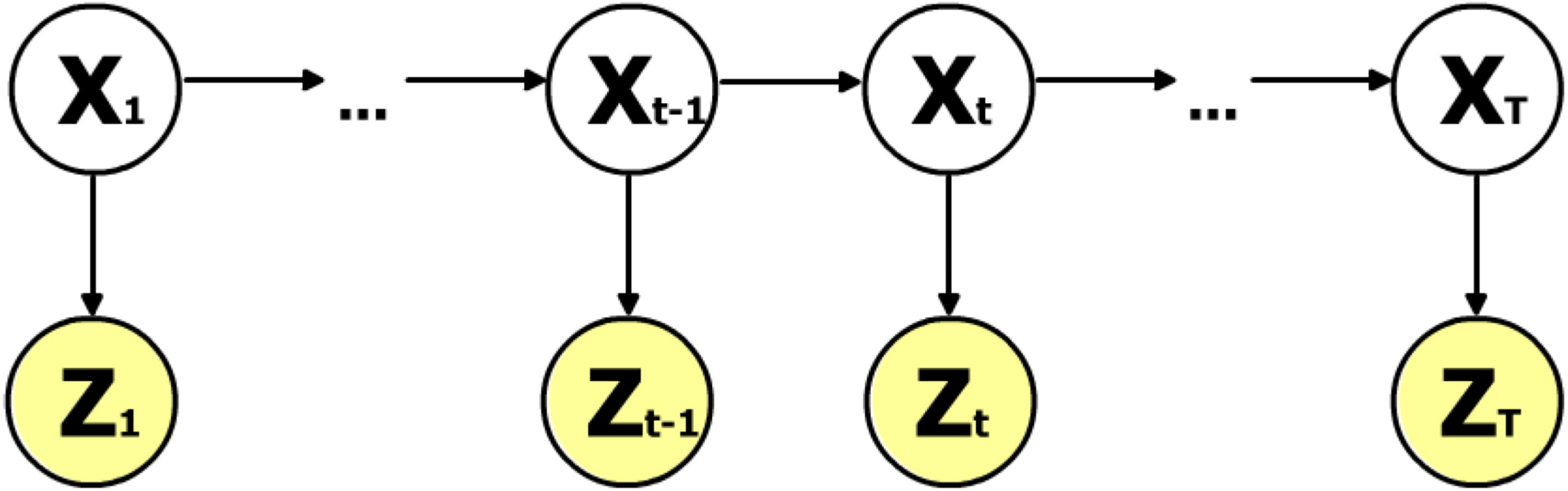


Multi-Object Configuration - the unknown (or hidden) state of the objects, i.e. position, size, etc.

Observations - information taken from the image such as color, motion, texture, etc.

[K. Smith, “Bayesian methods for visual multi-object tracking with applications to human activity recognition,” 2007.]

Probabilistic Formulation



Graphical model for the multi-object tracking problem with *T* time steps

[K. Smith, “Bayesian methods for visual multi-object tracking with applications to human activity recognition,” 2007.]

Probabilistic Formulation

$$p(\mathbf{X}_t|\mathbf{Z}_{1:t}) = \frac{p(\mathbf{Z}_t|\mathbf{X}_t)p(\mathbf{X}_t|\mathbf{Z}_{1:t-1})}{p(\mathbf{Z}_t|\mathbf{Z}_{1:t-1})}$$

$$p(\mathbf{X}_t|\mathbf{Z}_{1:t-1}) = \int_{\mathbf{X}_{t-1}} p(\mathbf{X}_t|\mathbf{X}_{t-1})p(\mathbf{X}_{t-1}|\mathbf{Z}_{1:t-1})d\mathbf{X}_{t-1}$$

$$p(\mathbf{X}_t|\mathbf{Z}_{1:t}) = C^{-1}p(\mathbf{Z}_t|\mathbf{X}_t) \times \int_{\mathbf{X}_{t-1}} p(\mathbf{X}_t|\mathbf{X}_{t-1})p(\mathbf{X}_{t-1}|\mathbf{Z}_{1:t-1})d\mathbf{X}_{t-1}$$

[K. Smith, “Bayesian methods for visual multi-object tracking with applications to human activity recognition,” 2007.]

Target Dynamics

Discrete Time LDS

- Continuous model are difficult to realize
  - Algorithms work at discrete time steps
  - Measurements are acquired with certain rates

- In practice, **discrete models** are employed

- Discrete-time LDS are governed by

$$x(k+1) = F(k)x(k) + G(k)u(k) + \xi(k)$$

  - $F \in \mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$ is the state transition matrix
  - $G \in \mathbb{R}^{n_x} \times \mathbb{R}^{n_u}$ is the discrete-time input gain

- Same observation function of continuous models

[G. Grisetti, C. Stachniss, K. Arras, and W. Burgard, Univ. of Freiburg Course on Robotics & Target Tracking]

Discrete Time LDS

- Continuous model are difficult to realize
  - Algorithms work at discrete time steps
  - Measurements are acquired with certain rates

- In practice, **discrete models** are employed

- Discrete-time LDS are governed by

$$x(k+1) = F(k)x(k) + G(k)u(k) + \xi(k)$$

In target tracking, the input is unknown!

  - $F \in \mathbb{R}^{n_x} \times \mathbb{R}^{n_x}$ is the state transition matrix
  - $G \in \mathbb{R}^{n_x} \times \mathbb{R}^{n_u}$ is the discrete-time input gain

- Same observation function of continuous models

LDS Example – Throwing ball

- We want to throw a ball and **compute its trajectory**
- This can be **easily done with an LDS**
- The ball's **state** shall be represented as

$$\mathbf{x} = \begin{bmatrix} x & y & \dot{x} & \dot{y} \end{bmatrix}^T$$

- We ignore winds but consider the **gravity force** g

$$\mathbf{u} = -g$$

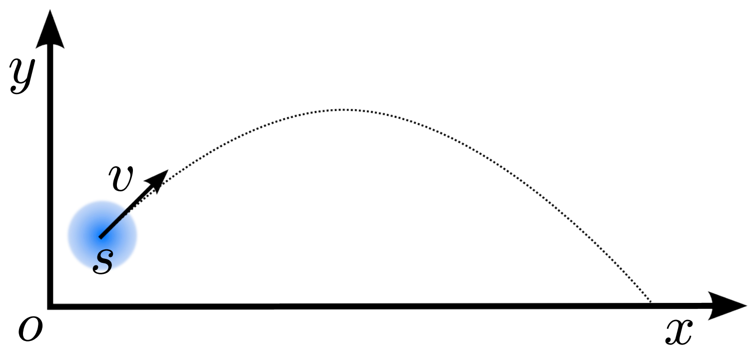
- No floor constraints
- We **observe** the ball with a noise-free position sensor

$$\mathbf{z} = \begin{bmatrix} x & y \end{bmatrix}^T$$

[G. Grisetti, C. Stachniss, K. Arras, and W. Burgard, Univ. of Freiburg Course on Robotics & Target Tracking]

LDS Example – Throwing ball

- Throwing a ball from s with initial velocity v
- Consider only the gravity force, g, of the ball

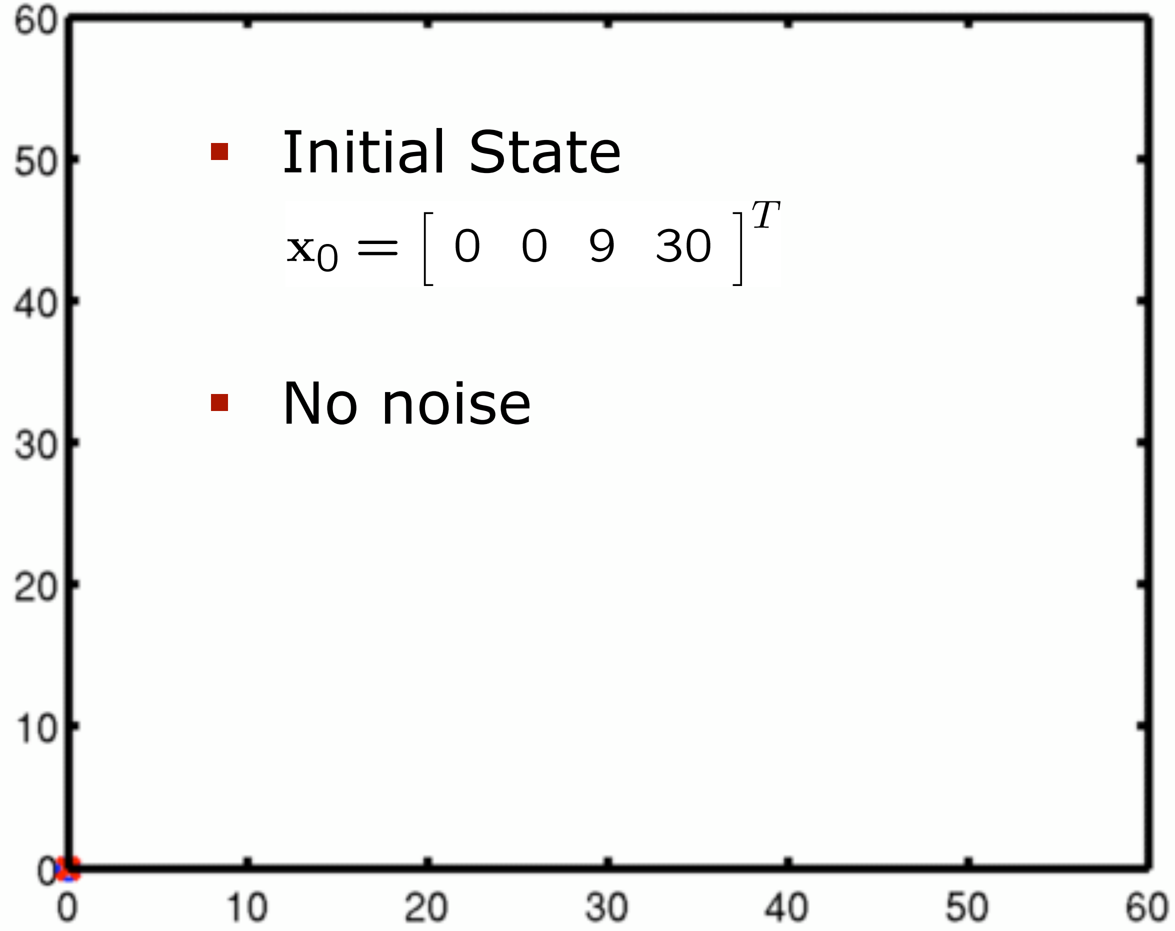


- State vector
$$\mathbf{x} = \begin{bmatrix} x & y & \dot{x} & \dot{y} \end{bmatrix}^T$$
- Initial state
$$\mathbf{x}_0 = \begin{bmatrix} s_x & s_y & v_x & v_y \end{bmatrix}^T$$
- Input vector (scalar)
$$\mathbf{u} = -g$$
- Measurement vector
$$\mathbf{z} = \begin{bmatrix} x & y \end{bmatrix}^T$$

- Process matrices
$$F = \begin{bmatrix} 1 & 0 & T & 0 \\ 0 & 1 & 0 & T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$
$$G = \begin{bmatrix} 0 & \frac{T^2}{2} & 0 & T \end{bmatrix}^T$$
- Measurement matrix
$$H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

[G. Grisetti, C. Stachniss, K. Arras, and W. Burgard, Univ. of Freiburg Course on Robotics & Target Tracking]

LDS Example – Throwing ball



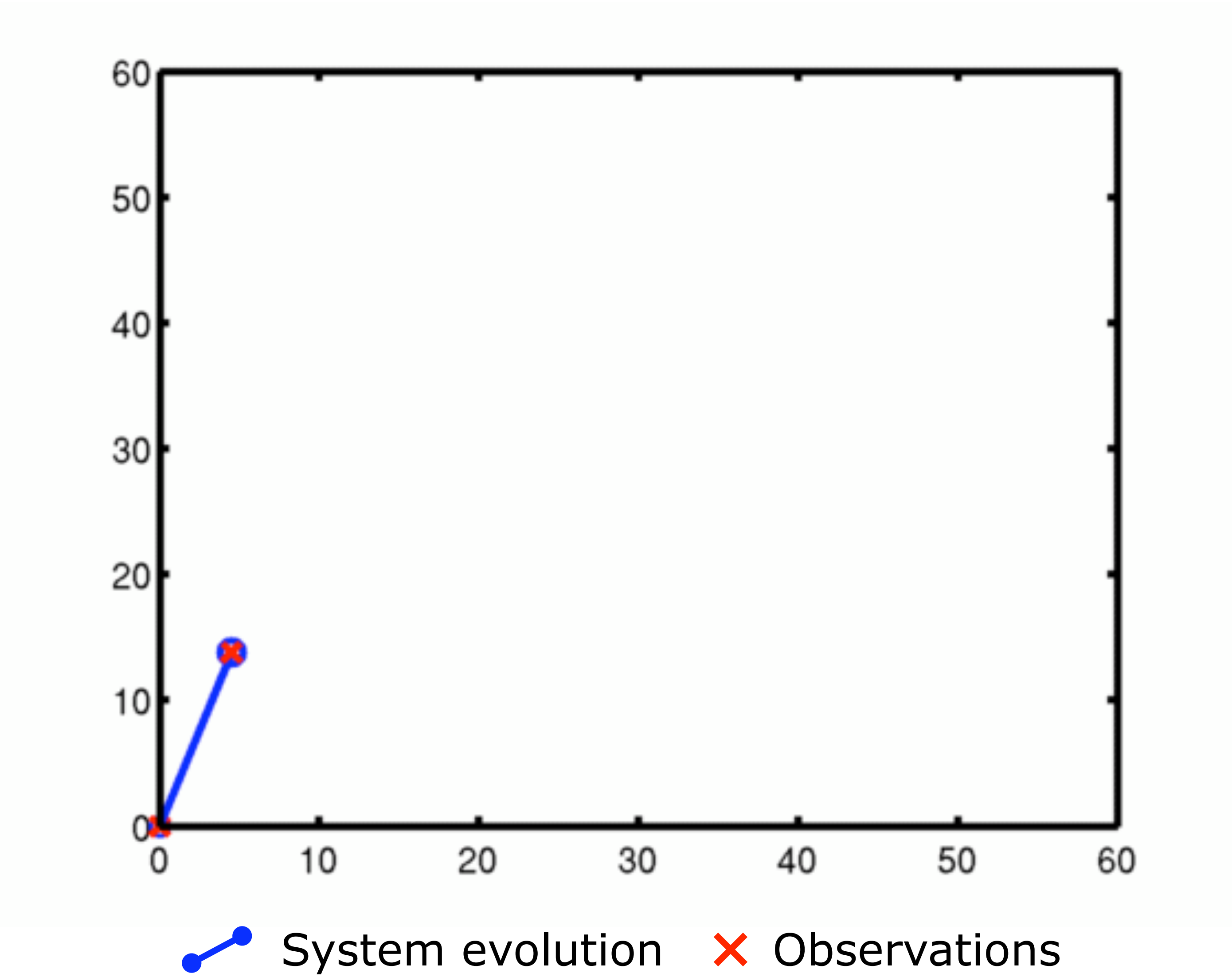
- Initial State

$$\mathbf{x}_0 = \begin{bmatrix} 0 & 0 & 9 & 30 \end{bmatrix}^T$$

- No noise

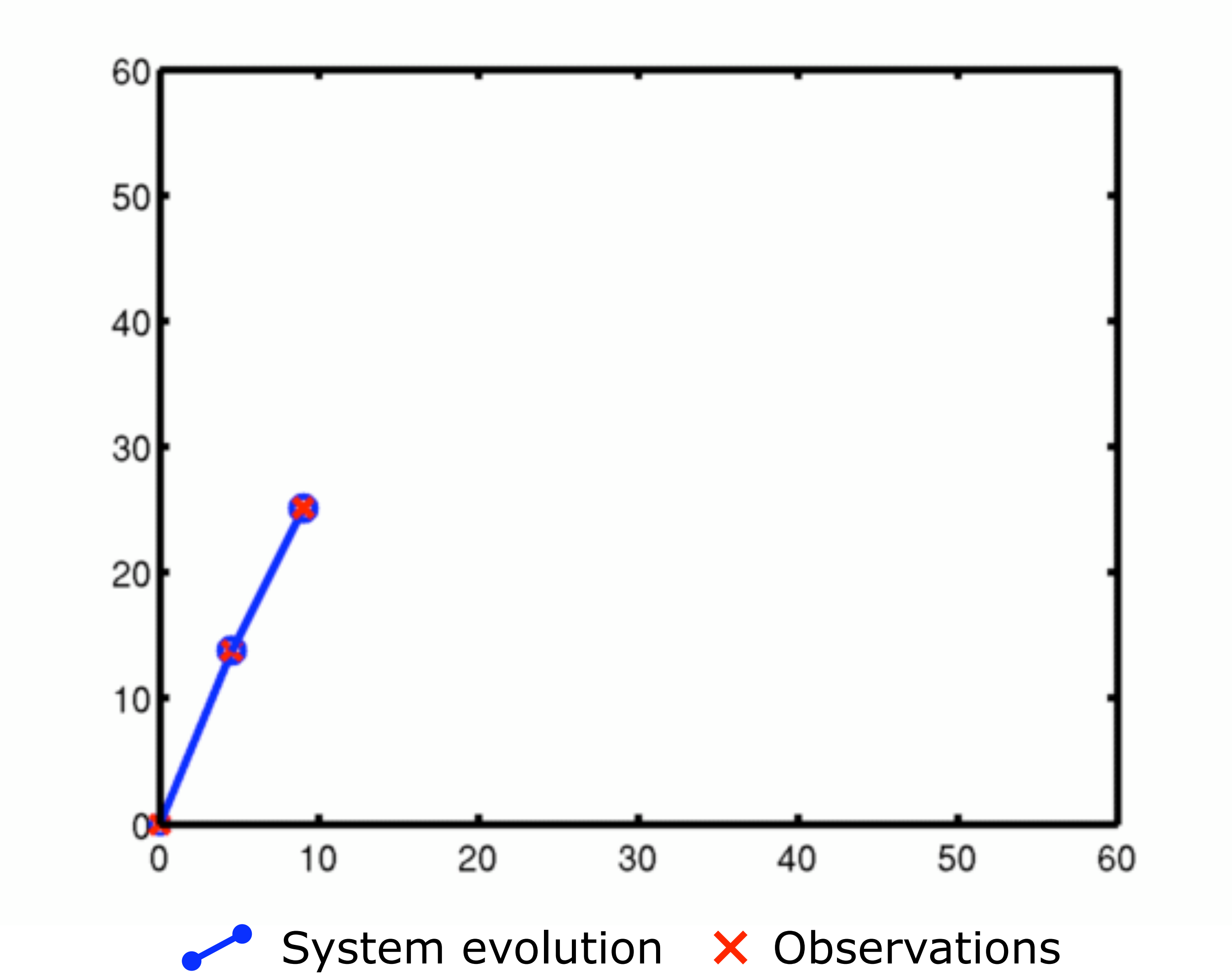
[G. Grisetti, C. Stachniss, K. Arras, and W. Burgard, Univ. of Freiburg Course on Robotics & Target Tracking]

