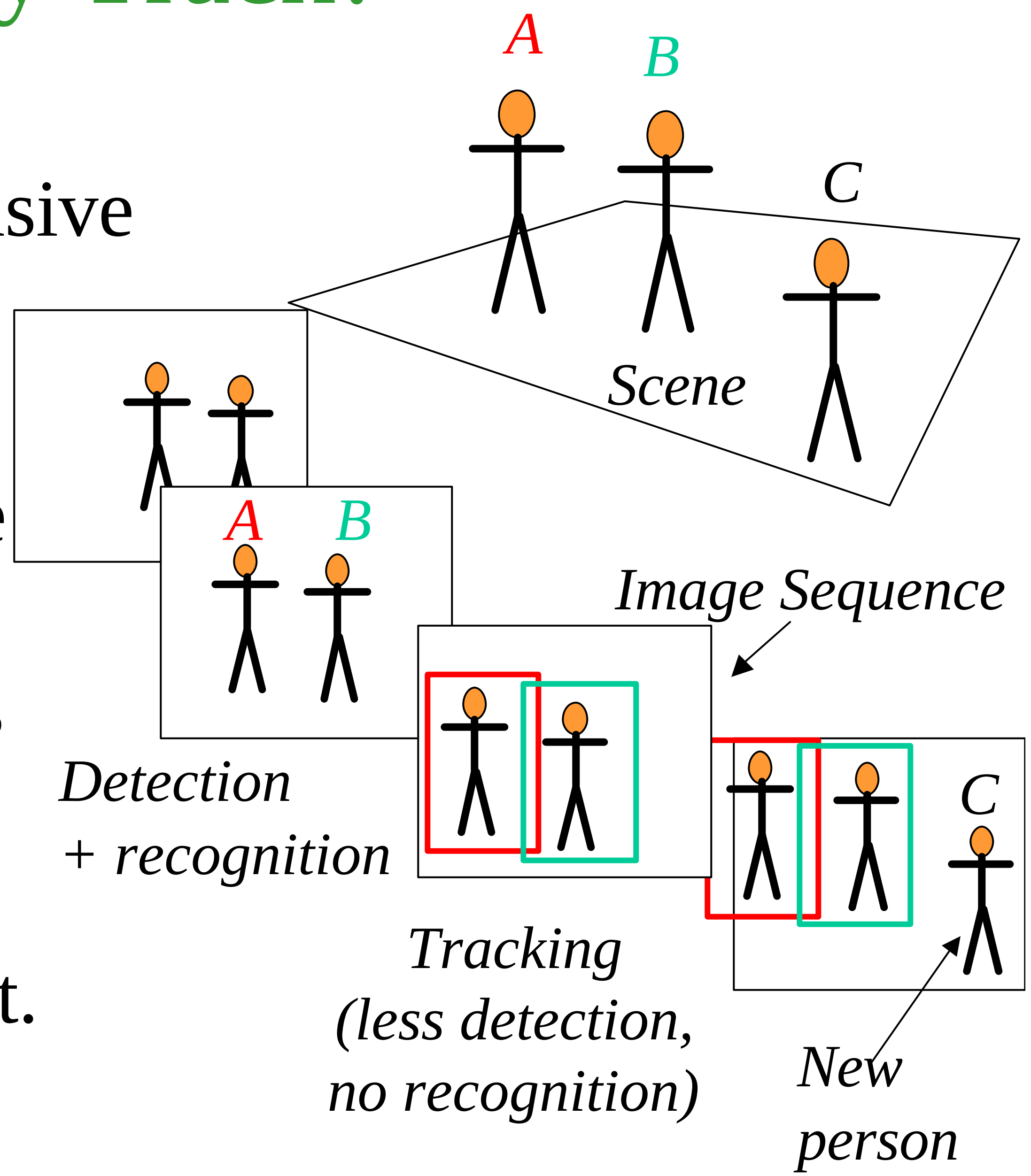
Tracking

Definition of Tracking

- Tracking:
  - Generate some conclusions about the motion of the scene, objects, or the camera, given a sequence of images.
  - Knowing this motion, predict where things are going to project in the next image, so that we don’t have so much work looking for them.

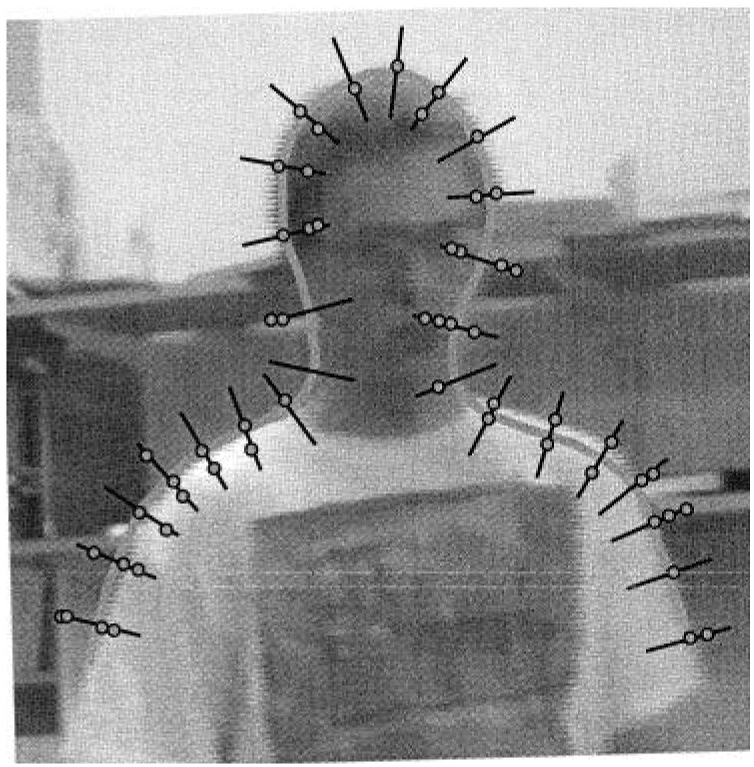
Why Track?

