Statistical Techniques in Robotics (16-831, F12) Lecture#17 (Wednesday October 31)

Kalman Filters

Lecturer: Drew Bagnell *Scribe:Greydon Foil* [1]

1 Gauss Markov Model

Consider $X_1, X_2, ...X_t, X_{t+1}$ to be the state variables and $Y_1, Y_2, ...Y_t, Y_{t+1}$ be the sequence of corresponding observations. As in Hidden Markov models, conditional independencies (see Figure 1) dictate that past and future states are uncorrelated given the current state, X_t at time t. For example, if we know what X_2 is, then no information about X_1 can possibly help us to reason about what X_3 should be.

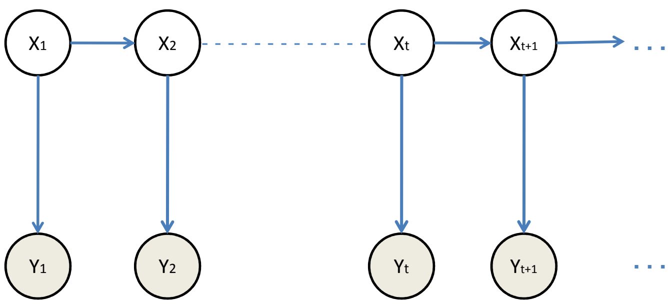

Figure 1: The Independence Diagram of a Gauss-Markov model

The update for state variable X_{t+1} is given by:

$$X_{t+1} = AX_t + \epsilon$$

where

$$X_0 \sim N(\mu_0, \Sigma_0)$$
$$\epsilon \sim N(0, Q)$$
$$X_{t+1}|X_t \sim N(AX_t, Q)$$

The corresponding observation Y_{t+1} is given by equation:

$$Y_{t+1} = CX_{t+1} + \delta$$

where

$$Y_0 \sim N(\mu_0, \epsilon_0)$$
$$\delta \sim N(0, R)$$

Each component is defined as follow:

[1] Some content adapted from previous scribes: Ammar Husain, Heather Justice, Kiho Kwak, Siddharth Mehrotra

- A_t: Matrix $(n \times n)$ that describes how the state evolves from t to t-1 without controls or noise.
- C_t: Matrix $(k \times n)$ that describes how to map the state X_t to an observation Y_t, where k is the number of observations.
- ϵ_t, δ_t: Random variables representing the process and measurement noise that are assumed to be independent and normally distributed with $n \times n$ noise covariances R_t and Q_t respectively.

We want to find $x_t|y_{1 \ldots t}$, so we need to calculate μ_{x+t} and Σ_{x_t}. Because a Gaussian will try to fit itself to all of the data, in a real situation we would first try to remove all outliers to achieve a more stable result.

Note that this parameterization is directly related to Bayes Linear Regression if it meets the following conditions:

- X here is equivalent to θ in BLR and Y here is equivalent to Y in BLR.
- The motion model is just the identity matrix.
- Q is going to 0 as $t \to \infty$. It is nonzero if the noise might be changing as a function of time.
- C is the vector x_t from BLR, different here at every timestep.
- $\delta \sim N(0, \sigma^2)$.

2 What can you do with Gaussians?

There are two common parameterizations for Gaussians, the moment parameterization and the natural parameterization. It is often most practical to switch back and forth between representations, depending on which calculations are needed. The moment parameterization is more convenient for visualization (simply draw a Gaussian centered around the mean with width determined by the variance), calculating expected value, and computing marginals. The natural parameterization is more convenient for multiplying Gaussians and for conditioning on known variables. While it is often convenient to switch between the two parameterizations, it is not always efficient, as we will discuss later.

2.1 Moment Parameterization

Recall that the moment parameterization of a Gaussian is:

$$\mathcal{N}(\mu, \Sigma) = p(\theta) = \frac{1}{z} \exp\left(-\frac{1}{2} (\theta - \mu)^T \Sigma^{-1} (\theta - \mu) \right)$$

Given:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right)$$

Marginal: computing $p(x_2)$

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

We find both of these by the definition of moments, specifically the fact that the moments of x_2 don't change if x_1 is removed.

Conditional: computing $p(x_1|x_2)$

$$\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$$

$(x_2 - \mu_2)$ is the distance x_2 is from its mean. We then multiply it by its uncertainty (Σ_{22}), and convert that value into the frame of x_1 using Σ_{12}, adding it to our best guess for x_1, μ_1.

$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

Here we start with the uncertainty in x_1, Σ_{11}, and subtract out the uncertainty in x_2 and between x_1 and x_2, again mapping it to the frame of x_1 using Σ_{12}.