LDS Example – Throwing ball



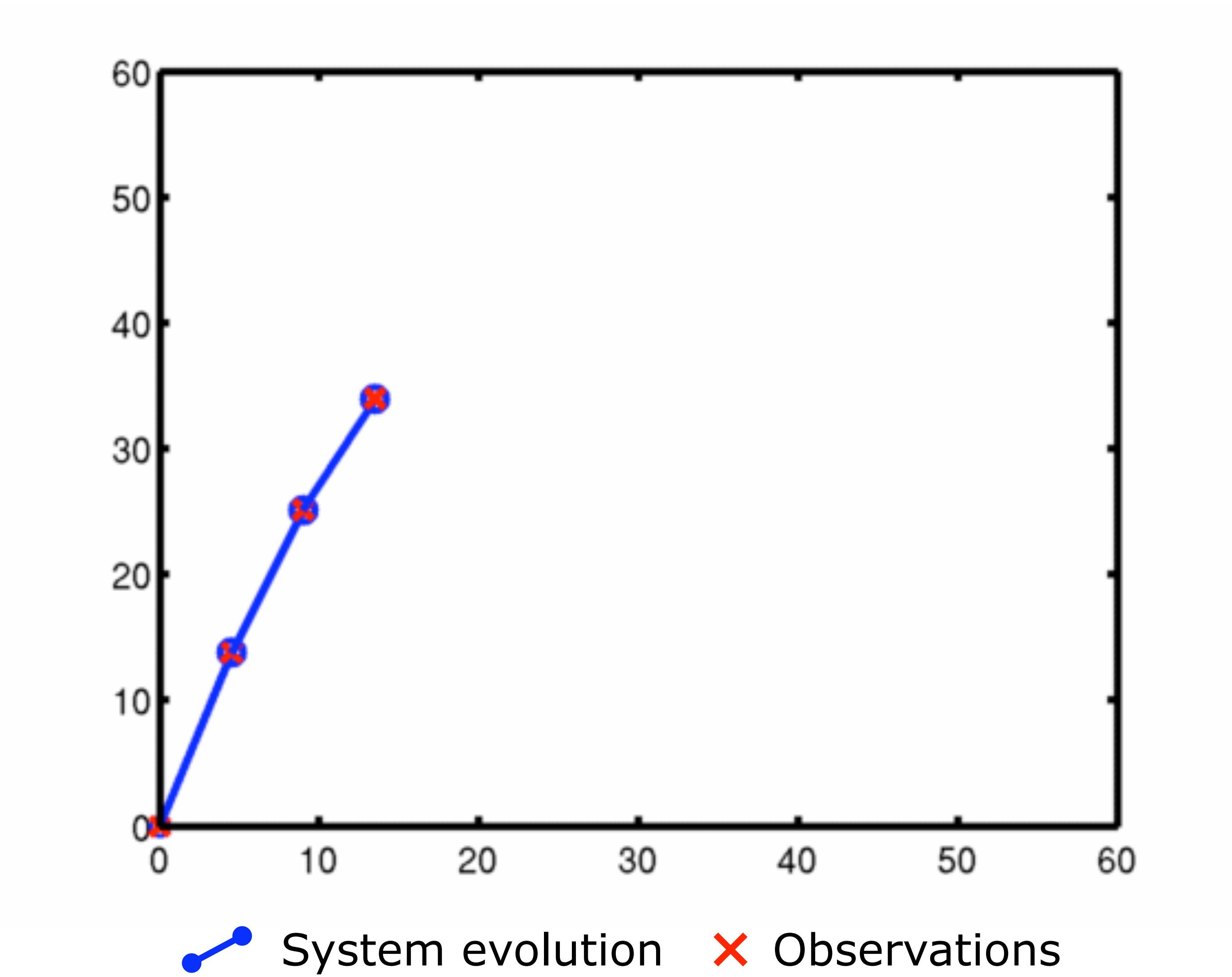
[G. Grisetti, C. Stachniss, K. Arras, and W. Burgard, Univ. of Freiburg Course on Robotics & Target Tracking]

LDS Example – Throwing ball



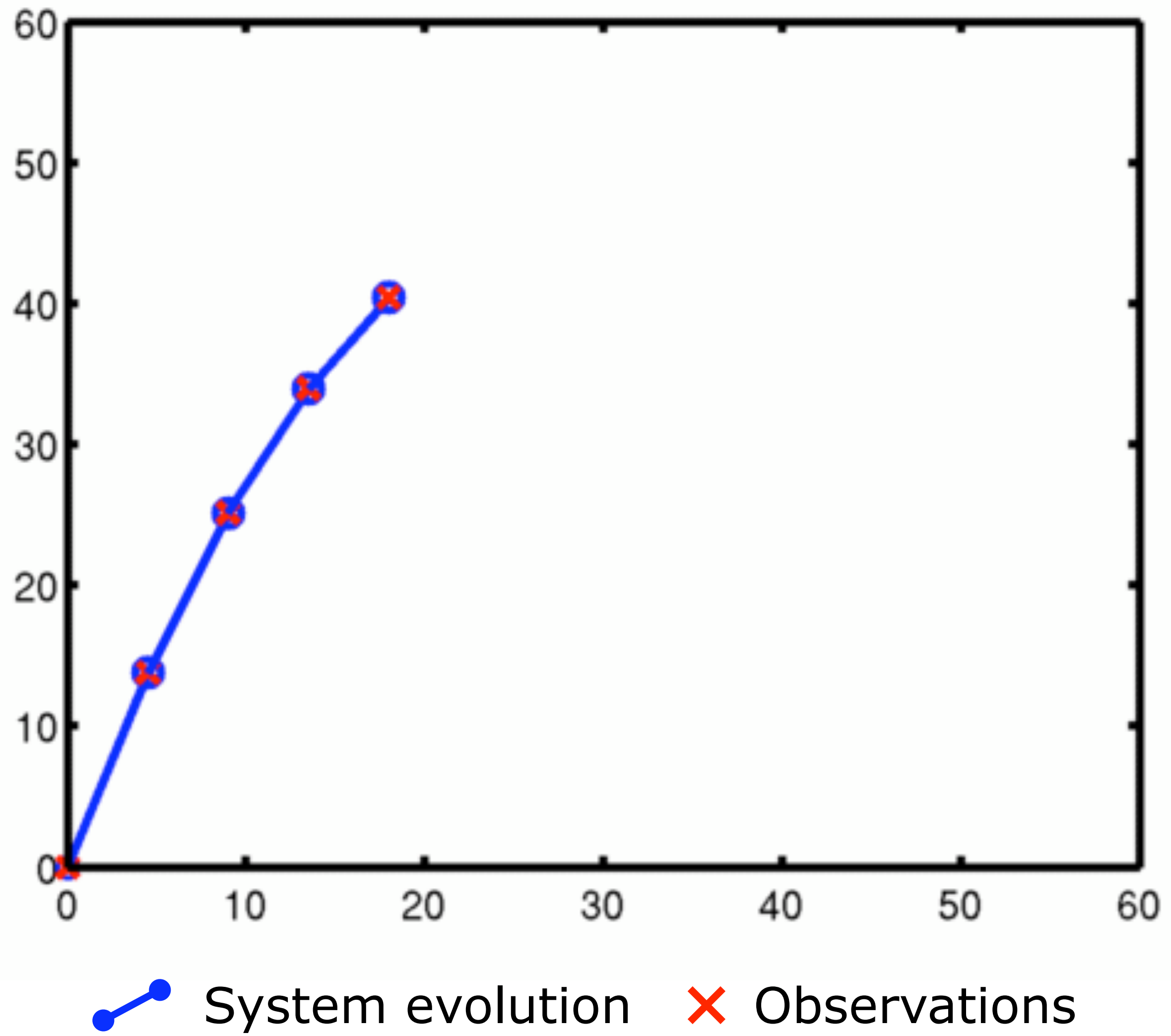
[G. Grisetti, C. Stachniss, K. Arras, and W. Burgard, Univ. of Freiburg Course on Robotics & Target Tracking]

LDS Example – Throwing ball



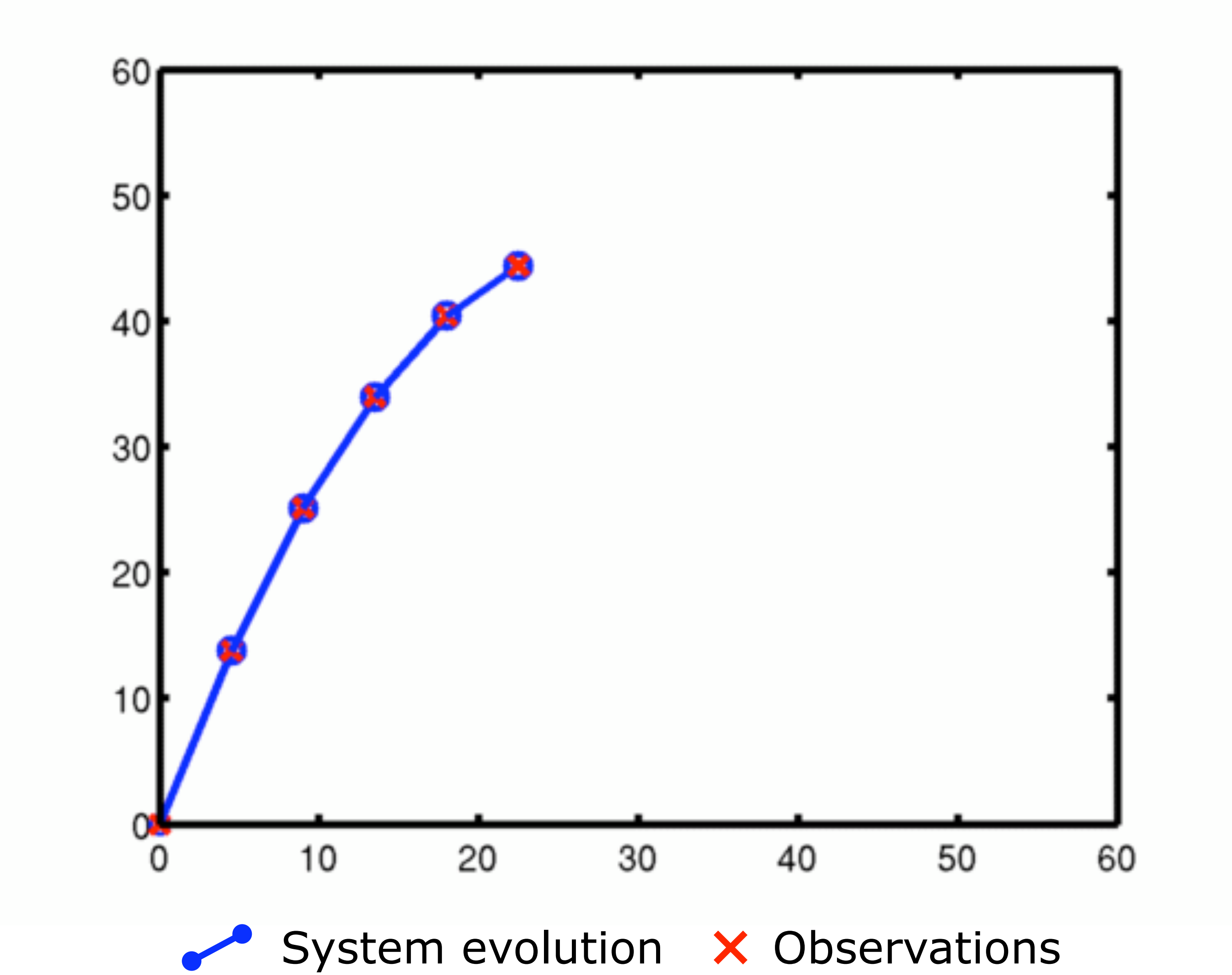
[G. Grisetti, C. Stachniss, K. Arras, and W. Burgard, Univ. of Freiburg Course on Robotics & Target Tracking]

LDS Example – Throwing ball



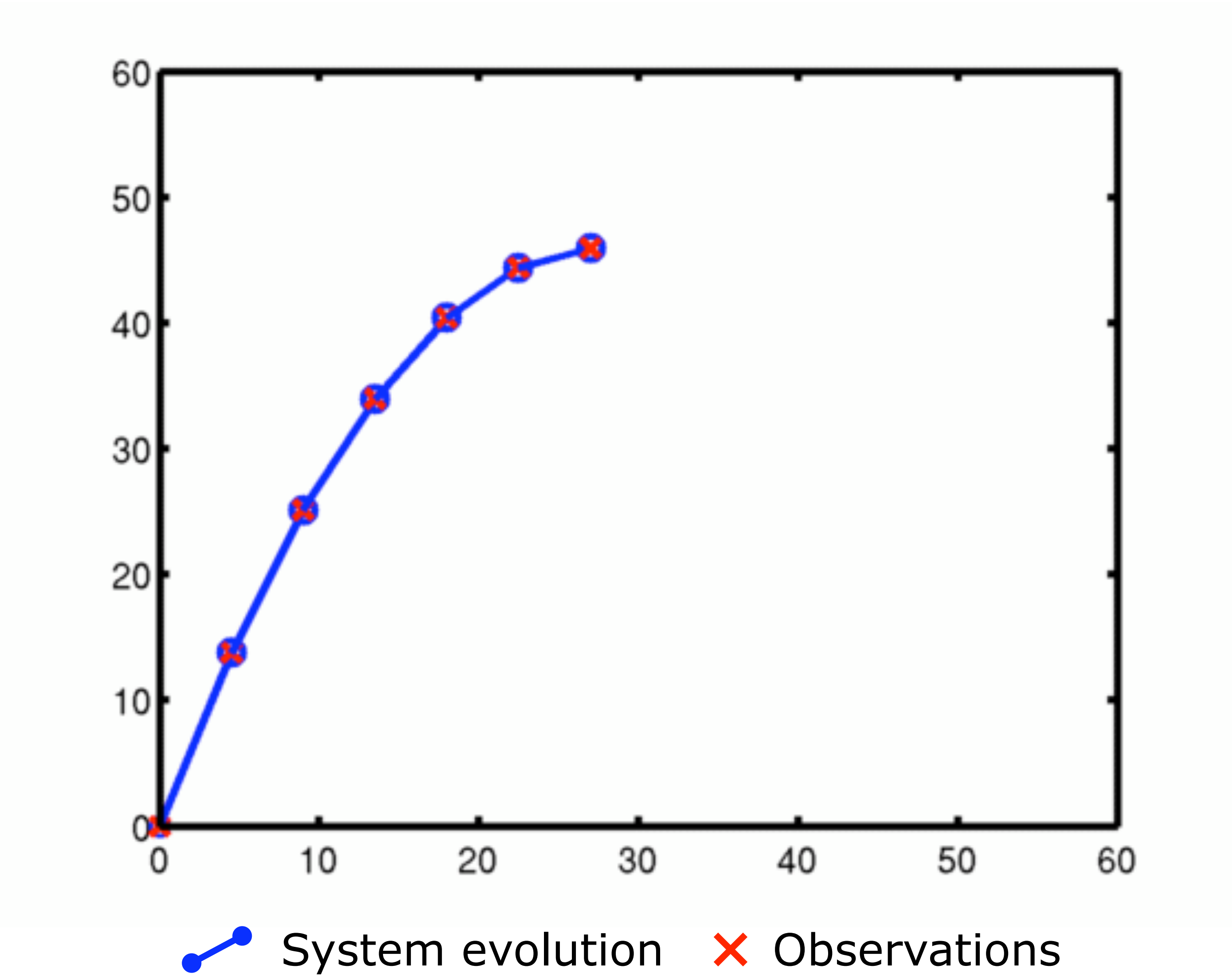
[G. Grisetti, C. Stachniss, K. Arras, and W. Burgard, Univ. of Freiburg Course on Robotics & Target Tracking]

LDS Example – Throwing ball



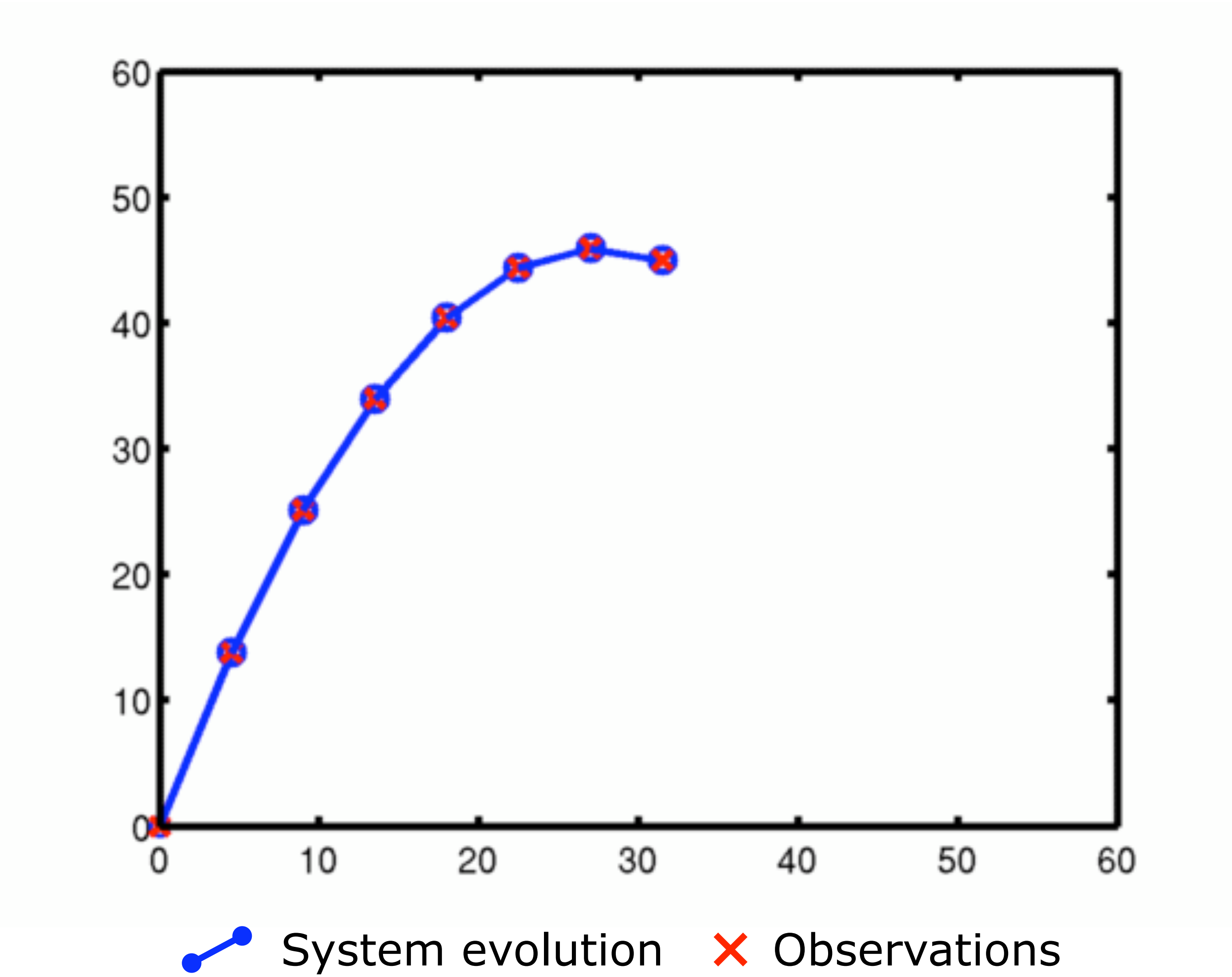
[G. Grisetti, C. Stachniss, K. Arras, and W. Burgard, Univ. of Freiburg Course on Robotics & Target Tracking]

LDS Example – Throwing ball



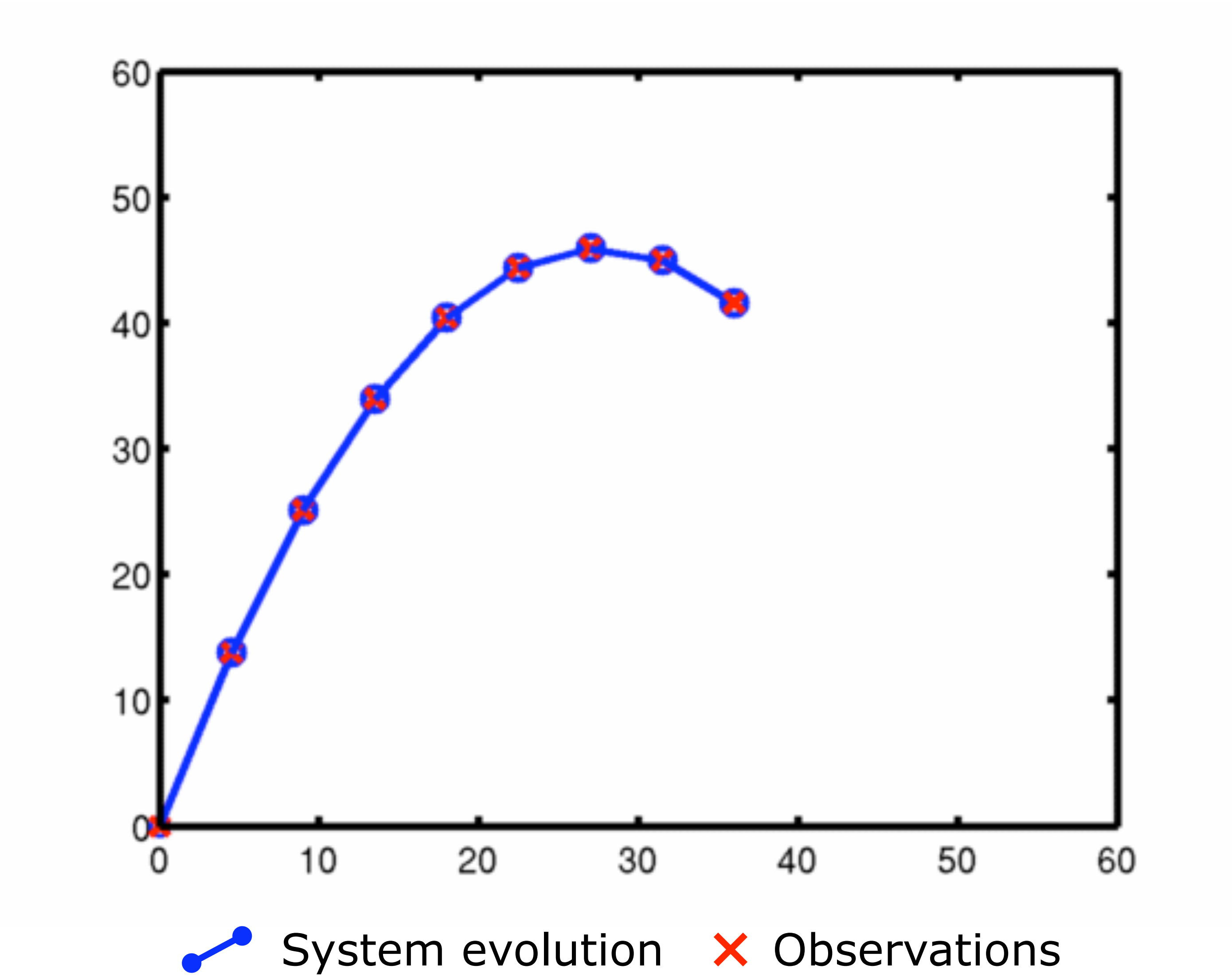
[G. Grisetti, C. Stachniss, K. Arras, and W. Burgard, Univ. of Freiburg Course on Robotics & Target Tracking]

