Kalman Filter

Lee Wee Sun

Semester 2, 2005

Multivariate Gaussian

- For $x \in \Re^n$, the multivariate Gaussian distribution is usually parameterized using a $n \times 1$ vector μ and an $n \times n$ matrix Σ.

$$p(x|\mu, \Sigma) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right\}.$$

- The parameters μ and Σ are the moments of the Gaussian distribution

$$\begin{aligned} \mu &= E(x) \\ \Sigma &= E(x-\mu)(x-\mu)^T \end{aligned}$$

- Suppose that we partition x into a $p \times 1$ subvector x_1 and a $q \times 1$ subvector x_2, such that

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

Then

$$p(x|\mu, \Sigma) = \frac{1}{(2\pi)^{(p+q)/2}|\Sigma|^{1/2}}$$
$$\exp\left\{-\frac{1}{2}\begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix}^T \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}^{-1} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix}\right\}$$

Partitioned Matrices

- Consider an invertable partitioned matrix

$$M = \begin{bmatrix} E & F \\ G & H \end{bmatrix},$$

where E and H are invertible. We wish to *block diagonalize* this matrix.

- To zero out the upper right hand corner of M, we premultiply by block matrix having elements I and $-FH^{-1}$. To zero out the lower left hand corner, we postmultiply by a block matrix having elements I and $-H^{-1}G$.

$$\begin{bmatrix} I & -FH^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} E & F \\ G & H \end{bmatrix} \begin{bmatrix} I & 0 \\ -H^{-1}G & I \end{bmatrix} = \begin{bmatrix} E - FH^{-1}G & 0 \\ 0 & H \end{bmatrix}$$

- Denote the *Shur complement* of M with respect to H, M/H as $E - FH^{-1}G$. M/H is invertible.
- Taking inverse of both sides

$$\begin{aligned}
\begin{bmatrix} E & F \\ G & H \end{bmatrix}^{-1} &= \begin{bmatrix} I & 0 \\ -H^{-1}G & I \end{bmatrix} \begin{bmatrix} (M/H)^{-1} & 0 \\ 0 & H^{-1} \end{bmatrix} \begin{bmatrix} I & -FH^{-1} \\ 0 & I \end{bmatrix} \\
&= \begin{bmatrix} (M/H)^{-1} & -(M/H)^{-1}FH^{-1} \\ -H^{-1}G(M/H)^{-1} & H^{-1} + H^{-1}G(M/H)^{-1}FH^{-1} \end{bmatrix}
\end{aligned}$$

- Applying the determinant operator to both sides, we obtain

$$|M| = |M/H||H|.$$

- Similarly

$$\begin{bmatrix} E & F \\ G & H \end{bmatrix}^{-1} = \begin{bmatrix} E^{-1} + E^{-1}F(M/E)^{-1}GE^{-1} & -E^{-1}F(M/E)^{-1} \\ -(M/E)^{-1}GE^{-1} & (M/E)^{-1} \end{bmatrix}$$

where $M/E = H - GE^{-1}F$.

- Equating the blocks, we get the matrix inversion lemma

$$(E - FH^{-1}G)^{-1} = E^{-1} + E^{-1}F(H - GE^{-1}F)^{-1}GE^{-1}$$

and

$$(E - FH^{-1}G)^{-1}FH^{-1} = E^{-1}F(H - GE^{-1}F)^{-1}.$$

Marginalization and Conditioning

- We would like to factorize $p(x_1, x_2) = p(x_1|x_2)p(x_2)$

$$
\begin{aligned}
& \exp\left\{-\frac{1}{2}\begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix}^T \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}^{-1} \begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix}\right\} \\
= \quad & \exp\left\{-\frac{1}{2}\begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix}^T \begin{bmatrix} I & 0 \\ -\Sigma_{22}^{-1}\Sigma_{21} & I \end{bmatrix}\right. \\
& \left.\begin{bmatrix} (\Sigma/\Sigma_{22})^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{bmatrix} \begin{bmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{bmatrix} \begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix}\right\} \\
= \quad & \exp\left\{-\frac{1}{2}(x_1-\mu_1-\Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2))^T(\Sigma/\Sigma_{22})^{-1}(x_1-\mu_1\right. \\
& \left.-\Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2))\right\}\exp\left\{-\frac{1}{2}(x_2-\mu_2)^T\Sigma_{22}^{-1}(x_2-\mu_2)\right\}
\end{aligned}
$$