- Detection and recognition are expensive
- If we get an idea of where an object is in the image because we have an idea of the motion from previous images, we need less work detecting or recognizing the object.



Tracking a Silhouette by Measuring Edge Positions

- Observations are positions of edges along normals to tracked contour



Why not Wait and Process the Set of Images as a Batch?

- In a car system, detecting and tracking pedestrians in real time is important.
- Recursive methods require less computing

Implicit Assumptions of Tracking

- Physical cameras do not move instantly from a viewpoint to another
- Object do not teleport between places around the scene
- Relative position between camera and scene changes incrementally
- We can model motion

Related Fields

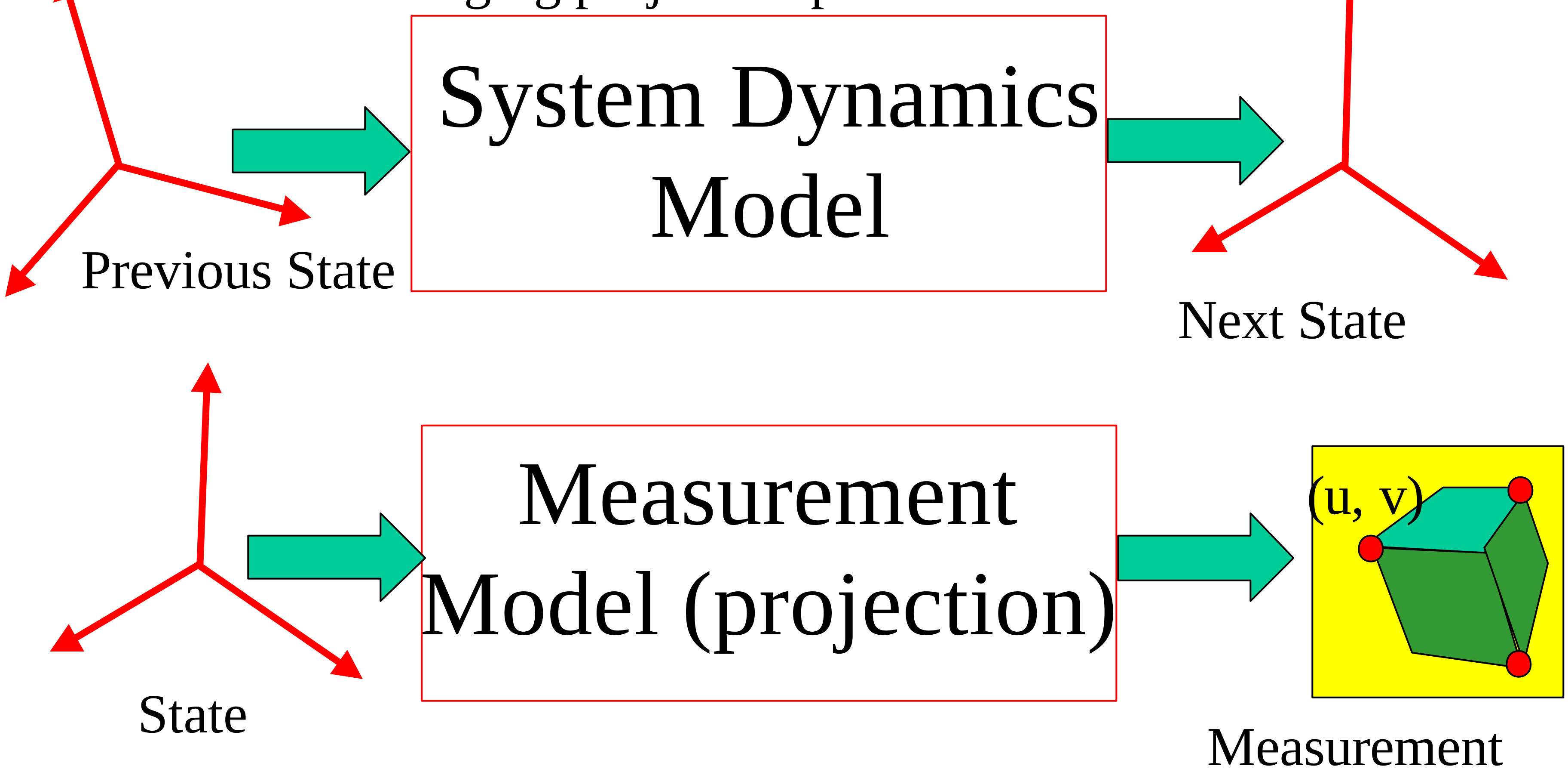
- Signal Detection and Estimation
- Radar technology