LDS Example – Throwing ball



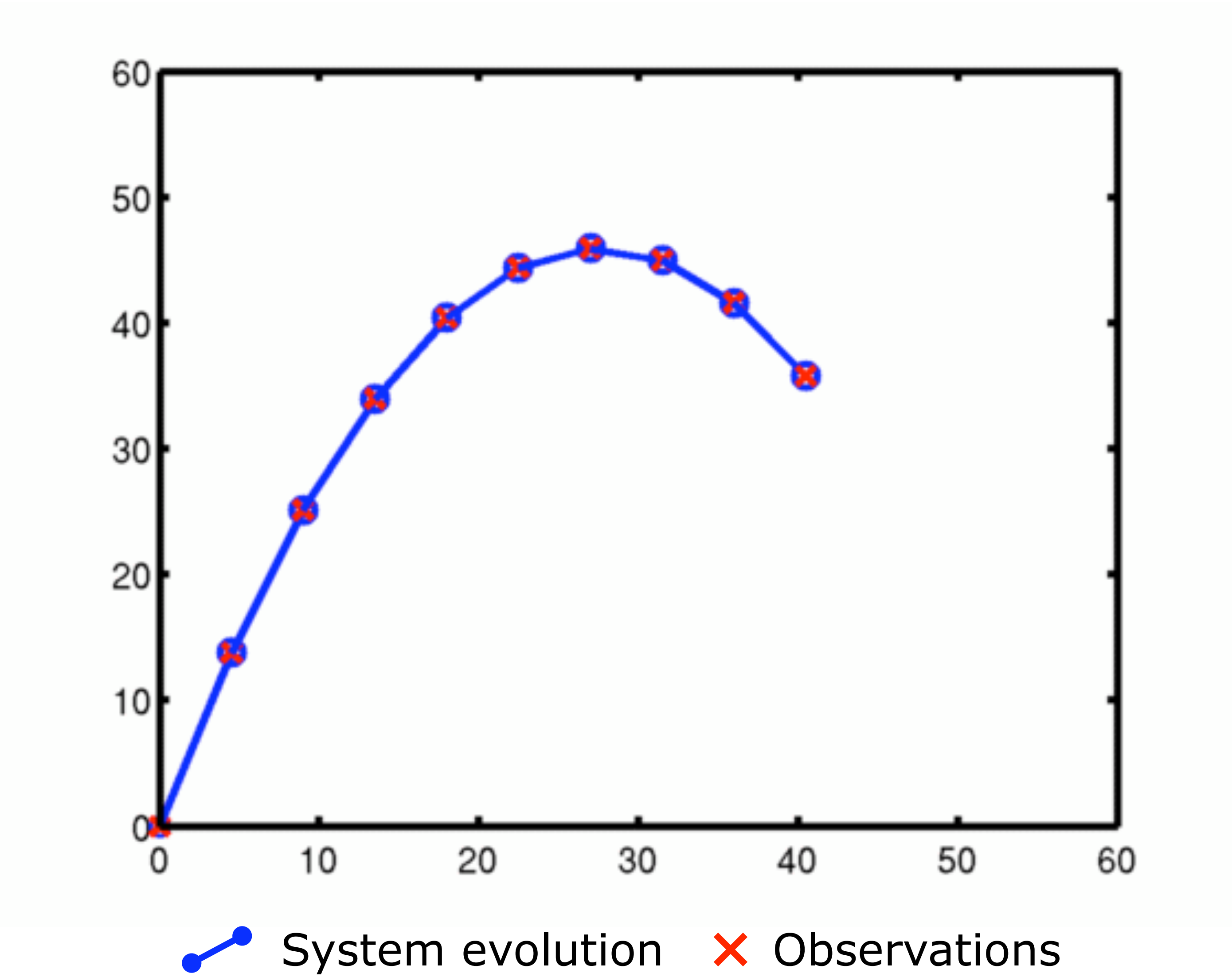
[G. Grisetti, C. Stachniss, K. Arras, and W. Burgard, Univ. of Freiburg Course on Robotics & Target Tracking]

LDS Example – Throwing ball



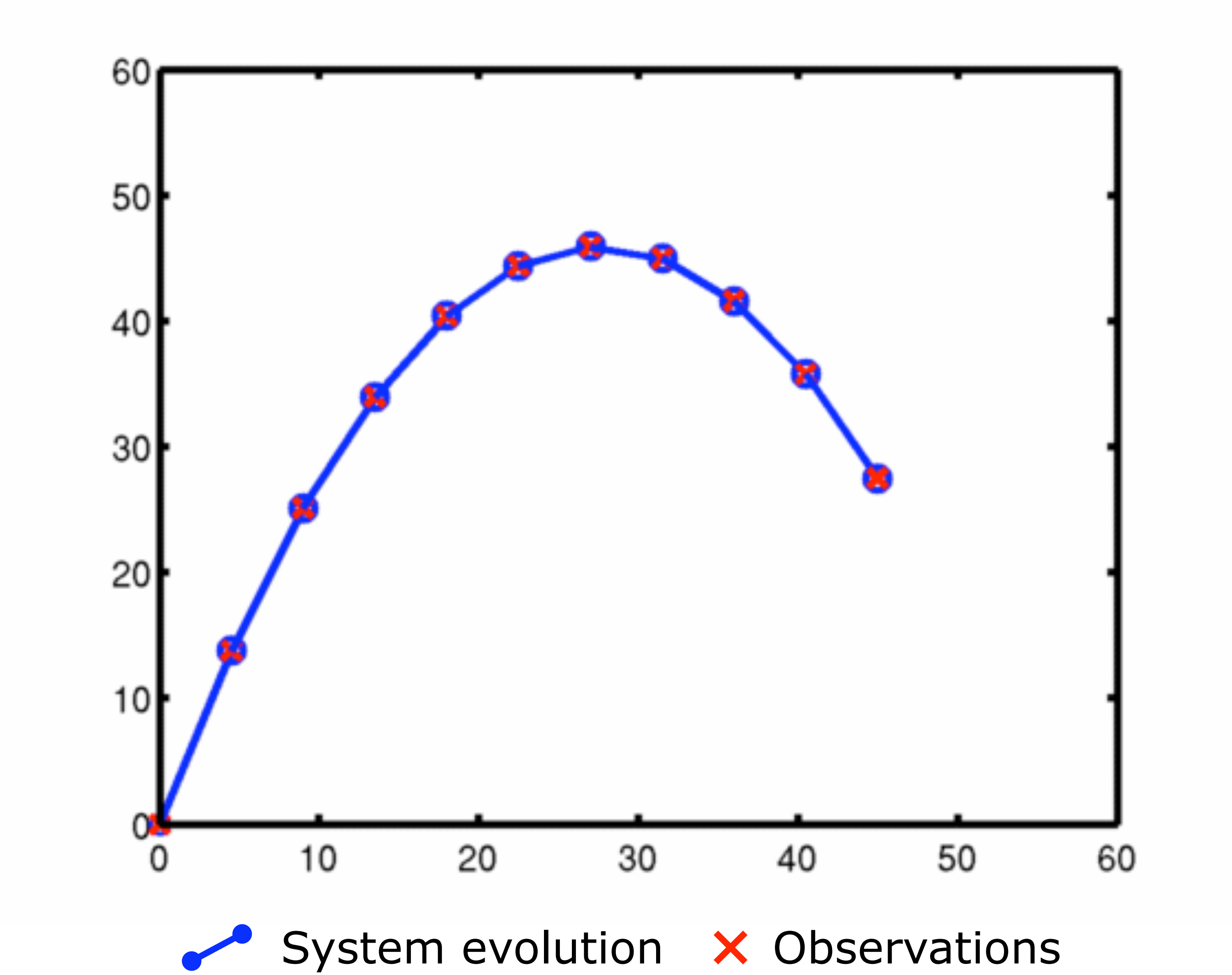
[G. Grisetti, C. Stachniss, K. Arras, and W. Burgard, Univ. of Freiburg Course on Robotics & Target Tracking]

LDS Example – Throwing ball



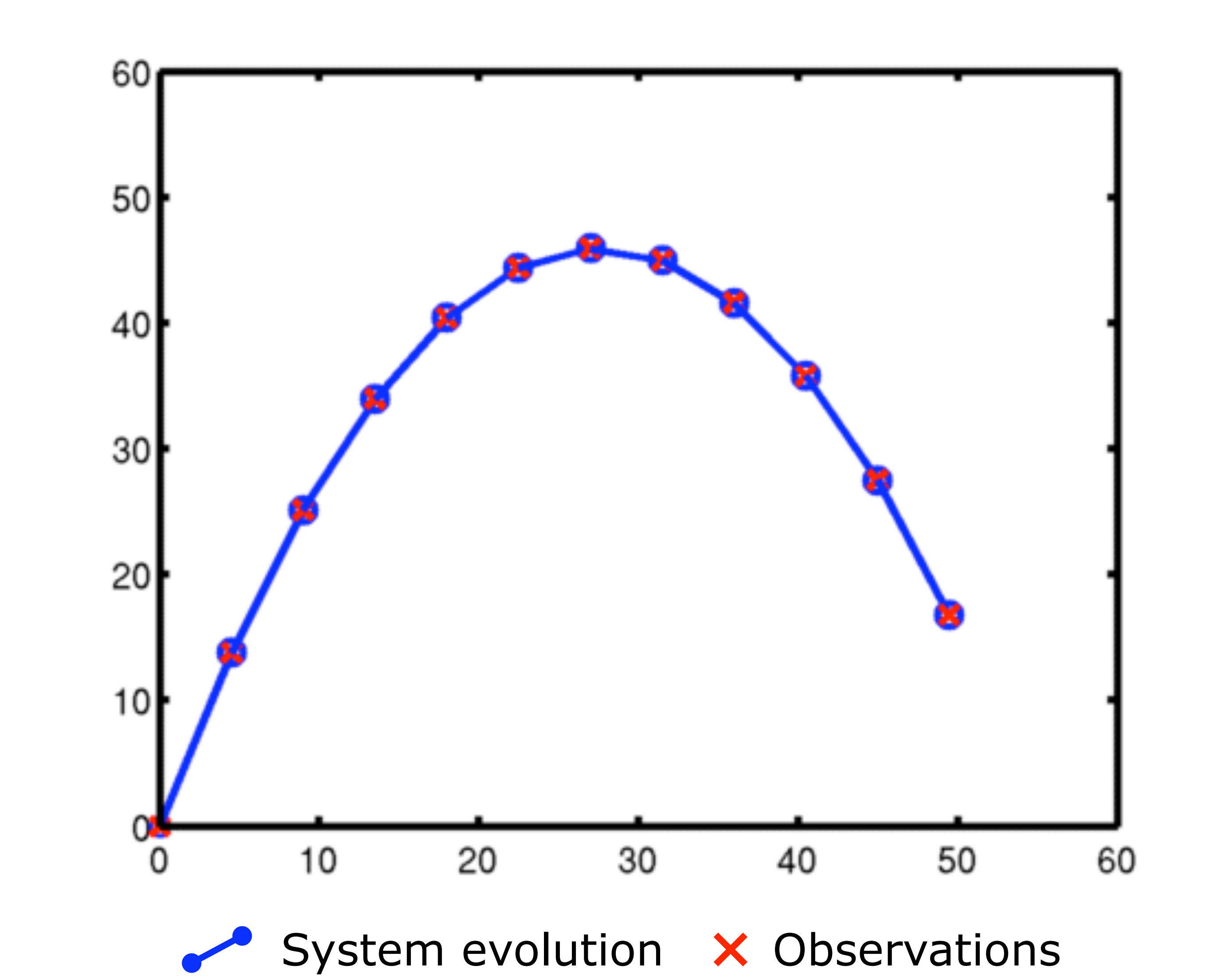
[G. Grisetti, C. Stachniss, K. Arras, and W. Burgard, Univ. of Freiburg Course on Robotics & Target Tracking]

LDS Example – Throwing ball



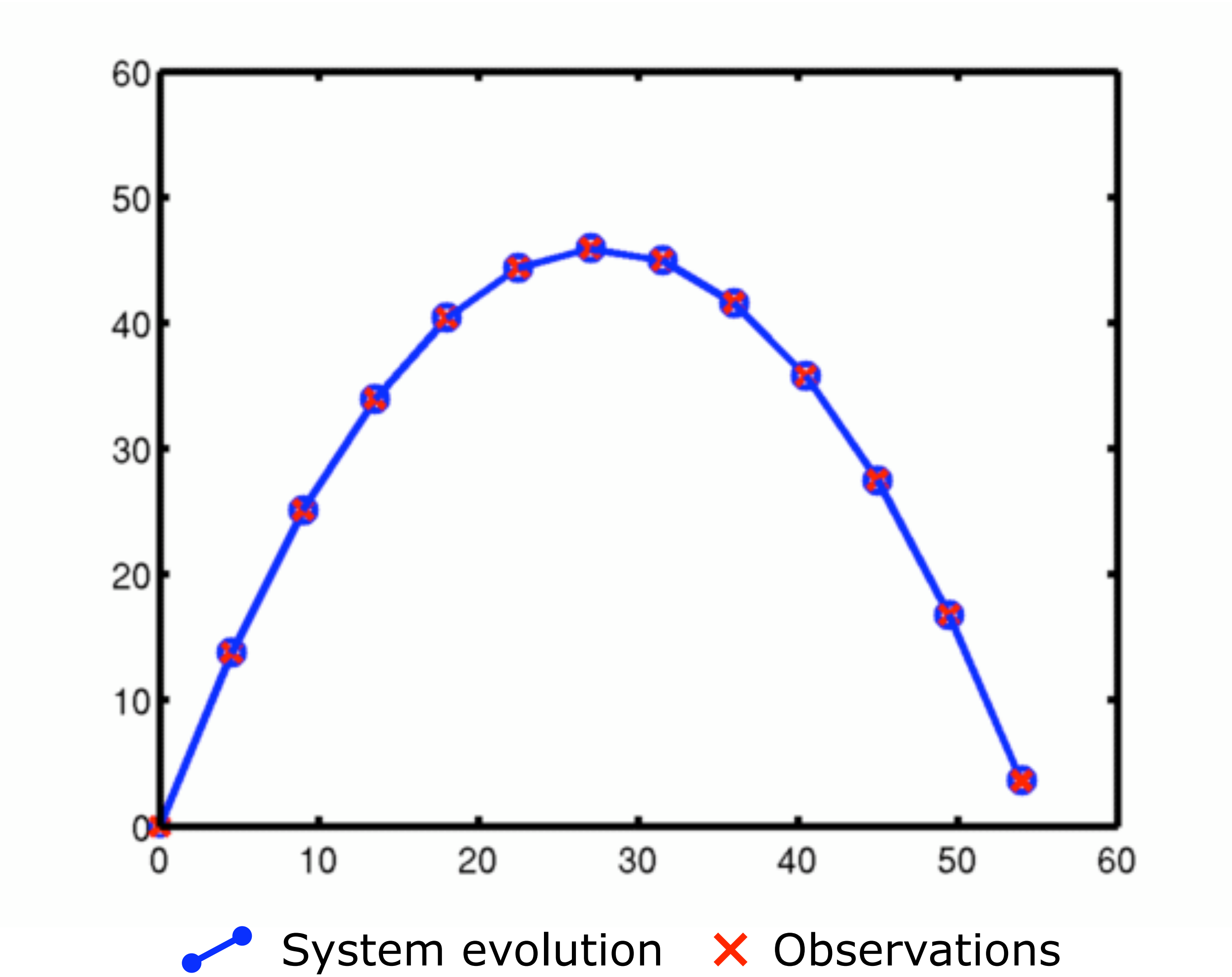
[G. Grisetti, C. Stachniss, K. Arras, and W. Burgard, Univ. of Freiburg Course on Robotics & Target Tracking]

LDS Example – Throwing ball



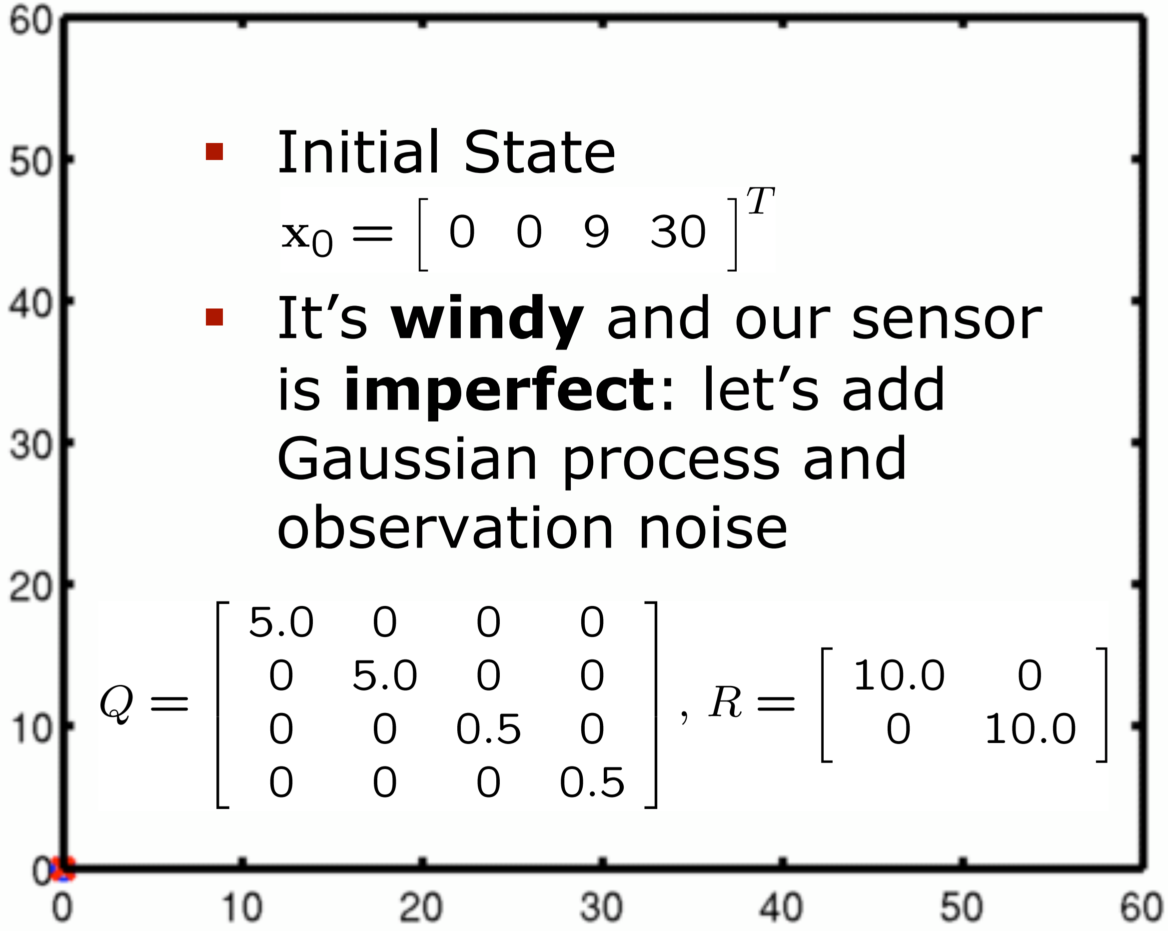
[G. Grisetti, C. Stachniss, K. Arras, and W. Burgard, Univ. of Freiburg Course on Robotics & Target Tracking]

LDS Example – Throwing ball



[G. Grisetti, C. Stachniss, K. Arras, and W. Burgard, Univ. of Freiburg Course on Robotics & Target Tracking]

