The Kalman Filter

D. A. Forsyth, UIUC

Linear dynamics and measurement

- State changes as:

Square matrix of full rank

$$\mathbf{x}_i = \mathcal{D}_i \mathbf{x}_{i-1} + \xi$$

This is a normal random variable with zero mean and known covariance

- Measurements are:

Any matrix whose dimensions are OK

$$\mathbf{y}_i = \mathcal{M}_i \mathbf{x}_i + \zeta$$

This is a (different!) normal random variable with zero mean and known covariance

Other notation

Read this as: x_i is normally distributed. The mean is a linear function of x_i-1 and whose variance is known (and can depend on i).

$$\mathbf{x}_i \sim N(\mathcal{D}_i \mathbf{x}_{i-1}; \Sigma_{di})$$

$$\mathbf{y}_i \sim N(\mathcal{M}_i \mathbf{x}_i; \Sigma_{mi})$$

Read this as: y_i is normally distributed. The mean is a linear function of x_i and whose variance is known (and can depend on i)

Examples

- Dynamical models
  - Drifting points
    - new state = old state + gaussian noise
  - Points moving with constant velocity
    - new position=old position + (dt) old velocity + gaussian noise
    - new velocity= old velocity+gaussian noise
  - Points moving with constant acceleration
    - etc
- Measurement models
  - state=position; measurement=position+gaussian noise
  - state=position and velocity; measurement=position+gaussian noise
    - but we could infer velocity
  - state=position and velocity and acceleration; measurement=position+gaussian noise

Key point

- $P(\mathbf{y}_i|\mathbf{x}_i)$ is normal.
- If $P(\mathbf{x}_{i-1}|\mathbf{y}_0, \ldots, \mathbf{y}_{i-1})$ is normal, then

$$P(\mathbf{x}_i|\mathbf{y}_0, \ldots, \mathbf{y}_{i-1})$$

$$P(\mathbf{x}_i|\mathbf{y}_0, \ldots, \mathbf{y}_i)$$

are both normal

Checking…

- Probability distribution is normal iff it has the form:

$$\log p(\mathbf{x}) = -\frac{1}{2}\left[(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\right] + K$$

- and you can check this for each of the relevant dists.

The Kalman Filter

- Dynamic Model

$$\mathbf{x}_i \sim N(\mathcal{D}_i \mathbf{x}_{i-1}, \Sigma_{d_i})$$

$$\mathbf{y}_i \sim N(\mathcal{M}_i \mathbf{x}_i, \Sigma_{m_i})$$

- Notation

$$\text{mean of } P(X_i|y_0, \ldots, y_{i-1}) \text{ as } \bar{X}_i^-$$

$$\text{mean of } P(X_i|y_0, \ldots, y_i) \text{ as } \bar{X}_i^+$$

$$\text{covar of } P(X_i|y_0, \ldots, y_{i-1}) \text{ as } \Sigma_i^-$$

$$\text{covar of } P(X_i|y_0, \ldots, y_i) \text{ as } \Sigma_i^+$$

Prediction

- We have:

$$\mathbf{x}_{i-1} \sim N(\bar{X}_{i-1}^{+}, \Sigma_{i-1}^{+})$$

$$\mathbf{x}_i \sim N(\mathcal{D}_i \mathbf{x}_{i-1}, \Sigma_{d_i})$$

$$\mathbf{x}_i = \mathcal{D}_i \mathbf{x}_{i-1} + \zeta$$

This is a normal random variable with zero mean and known covariance

$$\mathrm{mean}(\mathbf{x}_i) = \mathcal{D}_i \mathrm{mean}(\mathbf{x}_{i-1})$$

$$\mathrm{cov}(\mathbf{x}_i) = \mathcal{D}_i \mathrm{cov}(\mathbf{x}_{i-1}) \mathcal{D}_i^T + \mathrm{cov}(\zeta)$$