The Problem: Signal Estimation

- We have a system with parameters
  – Scene structure, camera motion, automatic zoom
  – System state is unknown (“hidden”)
- We have measurements
  – Components of stable “feature points” in the images.
  – “Observations”, projections of the state.
- We want to recover the state components from the observations

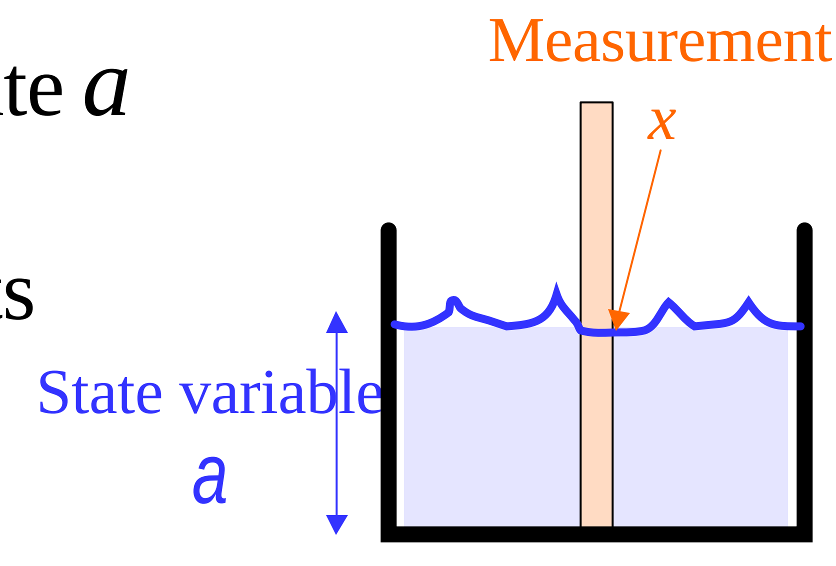
Necessary Models

- We use models to describe a priori knowledge about
  - the world (including external parameters of camera)
  - the imaging projection process



A Simple Example of Estimation by Least Square Method

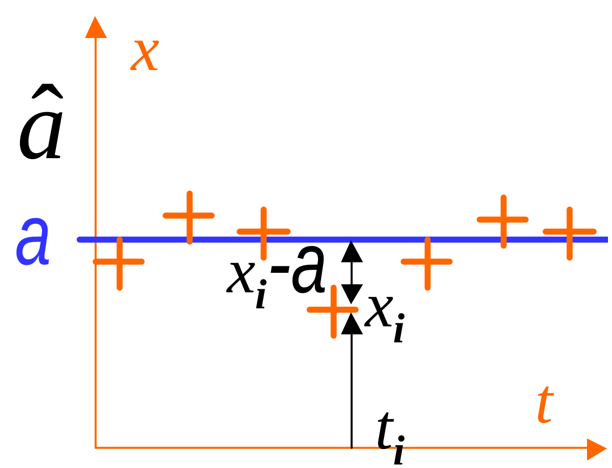
- Goal: Find estimate $\hat{a}$ of state a such that the least square error between measurements and the state is minimum



$$C = \frac{1}{2}\sum_{i=1}^{n}(x_i - a)^2$$

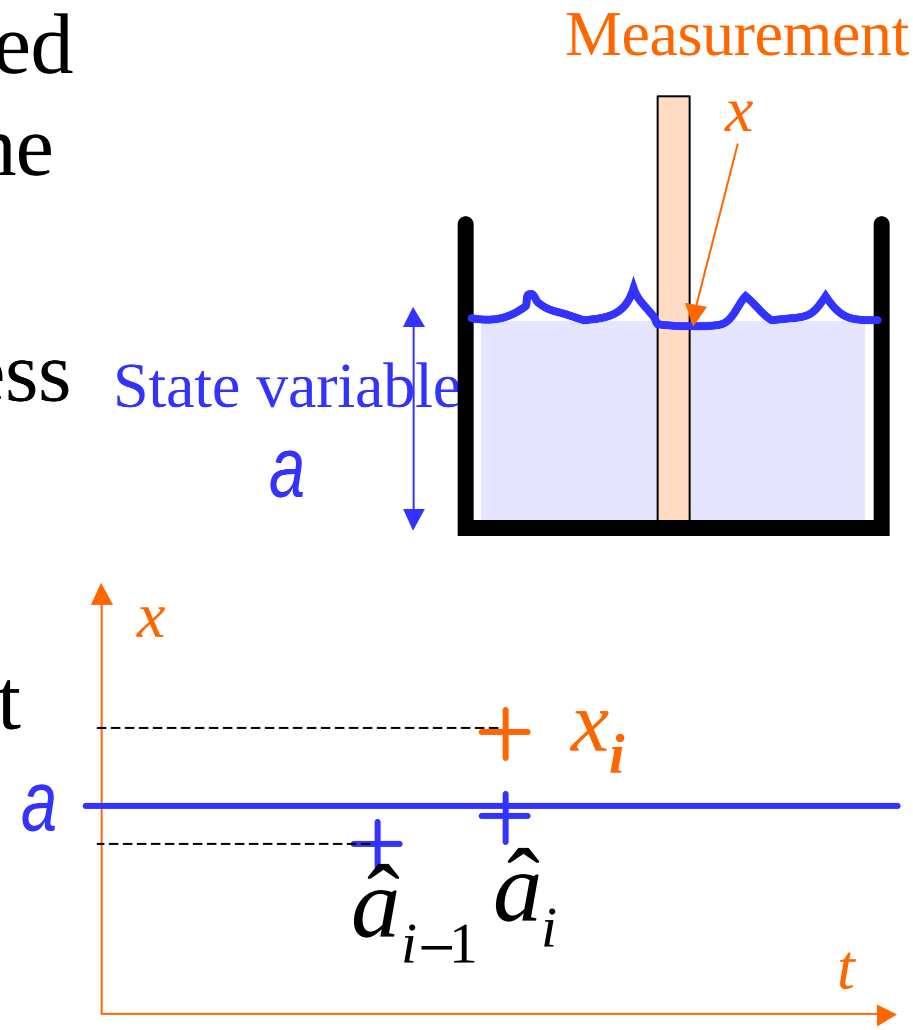
$$\frac{\partial C}{\partial a} = 0 = \sum_{i=1}^{n}(x_i - \hat{a}) = \sum_{i=1}^{n} x_i - n\,\hat{a}$$