LDS Example – Throwing ball



- Initial State

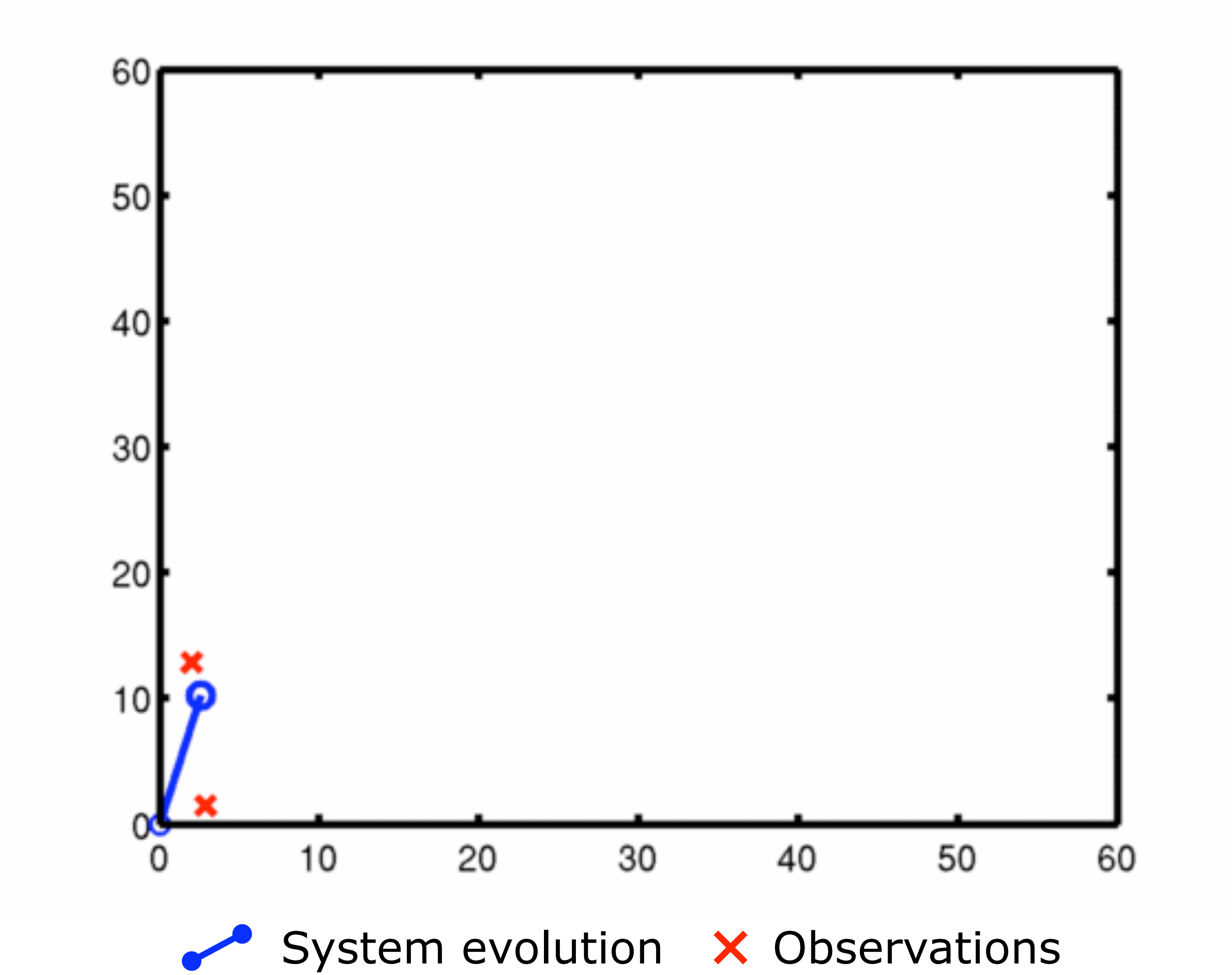
$$\mathbf{x}_0 = \begin{bmatrix} 0 & 0 & 9 & 30 \end{bmatrix}^T$$

- It's **windy** and our sensor is **imperfect**: let's add Gaussian process and observation noise

$$Q = \begin{bmatrix} 5.0 & 0 & 0 & 0 \\ 0 & 5.0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0.5 \end{bmatrix}, \quad R = \begin{bmatrix} 10.0 & 0 \\ 0 & 10.0 \end{bmatrix}$$

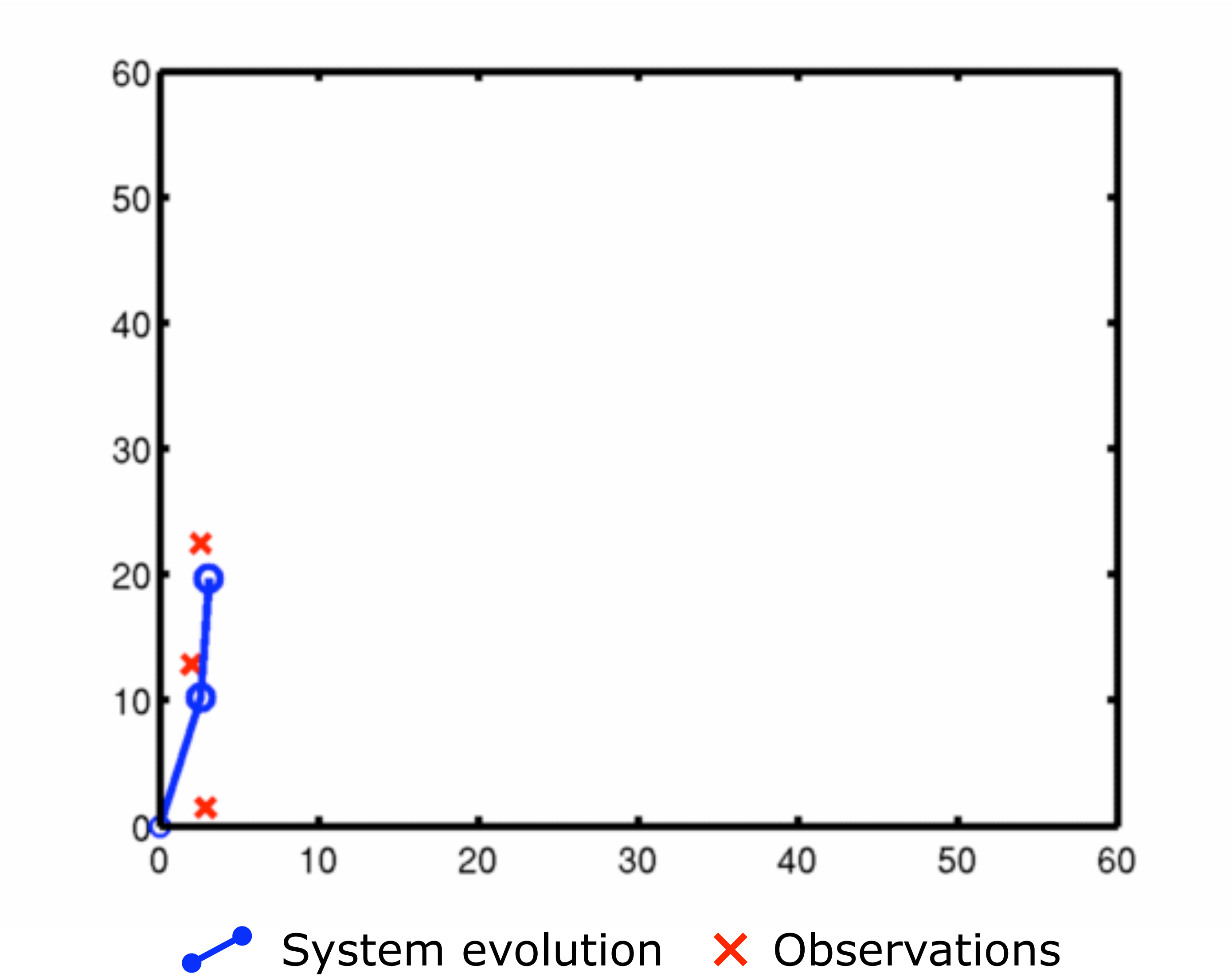
[G. Grisetti, C. Stachniss, K. Arras, and W. Burgard, Univ. of Freiburg Course on Robotics & Target Tracking]

LDS Example – Throwing ball



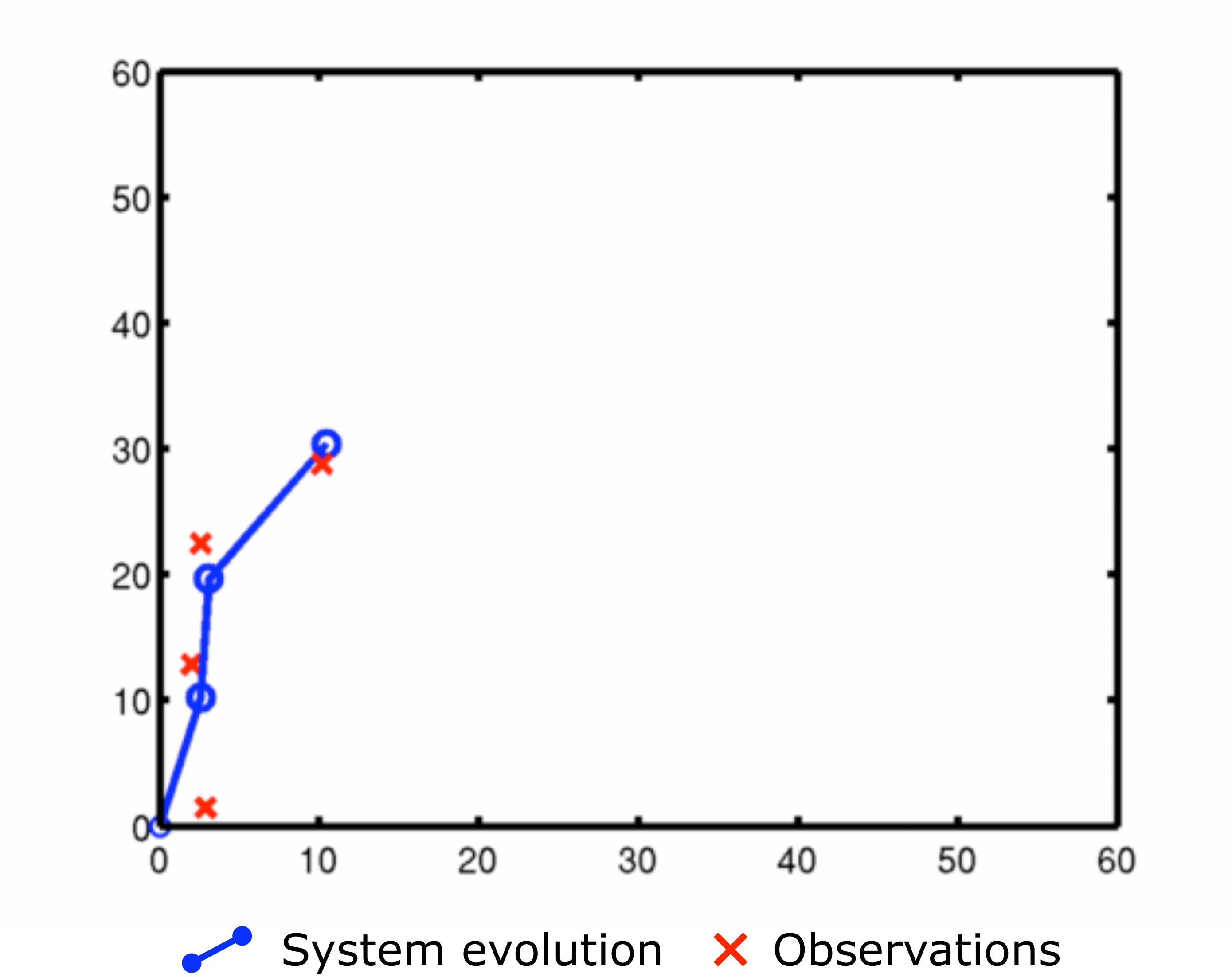
[G. Grisetti, C. Stachniss, K. Arras, and W. Burgard, Univ. of Freiburg Course on Robotics & Target Tracking]

LDS Example – Throwing ball



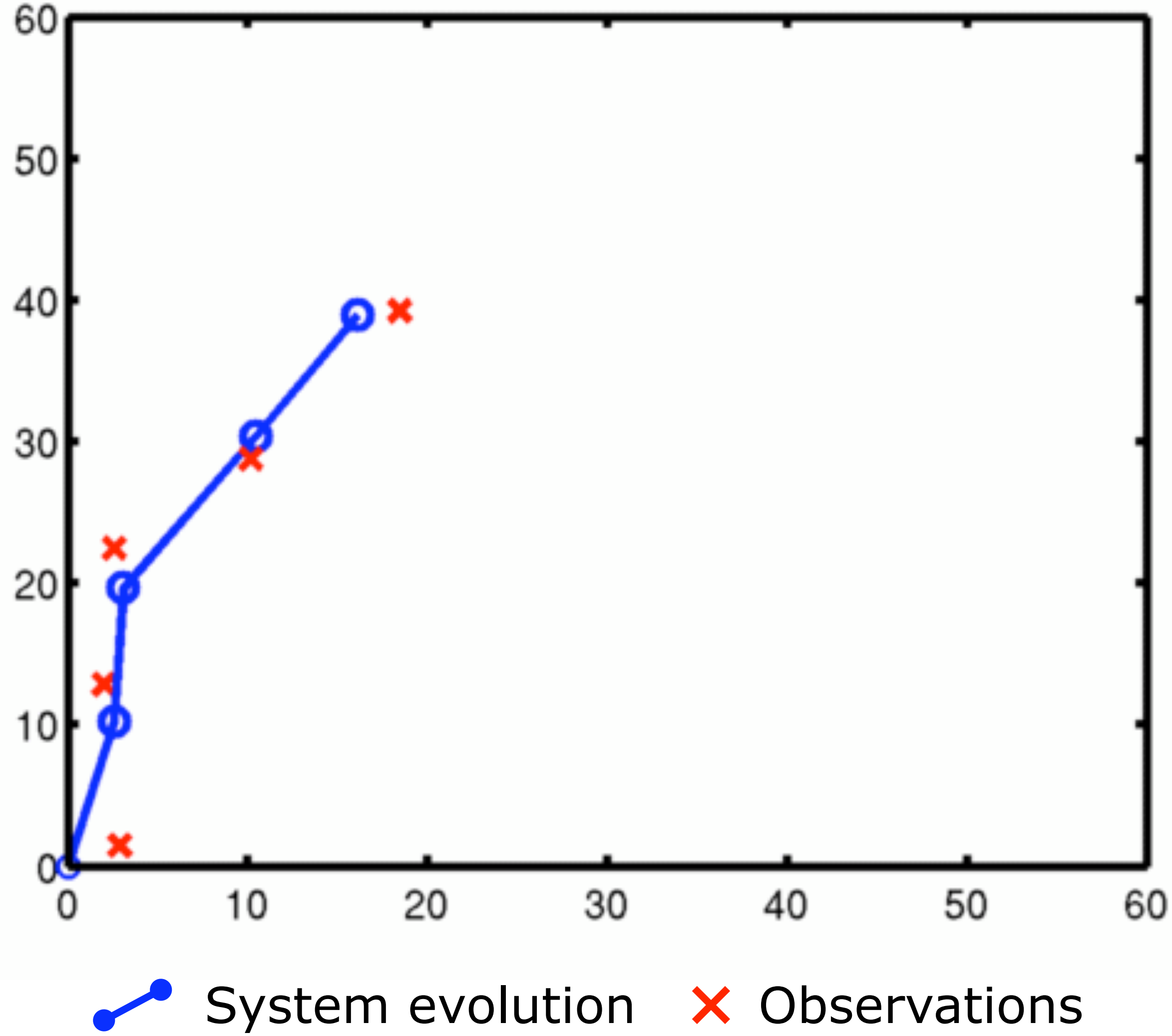
[G. Grisetti, C. Stachniss, K. Arras, and W. Burgard, Univ. of Freiburg Course on Robotics & Target Tracking]

LDS Example – Throwing ball



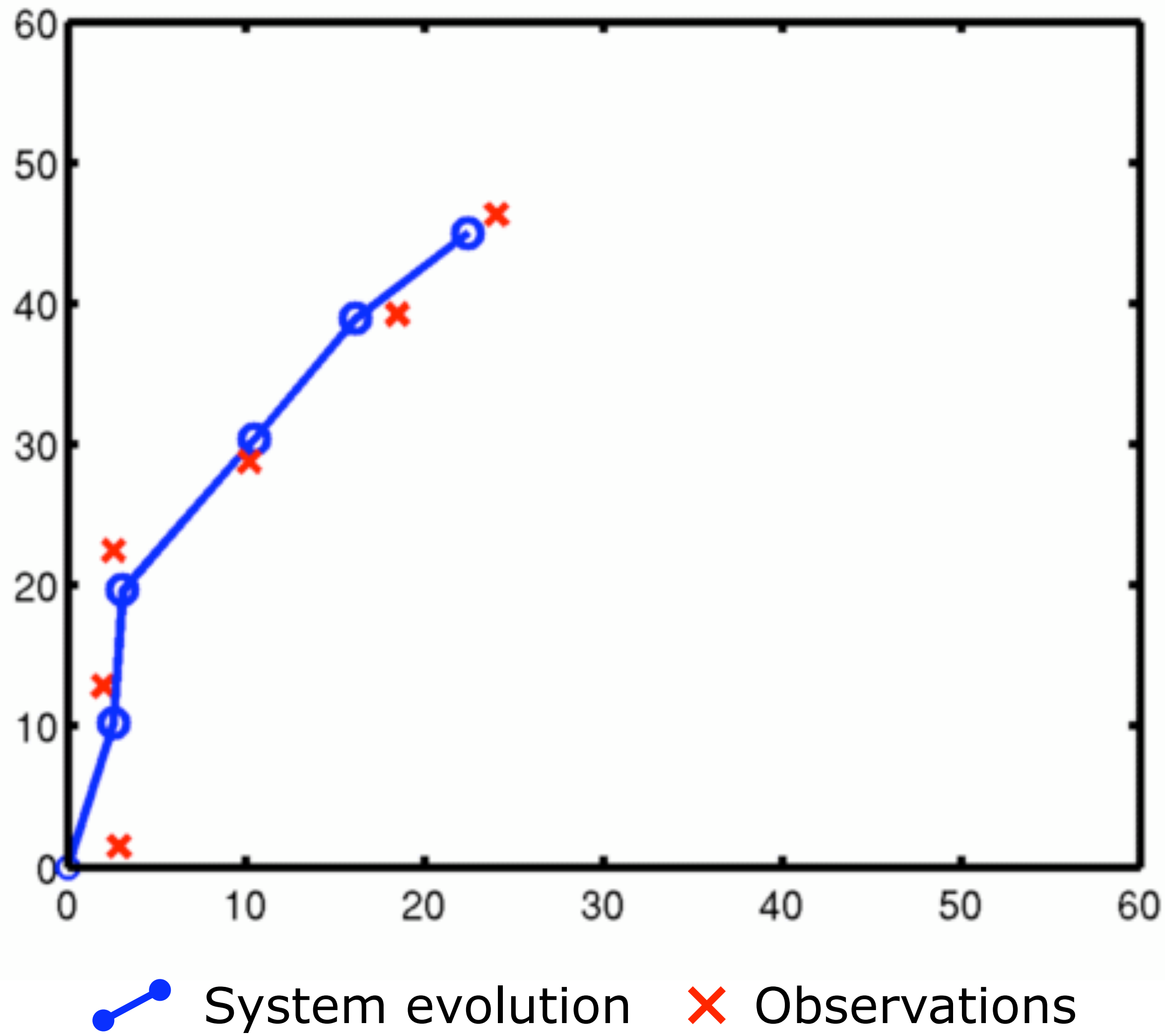
[G. Grisetti, C. Stachniss, K. Arras, and W. Burgard, Univ. of Freiburg Course on Robotics & Target Tracking]