Prediction

- We have:

$$\mathbf{x}_{i-1} \sim N(\bar{X}_{i-1}^{+}, \Sigma_{i-1}^{+})$$

$$\mathbf{x}_i \sim N(\mathcal{D}_i \mathbf{x}_{i-1}, \Sigma_{d_i})$$

$$\mathbf{x}_i = \mathcal{D}_i \mathbf{x}_{i-1} + \zeta$$

This is a normal random variable with zero mean and known covariance

Which yields….

$$\overline{X}_i^{-} = \mathcal{D}_i \overline{X}_{i-1}^{+}$$

$$\Sigma_i^{-} = \Sigma_{d_i} + \mathcal{D}_i \Sigma_{i-1}^{-} \mathcal{D}_i^T$$

Useful Fact: 9.2 *The parameters of a normal posterior with a single measurement*

Assume we wish to estimate a parameter θ. The prior distribution for θ is normal, with known mean μ_π and known standard deviation σ_π. We receive a single data item x_1 and a scale c_1. The likelihood of x_1 is normal with mean $c_1\theta$ and standard deviation $\sigma_{m,1}$, where $\sigma_{m,1}$ is known. Then the posterior, $p(\theta|x_1, c_1, \sigma_{m,1}, \mu_\pi, \sigma_\pi)$, is normal, with mean

$$\mu_1 = \frac{c_1 x_1 \sigma_\pi^2 + \mu_\pi \sigma_{m,1}^2}{\sigma_{m,1}^2 + c_1^2 \sigma_\pi^2}$$

and standard deviation

$$\sigma_1 = \sqrt{\frac{\sigma_{m,1}^2 \sigma_\pi^2}{\sigma_{m,1}^2 + c^2 \sigma_\pi^2}}.$$

posterior mean is weighted combo of prior mean and measurement

posterior covar is weighted combo of prior covar, measurement matrix and measurement covar

$$\mathcal{K}_i = \Sigma_i^- \mathcal{M}_i^T \left[\mathcal{M}_i \Sigma_i^- \mathcal{M}_i^T + \Sigma_{m_i}\right]^{-1}$$

$$\bar{X}_i^+ = \bar{X}_i^- + \mathcal{K}_i \left[\mathbf{y}_i - \mathcal{M}_i \bar{X}_i^-\right]$$

posterior mean is weighted combo of prior mean and measurement

$$\Sigma_i^+ = \left[\mathcal{I} - \mathcal{K}_i \mathcal{M}_i\right] \Sigma_i^-$$

posterior covar is weighted combo of prior covar, measurement matrix and measurement covar

The steps:

Have: Mean and covariance of posterior after i-1'th measurement

Construct: Mean and covariance of predictive distribution just before i'th measurement

Measurement arrives:

Now construct: Mean and covariance of posterior distribution just before i'th measurement

posterior mean is weighted combo of prior mean and measurement

posterior covar is weighted combo of prior covar, measurement matrix and measurement covar

The steps:

Have:

$$\overline{X}^{+}_{i-1} \qquad \Sigma^{+}_{i-1}$$

Construct:

$$\overline{X}^{-}_{i} = \mathcal{D}_i \overline{X}^{+}_{i-1} \qquad \Sigma^{-}_{i} = \Sigma_{d_i} + \mathcal{D}_i \Sigma^{-}_{i-1} \mathcal{D}^{T}_{i}$$

Measurement arrives: $\mathbf{y}_i \sim N(\mathcal{M}_i \mathbf{x}_i; \Sigma_{mi})$

Now construct:

$$\bar{X}^{+}_{i} = \bar{X}^{-}_{i} + \mathcal{K}_i \left[\mathbf{y}_i - \mathcal{M}_i \bar{X}^{-}_{i} \right] \qquad \Sigma^{+}_{i} = \left[\mathcal{I} - \mathcal{K}_i \mathcal{M}_i \right] \Sigma^{-}_{i}$$

Where:

$$\mathcal{K}_i = \Sigma^{-}_{i} \mathcal{M}^{T}_{i} \left[\mathcal{M}_i \Sigma^{-}_{i} \mathcal{M}^{T}_{i} + \Sigma_{m_i} \right]^{-1}$$