$$\hat{a} = \frac{1}{n}\sum_{i=1}^{n} x_i$$



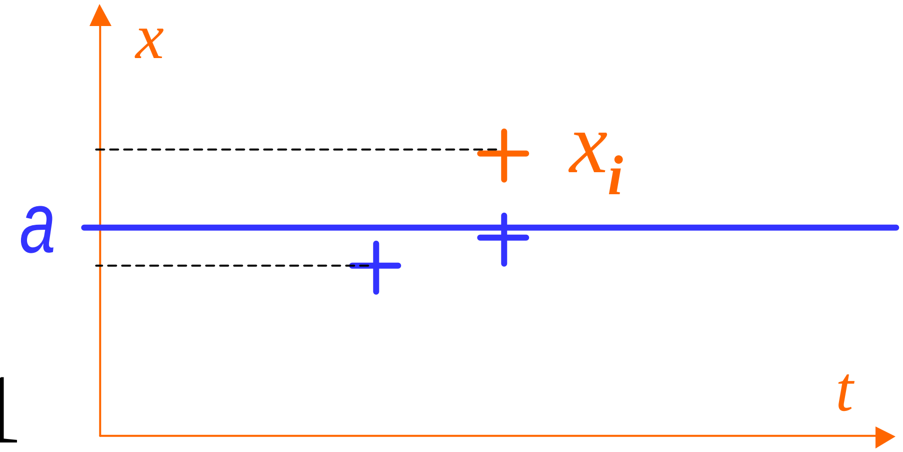
Recursive Least Square Estimation

- We don't want to wait until all data have been collected to get an estimate $\hat{a}$ of the depth
- We don't want to reprocess old data when we make a new measurement
- Recursive method: data at step *i* are obtained from data at step i-1



Recursive Least Square Estimation 2

- Recursive method: data at step i are obtained from data at step $i-1$



$$\hat{a}_i = \frac{1}{i}\sum_{k=1}^{i} x_k = \frac{1}{i}\sum_{k=1}^{i-1} x_k + \frac{1}{i} x_i$$

$$\hat{a}_{i-1} = \frac{1}{i-1}\sum_{k=1}^{i-1} x_k$$

$$\hat{a}_i = \frac{i-1}{i}\hat{a}_{i-1} + \frac{1}{i} x_i$$

$$\boxed{\hat{a}_i = \hat{a}_{i-1} + \frac{1}{i}(x_i - \hat{a}_{i-1})}$$

Recursive Least Square Estimation 3

Gain

Actual measure

Predicted measure

$$\hat{a}_i = \hat{a}_{i-1} + \frac{1}{i} \underbrace{\left(x_i - \hat{a}_{i-1} \right)}_{\text{Innovation}}$$