LDS Example – Throwing ball



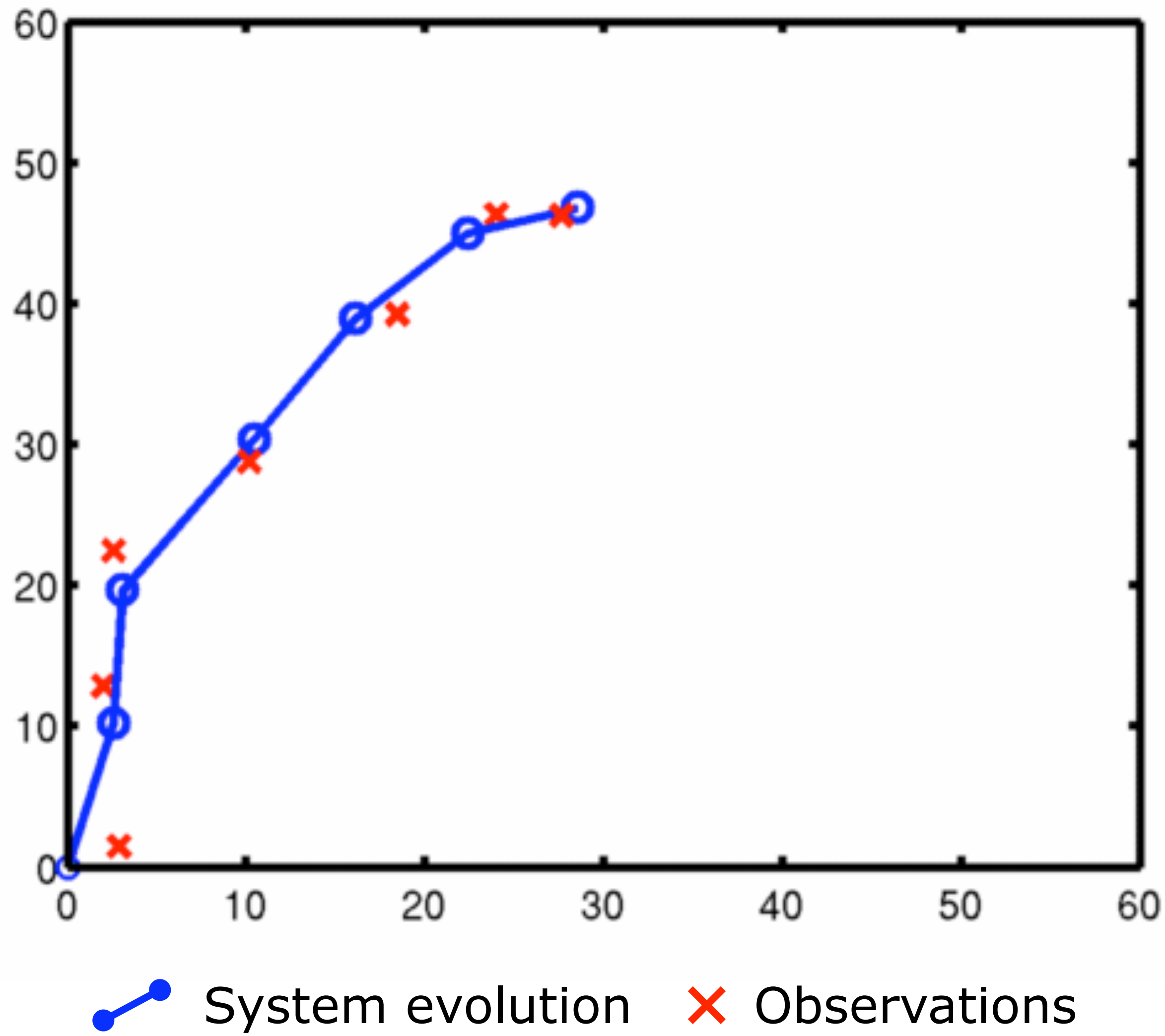
[G. Grisetti, C. Stachniss, K. Arras, and W. Burgard, Univ. of Freiburg Course on Robotics & Target Tracking]

LDS Example – Throwing ball



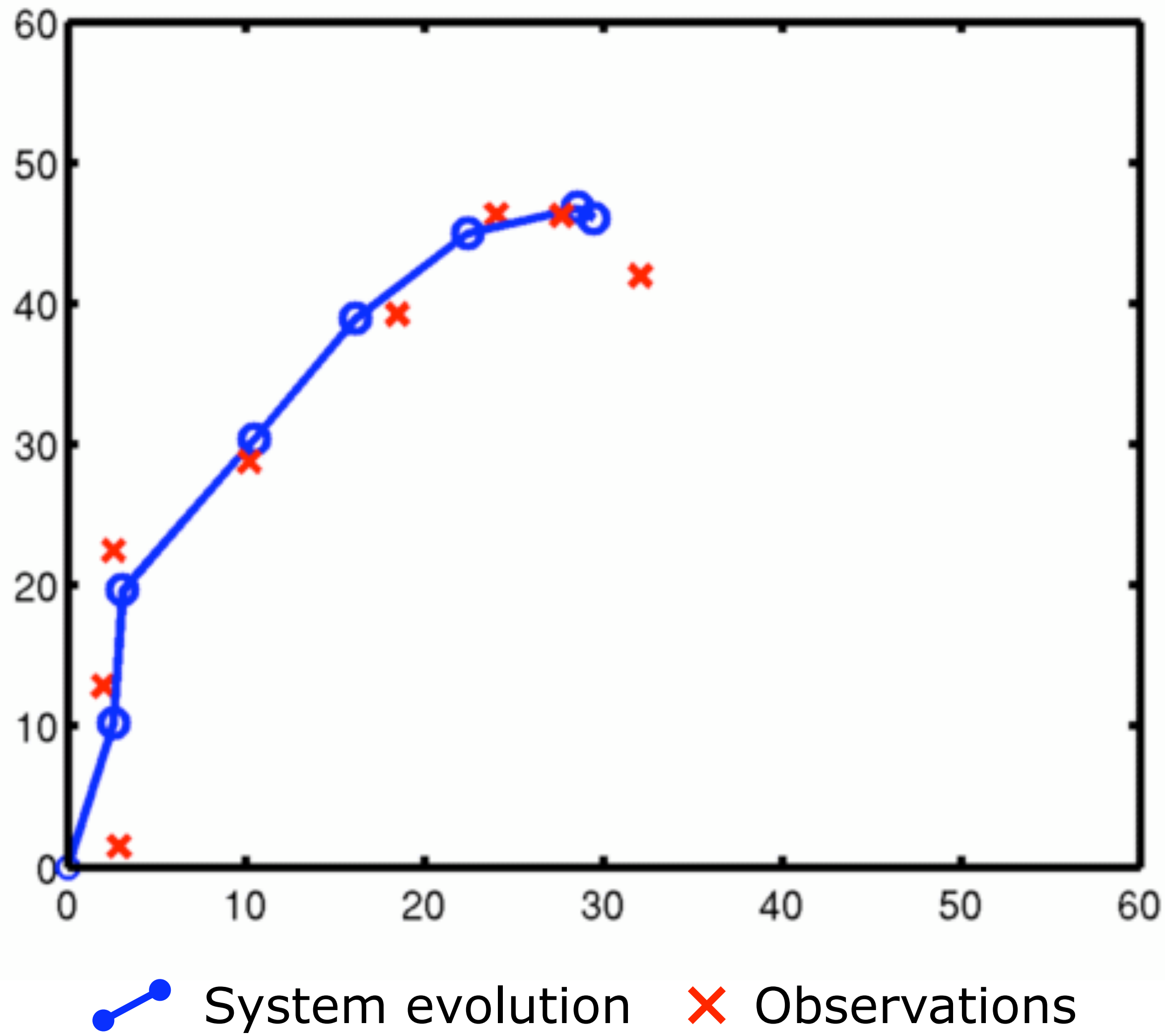
[G. Grisetti, C. Stachniss, K. Arras, and W. Burgard, Univ. of Freiburg Course on Robotics & Target Tracking]

LDS Example – Throwing ball



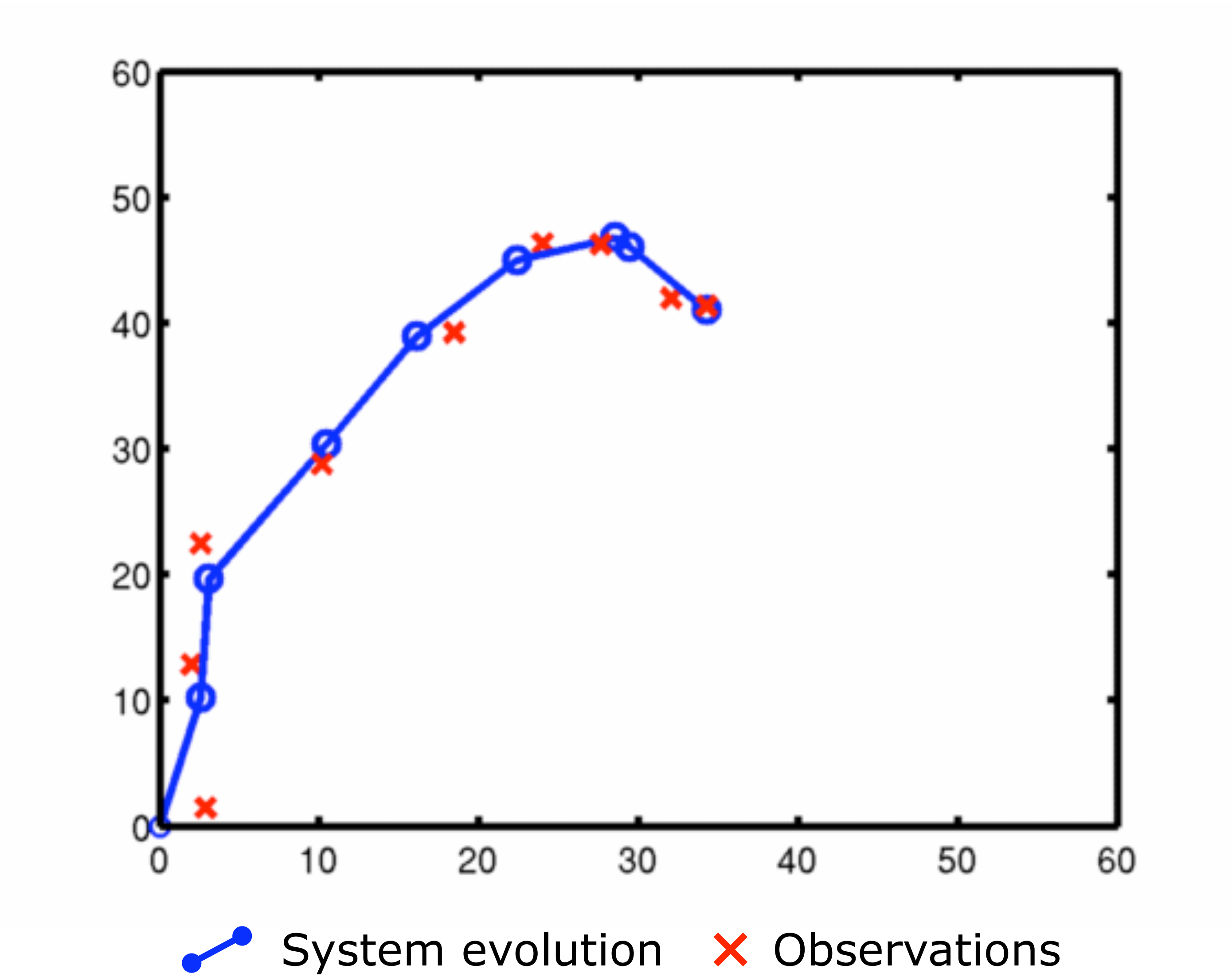
[G. Grisetti, C. Stachniss, K. Arras, and W. Burgard, Univ. of Freiburg Course on Robotics & Target Tracking]

LDS Example – Throwing ball



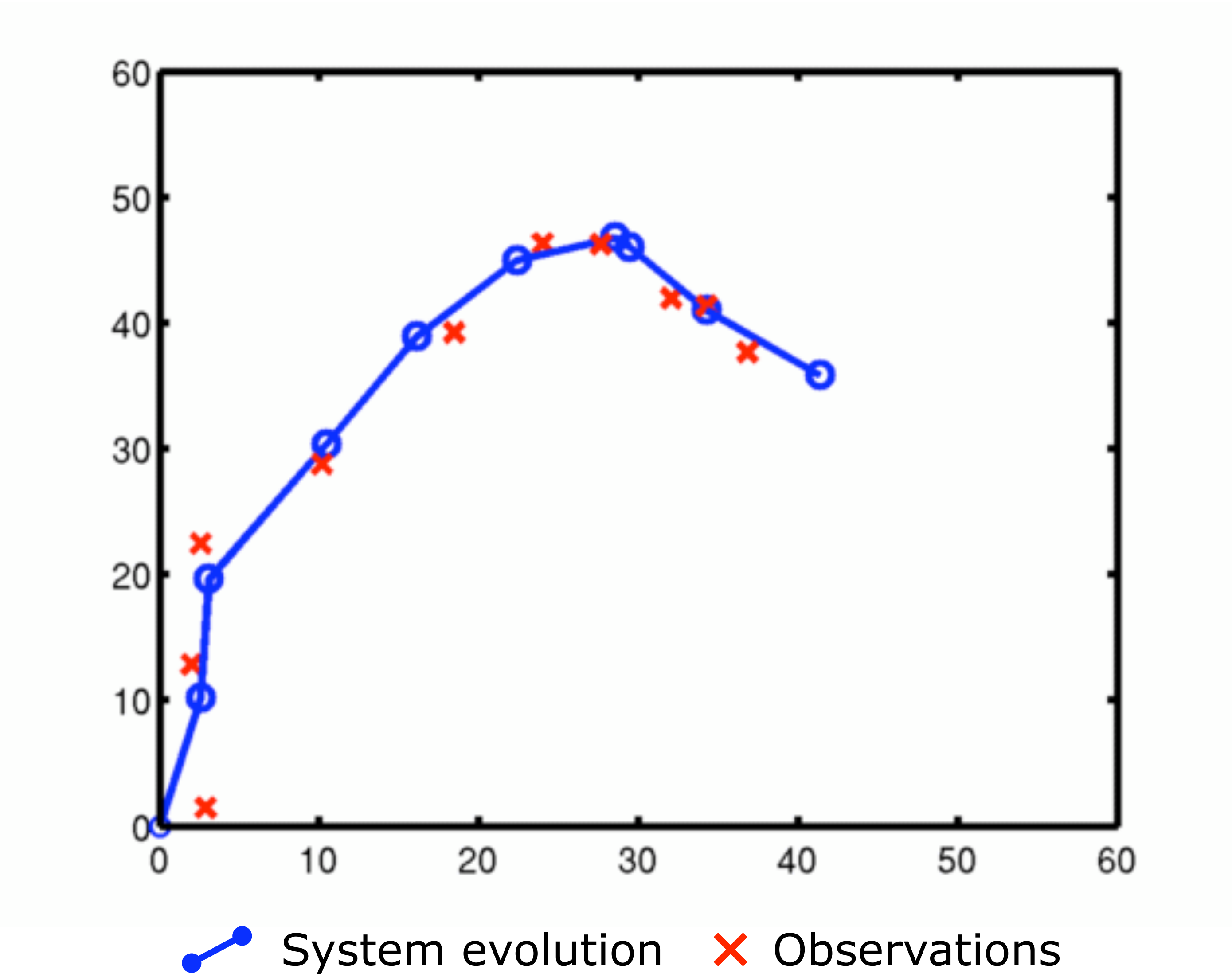
[G. Grisetti, C. Stachniss, K. Arras, and W. Burgard, Univ. of Freiburg Course on Robotics & Target Tracking]

LDS Example – Throwing ball



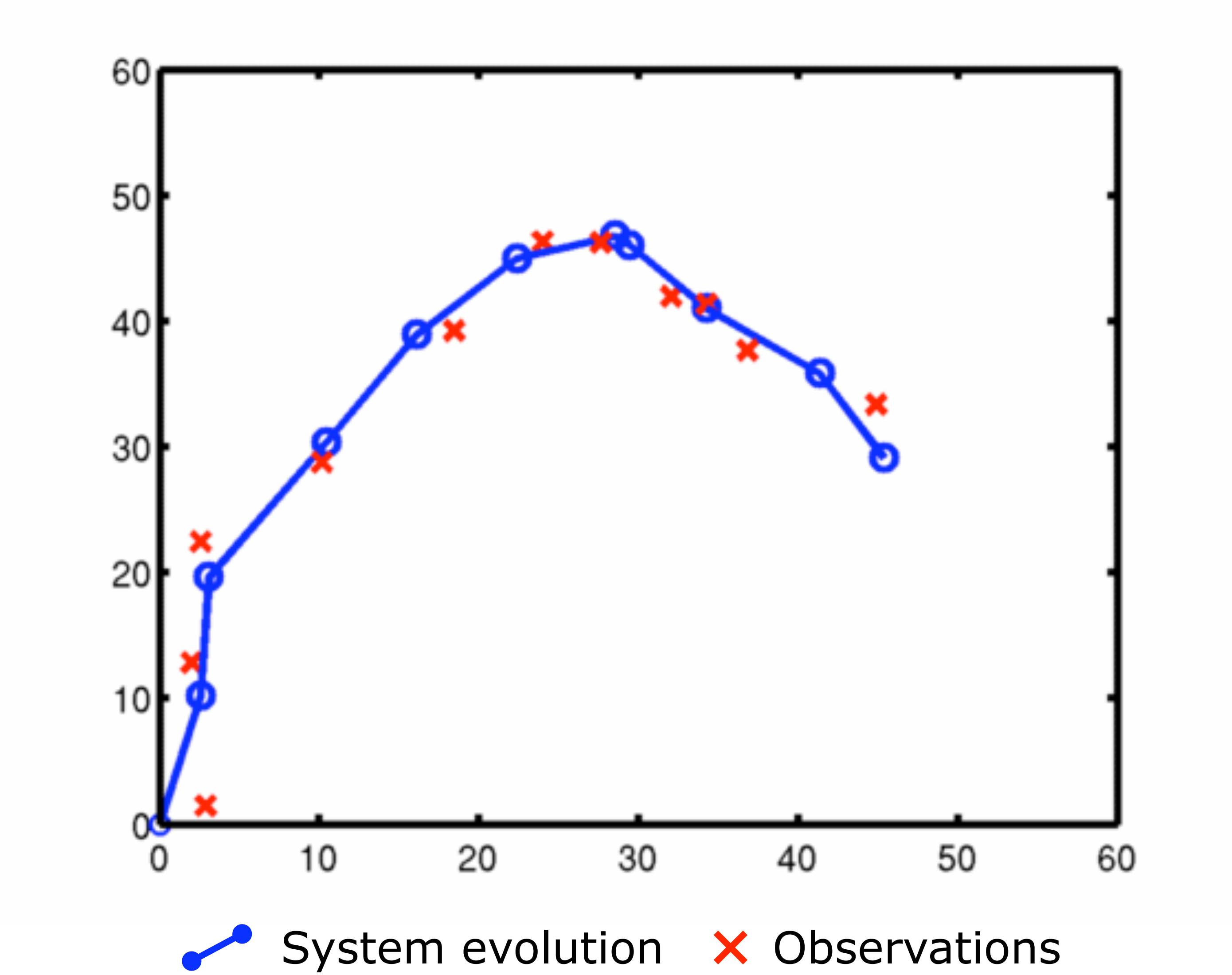
[G. Grisetti, C. Stachniss, K. Arras, and W. Burgard, Univ. of Freiburg Course on Robotics & Target Tracking]

LDS Example – Throwing ball



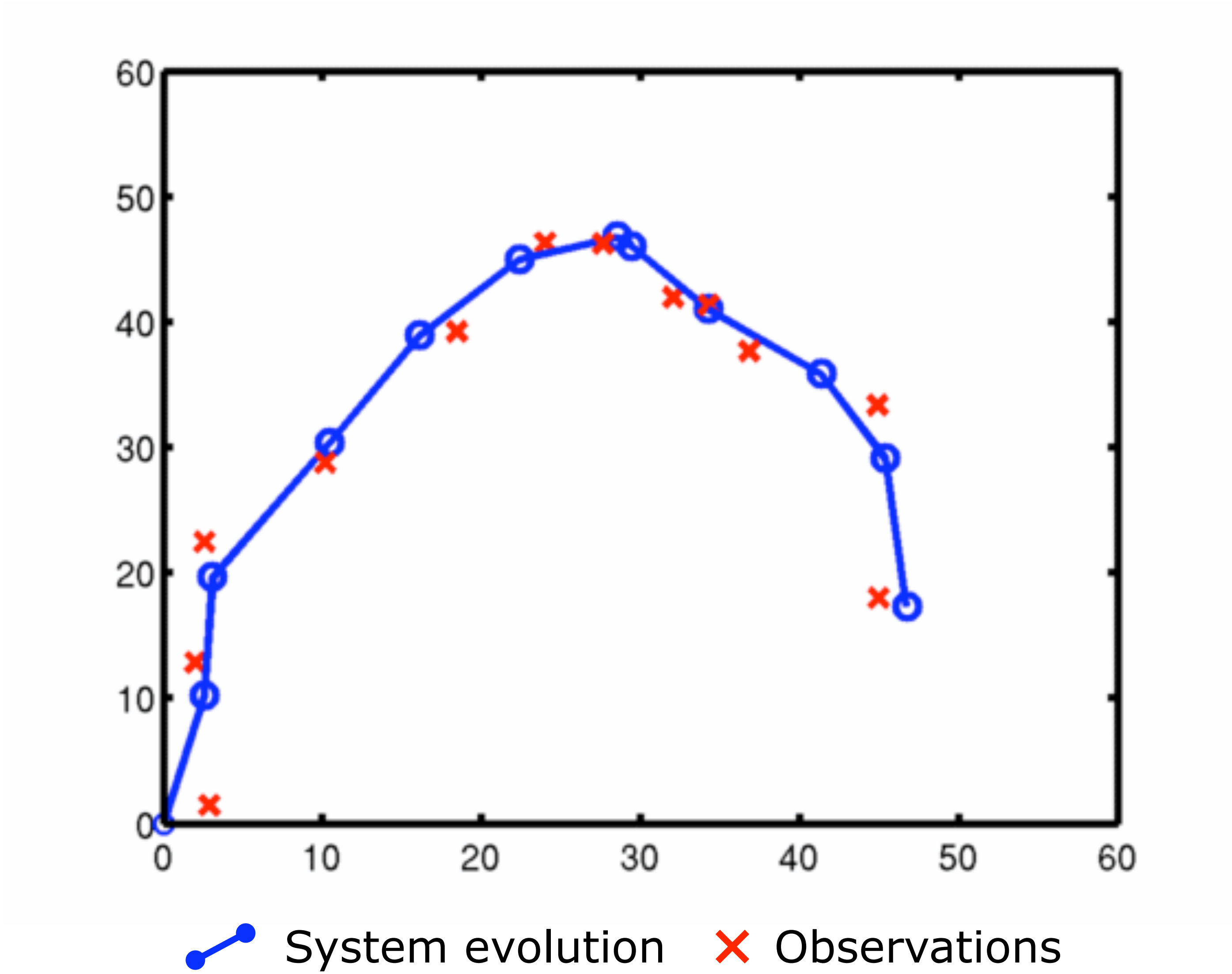
[G. Grisetti, C. Stachniss, K. Arras, and W. Burgard, Univ. of Freiburg Course on Robotics & Target Tracking]

