Tracking Boundaries

We are given a set of points $\Gamma_1$, along a shape boundary, with coordinates $\Gamma_1 = \{x_1, x_2, \ldots, x_N\}$ at the initial frame $\tau=1$, were the image is $I_{\tau=1}$.

We are interested in tracking this contour through several image frames, say through $T$ image frames given by

$$\tilde{I}_T = \{I_T, I_{T-1}, \ldots, I_1\}$$

We denote each of these contours by $\Gamma_\tau$ and so $\Gamma_1 = \Gamma_{\tau=1}$. We track each of the coordinates of the initial shape boundary, i.e., we focus on the tracking of the $N$ initial coordinates. We ask “Which state (coordinates) the shape boundary will be at a time $\tau+1$, given all the image observations $\tilde{I}_{\tau+1}$?”

In a Bayesian framework we attempt to obtain the probability

$$p(\Gamma_{\tau+1} \mid \tilde{I}_{\tau+1}) \quad (1)$$
Propagating Probability Density

An interesting expansion of (1) is given by:

\[
P(\Gamma_{\tau+1} \mid \bar{I}_{\tau+1}) = P(\Gamma_{\tau+1} \mid I_{\tau+1}, \bar{I}_{\tau})
\]

\[
= \frac{p(I_{\tau+1} \mid \Gamma_{\tau+1}, \bar{I}_{\tau})}{p(I_{\tau+1} \mid \bar{I}_{\tau})} p(\Gamma_{\tau+1} \mid \bar{I}_{\tau})
\]

\[
= \frac{p(I_{\tau+1} \mid \Gamma_{\tau+1}, \bar{I}_{\tau})}{p(I_{\tau+1} \mid \bar{I}_{\tau})} \int_{\Gamma_{\tau}} p(\Gamma_{\tau+1}, \Gamma_{\tau} \mid \bar{I}_{\tau}) d\Gamma_{\tau}
\]

Baye's (conditional on C): \(P(A \mid B, C) = \frac{P(B \mid A, C)}{P(B \mid C)} P(A \mid C)\)

Marginal Probability: \(P(A \mid C) = \sum_B P(A, B \mid C)\)

\[
P(A, B \mid C) = P(A \mid B, C) P(B \mid C)
\]

\[
= \frac{p(I_{\tau+1} \mid \Gamma_{\tau+1}, \bar{I}_{\tau})}{p(I_{\tau+1} \mid \bar{I}_{\tau})} \int_{\Gamma_{\tau}} p(\Gamma_{\tau+1} \mid \Gamma_{\tau}, \bar{I}_{\tau}) P(\Gamma_{\tau} \mid \bar{I}_{\tau}) d\Gamma_{\tau}
\]

(2)

Assuming we can estimate \(p(I_{\tau+1} \mid \Gamma_{\tau+1}, \bar{I}_{\tau}), p(\Gamma_{\tau+1} \mid \Gamma_{\tau}, \bar{I}_{\tau}), p(I_{\tau+1} \mid \bar{I}_{\tau})\), we have a recursive method to estimate \(P(\Gamma_{\tau+1} \mid \bar{I}_{\tau+1})\).
Assumptions

\[ p(I_{\tau+1} | \Gamma_{\tau+1}, \bar{I}_\tau) \]  Independence assumption on data generation, i.e., the current state completely defines the probability of the measurements/data (often the previous data is also neglected)

\[ p(I_{\tau+1} | \Gamma_{\tau+1}) \]  or

\[ p(\Gamma_{\tau+1} | \Gamma_\tau, \bar{I}_\tau) \]  First order Markov process, i.e., the current state of the system is completely described by the immediate previous state, and possibly the data (often the data is also neglected).

\[ p(\Gamma_{\tau+1} | \Gamma_\tau) \]  or

\[ p(I_{\tau+1} | \bar{I}_\tau) \]  This is a normalization constant that can always be computed by normalizing the recurrence equation (2)
Linear Dynamics

\[ p(\Gamma_{\tau} \mid \Gamma_{\tau-1}) = N(D_{\tau-1} \Gamma_{\tau-1}; \Sigma_{d_{\tau}}) \]
\[ p(I_{\tau} \mid \Gamma_{\tau}) = N(M_{\tau} \Gamma_{\tau}; \Sigma_{m_{\tau}}) \]

where \( N(\mu; \Sigma) \) is the multivariable normal distribution, i.e., a Gaussian with mean vector \( \mu \) and covariance matrix \( \Sigma \). \( M_{\tau} \) and \( D_{\tau} \) are linear matrices.