Estimate at step *i*

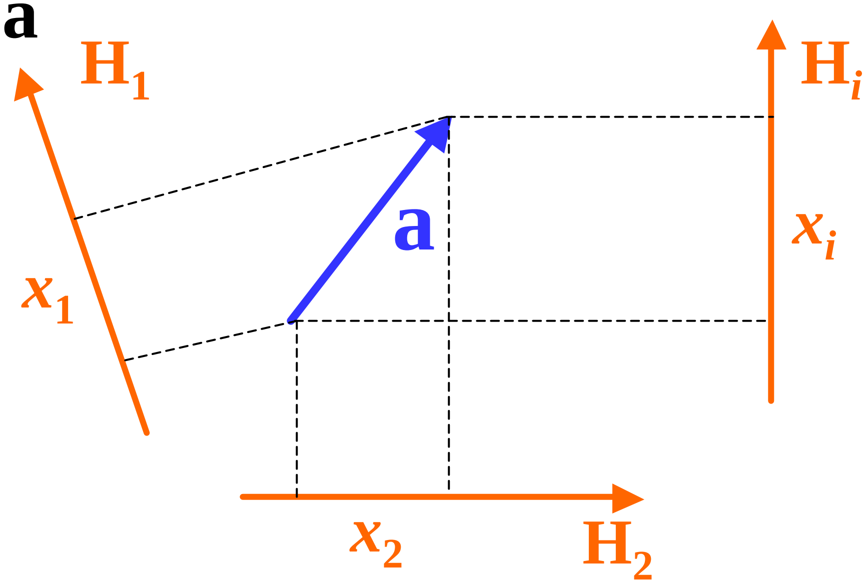
Gain specifies how much
do we pay attention
to the difference
between what we expected
and what we actually get

Least Square Estimation of the State Vector of a Static System

1. Batch method

$$x_i = \mathbf{H}_i \cdot \mathbf{a} \quad \Rightarrow \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix} = \begin{bmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \\ \cdots \\ \mathbf{H}_n \end{bmatrix} \mathbf{a}$$

measurement equation



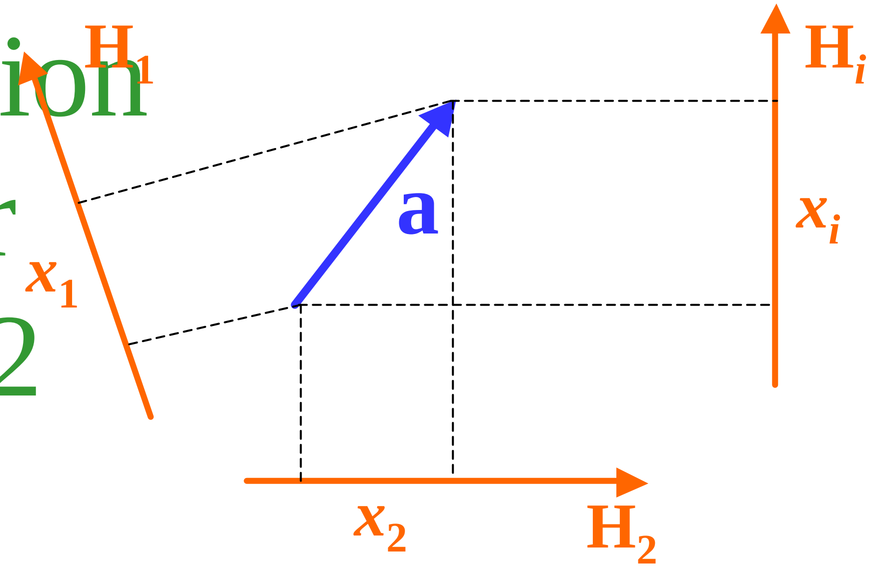
$$\Rightarrow \mathbf{x} = \mathbf{H}\,\mathbf{a}$$

Find estimate $\hat{\mathbf{a}}$ that minimizes

$$C = \frac{1}{2}(\mathbf{X} - \mathbf{H}\,\mathbf{a})^{\mathrm{T}}(\mathbf{X} - \mathbf{H}\,\mathbf{a})$$

We find $\hat{\mathbf{a}} = (\mathbf{H}^{\mathrm{T}}\mathbf{H})^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{X}$

Least Square Estimation of the State Vector of a Static System 2



2. Recursive method

Calculation is similar to calculation of running average

We had: $\hat{a}_i = \hat{a}_{i-1} + \frac{1}{i}(x_i - \hat{a}_{i-1})$

Predicted measure

Now we find: $\mathbf{\hat{a}_i} = \mathbf{\hat{a}_{i\text{-}1}} + \mathbf{K_i}(\mathbf{X_i} - \mathbf{H_i}\mathbf{\hat{a}_{i\text{-}1}})$

Gain matrix

Innovation

with $\mathbf{K_i} = \mathbf{P_i}\,\mathbf{H_i}^{\mathbf{T}}$

$$\mathbf{P_i} = (\mathbf{H_i}\,\mathbf{H_i}^{\mathbf{T}})^{-1}$$

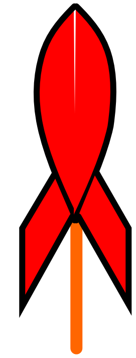
Dynamic System

$$X_i = X_{i-1} + V_{i-1}\,\Delta t$$

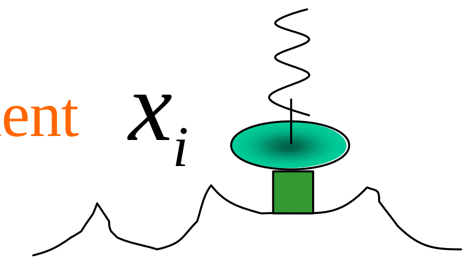
$$V_i = V_{i-1} + A_{i-1}\,\Delta t$$

$$A_i = A_{i-1} + w$$

State of rocket $\mathbf{a}_i \left| \begin{matrix} X_i \\ V_i \\ A_i \end{matrix} \right.$