- Again utilizing the result of the partitioning

$$
\begin{aligned}
\frac{1}{(2\pi)^{(p+q)/2}|\Sigma|^{1/2}} &= \frac{1}{(2\pi)^{(p+q)/2}(|\Sigma/\Sigma_{22}||\Sigma_{22}|)^{1/2}} \\
&= \left(\frac{1}{(2\pi)^{p/2}|\Sigma/\Sigma_{22}|^{1/2}}\right)\left(\frac{1}{(2\pi)^{q/2}|\Sigma_{22}|^{1/2}}\right)
\end{aligned}
$$

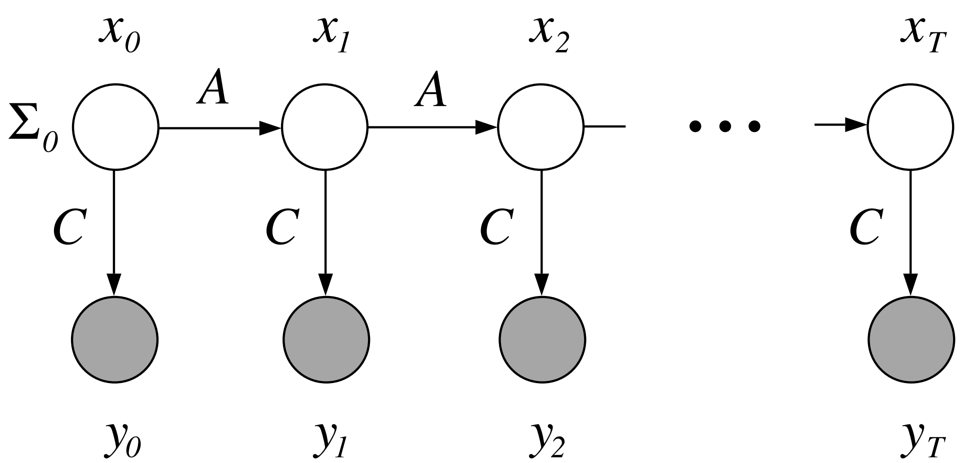
- As a result, we can express the following on marginalization

$$
\begin{aligned}
\mu_2^m &= \mu_2 \\
\Sigma_2^m &= \Sigma_{22}
\end{aligned}
$$

and on conditioning

$$
\begin{aligned}
\mu_{1|2}^c &= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2) \\
\Sigma_{1|2}^c &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}
\end{aligned}
$$

Kalman Filter



Kalman filter has the same structure as a graphical model as a Hidden Markov Model as shown above. The different algorithms are simply a reflection of the differences between the multinomial distribution and the Gaussian distribution.

The State Space Model

- We assume that the mean of the state at time $t+1$ is a linear function of the state at time t and write

$$x_{t+1} = Ax_t + Gw_t$$

  where w_t is a zero mean Gaussian noice that is independent of w_s for $s < t$ and independent of x_t.

- The noise term w_t has zero mean and covariance matrix Q.

- Conditioned on x_t, the mean of x_{t+1} is Ax_t and its covariance is GQG^T.

- The output is endowed with a distribution with a mean that is a linear function of the state

$$y_t = Cx_t + v_t$$

where v_t is a zero mean Gaussian random variable with covariance matrix R.

- Conditioned on x_t, y_t is a Gaussian with mean Cx_t and covariance R.

- Finally, we need an initial state x_0 which is a zero mean Gaussian distribution with covariance Σ_0.

Filtering

- The problem is to calculate an estimate of the state x_t given the sequence of observations so far $y_0, \ldots, y_t$ i.e. $P(x_t | y_0, \ldots, y_t)$.
- Sum of Gaussian variables are Gaussian. Conditional of Gaussians are Gaussian, thus we only need to calculate a mean vector and a covariance matrix to specify $P(x_t | y_0, \ldots, y_t)$.
- We write $\hat{x}_{t|t}$ to denote the mean of x_t conditioned on $y_0, \ldots, y_t$ and $P_{t|t}$ to denote the covariance matrix of x_t conditioned on $y_0, \ldots, y_t$

$$\begin{aligned}
\hat{x}_{t|t} &= E[x_t | y_0, \ldots, y_t] \\
P_{t|t} &= E[(x_t - \hat{x}_{t|t})(x_t - \hat{x}_{t|t})^T | y_0, \ldots, y_t].
\end{aligned}$$