2.2 Natural Parameterization

Recall that the natural parameterization of a Gaussian is:

$$\tilde{\mathcal{N}}(J, P) = \tilde{p}(\theta) = \frac{1}{z}\exp\left(J^T\theta - \frac{1}{2}\theta^T P\theta\right)$$

where

$$\begin{aligned}P_0 &= \Sigma_0^{-1} \\ J_0 &= P_0\mu_0\end{aligned}$$

Given:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} J_1 \\ J_2 \end{bmatrix}, \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}\right)$$

Marginal: computing $p(x_2)$

$$\begin{aligned}J_2^{\text{marg}} &= J_2 - P_{21}P_{11}^{-1}J_1 \\ P_2^{\text{marg}} &= P_{22} - P_{21}P_{11}^{-1}P_{12}\end{aligned}$$

These are most easily calculated by deriving the marignals in moment parameterization and converting to natural parameterization.

Conditional: computing $p(x_1|x_2)$

$$\begin{aligned}J_{1|2} &= J_1 - P_{12}x_2 \\ P_{1|2} &= P_{11}\end{aligned}$$

Derivation of these conditionals is a straightforward expansion of the full moment parameterization, and was said to be a possible test question. I encourage you to read page 7 of [1] for the full derivation. Also note that the moment parameterization is often called the *canonical parameterization.*

3 Lazy Gauss Markov Filter

Motion Model (Prediction step):

Before the observation is taken:

$$X_{t+1} \sim \mu_{t+1}^{-} = A\mu_t$$

Proof:

Mean:

$$\begin{aligned} E[X_{t+1}] &= E[AX_t + \epsilon] \\ &= E[AX_t] + E[\epsilon] \\ &= AE[X_t] \text{(since the mean of } \epsilon \text{ is 0)} \\ &= A\mu_t \end{aligned}$$

Variance:

$$\begin{aligned} \Sigma_{t+1}^{-} &= E[X_{t+1} * X_{t+1}{}^T] \\ &= E[(AX_t + \epsilon)(AX_t + \epsilon)^T] \\ &= E[(AX_t)(AX_t)^T] + Var(\epsilon) \\ &= AE[(X_t)(X_t)^T]A^T + Q \\ &= A\Sigma_t A^T + Q \end{aligned}$$

Therefore the motion update becomes:

$$\mu_{t+1}^{-} = A\mu_t$$

$$\Sigma_{t+1}^{-} = A\Sigma_t A^T + Q$$

3.1 Observation Model (Correction step):

For the observation model the natural parameterization is more suitable as it involves multiplication of terms. The model equation in terms of Natural Parameters J and P is given by:

$$\begin{aligned} P(y_{t+1}|x_{t+1})P(x_{t+1}) &\propto e^{(J^{-T}x_{t+1} - \frac{1}{2}x_{t+1}^T P x_{t+1})} * e^{-\frac{1}{2}(y_{t+1} - Cx_{t+1})^T R^{-1}(y_{t+1} - Cx_{t+1})} \\ &= e^{-\frac{1}{2}[-2y_{t+1}^T R^{-1} C x_{t+1} + x_{t+1}^T C^T R^{-1} C x_{t+1} + y_{t+1}^T R^{-1} y_{t+1}]} \\ &= e^{-\frac{1}{2}[-2y_{t+1}^T R^{-1} C x_{t+1} + x_{t+1}^T C^T R^{-1} C x_{t+1}]} \end{aligned}$$

The last term in the second line is constant with respect to x_{t+1}, so it can be added to the the marginalization term. Therefore the observation update is:

$$\begin{aligned} J_{t+1}^{+} &= J_{t+1}^{-} + (y_{t+1}^T R^{-1} C)^T \\ P_{t+1}^{+} &= P_{t+1}^{-} + C^{-1} R^{-1} C \end{aligned}$$

3.2 Performance

Lazy Gauss Markov can be expressed in two forms:

- When expressed in terms of moment parameters, μ and Σ, it acts as **Kalman Filter**.
- When expressed in terms of natural parameters, J and P, it acts as **Information Filter**.

Kalman filters, as we will see, require matrix multiplications, approximately $O(n^2)$ time, to do a prediction step, yet require matrix inversions, approximately $O(n^{2.8})$ time, to perform the observation update. Information filters are the exact opposite, requiring matrix inversions for the prediction step and matrix multiplications for the observation update. As mentioned above, the conversion between moment and natural parameterization requires an inversion of the covariance matrix, so switching between the two can be costly. Depending on the ratio of motion model updates to observation model updates one filter may be faster than the other.

References

[1] Michael I. Jordan and Christopher M. Bishop, "An Introduction to Graphical Models." July 10, 2001. `http://people.csail.mit.edu/yks/documents/classes/mlbook/pdf/chapter12.pdf`.