Measurement x_i



$$\begin{bmatrix} X_i \\ V_i \\ A_i \end{bmatrix} = \begin{bmatrix} 1 & \Delta t & 0 \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_{i-1} \\ V_{i-1} \\ A_{i-1} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ w \end{bmatrix} \Rightarrow \mathbf{a}_\mathbf{i} = \Phi\, \mathbf{a}_\mathbf{i-1} + \mathbf{w}$$

Tweak factor

State equation for rocket

$$x_i = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} X_i \\ V_i \\ A_i \end{bmatrix} + V \Rightarrow x_i = \mathbf{H}\,\mathbf{a}_\mathbf{i} + V$$

Noise

Measurement equation

Recursive Least Square Estimation for a Dynamic System (Kalman Filter)

State equation

$$\mathbf{a}_i = \Phi_i \, \mathbf{a}_{i-1} + \mathbf{w}_{i-1}$$

Tweak factor for model

$$\mathbf{w}_i \sim N(0, \mathbf{Q}_i)$$

Measurement equation

$$\mathbf{x}_i = \mathbf{H}_i \, \mathbf{a}_i + \mathbf{v}_i$$

Measurement noise

$$\mathbf{v}_i \sim N(0, \mathbf{R}_i)$$

Prediction for $\mathbf{a}_i$

Prediction for $\mathbf{x}_i$

$$\hat{\mathbf{a}}_i = \Phi_i \, \hat{\mathbf{a}}_{i-1} + \mathbf{K}_i(\mathbf{x}_i - \mathbf{H}_i \Phi_i \hat{\mathbf{a}}_{i-1})$$

$$\mathbf{K}_i = \mathbf{P}'_i \, \mathbf{H}_i^{T}(\mathbf{H}_i \, \mathbf{P}'_i \, \mathbf{H}_i^{T} + \mathbf{R}_i)^{-1}$$

Gain

$$\mathbf{P}'_i = \Phi_i \mathbf{P}_{i-1} \Phi_i^{T} + \mathbf{Q}_{i-1}$$

Covariance matrix for prediction error

$$\mathbf{P}_{i-1} = (\mathbf{I} - \mathbf{K}_{i-1}\mathbf{H}_{i-1})\mathbf{P}'_{i-1}$$

Covariance for estimation error

Estimation when System Model is Nonlinear (Extended Kalman Filter)

State equation

$$\mathbf{a}_i = \mathbf{f}(\mathbf{a}_{i\text{-}1}) + \mathbf{w}_{i\text{-}1}$$

Measurement equation

$$\mathbf{x}_i = \mathbf{H}_i \mathbf{a}_i + \mathbf{v}_i$$

$$\hat{\mathbf{a}}_i = \mathbf{f}(\hat{\mathbf{a}}_{i\text{-}1}) + \mathbf{K}_i(\mathbf{x}_i - \mathbf{H}_i \mathbf{f}(\hat{\mathbf{a}}_{i\text{-}1}))$$

$$\mathbf{K}_i = \mathbf{P'}_i \mathbf{H}_i^{\mathbf{T}} (\mathbf{H}_i \mathbf{P'}_i \mathbf{H}_i^{\mathbf{T}} + \mathbf{R}_i)^{\mathbf{-1}}$$

$$\mathbf{P'}_i = \frac{\partial \mathbf{f}}{\partial \mathbf{a}_i} \mathbf{P}_{i\text{-}1} \left(\frac{\partial \mathbf{f}}{\partial \mathbf{a}_i}\right)^{\mathbf{T}} + \mathbf{Q}_{i\text{-}1}$$

Jacobian Matrix

Differences compared to regular Kalman filter are circled in red

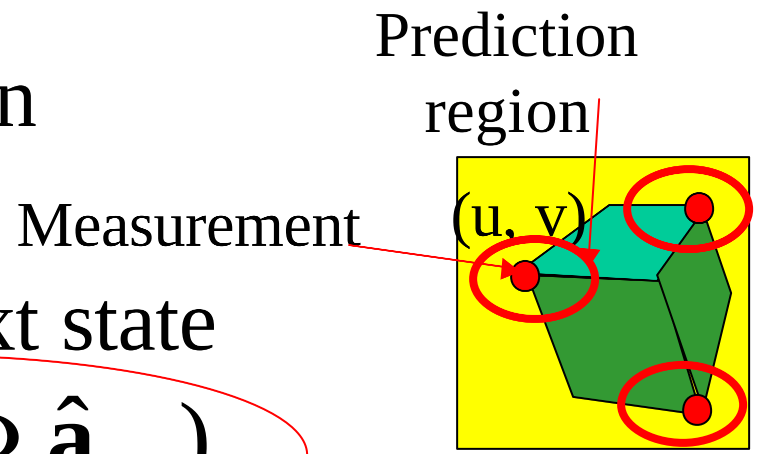
$$\mathbf{P}_{i\text{-}1} = (\mathbf{I} - \mathbf{K}_{i\text{-}1} \mathbf{H}_{i\text{-}1}) \mathbf{P'}_{i\text{-}1}$$