LDS Example – Throwing ball



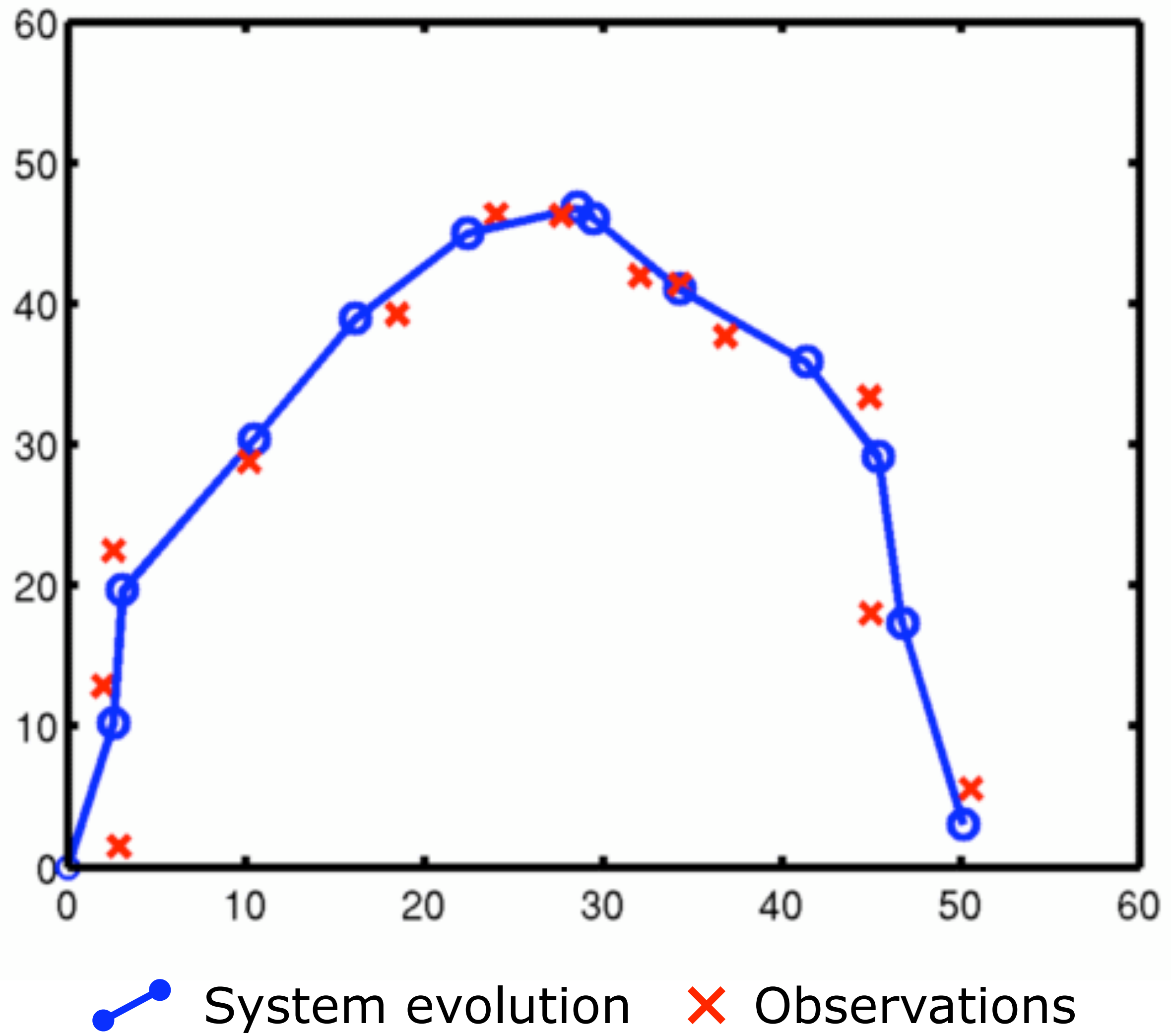
[G. Grisetti, C. Stachniss, K. Arras, and W. Burgard, Univ. of Freiburg Course on Robotics & Target Tracking]

LDS Example – Throwing ball



[G. Grisetti, C. Stachniss, K. Arras, and W. Burgard, Univ. of Freiburg Course on Robotics & Target Tracking]

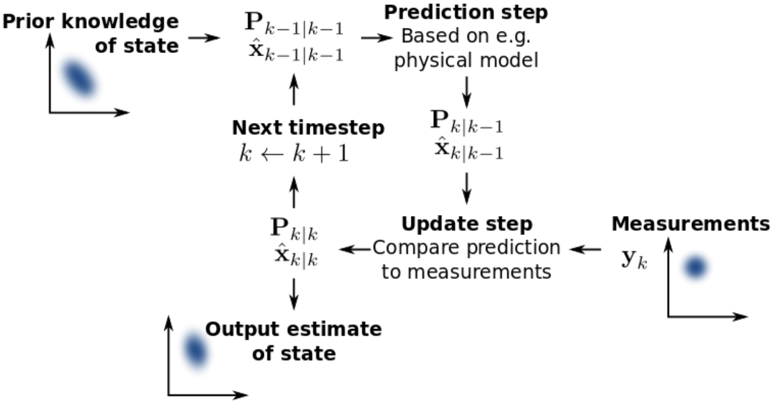
LDS Example – Throwing ball



[G. Grisetti, C. Stachniss, K. Arras, and W. Burgard, Univ. of Freiburg Course on Robotics & Target Tracking]

Tracking Is Filtering

Kalman Filter



source: `https://en.wikipedia.org/wiki/Kalman_filter`

CONDENSATION

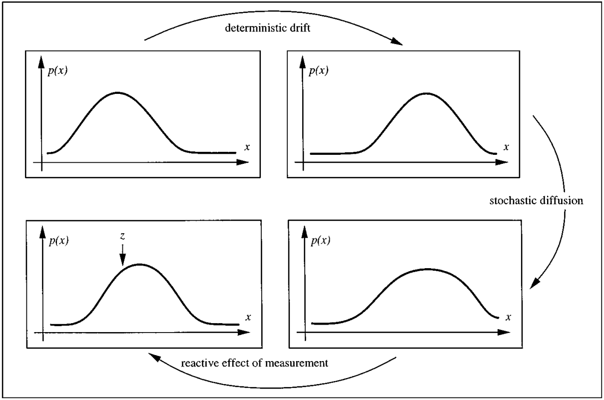


Figure 1. Kalman filter as density propagation: in the case of Gaussian prior, process and observation densities, and assuming linear dynamics, the propagation process of Fig. 2 reduces to a diffusing Gaussian state density, represented completely by its evolving (multivariate) mean and variance—precisely what a Kalman filter computes.

[M. Isard and A. Blake, "CONDENSATION: Conditional Density Propagation for Visual Tracking," International Journal of Computer Vision, 1998]

CONDENSATION

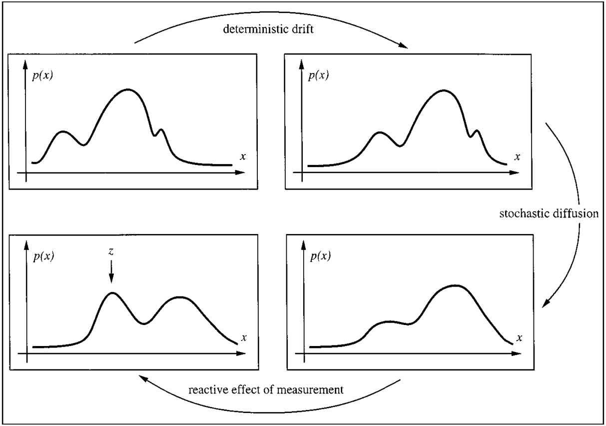


Figure 2. Probability density propagation: propagation is depicted here as it occurs over a discrete time-step. There are three phases: drift due to the deterministic component of object dynamics; diffusion due to the random component; reactive reinforcement due to observations.

[M. Isard and A. Blake, "CONDENSATION: Conditional Density Propagation for Visual Tracking," International Journal of Computer Vision, 1998]

CONDENSATION

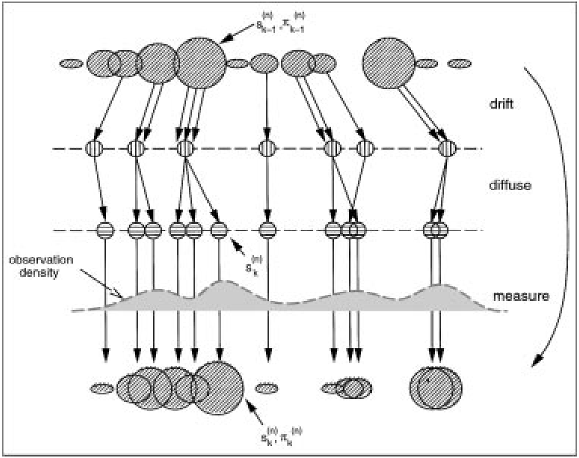
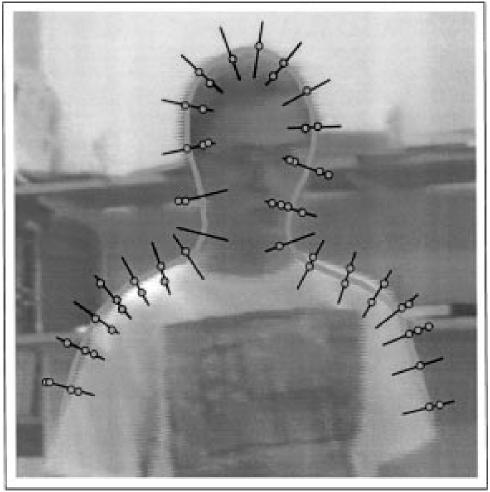


Figure 5. One time-step in the CONDENSATION algorithm: Each of the three steps—drift-diffuse-measure—of the probabilistic propagation process of Fig. 2 is represented by steps in the CONDENSATION algorithm.

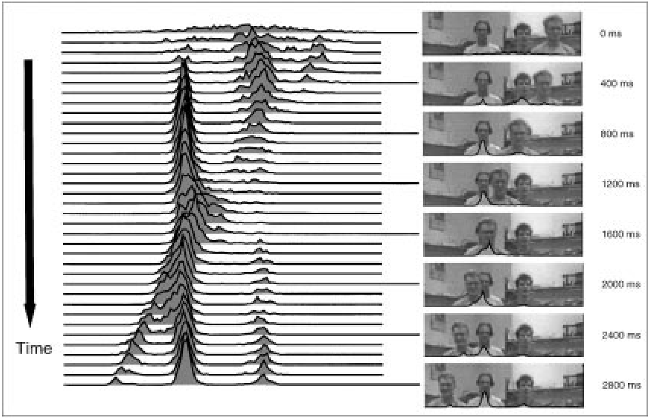
[M. Isard and A. Blake, "CONDENSATION: Conditional Density Propagation for Visual Tracking," International Journal of Computer Vision, 1998]

CONDENSATION



[M. Isard and A. Blake, “CONDENSATION: Conditional Density Propagation for Visual Tracking,” International Journal of Computer Vision, 1998]

CONDENSATION



[M. Isard and A. Blake, "CONDENSATION: Conditional Density Propagation for Visual Tracking," International Journal of Computer Vision, 1998]

Multiple-Hypothesis Tracking (MHT)

Bibliography

D.B. Reid, “An algorithm for tracking multiple targets,” IEEE Trans. on Automatic Control, vol. 24, no. 6, pp. 843-854, Dec. 1979.

Y. Bar-Shalom and T.E. Fortmann, “Tracking and Data Association.” Academic Press, 1988

I. Cox and S.L. Hingorani, “An Eficient Implementation of Reid’s Multiple Hypothesis Tracking Algorithm and Its Evaluation for the Purpose of Visual Tracking,” IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 18(2), 1996

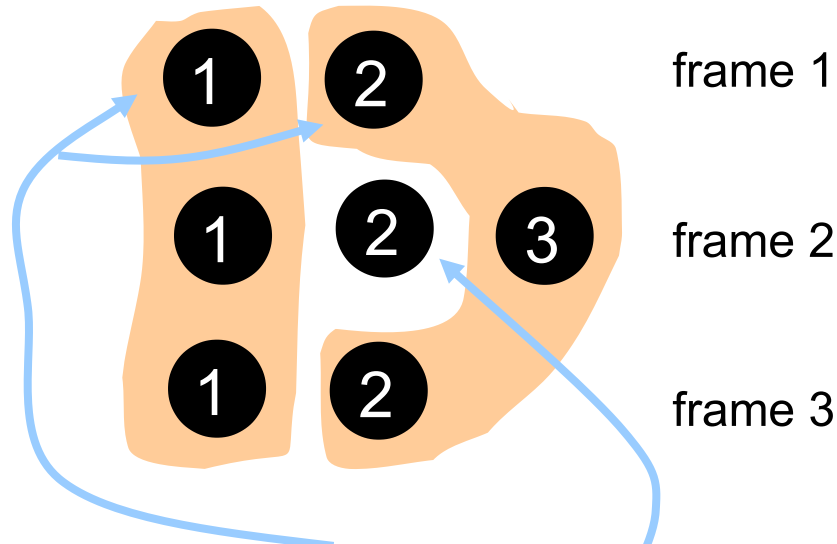
Multiple-Hypothesis Tracking (MHT)

Problem:

- data association
  (partitioning of observations)
- estimation of target tracks

Assumptions

- A1: one observation comes from one target or clutter
- A2: one target yields zero or one observation
- ↕ A3: one target yields zero or several observations



$$Z^k = \{Z(1), \ldots, Z(k)\}$$

partition $\omega \triangleq \{t_0, \ldots, t_M\}$

seek most probable partition

$Z^k \triangleq$ all obs. $Z(k) \triangleq$ obs. current frame $t_0 \triangleq$ false alarms $t_m \triangleq$ track *m*

Multiple-Hypothesis Tracking (MHT)

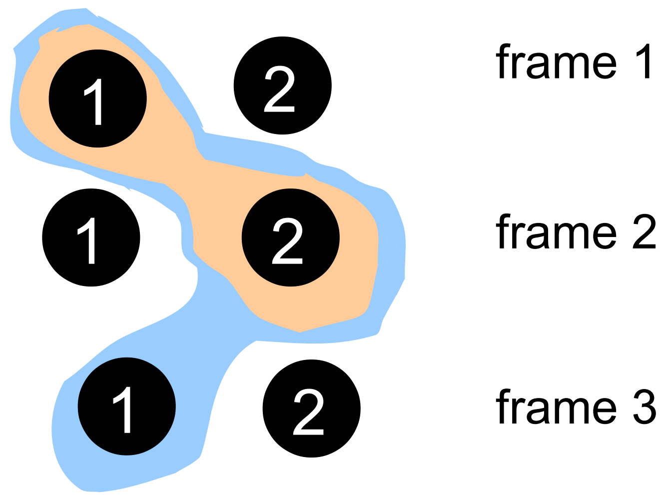
Key idea: hypothesis tree

parent hypothesis $\Theta^{k-1}_{\wp(l)}$

$\Theta^k_1 \quad \Theta^k_2 \quad \Theta^k_l$

current hypothesis

$$\Theta^k_l = \left\{\Theta^{k-1}_{\wp(l)}, \theta_l(k)\right\}$$



MHT integrates

- Track initiation and termination
- Track update with (or without) observation
- Accounting for false alarms
- Enforcement of assumptions (A1, A2)

$\theta_l(k) \triangleq$ assignment l at frame k $\quad \Theta^{k-1}_{\wp(l)} \triangleq$ $\wp$arent hypothesis

Hypothesis Probability

Given *N* frames, enumerate finite set of hypotheses

Compute probability for each element of this set

$$P\left(\Theta_l^k \mid Z^k\right) \propto$$

$$P\left\{Z(k) \mid \theta_l(k), \Theta_{\wp(l)}^{k-1}, Z^{k-1}\right\}$$

P(current obs. | current assignment)

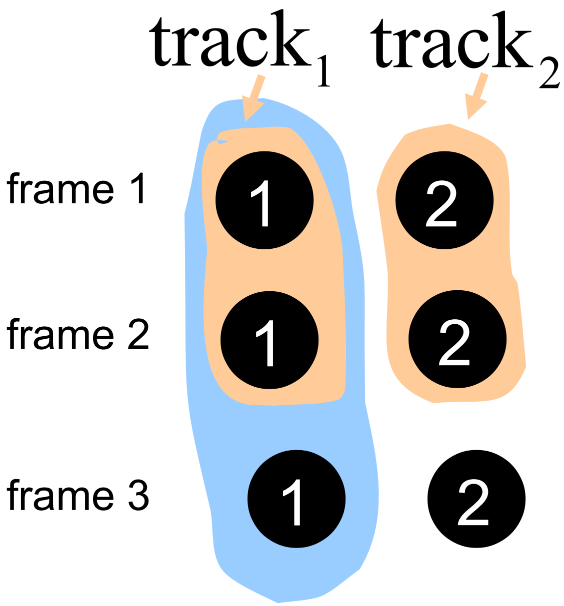
$$P\left\{\theta_l(k) \mid \Theta_{\wp(l)}^{k-1}, Z^{k-1}\right\}$$

× *P*(current assignment | parent hypothesis)

$$P\left\{\Theta_{\wp(l)}^{k-1} \mid Z^{k-1}\right\}$$

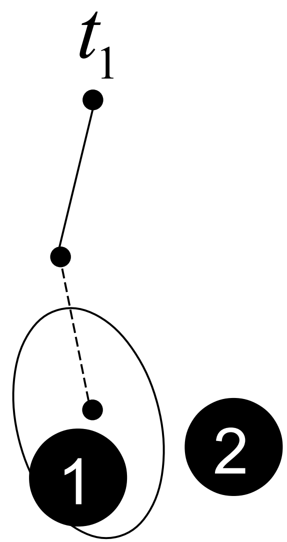
× *P*(parent hypothesis | prior obs.)

Probability of Observations given Assignment



$$m_3 = 2$$

$$\tau_1 = 1 \qquad \tau_2 = 0$$



$$p\left\{Z(k) \mid \theta_l(k), \Theta^{k-1}_{\wp(l)}, Z^{k-1}\right\} = \prod_{i=1}^{m_k} \left(\mathcal{N}_{t_i}\left(z_i(k)\right)\right)^{\tau_i} \left(\frac{1}{V}\right)^{1-\tau_i}$$

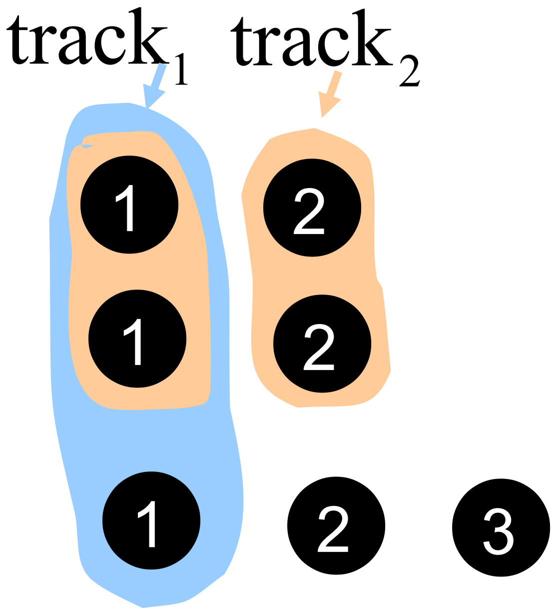
observations are independent

obs. from track t_i Normally distributed

false alarms uniformly distributed in obs. area V

$m_k \triangleq$ num obs. frame k

Probability of Assignment (1)

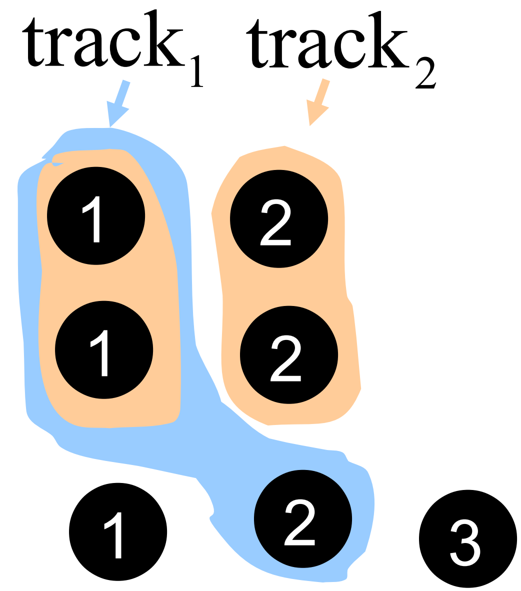


Hypothesis 1

similarities:

no assignment for t_2

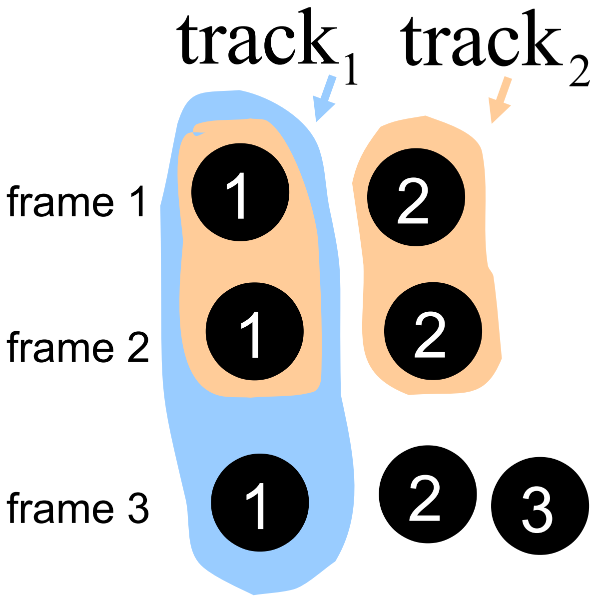
2 false alarms



Hypothesis 2

We can sum probabilities for these hypotheses !