Very simple example

- Car is translating
  - we supply a known demand to the accelerator,
    - changing at each time step
  - it sees 2 beacons (which are in its coordinate system)
    - beacon 1 measured in car x but not y
    - beacon 2 measured in car y but not x
- Q:
  - recover filtered estimates of:
    - position, velocity and acceleration in world coords

Dynamical model

- We supply a demand to the accelerator
  - acceleration updates as noise (measured to be about the same as demand!)

$$\mathbf{a}_{i+1} = \mathbf{a}_i + \text{noise}$$

  - velocity by integrating acceleration

$$\mathbf{v}_{i+1} = \mathbf{v}_i + \delta t \mathbf{a}_i + \text{noise}$$

  - position by integrating velocity

$$\mathbf{c}_{i+1} = \mathbf{c}_i + \delta t \mathbf{v}_i + \text{noise}$$

Stack the vectors to get:
$$\mathbf{x}_i = \begin{bmatrix} \mathbf{c}_i \\ \mathbf{v}_i \\ \mathbf{a}_i \end{bmatrix}$$

Which gives:

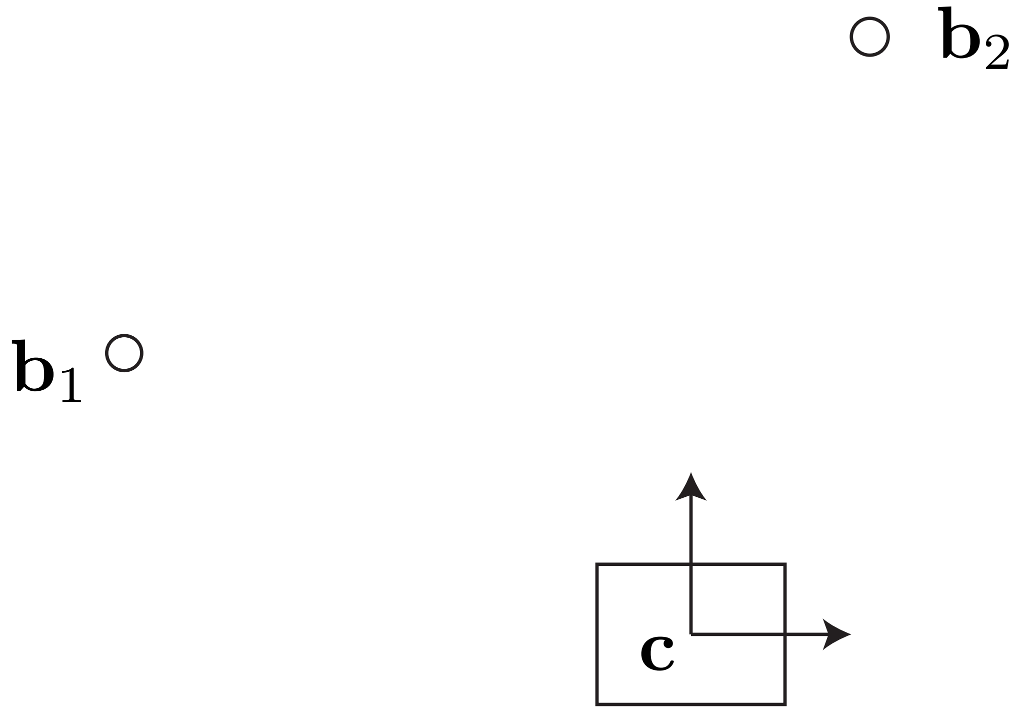
$$\mathbf{x}_{i+1} = \begin{bmatrix} \mathbf{c}_{i+1} \\ \mathbf{v}_{i+1} \\ \mathbf{a}_{i+1} \end{bmatrix} = \begin{bmatrix} \mathcal{I} & \delta t\mathcal{I} & 0 \\ 0 & \mathcal{I} & \delta t\mathcal{I} \\ 0 & 0 & \mathcal{I} \end{bmatrix} \begin{bmatrix} \mathbf{c}_i \\ \mathbf{v}_i \\ \mathbf{a}_i \end{bmatrix} + \xi_i$$