Tracking Steps

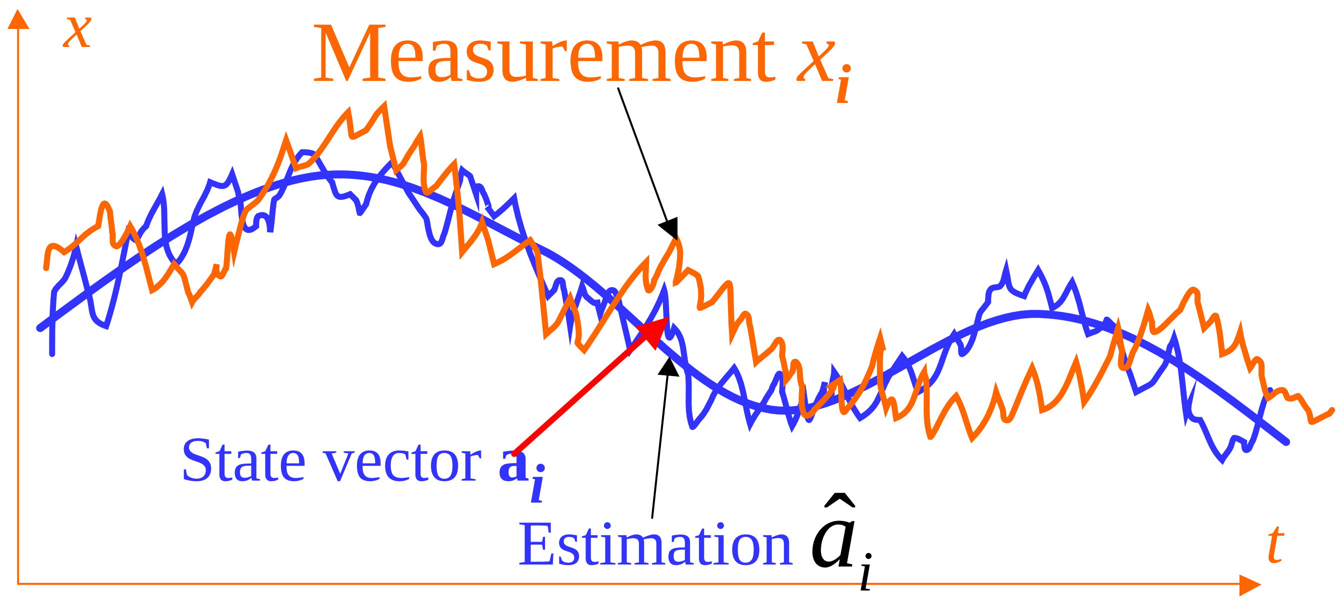
- Predict next state as $\Phi_i \hat{\mathbf{a}}_{i-1}$ using previous step and dynamic model
- Predict regions $N(\mathbf{H}_i \Phi_i \hat{\mathbf{a}}_{i-1}, \mathbf{P}'_i)$ of next measurements using measurement model and uncertainties
- Make new measurements $\mathbf{x}_i$ in predicted regions
- Compute best estimate of next state

$$\hat{\mathbf{a}}_i = \Phi_i \hat{\mathbf{a}}_{i-1} + \mathbf{K}_i(\mathbf{x}_i - \mathbf{H}_i \Phi_i \hat{\mathbf{a}}_{i-1})$$

"Correction" of predicted state



Recursive Least Square Estimation for a Dynamic System (Kalman Filter)



Tracking as a Probabilistic Inference Problem

- Find distributions for state vector $\mathbf{a_i}$ and for measurement vector $\mathbf{x_i}$. Then we are able to compute the expectations $\hat{\mathbf{a}}_\mathbf{i}$ and $\hat{\mathbf{x}}_\mathbf{i}$
- Simplifying assumptions (same as for HMM)

$$P(\mathbf{a_i} \mid \mathbf{a_1}, \mathbf{a_2}, \cdots, \mathbf{a_{i\text{-}1}}) = P(\mathbf{a_i} \mid \mathbf{a_{i\text{-}1}})$$

(Only immediate past matters)

$$P(\mathbf{x_i}, \mathbf{x_j}, \ldots, \mid \mathbf{a_i}) = P(\mathbf{x_i} \mid \mathbf{a_i}) P(\mathbf{x_j} \mid \mathbf{a_i}) \ldots P(\mathbf{x_k} \mid \mathbf{a_i})$$

(Conditional independence of measurements given a state)

Tracking as Inference

- Prediction

$$P(\mathbf{a}_{\mathrm{i}} \mid \mathbf{x}_1, \cdots, \mathbf{x}_{\mathrm{i-1}}) = \int P(\mathbf{a}_{\mathrm{i}} \mid \mathbf{a}_{\mathrm{i-1}}) P(\mathbf{a}_{\mathrm{i-1}} \mid \mathbf{x}_1, \cdots, \mathbf{x}_{\mathrm{i-1}}) d\mathbf{a}_{\mathrm{i-1}}$$