- It is useful to do the calculation in two steps, first calculating the conditional distribution with mean $\hat{x}_{t|t-1}$ and covariance matrix $P_{t|t-1}$ before y_t is observed and then updating it after y_t is observed.
- Recall that

$$x_{t+1} = Ax_t + Gw_t$$

- Taking the conditional expectation on both sides, we have

$$\hat{x}_{t+1|t} = A\hat{x}_{t|t}.$$

- Taking the conditional covariance, we have

$$\begin{aligned}
P_{t+1|t} &= E[(x_{t+1} - \hat{x}_{t+1|t})(x_{t+1} - \hat{x}_{t+1|t})^T | y_0, \ldots, y_t] \\
&= E[(Ax_t + Gw_t - A\hat{x}_{t|t})(Ax_t + Gw_t - A\hat{x}_{t|t})^T | y_0, \ldots, y_t] \\
&= AP_{t|t}A^T + GQG^T
\end{aligned}$$

because w_t has zero mean and independent of x_t.

- We now update the conditional distribution after y_{t+1} is observed. We first calculate the joint distribution of x_{t+1} and y_{t+1} conditioned on $y_0, \ldots, y_t$. We need the mean and covariance matrix:

$$\begin{aligned} E[y_{t+1}|y_0, \ldots, y_t] &= E[Cx_{t+1} + v_{t+1}|y_0, \ldots, y_t] \\ &= C\hat{x}_{t+1|t} \end{aligned}$$

$$\begin{aligned} &E[(y_{t+1|t} - \hat{y}_{t+1|t})(y_{t+1|t} - \hat{y}_{t+1|t})^T|y_0, \ldots, y_t] \\ &= E[(Cx_{t+1} + v_{t+1} - C\hat{x}_{t+1|t})(Cx_{t+1} + v_{t+1} - C\hat{x}_{t+1|t})^T|y_o, \ldots, y_t] \\ &= CP_{t+1|t}C^T + R \end{aligned}$$

$$\begin{aligned} &E[(y_{t+1} - \hat{y}_{t+1|t})(x_{t+1} - \hat{x}_{t+1|t})^T|y_0, \ldots, y_t] \\ &= E[(Cx_{t+1} + v_{t+1} - \hat{y}_{t+1|t})(x_{t+1} - \hat{x}_{t+1|t})^T|y_0, \ldots, y_t] \\ &= CP_{t+1|t}. \end{aligned}$$

- Conditioned on $y_0, \ldots, y_t$, x_{t+1} and y_{t+1} have a joint Gaussian distribution with mean and covariance matrix:

$$\begin{bmatrix} \hat{x}_{t+1|t} \\ C\hat{x}_{t+1|t} \end{bmatrix} \text{ and } m \begin{bmatrix} P_{t+1|t} & P_{t+1|t}C^T \\ CP_{t+1|t} & CP_{t+1|t}C^T + R \end{bmatrix}$$

- Using the property of multivariate Gaussians, the conditional probability of x_t is

$$\hat{x}_{t+1|t+1} = \hat{x}_{t+1|t} + P_{t+1|t}C^T(CP_{t+1|t}C^T + R)^{-1}(y_{t+1} - C\hat{x}_{t+1|t})$$
$$P_{t+1|t+1} = P_{t+1|t} - P_{t+1|t}C^T(CP_{t+1|t}C^T + R)^{-1}CP_{t+1|t}.$$

Kalman Filter

- Kalman filter is defined by the following recursive equations, initialized by $\hat{x}_{0|-1} = 0$ and $P_{0|-1} = P_0$.

$$\hat{x}_{t+1|t} = A\hat{x}_{t|t}$$
$$P_{t+1|t} = AP_{t|t}A^T + GQG^T$$
$$\hat{x}_{t+1|t+1} = \hat{x}_{t+1|t} + P_{t+1|t}C^T(CP_{t+1|t}C^T + R)^{-1}(y_{t+1} - C\hat{x}_{t+1|t})$$
$$P_{t+1|t+1} = P_{t+1|t} - P_{t+1|t}C^T(CP_{t+1|t}C^T + R)^{-1}CP_{t+1|t}.$$