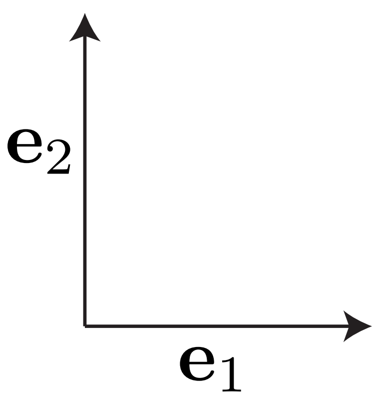
Where:

$$\xi_i \sim N(\mathbf{0}; \Sigma_{d,i})$$

Measurement model

- The acceleration at i should be demand
  - +noise
- Beacons are in car coordinate system)
  - beacon 1 measured in car x but not y
  - beacon 2 measured in car y but not x





In world coordinates, car is at: $\mathbf{c}$

In car coordinates, beacon 1 measurement is: $\mathbf{e}_1^T(\mathbf{b}_1 - \mathbf{c})$

In car coordinates, beacon 2 measurement is: $\mathbf{e}_2^T(\mathbf{b}_2 - \mathbf{c})$

The acceleration demand

$$\mathbf{y}_i = \begin{bmatrix} \mathbf{d}_i \\ \mathbf{e}_1^T \mathbf{b}_1 - b_1 \\ \mathbf{e}_2^T \mathbf{b}_2 - b_2 \end{bmatrix} + \text{noise}$$

These are known constants

Measurements from the beacons

$$\mathbf{y}_i = \begin{bmatrix} 0 & 0 & \mathcal{I} \\ \mathbf{e}_1^T & 0 & 0 \\ \mathbf{e}_2^T & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{c}_i \\ \mathbf{v}_i \\ \mathbf{a}_i \end{bmatrix} + \text{noise} = \begin{bmatrix} 0 & 0 & \mathcal{I} \\ \mathbf{e}_1^T & 0 & 0 \\ \mathbf{e}_2^T & 0 & 0 \end{bmatrix} \mathbf{x}_i + \zeta_i$$

$$\zeta_i \sim N(0; \Sigma_{m,i})$$

The steps:

Have:

$$\overline{X}_{i-1}^{+} \qquad \Sigma_{i-1}^{+}$$

Construct:

$$\overline{X}_{i}^{-} = \mathcal{D}_i \overline{X}_{i-1}^{+} \qquad \Sigma_{i}^{-} = \Sigma_{d_i} + \mathcal{D}_i \Sigma_{i-1}^{-} \mathcal{D}_i^{T}$$

Measurement arrives: $\mathbf{y}_i \sim N(\mathcal{M}_i \mathbf{x}_i; \Sigma_{mi})$

Now construct:

$$\bar{X}_i^{+} = \bar{X}_i^{-} + \mathcal{K}_i \left[\mathbf{y}_i - \mathcal{M}_i \bar{X}_i^{-} \right] \qquad \Sigma_i^{+} = \left[\mathcal{I} - \mathcal{K}_i \mathcal{M}_i \right] \Sigma_i^{-}$$

Where:

$$\mathcal{K}_i = \Sigma_i^{-} \mathcal{M}_i^{T} \left[\mathcal{M}_i \Sigma_i^{-} \mathcal{M}_i^{T} + \Sigma_{m_i} \right]^{-1}$$