Examples

Random Walk

Points moving with constant velocity

Points moving with constant acceleration

Periodic motion
Random Walk of a Point

**Drift:** New position is the previous one plus noise ($\sigma_d$)

\[
\begin{align*}
\vec{x}_\tau &= \begin{pmatrix} x_\tau \\ y_\tau \end{pmatrix} \\
p(\vec{x}_\tau | \vec{x}_{\tau-1}) &= N(\vec{x}_\tau; D_{\tau-1} \vec{x}_{\tau-1}; \sigma_d) \\
D_{\tau-1} &= \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \\
p(m_\tau | \vec{x}_\tau) &= N(m_\tau; M_{\tau} \vec{x}_\tau; \sigma_{m_\tau}) \\
M_{\tau} &= \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \sigma_{m_\tau} = 0
\end{align*}
\]
Point moving with constant velocity

Motion equation: \[ \frac{dx}{dt} = v \quad \rightarrow \quad x(t + dt) = x(t) + v \, dt \]

\[ \bar{u}_\tau = \begin{pmatrix} x_\tau \\ v_\tau \end{pmatrix} \xrightarrow{\frac{dx}{dt} = v} \bar{u}_\tau = \begin{pmatrix} 1 & \Delta t \\ 0 & 1 \end{pmatrix} \bar{u}_{\tau-1} \]

\[ \begin{align*}
\mathcal{P}(\bar{u}_\tau | \bar{u}_{\tau-1}) &= N_{\bar{x}_\tau} (D_{\tau-1} \bar{u}_{\tau-1}; \sigma_{d_\tau}) \\
D_{\tau-1} &= \begin{pmatrix} 1 & \Delta t \\ 0 & 1 \end{pmatrix} \\
\mathcal{P}(m_\tau | \bar{u}_\tau) &= N_{m_\tau} (M_\tau \bar{u}_\tau; \sigma_{m_\tau}) \\
M_\tau &= \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}
\end{align*} \]

Note: \( \mathcal{P}(\bar{u}_\tau | \bar{u}_{\tau-1}) \) is equivalent to a second order model

\[ \mathcal{P}(x_\tau | x_{\tau-1}, x_{\tau-2}) = N(x_{\tau-1} + \Delta t(x_{\tau-1} - x_{\tau-2}); \sigma_{d_\tau}) \]
Point moving with constant acceleration

Motion with constant acceleration: \( v(t + dt) = v(t) + a \, dt \)

\[
\begin{pmatrix}
  x_t \\
v_t \\
a_t
\end{pmatrix}
\] \( \rightarrow \)

\[
p(\vec{u}_r | \vec{u}_{r-1}) = N_{\vec{u}_r} (D_{r-1} \vec{u}_{r-1}; \sigma_r)
\]

\[
D_{r-1} = \begin{pmatrix}
1 & \Delta t & 0 \\
0 & 1 & \Delta t \\
0 & 0 & 1
\end{pmatrix}
\]

\[
p(m_r | \vec{u}_r) = N_{m_r} (M_r \vec{u}_r; \sigma_r)
\]

\[
M_r = \begin{pmatrix}
1 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0
\end{pmatrix}
\]

Note: \( p(\vec{u}_r | \vec{u}_{r-1}) \) is equivalent to a third order model, using \( x(t) = x(t) + v(t)dt + \frac{1}{2} a(dt)^2 \),

\[
p(x_r | x_{r-1}, x_{r-2}, x_{r-3}) = N \left( x_{r-1} + \Delta t (x_{r-1} - x_{r-2}) + \frac{1}{2} (\Delta t)^2 (x_{r-1} + x_{r-3} - 2x_{r-2}); \sigma_r \right)
\]
Point moving with periodic motion

Periodic Motion: \( \frac{d^2x}{dt^2} = -x(t) \rightarrow \) periodic solution \( x(t) = A \cos wt \)

equivalent to

\[
\begin{align*}
\frac{dx}{dt} &= v(t) \quad \rightarrow \quad x(t + dt) &= x(t) + v(t)dt \\
\frac{dv}{dt} &= -x(t) \quad \rightarrow \quad v(t + dt) &= v(t) - x(t)dt
\end{align*}
\]

\[
\begin{pmatrix}
\bar{u}_\tau \\
\bar{v}_\tau
\end{pmatrix}
= \begin{pmatrix}
x_\tau \\
v_\tau
\end{pmatrix}
\rightarrow
\begin{align*}
p(\bar{u}_\tau | \bar{u}_{\tau-1}) &= N_{\bar{u}_\tau}(D_{\tau-1}\bar{u}_{\tau-1}; \sigma_{d_\tau}) \\
D_{\tau-1} &= \begin{pmatrix} 1 & \Delta t \\ -\Delta t & 1 \end{pmatrix}
\end{align*}
\]

\[
p(m_\tau | \bar{u}_\tau) &= N_{m_\tau}(M_\tau \bar{u}_\tau; \sigma_{m_\tau}) \\
M_\tau &= \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}
\]
Tracking one point (dynamic programming like)

