

## Kalman Filter

*Lecturer: Drew Bagnell*

*Scribe: Akanksha Saran<sup>1</sup>*

### 1 Gauss Markov Model

Last time we looked at Baye’s linear regression, where given an input data point the outcome is a single output. Today we look at a special case of filtering algorithms, where the outcome is a vector of outputs.

Consider  $X_1, X_2, \dots, X_t, X_{t+1}$  to be the state variables and  $Y_1, Y_2, \dots, Y_t, Y_{t+1}$  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.

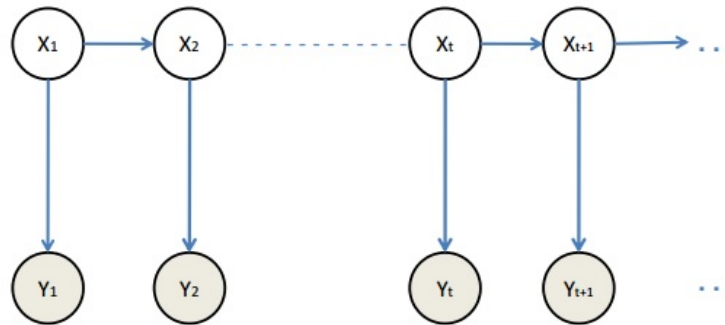


Figure 1: The Independence Diagram of a Gauss-Markov model

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:

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

---

<sup>1</sup>Content adapted from previous scribes: Greydon Foil, Ammar Husain, Heather Justice, Kiho Kwak.

where

$$Y_0 \sim N(\mu_0, \epsilon_0)$$

$$\delta \sim N(0, R)$$

Each component is defined as follow:

- $A_t$ : 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.
- $\epsilon_t, \delta_t$ : 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  $x_t|y_{1...t}$ , so we need to calculate  $\mu_{x+t}$  and  $\Sigma_{x_t}$ . 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 parametrization is directly related to Bayes Linear Regression if it meets the following conditions:

- $X$  here is equivalent to  $\theta$  in BLR and  $Y$  here is equivalent to  $Y$  in BLR.
- The motion model ( $A$ ) is just the identity matrix.
- $Q$  is going to 0 as  $t \rightarrow \infty$ . 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.
- $\delta \sim N(0, \sigma^2)$ . 0 here represents a zero vector.

## 2 What can you do with Gaussians?

$X$  and  $Y$  are jointly a Gaussian distribution. The state transition is linear and the transition between every pair of state and observation nodes is also linear. These linear transitions preserve everything to be Gaussian.

There are two common parametrizations for Gaussians, the moment parametrization and the natural parametrization. It is often most practical to switch back and forth between representations, depending on which calculations are needed. The moment parametrization 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 parametrization is more convenient for multiplying Gaussians and for conditioning on known variables. While it is often convenient to switch between the two parametrizations, it is not always efficient, as we will discuss later.

## 2.1 Moment Parametrization

Recall that the moment parametrization 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 ( $\Sigma_{22}$ ), and convert that value into the frame of  $x_1$  using  $\Sigma_{12}$ , adding it to our best guess for  $x_1$ ,  $\mu_1$ .

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

Here we start with the uncertainty in  $x_1$ ,  $\Sigma_{11}$ , 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  $\Sigma_{12}$ .

### Convenience of Marginal Parametrization

- Linear operations are easy.

$$\begin{aligned} x &\sim \mathcal{N}(\mu, \Sigma) \\ y &= Ax \\ y &\sim \mathcal{N}(A\mu, A\Sigma) \end{aligned}$$

- Marginals are easy to compute.

$$x_1 \sim \mathcal{N}(\mu_1, \Sigma_{11})$$

## 2.2 Natural Parametrization

Recall that the natural parametrization 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 marginals in moment parametrization and converting to natural parametrization.

**Conditional:** computing  $p(x_1|x_2)$

$$p(x_1|x_2) = \frac{p(x_1, x_2)}{p(x_2)} \propto p(x_1, x_2)$$

$$\begin{bmatrix} J_1 \\ J_2 \end{bmatrix}^T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^T \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = J^T x - \frac{1}{2} \begin{bmatrix} x_1^T & x_2^T \end{bmatrix} \begin{bmatrix} P_{11}x_1 & P_{12}x_2 \\ P_{21}x_1 & P_{22}x_2 \end{bmatrix}$$

$$x_1|x_2 \sim \tilde{\mathcal{N}}(J_1 - P_{12}x_2, P_{11})$$

I also encourage you to read page 7 of [1] for the full derivation. Also note that the moment parametrization is often called the *canonical parametrization*.

**Convenience of Natural Parametrization** Unlike marginal parametrization, linear operations and computing marginals are not direct with natural parametrization. But there are other advantages.

- Multiplication is easy (addition of parameters).
- Conditionals are easy to compute.

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

### 3 Lazy Gauss Markov Filter

Lets take the example of a robot in a plane.

$$X_{t+1} = \begin{bmatrix} pos \\ vel \end{bmatrix}_{t+1} = \begin{bmatrix} I & (\Delta t)I \\ 0 & I \end{bmatrix} X_t + \epsilon$$

$$\epsilon \sim \mathcal{N}\left(0, \begin{pmatrix} \gamma^2 I & 0 \\ 0 & \sigma^2 I \end{pmatrix}\right)$$

The observation or GPS measurement is given as

$$Y_{t+1} = \begin{bmatrix} I & 0 \end{bmatrix} \begin{bmatrix} X_{t+1} \end{bmatrix} + \delta$$

$$\delta \sim \mathcal{N}\left(0, \begin{pmatrix} 3I & 0 \\ 0 & 3I \end{pmatrix}\right)$$

Now given a series of measurements we want to solve the problem of determining the robot's pose and velocity.

#### 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:

$$\begin{aligned} \mu_{t+1}^- &= A\mu_t \\ \Sigma_{t+1}^- &= A\Sigma_t A^T + Q \end{aligned}$$

### 3.1 Observation Model (Correction step):

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

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

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,  $\mu$  and  $\Sigma$ , 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 parametrization 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>.