Probability of Assignment (2)



$$m_k = 3 \quad \phi = 2$$
$$\nu = 0$$
$$\chi_1 = 0 \; \chi_2 = 0$$

$$P\left\{\theta_l(k) \mid \Theta_{\wp(l)}^{k-1}, Z^{k-1}\right\} =$$

$$\frac{\phi!\nu!}{m_k!} \times \mu_F(\phi)\mu_N(\nu)$$

sum over similar hypotheses

priors on number of false alarms & new tracks

$$\times \prod_j (P_D)^{\delta_j}(1-P_D)^{1-\delta_j}(P_\chi)^{\chi_j}(1-P_\chi)^{1-\chi_j}$$

prob. track ends

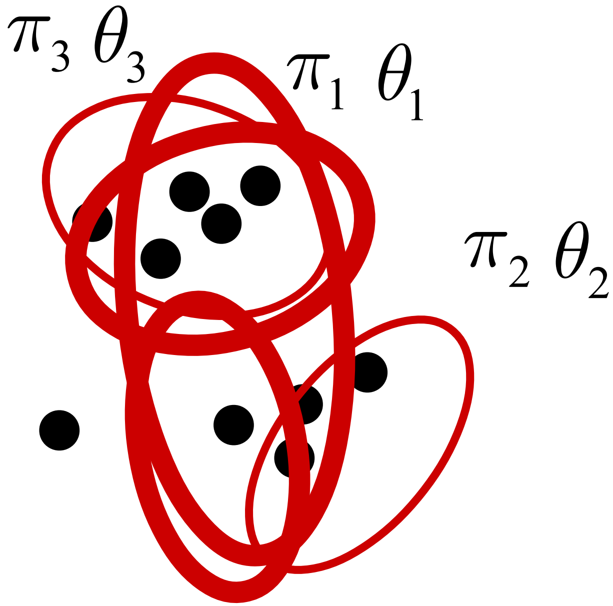
product over all existing tracks

prob. detection

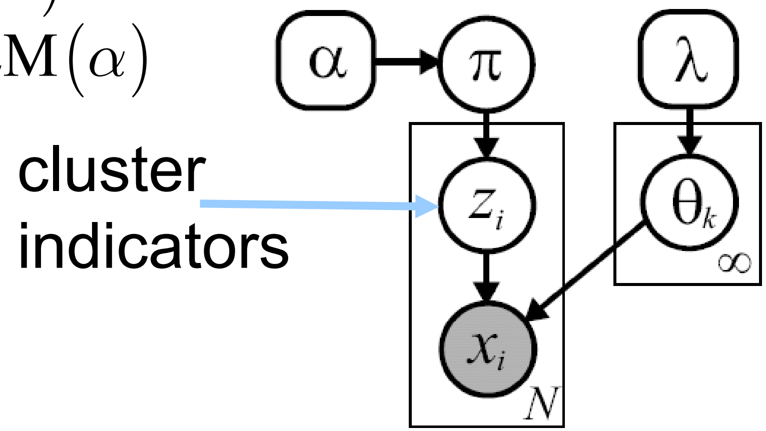
$\phi \triangleq$ num. false alarms $\nu \triangleq$ num. new trks. $\chi_i \triangleq$ 1 if trk. i ends

Multi-Target Tracking as Bayesian Clustering

Bayesian Clustering



$$\theta \sim H(\lambda)$$
$$\pi \sim \mathrm{GEM}(\alpha)$$



Goal: sample from $p(\pi, \theta \mid x_1, \ldots, x_N, \alpha, \lambda)$

$\theta \triangleq$ cluster params. $\pi \triangleq$ cluster weights α, λ : known, fixed

Modeling Infinitely Many Clusters

How to model infinite mixture ?

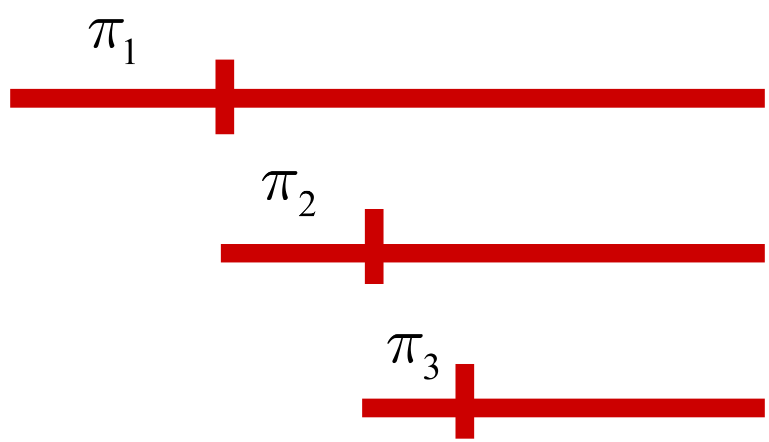
mixture weights

Via Dirichlet process mixture: $p(x \mid \theta_1, \theta_2, \ldots) = \sum_{k=1}^{\infty} \boxed{\pi_k} f(x \mid \theta_k)$

$$\pi \sim \mathrm{GEM}(\alpha)$$

$$\beta_k \sim \mathrm{Beta}(1, \alpha)$$

$$\pi_k = \beta_k \prod_{\ell=1}^{k-1} (1 - \beta_k)$$



Key Densities

assignment — all other assignments

$$p(z_i \mid z_{\setminus i}, \alpha) = \frac{1}{\alpha + N - 1}\left(\alpha\delta(z_i, K+1) + \sum_{k=1}^{K} N_k^{-i}\,\delta(z_i, k)\right)$$

new cluster — num. obs. for cluster *k*

$$p(x_i \mid z_i = k, z_{\setminus i}, x_{\setminus i}, \lambda) = p\left(x_i \mid \left\{x_j \mid z_j = k, j \neq i\right\}, \lambda\right)$$

obs. — all assignments, other obs. — obs. — { obs. in cluster *k* })

Sampling the Posterior Distribution

p (assignment | all other assignments, obs. , α, λ) $\propto$

p (assignment | all other assignments, α) $\times$

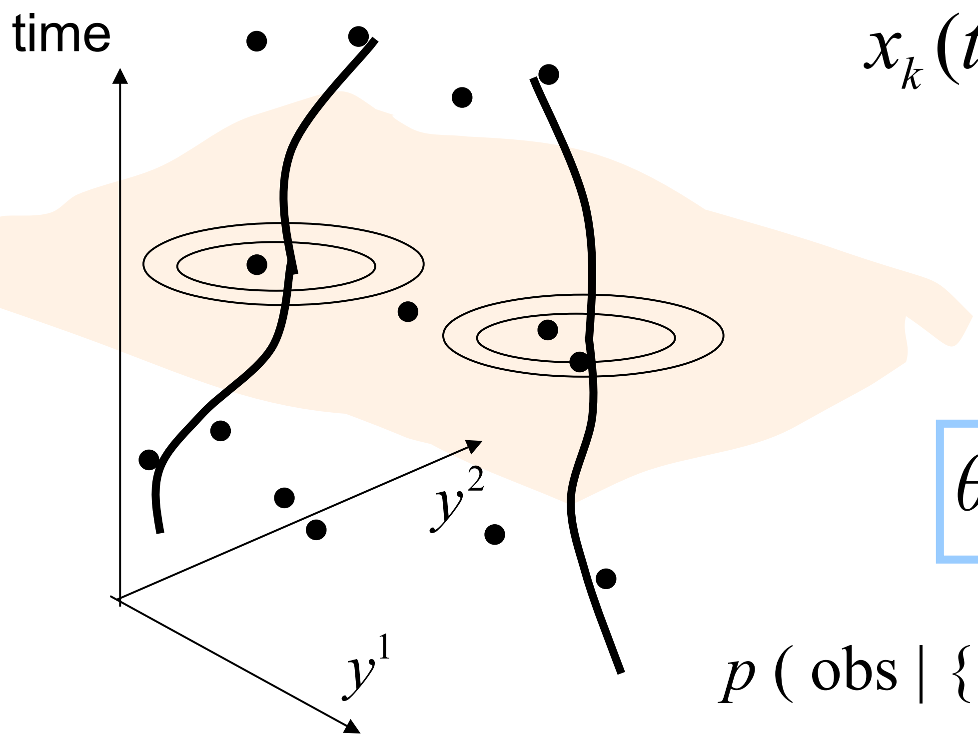
p (obs. | all obs. assignments, λ)

Algorithm: Rao-Blackwellized Gibbs Sampler

Given prior cluster assignments and cluster statistics

1. sample random permutation { 1,…,N }
   (a) For each obs. sample its cluster assignment
   (b) Update that cluster's statistics
2. delete empty clusters

Data Association as Clustering



$$x_k(t) = Ax_k(t-1) + Bu_k(t-1)$$

$$y_k(t) = Cx_k(t) + w_k(t)$$

cluster parameters

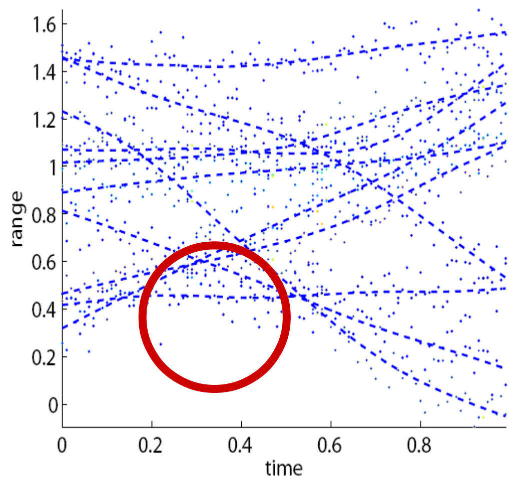
$$\theta_k \triangleq \left[x_k(0), u_k(1), \ldots, u_k(T)\right]$$

p (obs | { obs in cluster k}) =

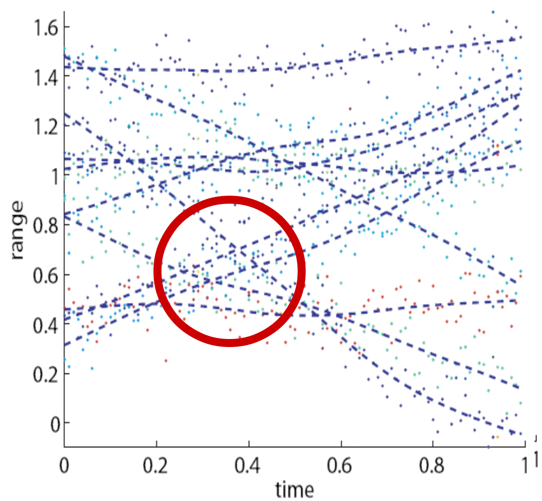
p (obs | { smoothed track k})

$k \triangleq$ track id $u \sim \mathcal{N}(0, \Lambda_u)$ $w \sim \mathcal{N}(0, \Lambda_w)$ A, B, C : fixed

Assessment



true tracks



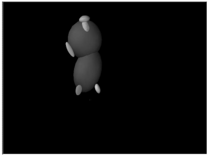
20,000 iterations

7th most frequent assignment

Pfinder

(a)

(b)



(c)

Fig. 1. (a) Video input (n.b. color image, shown here in grayscale). (b) segmentation. (c) A 2D representation of the blob statistics.

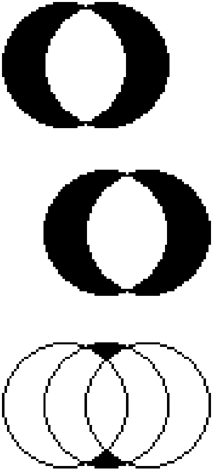
[C.R. Wren et al., "Pfinder: Real-Time Tracking of the Human Body," PAMI 1997]

Fig. 3. (a) Chris Wren playing with Bruce Blumberg's virtual dog in the ALIVE space. (b) Playing SURVIVE. (c) Real-time reading of American Sign Language (with Thad Starner doing the signing). (d) Trevor Darrell demonstrating vision-driven avatars.

[C.R. Wren et al., "Pfinder: Real-Time Tracking of the Human Body," PAMI 1997]

Background Modeling

Challenges of Finding Moving Regions



[K. Toyama et al., "Wallflower: Principles and Practice of Background Maintenance," ICCV 1999]

Pixel-Level Model

Wiener filter

$$\tilde{s}_t = -\sum_{k=1}^{p} a_k \, s_{t-k}$$

$$E[e_t^2] = E[s_t^2] + \sum_{k=1}^{p} a_k \, E[s_t \, s_{t-k}]$$

[K. Toyama et al., "Wallflower: Principles and Practice of Background Maintenance," ICCV 1999]

Region Level

As each new pair of raw and foreground-marked images, I_t and F_t, arrives,

1. Compute image differences (Figures 1a and b):

$$J_t(\mathbf{x}) = \begin{cases} 1, & \text{if } |I_t(\mathbf{x}) - I_{t-1}(\mathbf{x})| > k_{\text{motion}}, \\ 0, & \text{otherwise.} \end{cases}$$

2. Compute the subset of pixels which occur at the intersection of adjacent pairs of differenced images [1] and the previous foreground image (Figure 1c):

$$K_t(\mathbf{x}) = J_t(\mathbf{x}) \wedge J_{t-1}(\mathbf{x}) \wedge F_{t-1}(\mathbf{x}).$$

3. Find 4-connected regions, $\mathbf{R}_i$, in K_t, discarding regions consisting of less than $k_{\min}$ pixels [2].
4. Compute H_i, the normalized histogram of each $\mathbf{R}_i$, as projected onto the image I_{t-1} (s is a pixel value):

$$H_i(s) = \frac{|\{\mathbf{x} : \mathbf{x} \in \mathbf{R}_i \text{ and } I_{t-1}(\mathbf{x}) = s\}|}{|\mathbf{R}_i|}$$

5. Backproject histograms in I_t: For each $\mathbf{R}_i$, compute $F_{t-1} \wedge \mathbf{R}_i$, and from each point in the intersection, grow L_i, the 4-connected regions in the image,

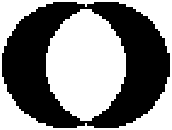
$$L_i(\mathbf{x}) = \begin{cases} 1, & \text{if } H_i(I_t(\mathbf{x})) > \varepsilon, \\ 0, & \text{otherwise.} \end{cases}$$

where we use $k_{\text{motion}} = 16$, $k_{\min} = 8$, $\varepsilon = 0.1$.

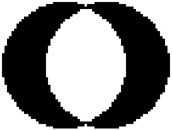
[K. Toyama et al., "Wallflower: Principles and Practice of Background Maintenance," ICCV 1999]

Challenges of Finding Moving Regions

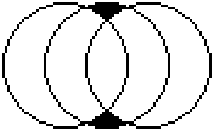
(a) A homogeneous disk moves to the right. Change is visible in the black regions only (J_{t-1} in text).



(b) The same thing happens one frame later (J_t).

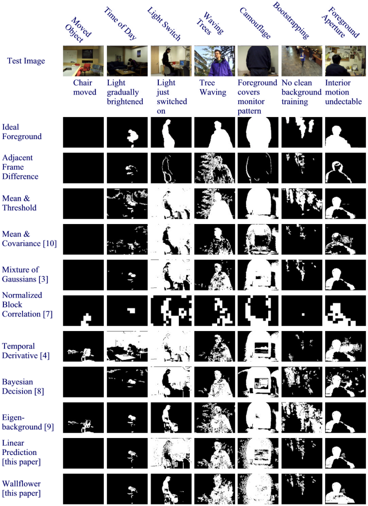


(c) Only the intersection ($J_{t-1} \wedge J_t$) is certain to be foreground in the middle image.



[K. Toyama et al., "Wallflower: Principles and Practice of Background Maintenance," ICCV 1999]

Comparison to Prior Methods



[K. Toyama et al., "Wallflower: Principles and Practice of Background Maintenance," ICCV 1999]

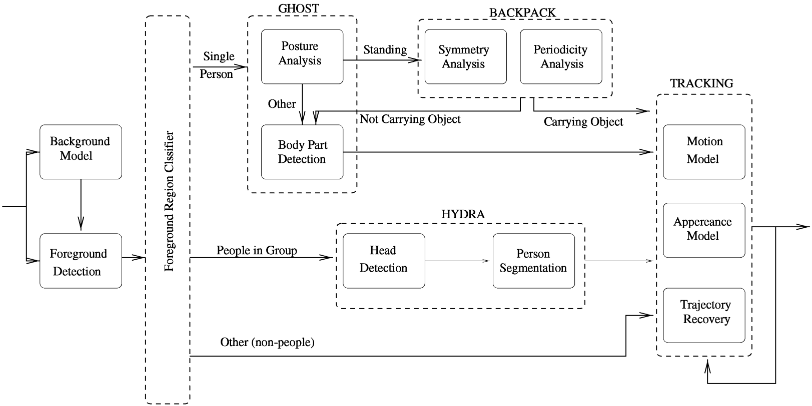
W^4 System

(a)

(b)

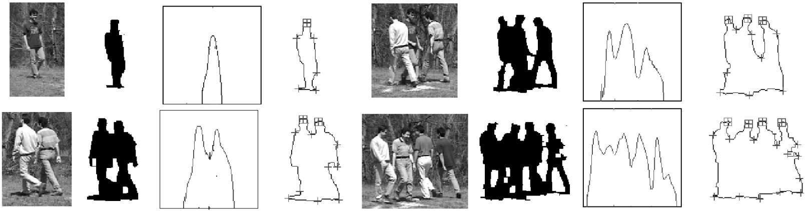
[I. Haritaoglu, D. Harwood, and L.S. Davis, "W^4: Real-Time Surveillance of People and Their Activities," PAMI 2000]

W^4 System



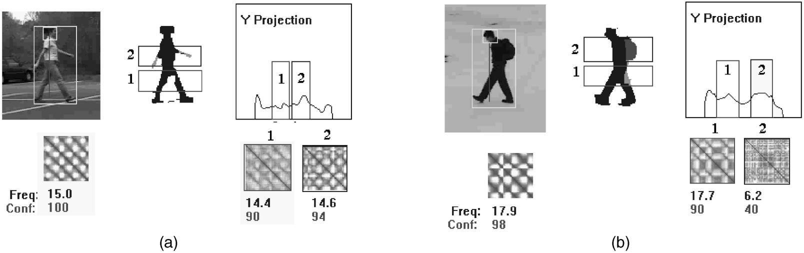
[I. Haritaoglu, D. Harwood, and L.S. Davis, "W^4: Real-Time Surveillance of People and Their Activities," PAMI 2000]

W^4 System



[I. Haritaoglu, D. Harwood, and L.S. Davis, "W^4: Real-Time Surveillance of People and Their Activities," PAMI 2000]

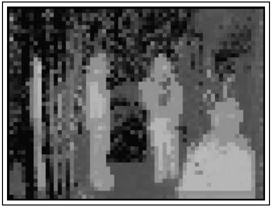
W^4 System

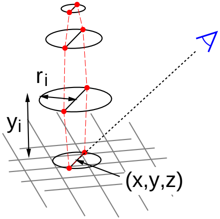


[I. Haritaoglu, D. Harwood, and L.S. Davis, "W^4: Real-Time Surveillance of People and Their Activities," PAMI 2000]