Equation (2) is written as

\[ P(x_\tau | \bar{I}_\tau) = \frac{p(I_\tau | x_\tau, \bar{I}_{\tau-1})}{p(I_\tau | \bar{I}_{\tau-1})} \int p(x_\tau | x_{\tau-1}, \bar{I}_{\tau-1}) P(x_{\tau-1} | \bar{I}_{\tau-1}) dx_{\tau-1} \]

\[ F(x_\tau | \bar{I}_\tau) = p(I_\tau | x_\tau, \bar{I}_{\tau-1}) \sum_{x_{\tau-1}} p(x_\tau | x_{\tau-1}, \bar{I}_{\tau-1}) P(x_{\tau-1} | \bar{I}_{\tau-1}) \]

\[ P(x_\tau | \bar{I}_\tau) = \frac{F(x_\tau | \bar{I}_\tau)}{\sum_{x_\tau} F(x_\tau | \bar{I}_\tau)} \]

note that: \[ p(I_\tau | \bar{I}_{\tau-1}) = \sum_{x_\tau} F(x_\tau | \bar{I}_\tau) \]
Kalman Filter

Simplify calculation of the probabilities $P(\Gamma_t \mid \tilde{I}_t)$ due to good Gaussian properties (applied to $p(I_t \mid \Gamma_t, \tilde{I}_{t-1})$ and $p(\Gamma_t \mid \tilde{I}_{t-1}, \tilde{I}_{t-1})$)

Reminder: From equation (2) we have

$$P(\Gamma_t \mid \tilde{I}_t) = \frac{p(I_t \mid \Gamma_t, \tilde{I}_{t-1})}{p(I_t \mid \tilde{I}_{t-1})} \int p(\Gamma_t \mid \Gamma_{t-1}, \tilde{I}_{t-1}) P(\Gamma_{t-1} \mid \tilde{I}_{t-1}) d\Gamma_{t-1}$$

**1D case:** For one dimension point and Gaussian distributions

$$P(x_t \mid \bar{y}_t) = \frac{N_{y_t}(m_{x_t}, \sigma_{m_{x_t}})}{p(y_t \mid \bar{y}_{t-1})} \int N_{x_t}(d_{x_t}, \sigma_{d_{x_t}}) P(x_{t-1} \mid \bar{y}_{t-1}) dx_{t-1}$$

Kalman observed the recurrence: if $P(x_{t-1} \mid \bar{y}_{t-1}) = N_{x_{t-1}}(\mu_{t-1}; \sigma^+_{t-1})$ then $P(x_t \mid \bar{y}_t) = N_{x_t}(\mu_{t}; \sigma^+_{t})$

where $\sigma^+ = \sqrt{\frac{(\sigma_{m_{x_t}} \sigma^-_{x_t})^2}{(\sigma_{m_{x_t}})^2 + (\sigma^-_{x_t})^2}}$, $\sigma^- = \sqrt{\sigma^2_{d_{x_t}} + (d_{x_t} \sigma^+_{x_{t-1}})^2}$ and $\mu_t = \frac{m_{x_t} \sigma^-_{x_t} + d_{x_t} \mu_{t-1} \sigma^+_{m_{x_t}}}{(m_{x_t} \sigma^-_{x_t})^2 + \sigma^2_{m_{x_t}}}$
Kalman Filter and Gaussian Properties

Proof: Assume $P(x_{r-1} \mid \tilde{y}_{r-1}) = \mathcal{N}_{x_{r-1}}(\mu_{r-1}; \sigma^+_{r-1})$