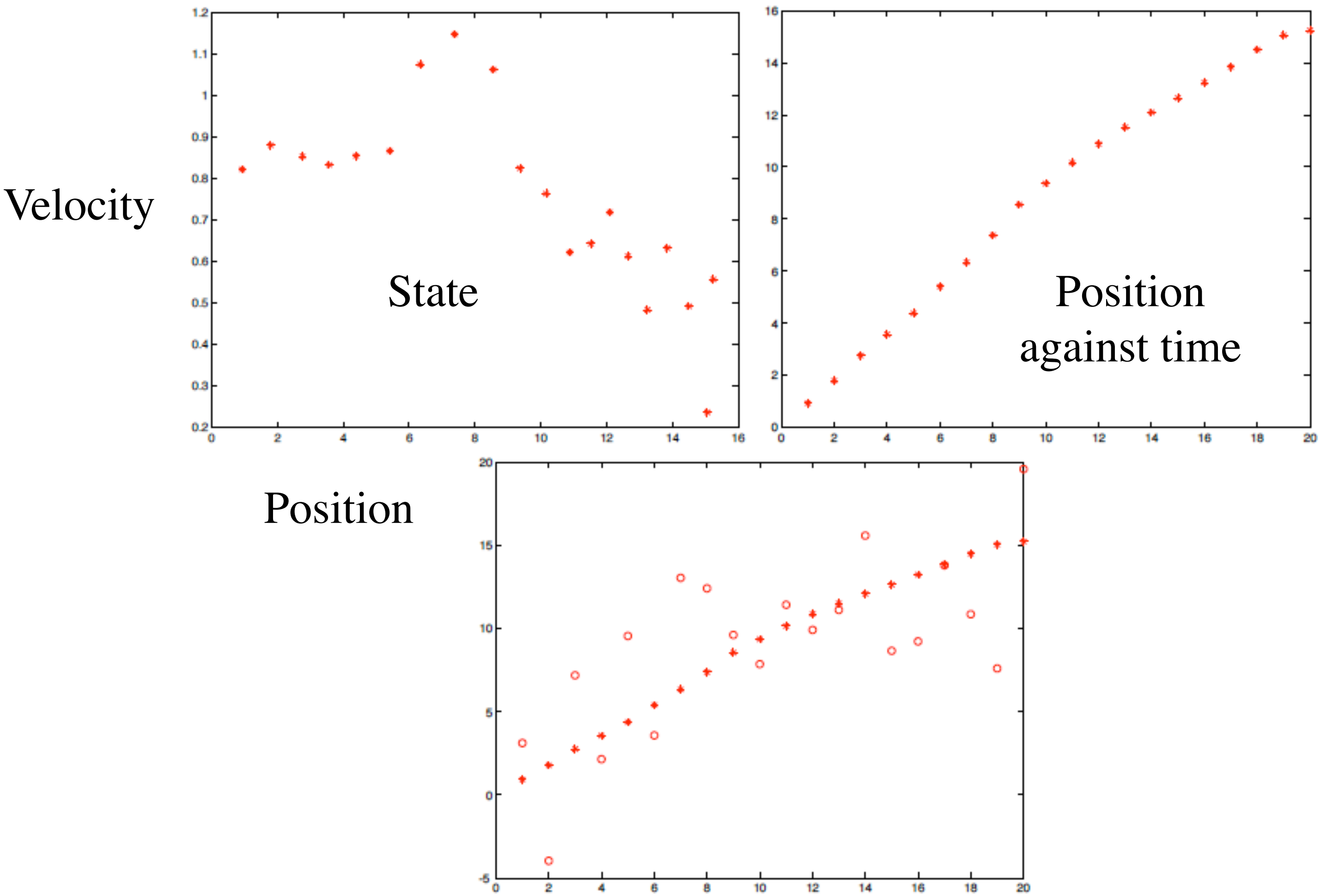
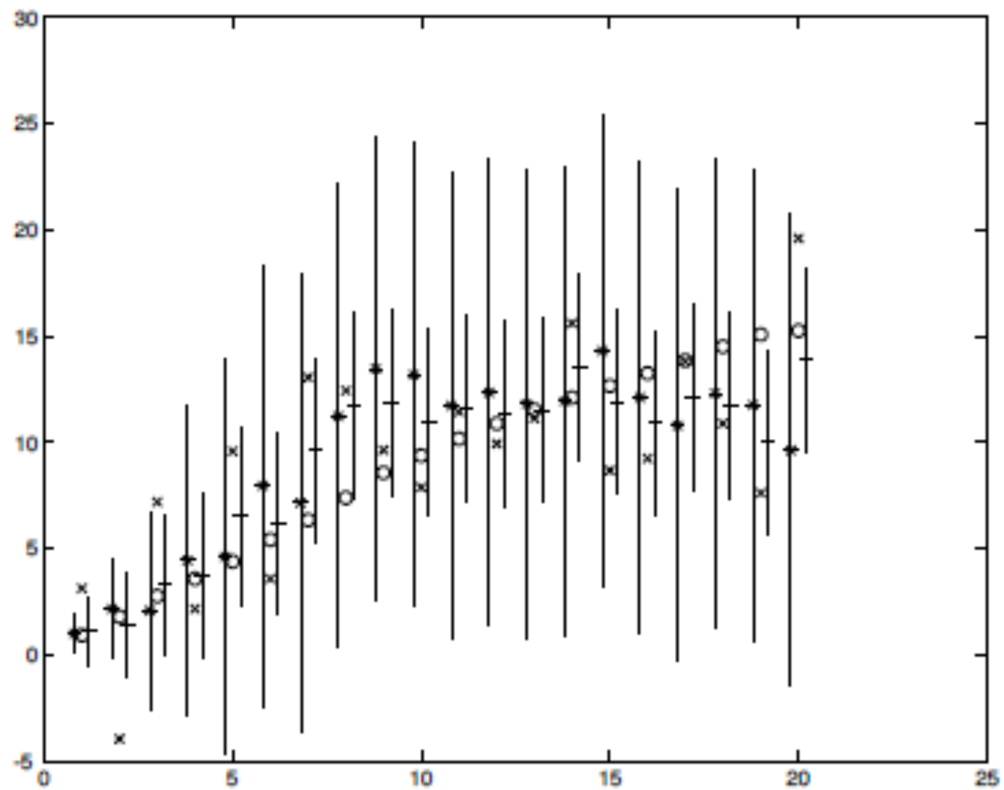