- Correction

$$P(\mathbf{a}_{\mathrm{i}} \mid \mathbf{x}_1, \cdots, \mathbf{x}_{\mathrm{i}}) = \frac{P(\mathbf{x}_{\mathrm{i}} \mid \mathbf{a}_{\mathrm{i}}) P(\mathbf{a}_{\mathrm{i}} \mid \mathbf{x}_1, \cdots, \mathbf{x}_{\mathrm{i-1}})}{\int P(\mathbf{x}_{\mathrm{i}} \mid \mathbf{a}_{\mathrm{i}}) P(\mathbf{a}_{\mathrm{i}} \mid \mathbf{x}_1, \cdots, \mathbf{x}_{\mathrm{i-1}}) d\mathbf{a}_{\mathrm{i}}}$$

- Produces same results as least square approach if distributions are Gaussians: Kalman filter
- See Forsyth and Ponce, Ch. 19

Kalman Filter for 1D Signals

State equation

$$a_i = f\, a_{i\text{-}1} + w_{i\text{-}1}$$

Tweak factor for model

$$w_i \sim N(0, q)$$

Measurement equation

$$x_i = h\, a_i + v_i$$

Measurement noise

$$v_i \sim N(0, r)$$

Prediction for a_i (*a priori* estimate)

Prediction for x_i

$$\hat{a}_i = f\, \hat{a}_{i\text{-}1} + K_i (x_i - h\, f\, \hat{a}_{i\text{-}1})$$

Gain

$$K_i = p'_i\, h (h^2\, p'_i + r)^{-1}$$

Standard deviation for prediction error

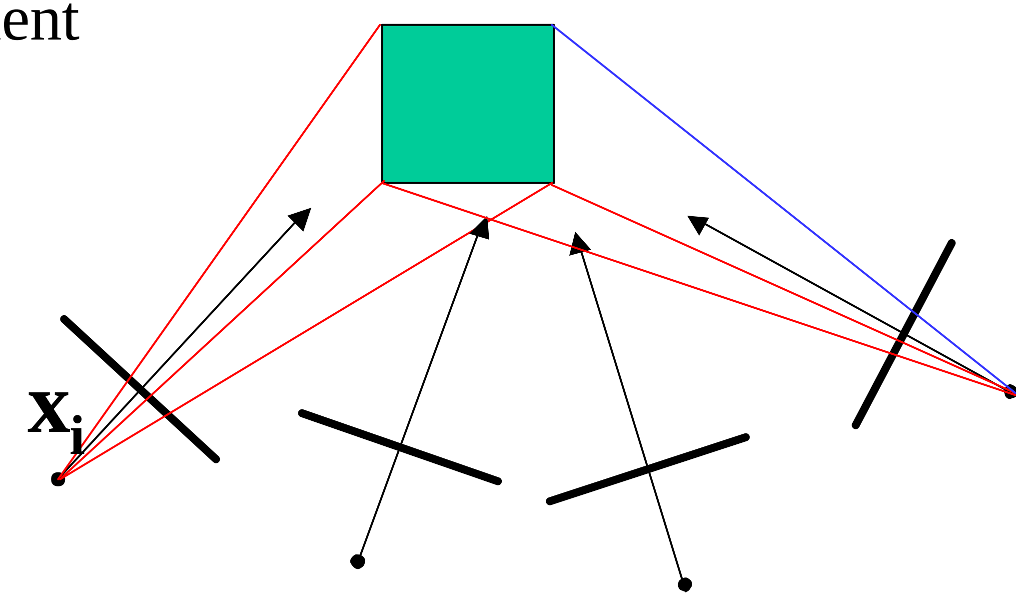
$$p'_i = f^2 p_{i-1} + q$$

St.d. for estimation error

$$p_{i\text{-}1} = (1 - K_{i-1}\, h)\, p'_{i\text{-}1}$$

Applications: Structure from Motion

- Measurement vector components:
  - Coordinates of corners, “salient points”
- State vector components:
  - Camera motion parameters
  - Scene structure
- Is there enough equations?
  - N corners, 2N measurements
  - N unknown state components from structure (distances from first center of projection to 3D points)
  - 6 unknown state components from motion (translation and rotation)
  - More measurements than unknowns for every frame if N>6 (2N > N + 6)



- Batch methods
- Recursive methods (Kalman filter)

Problems with Tracking

- Initial detection
  - If it is too slow we will never catch up
  - If it is fast, why not do detection at every frame?
    - Even if raw detection can be done in real time, tracking saves processing cycles compared to raw detection. The CPU has other things to do.
- Detection is needed again if you lose tracking
- Most vision tracking prototypes use initial detection done by hand
  (see Forsyth and Ponce for discussion)

References

- Kalman, R.E., “A New Approach to Linear Prediction Problems”, Transactions of the ASME--Journal of Basic Engineering, pp. 35-45, March 1960.
- Sorenson, H.W., “Least Squares Estimation: from Gauss to Kalman”, IEEE Spectrum, vol. 7, pp. 63-68, July 1970.
- http://www.cs.unc.edu/~welch/kalmanLinks.html
- D. Forsyth and J. Ponce. Computer Vision: A Modern Approach, Chapter 19. http://www.cs.berkeley.edu/~daf/book3chaps.html
- O. Faugeras.. Three-Dimensional Computer Vision. MIT Press. Ch. 8, “Tracking Tokens over Time”.