Then $P(x_r \mid \tilde{y}_r) = c_r \mathcal{N}_{y_r}(m_r x_r; \sigma_{m_r}) \int \mathcal{N}_{x_r}(d_r x_{r-1}; \sigma_{d_r}) \mathcal{N}_{x_{r-1}}(\mu_{r-1}; \sigma^+_{r-1}) dx_r$

since

$$
\frac{1}{\sqrt{2\pi\sigma^2_{d_r}}} e^{-\frac{(x_{r-d_r,x_{r-1}})^2}{2\sigma^2_{d_r}}} = \frac{1}{\sqrt{2\pi\sigma^2_{d_r}}} e^{-\frac{d_r^2(x_{r-1} x_{r-1})^2}{2\sigma^2_{d_r}}} = \frac{d_r^{-1}}{\sqrt{2\pi(d_r^{-1} \sigma_{d_r})^2}} e^{-\frac{(x_{r-1} - d_r^{-1} x_r)^2}{2(d_r^{-1} \sigma_{d_r})^2}}
$$

$$
= c_r d_r^{-1} \mathcal{N}_{y_r}(m_r x_r; \sigma_{m_r}) \int \mathcal{N}_{x_r}(d_r^{-1} x_r; d_r^{-1} \sigma_{d_r}) \mathcal{N}_{x_{r-1}}(\mu_{r-1}; \sigma^+_{r-1}) dx_r
$$

and since

$$
\int_{x=\infty} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}} dx = e^{-\frac{(\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}} \quad c_r = c_r d_r^{-1}
$$

$$
= c_r \mathcal{N}_{y_r}(m_r x_r; \sigma_{m_r}) \mathcal{N}_{\mu_{r-1}}(d_r^{-1} x_r; d_r^{-1} \sigma_{r}^{-} = \sqrt{(d_r^{-1} \sigma_{d_r})^2 + (\sigma^+_{r-1})^2})
$$

$$
= c_r \mathcal{N}_{x_r}(m_r^{-1} y_r; m_r^{-1} \sigma_{m_r}) \mathcal{N}_{x_r}(d_r \mu_{r-1}; \sigma_{r}^{-} = \sqrt{(\sigma_{d_r})^2 + (d_r \sigma^+_{r-1})^2})
$$

One can interpret it as a product of two Gaussian models: (i) The "accumulated prior model" where the prediction is $\mu_r^\prime = d_r \mu_{r-1}$" with variance $(\sigma^{-}_{r})^2$ and (ii) The data measurement which gives prediction $m_r^{-1} y_r$ with variances $(m_r^{-1} \sigma_{m_r})^2$. 
Kalman Filter ...(continuing the proof)

\[ P(x_t | \bar{y}_t) = \ldots \]

\[ = c' \, N_{x_t} (m^{-1}_t, y_t; m^{-1}_t \sigma_{m_t}) \, N_{x_t} (d_t \mu_{t-1}; \sigma^-_t = \sqrt{(\sigma_{d_t})^2 + (d_t \sigma^{+}_{t-1})^2}) \]

and since

\[ e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \times e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}} = b(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = bN_x (\mu; \sigma), \]

where

\[ \mu = \frac{(\mu_1 \sigma_2^2 + \mu_2 \sigma_1^2)}{\sigma_2^2 + \sigma_1^2}, \quad \sigma^2 = \frac{\sigma_1^2 \sigma_2^2}{\sigma_2^2 + \sigma_1^2} \]

\[ = c' \, b \, N_{x_t} \left( \mu_t = \frac{(m^{-1}_t y_t (\sigma')^2 + d_t \mu_{t-1} (m^{-1}_t \sigma_{m_t})^2)}{(\sigma')^2 + (m^{-1}_t \sigma_{m_t})^2); \sigma^+_t = \sqrt{\frac{(m^{-1}_t \sigma_{m_t})^2 (\sigma')^2}{(\sigma')^2 + (m^{-1}_t \sigma_{m_t})^2}} \right) \]

\[ = N_{x_t} \left( \mu_t = \frac{m_t y_t (\sigma')^2 + d_t \mu_{t-1} \sigma_{m_t}^2}{(m_t \sigma')^2 + \sigma_{m_t}^2); \sigma^+_t = \sqrt{\frac{\sigma_{m_t}^2 (\sigma')^2}{(m_t \sigma')^2 + \sigma_{m_t}^2}} \right) \]

\[ = N_{x_t} (\mu_t; \sigma^+_t) \]
Using Kalman Filter

The best estimate of the point \( x \) at \( \tau - 1 \), given the data history \( \bar{y}_{\tau-1} \) is

\[
\langle x_{\tau-1} \rangle = \mu_{\tau-1} = \arg\max_{x_{\tau-1}} P(x_{\tau-1} \mid \bar{y}_{\tau-1}) = \arg\max_{x_{\tau-1}} N_{x_{\tau-1}} (\mu_{\tau-1}; \sigma_{\tau-1}^+) \]