FIGURE 11.7: A constant velocity dynamic model for a point on the line. In this case, the state space is two dimensional, with one coordinate for position, one for velocity. The figure on the **top left** shows a plot of the state; each asterisk is a different state. Notice that the vertical axis (velocity) shows some small change compared with the horizontal axis. This small change is generated only by the random component of the model, so the velocity is constant up to a random change. The figure on the **top right** shows the first component of state (which is position) plotted against the time axis. Notice we have something that is moving with roughly constant velocity. The figure on the **bottom** overlays the measurements (the circles) on this plot. We are assuming that the measurements are of position only, and are quite poor; as we see, this doesn't significantly affect our ability to track.



Notice how uncertainty in state grows with movement and is reduced with measurement.

FIGURE 11.9: The Kalman filter for a point moving on the line under our model of constant velocity (compare with Figure 11.7). The state is plotted with open circles as a function of the step i. The *s give $\overline{x}_i^-$, which is plotted slightly to the left of the state to indicate that the estimate is made before the measurement. The xs give the measurements, and the +s give $\overline{x}_i^+$, which is plotted slightly to the right of the state. The vertical bars around the *s and the +s are three standard deviation bars, using the estimate of variance obtained before and after the measurement, respectively. When the measurement is noisy, the bars don't contract all that much when a measurement is obtained (compare with Figure 11.10).

Tricks

- Smoothing
  - You can build a representation of $P(X_i|Y_0, \ldots Y_N)$
    - (i.e. incorporating future measurements)
    - run one filter forward, one backward
      - posterior of forward filter is normal
      - predictive for backward is normal
      - etc.
- Polishing
  - This means that, if I can endure latency, I can have two estimates
    - one at the time of the i'th measurement
    - one a few measurements later, that is more accurate

Data Association

- Nearest neighbours
  - choose the measurement with highest probability given predicted state
  - popular, but can lead to catastrophe
- Probabilistic Data Association
  - combine measurements, weighting by probability given predicted state
  - gate using predicted state