- The update equations are often written more compactly using the Kalman gain matrix:

$$K_{t+1} = P_{t+1|t}C^T(CP_{t+1|t}C^T + R)^{-1}.$$

Extended Kalman Filter

- When the process to be estimated or the measurement relationship to the process is non-linear, Kalman filter no longer applies.
- A Kalman filter that linearizes about the current mean and covariance is refered to as an extended Kalman filter (EKF). EKF is only an approximation but often works well.
- Assume that the process is governed by a non-linear stochastic difference equation

$$x_{t+1} = f(x_t, u_{t+1}, w_t),$$

and the measurement is

$$y_{t+1} = h(x_{t+1}, v_{t+1}),$$

where w_t and v_{t+1} represent process and measurement noise.

- The state vector is estimated as

$$\hat{x}_{t+1|t} = f(\hat{x}_{t|t}, u_{t+1}, 0)$$

and the measurement noise is estimated as

$$\hat{y}_{t+1} = h(\hat{x}_{t+1|t}, 0)$$

- We approximate the governing equations by linearizing

$$\begin{aligned} x_{t+1} &\approx \hat{x}_{t+1|t} + A(x_t - \hat{x}_{t|t}) + W w_t, \\ y_{t+1} &\approx \hat{y}_{t+1} + C(x_{t+1} - \hat{x}_{t+1|t}) + V v_{t+1}, \end{aligned}$$

where A is the Jacobian matrix of partial derivatives of f with respect to x, $A_{ij} = \frac{\partial f_i(\hat{x}_{t|t}, u_{t+1}, 0)}{\partial x_j}$, W is the Jacobian matrix of partial derivatives of f with respect to w, C is the Jacobian matrix of partial derivatives of h with respect to x and V is the Jacobian matrix of partial derivatives of h with respect to v.

- For simplicity we have not used the subscript $t+1$ with the Jacobians even though they are different at each time step.
- Denote the state prediction error as

$$\tilde{e}_{x_{t+1}} = x_{t+1} - \hat{x}_{t+1|t},$$

and the measurement residual as

$$\tilde{e}_{y_{t+1}} = y_{t+1} - \hat{y}_{t+1}.$$

- We can write the following equation that approximates the error process

$$\begin{aligned} \tilde{e}_{x_{t+1}} &\approx A(x_t - \hat{x}_{t|t}) + \epsilon_{t+1}, \\ \tilde{e}_{y_{t+1}} &\approx H\tilde{e}_{x_{t+1}} + \eta_{t+1}, \end{aligned} \tag{1}$$

where ϵ_{t+1} and η_{t+1} are independent random variables with zero mean and covariance matrices WQW^T and VRV^T respectively.

- Let $\hat{e}_{x_{t+1}} = \hat{x}_{t+1|t+1} - \hat{x}_{t+1|t}$.
- Motivated by equation (1), we use the Kalman filter equation to approximate $\hat{e}_{x_{t+1}}$ assuming that the state prediction is zero to get
$$\hat{e}_{x_{t+1}} = K_{t+1}\tilde{e}_{y_{t+1}}.$$
- Substituting back,
$$\begin{aligned} \hat{x}_{t+1|t+1} &= \hat{x}_{t+1|t} + K_{t+1}\tilde{e}_{y_{t+1}} \\ &= \hat{x}_{t+1|t} + K_{t+1}(y_{t+1} - \hat{y}_{t+1}) \end{aligned}$$

EKF Equations

$$
\begin{aligned}
\hat{x}_{t+1|t} &= f(\hat{x}_{t|t}, u_{t+1}, 0) \\
P_{t+1|t} &= A_{t+1} P_{t|t} A_{t+1}^T + W_{t+1} Q_{t+1} W_{t+1}^T \\
\hat{x}_{t+1|t+1} &= \hat{x}_{t+1|t} + K_{t+1}(y_{t+1} - h(\hat{x}_{t+1|t}, 0)) \\
P_{t+1|t+1} &= (I - K_{t+1} C_{t+1}) P_{t+1|t} \\
K_{t+1} &= P_{t+1|t} C_{t+1}^T (C_{t+1} P_{t+1|t} C_{t+1}^T + V_{t+1} R_{t+1} V_{t+1}^T)^{-1}.
\end{aligned}
$$