Note: the mean and the max values of a Gaussian distribution are the same

Assuming the estimate \( \mu_{\tau-1} \) and \( \sigma_{\tau-1}^+ \) to be known at \( \tau - 1 \) we estimate \( x \) at frame \( \tau \) as follows

1. \( (\sigma_{\tau}^-)^2 = (\sigma_{d_{\tau}})^2 + (d_{\tau} \sigma_{\tau-1}^+)^2 \)

2. \( (\sigma_{\tau}^+)^2 = \frac{(m_{\tau}^{-1} \sigma_{m_{\tau}})^2 (\sigma_{\tau}^-)^2}{(\sigma_{\tau}^-)^2 + (m_{\tau}^{-1} \sigma_{m_{\tau}})^2} \)

3. \( \mu_{\tau} = \frac{m_{\tau}^{-1} y_{\tau} (m_{\tau} \sigma_{\tau}^-)^2 + d_{\tau} \mu_{\tau-1} \sigma_{m_{\tau}}^2}{(m_{\tau} \sigma_{\tau}^-)^2 + \sigma_{m_{\tau}}^2} = \frac{m_{\tau}^{-1} y_{\tau} (\sigma_{\tau}^-)^2 + d_{\tau} \mu_{\tau-1} (m_{\tau}^{-1} \sigma_{m_{\tau}})^2}{(\sigma_{\tau}^-)^2 + (m_{\tau}^{-1} \sigma_{m_{\tau}})^2} \)

and that is our best estimate of \( x_{\tau} \). One can interpret it as a weighted average between the prediction from the (prior) dynamical model applied to the previous estimation \( \mu_{\tau}^- = d_{\tau} \mu_{\tau-1} \) and the "correction" from the new data \( m_{\tau}^{-1} y_{\tau} \). The weights are given by the variances of the data, \( (m_{\tau}^{-1} \sigma_{m_{\tau}})^2 \), and variance of the accumulated prior model, \( (\sigma_{\tau}^-)^2 \), respectively.
Kalman Filter (Generalization to N-D)

The generalization to many dimensions (N-D) is straightforward, though one has to be careful with matrix multiplications ...

\[ x_r \rightarrow X_r = [x_r^1, \ldots, x_r^N] \]

1. \[ (\sigma_r^-)^2 = (\sigma_{d_r})^2 + (d_r \sigma_{r-1}^+)^2 \]
   \[ \rightarrow \Sigma_r^- = \Sigma_{d_r} + D_r \Sigma_{r-1}^+ D_r \]

2. \[ (\sigma_r^+)^2 = \frac{(\sigma_{m_r})^2 (\sigma_r^-)^2 m_r^2}{(\sigma_r^-)^2 m_r^2 + (\sigma_{m_r})^2} \]
   \[ \rightarrow \Sigma_{r-1}^+ = \Sigma_{m_r} M_r \Sigma_r^- M_r^T \left[ M_r \Sigma_r^- M_r^T + \Sigma_{m_r} \right]^{-1} \]

3. \[ \mu_r = \frac{(\sigma_r^-)^2 m_r^2 y_r + (\sigma_{m_r})^2 d_r \mu_{r-1}}{(\sigma_r^-)^2 m_r^2 + (\sigma_{m_r})^2} \]
   \[ \rightarrow \bar{X}_r = \Sigma_r^- M_r^T \left[ M_r \Sigma_r^- M_r^T + \Sigma_{m_r} \right]^{-1} Y_r + \Sigma_{m_r} \left[ M_r \Sigma_r^- M_r^T + \Sigma_{m_r} \right]^{-1} D_r \bar{X}_{r-1} \quad (3) \]

sometimes the matrix \( K_r = \Sigma_r^- M_r^T \left[ M_r \Sigma_r^- M_r^T + \Sigma_{m_r} \right]^{-1} \) is called Kalman Gain and (3)

\[ \bar{X}_r = K_r Y_r + (I - K_r M_r) D_r \bar{X}_{r-1} = D_r \bar{X}_{r-1} + K_r (Y_r - M_r D_r \bar{X}_{r-1}) \]

Note: \( \lim_{\Sigma_{m_r} \to 0} K_r = M_r^{-1} \) and \( \lim_{\Sigma_r \to 0} K_r = 0 \)

as measurements become more reliable \( K_r \) weights the data residuals more, and as the prior covariance, \( \Sigma_r^- \), approaches 0, measurements are ignored.