Bayesian Methods for Multi-Target Tracking

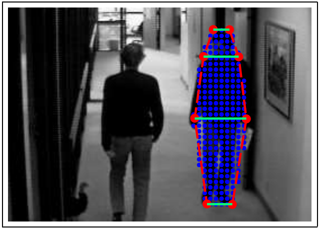
BraMBLe





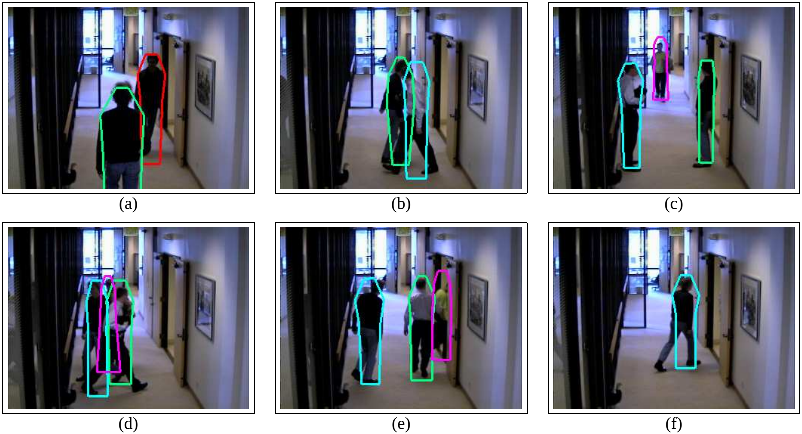
[M. Isard and J. MacCormick, "BraMBLe: A Bayesian Multiple-Blob Tracker," ICCV 2001]

BraMBLe



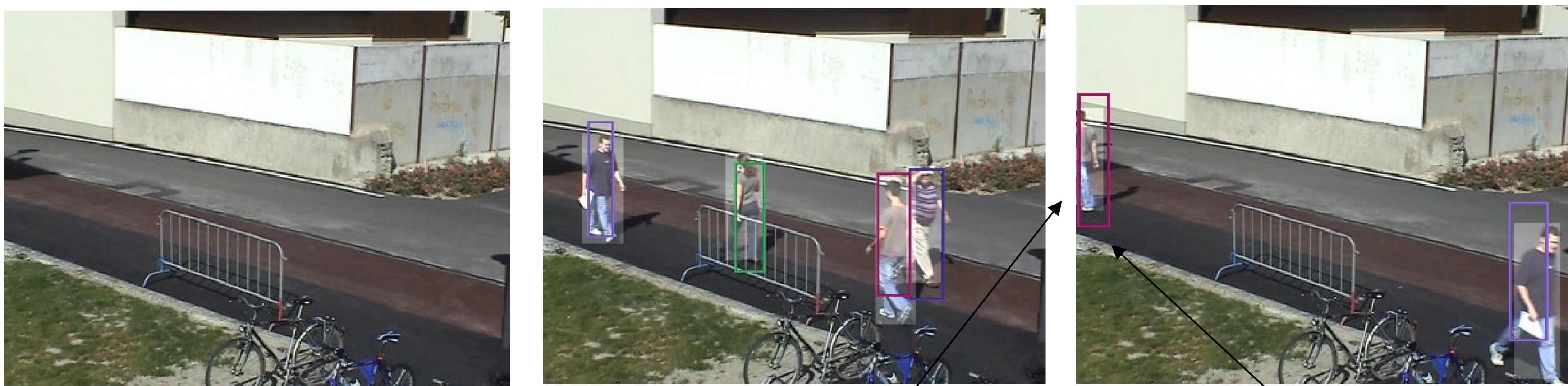
[M. Isard and J. MacCormick, "BraMBLe: A Bayesian Multiple-Blob Tracker," ICCV 2001]

BraMBLe



[M. Isard and J. MacCormick, “BraMBLe: A Bayesian Multiple-Blob Tracker,” ICCV 2001]

Using Particles to Track Varying Numbers of Interacting People



textured background

variable num targets

appearance (size) change

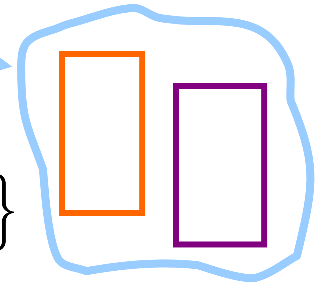
identity ambiguity (similar pants)

K. Smith, D. Gatica-Perez, and J-M. Odobez, “Using Particles to Track Varying Numbers of Interacting People,” CVPR 2005

State and Observation Model

Multi-object configuration

$$X_t = \{X_{1,t}, \ldots, X_{K,t}\}$$



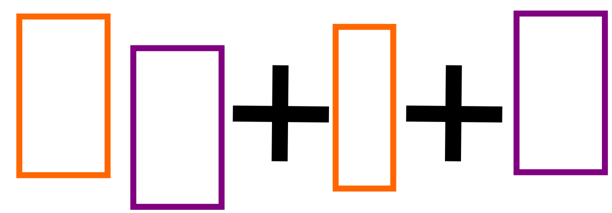
multi-obj transition

Filtering distribution

$$p(X_t \mid Z_{1:t}) \propto p(Z_t \mid X_t) \times \int_{X_{t-1}} p(X_t \mid X_{t-1})\, p(X_{t-1} \mid Z_{1:t-1})\, dX_{t-1}$$

Approximation by samples

$$p(X_t \mid Z_{1:t}) \approx \sum_{n=1}^{N} \delta_n \left(X_t, X_t^{(n)}\right)$$



Observation Z_t

K. Smith, D. Gatica-Perez, and J-M. Odobez, "Using Particles to Track Varying Numbers of Interacting People," CVPR 2005

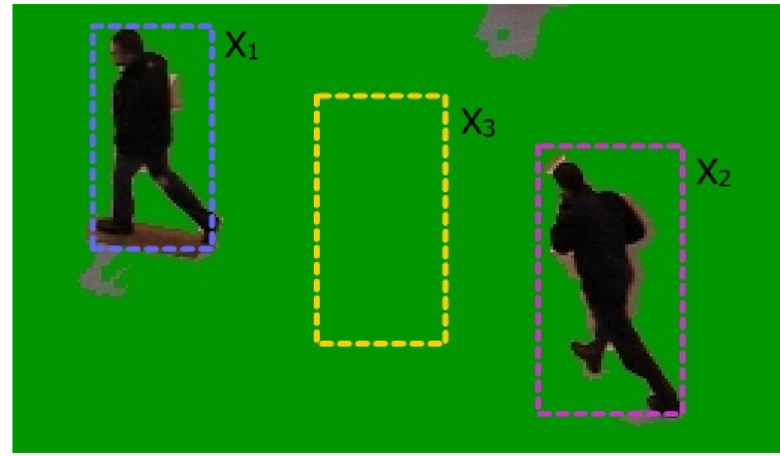
Global Binary Observation Model

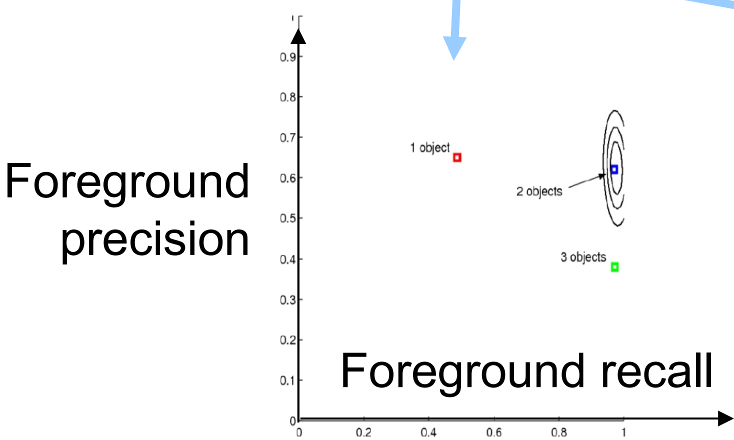
wrong:

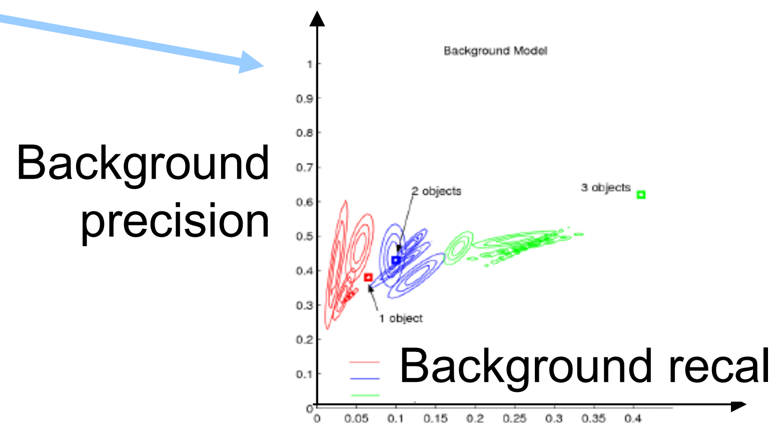
$$p\left(Z_t \mid X_t\right) = \prod_{k=1}^{K} p\left(Z_t \mid X_{k,t}\right)$$

reasonable:

$$p\left(Z_t \mid X_t\right) \triangleq p(Z_t^{\mathcal{F}} \mid X_t)\, p(Z_t^{\mathcal{B}} \mid X_t)$$







K. Smith, D. Gatica-Perez, and J-M. Odobez, “Using Particles to Track Varying Numbers of Interacting People,” CVPR 2005

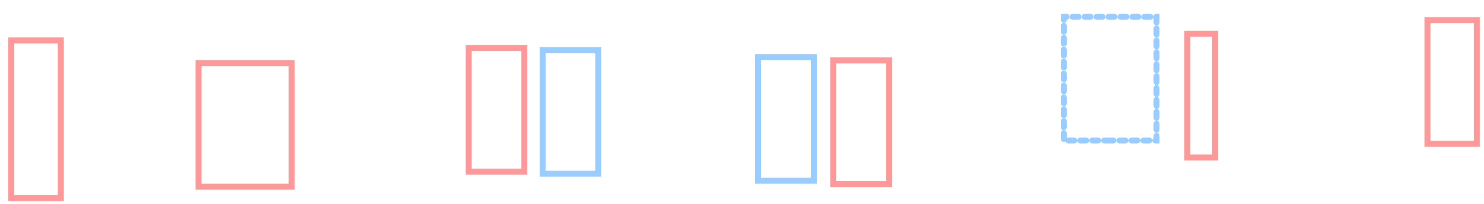
$p(X_t \mid Z_{1:t})$ Approximation via RJ MCMC

Samples within Markov chain for a given frame

frame $t-1$ $\quad X_{t-1}^{(1)} \longrightarrow X_{t-1}^{(2)} \longrightarrow \quad X_{t-1}^{(n)} \quad \longrightarrow X_{t-1}^{(N)}$

frame t $\quad X_{t}^{(1)} \longrightarrow X_{t}^{(2)} \longrightarrow \quad X_{t}^{(n)} \quad \longrightarrow X_{t}^{(N)}$

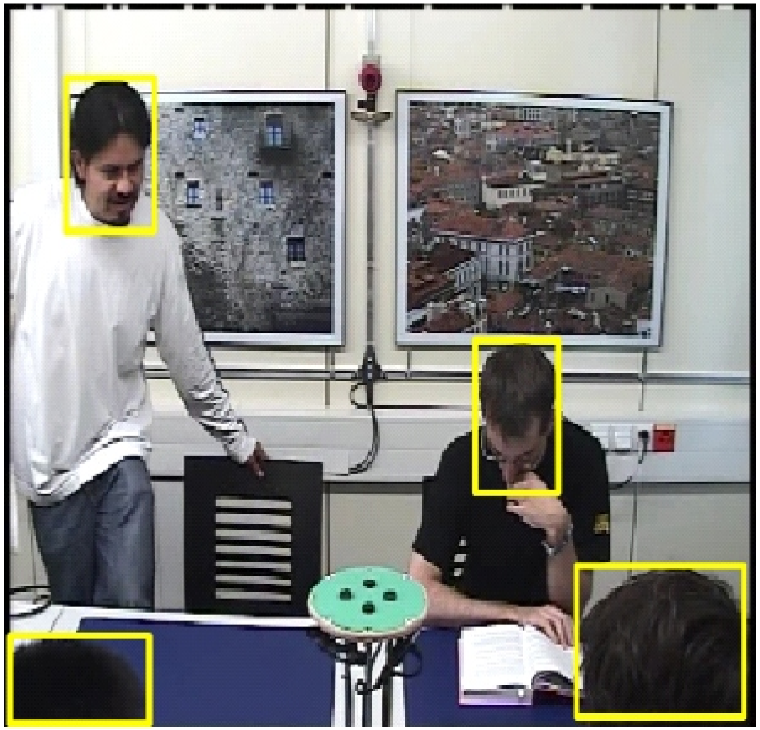
RJ MCMC moves: update birth swap death



$X_{1,t}^{(1)} \quad X_{1,t}^{(2)} \quad X_{1,t}^{(n)}, X_{2,t}^{(n)} \quad X_{1,t}^{(n+1)}, X_{2,t}^{(n+1)} \quad X_{1,t}^{(n+2)} \quad X_{1,t}^{(N)}$

K. Smith, D. Gatica-Perez, and J-M. Odobez, “Using Particles to Track Varying Numbers of Interacting People,” CVPR 2005

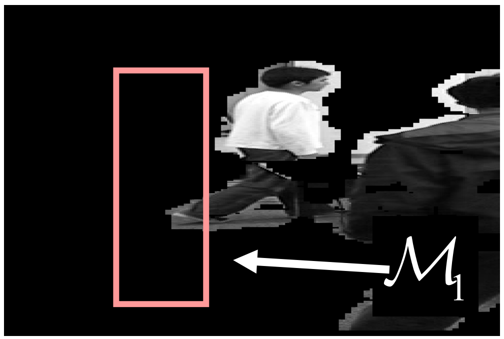
Assessment

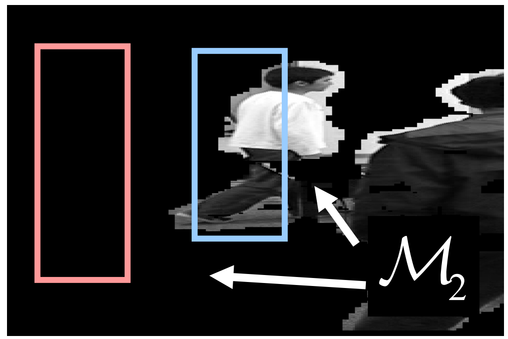


ranked 2nd compared with
KLT,
Active Shapes,
face detector

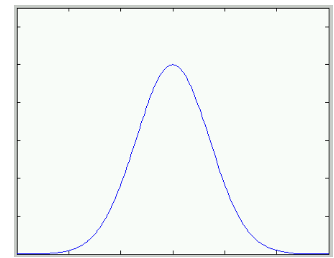
K. Smith, D. Gatica-Perez, and J-M. Odobez, “Using Particles to Track Varying Numbers of Interacting People,” CVPR 2005

RJ MCMC Model Selection (1)

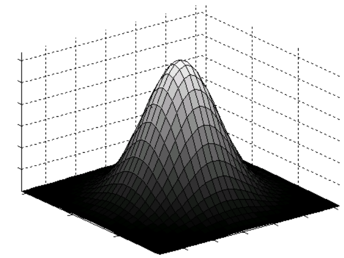




$\mathcal{M}_1$



$\mathcal{M}_2$



$$p(m, x_m \mid Z),\ x_m \in \mathcal{X}_m$$

$$\mathcal{M}_m\ ,\ m = 1, \ldots, M$$

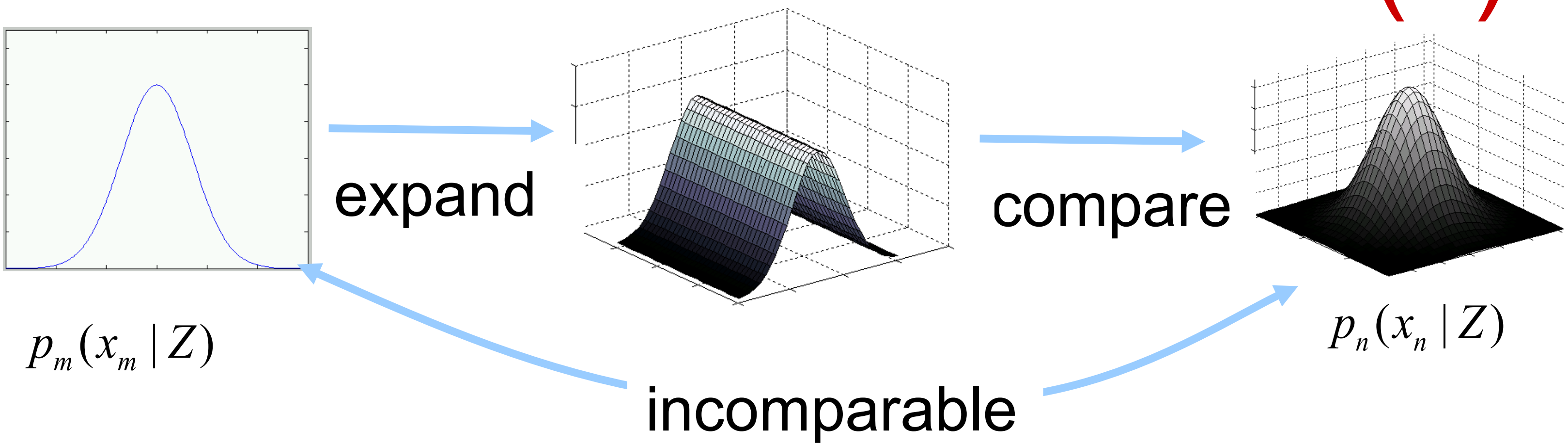
$$\{1, \ldots, M\} \times \prod_{m=1}^{M} \mathcal{X}_m \quad \text{vs.} \quad \bigcup_{m=1}^{M} \{m\} \times \mathcal{X}_m$$

product space

RJ MCMC

Carlin & Chib, Bayesian model choice via MCMC, Journal of the Royal Statistical Society 1995

RJ MCMC Model Selection (II)



$$\mathcal{X}_{n,m} \triangleq \mathcal{X}_n \times \mathcal{U}_{n,m}$$
$$\mathcal{X}_{m,n} \triangleq \mathcal{X}_m \times \mathcal{U}_{m,n}$$

$$f_{n\to m}\left(f_{m\to n}\left(x_m, u_{m,n}\right)\right) = \left(x_m, u_{m,n}\right)$$

deterministic

$$u_{n,m} \sim q_{n\to m}\left(\cdot \mid n, x_n\right)$$
$$u_{m,n} \sim q_{m\to n}\left(\cdot \mid m, x_m\right)$$

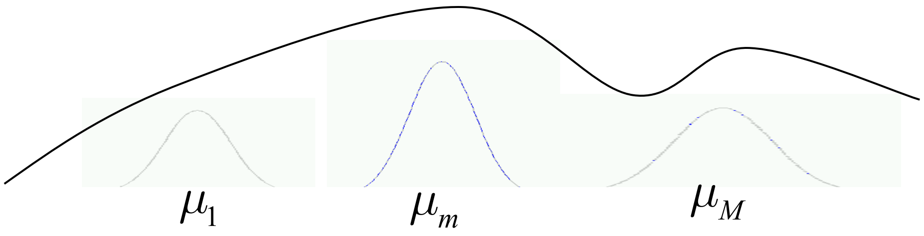
$$\mathcal{A}_{n\to m} = \min\left\{1, \frac{p\left(m, x_m^*\right)}{p\left(n, x_n\right)} \times \frac{q\left(n \mid m\right)}{q\left(m \mid n\right)} \times \frac{q_{m\to n}\left(u_{m,n} \mid m, x_m^*\right)}{q_{n\to m}\left(u_{n,m} \mid n, x_n\right)} \times \left|J_{f_{n\to m}}\right|\right\}$$

$u_{m,n} \triangleq$ auxiliary var. with proposal density $q_{m\to n}(\cdot \mid \cdot)$, $\left|J_f\right| \triangleq$ det. of Jacobian

RJ MCMC Example

RBF regression

how many kernels?



RJ MCMC moves: merge split birth death

$$\mu = \frac{\mu_1 + \mu_2}{2}$$

$$q(k \mid k+1) = \frac{1}{k+1}$$

$$\mu_1 = \mu - \beta u_{n,m} \qquad \mu_2 = \mu + \beta u_{n,m} \qquad u_{n,m} \sim \mathcal{U}_{[0,1]}$$

$$q(k+1 \mid k) = \frac{1}{k} \qquad J_{split} = \begin{bmatrix} 1 & 1 \\ -\beta & \beta \end{bmatrix}$$