

Kalman Filtering: A Bayesian Approach

Adam S. Charles

December 14, 2017

The Kalman Filtering process seeks to discover an underlying set of state variables $\{\mathbf{x}_k\}$ for $k \in [0, n]$ given a set of measurements $\{\mathbf{y}_k\}$. The process and measurement equations are both linear and given by

$$\mathbf{x}_{n+1} = \mathbf{F}_{n+1}\mathbf{x}_n + \boldsymbol{\nu}_{o,n+1} \quad (1)$$

$$\mathbf{y}_n = \boldsymbol{\Phi}_n\mathbf{x}_n + \boldsymbol{\nu}_{d,n}. \quad (2)$$

The Kalman filter wants to find, at each iteration, the most likely *cause* of the measurement \mathbf{y}_n given the approximation made by a flawed estimation (the linear dynamics \mathbf{F}_n). Figure 1 shows a 2-dimensional graphical depiction. What is important here is not only that we have the measurement and the prediction, but knowledge of *how* each is flawed. In the Kalman case, this knowledge is given by the covariance matrices (essentially fully describing the distribution of the measurement and prediction for the Gaussian case). In Figure 1, this knowledge is represented by the ovals surrounding each point. The power of the Kalman filter comes from it's ability not only to perform this estimation once (a simple Bayesian task), but to use both estimates and knowledge of their distributions to find a distribution for the updated estimate, thus iteratively calculating the best solution for state at each iteration.

While many derivations of the Kalman filter are available, utilizing the orthogonality principle or finding iterative updates to the Best Linear Unbiased Estimator (BLUE), I will derive the Kalman Filter here using a Bayesian approach, where 'best' is interpreted in the Maximum A-Posteriori (MAP) sense instead of an L_2 sense (which for Gaussian innovations and measurement noise is the same estimate). Bayesian analysis uses Bayes rule, $p(a|b)p(b) = p(b|a)p(a)$, to express the posterior probability in terms of the likelihood and the prior. In this case we want to optimize over all states \mathbf{x}_k :

$$\{\hat{\mathbf{x}}_k\}_{k \in [0, n]} = \arg \max \left[\left(\prod_{i=1}^n p(\mathbf{x}_i | \mathbf{x}_{i-1}) p(\mathbf{y}_i | \mathbf{x}_i) \right) p(\mathbf{y}_0 | \mathbf{x}_0) p(\mathbf{x}_0) \right] \quad (3)$$

$$= \arg \max \left[p(\mathbf{y}_n | \mathbf{x}_n) p(\mathbf{x}_n | \mathbf{x}_{n-1}) \left(\prod_{i=1}^{n-1} p(\mathbf{x}_i | \mathbf{x}_{i-1}) p(\mathbf{y}_i | \mathbf{x}_i) \right) p(\mathbf{y}_0 | \mathbf{x}_0) p(\mathbf{x}_0) \right] \quad (4)$$

In order to find a globally optimal solution at the n^{th} time-step only, a marginalization is performed by:

$$\hat{\mathbf{x}}_n = \arg \max_{\mathbf{x}_n} \left[\int_{\mathbb{R}^n} p(\mathbf{y}_n | \mathbf{x}_n) p(\mathbf{x}_n | \mathbf{x}_{n-1}) \left(\prod_{i=1}^{n-1} p(\mathbf{x}_i | \mathbf{x}_{i-1}) p(\mathbf{y}_i | \mathbf{x}_i) \right) p(\mathbf{y}_0 | \mathbf{x}_0) p(\mathbf{x}_0) d\{\mathbf{x}_l\}_{l \in [0, n-1]} \right] \quad (5)$$

$$= \arg \max_{\mathbf{x}_n} \left[p(\mathbf{y}_n | \mathbf{x}_n) \int_{\mathbb{R}^n} p(\mathbf{x}_n | \mathbf{x}_{n-1}) \left(\prod_{i=1}^{n-1} p(\mathbf{x}_i | \mathbf{x}_{i-1}) p(\mathbf{y}_i | \mathbf{x}_i) \right) p(\mathbf{y}_0 | \mathbf{x}_0) p(\mathbf{x}_0) d\{\mathbf{x}_l\}_{l \in [0, n-1]} \right] \quad (6)$$

Note that this integral is essentially the prior on \mathbf{x}_n . Since this prior is an integral of all Gaussian random variables, the result is a Gaussian random variable (Gaussian distributions are self conjugate, and

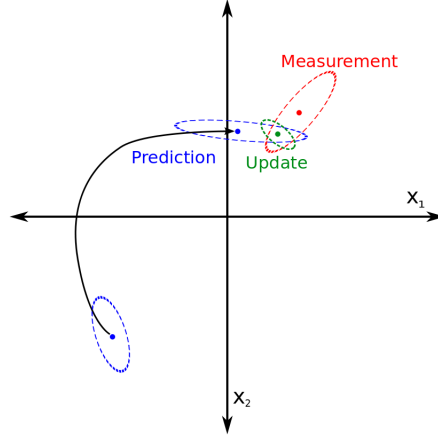


Figure 1: The Kalman filter uses the prediction of a current state based on a previous estimate (blue points) in conjunction with a current measurement (red point) to estimate the true current state (green point). The error in the dynamics (shown here by the blue ovals which represent the covariance) is a combination of the error in the past state and the error in the model of the system. This error in conjunction with the measurement error (the red ovals) allow the covariance of the state update (green oval) to be calculated, propagating forward the confidence of each update.

marginalizing over a Gaussian yields a Gaussian). Thus while only performing a temporally localized update, an updated distribution on \mathbf{x}_n is used so that Equation (4) can be written as

$$\hat{\mathbf{x}}_n = \arg \max_{\mathbf{x}_n} [p(\mathbf{y}_n | \mathbf{x}_n) p_{\hat{\mathbf{x}}_{n-1}}(\mathbf{x}_n)] \quad (7)$$

The updated distribution uses all past information to give in essence a likelihood $\mathbf{x}_n | \{\mathbf{y}_k\}_{k \in [0, n-1]}$. This estimate comes in the form of a probability distribution on the previous estimate $\hat{\mathbf{x}}_{n-1}$, and takes the place of the prior on \mathbf{x}_n .

The Kalman equations can then be derived by using a MAP estimate. Let the prior on the prediction, $p(\mathbf{x}_n | \mathbf{x}_{n-1})$, be determined by Equation (1). In the case of the regular Kalman Filter (a linear process), this is the sum of two multivariate Gaussian distributions. Since the Gaussian is α -stable, this sum is itself a multivariate Gaussian distribution, and can thus be described completely by finding the mean and covariance matrix. The prior on $\hat{\mathbf{x}}_n$ takes the form $\mathcal{N}(\mathbf{F}_n \hat{\mathbf{x}}_{n-1}, \mathbf{F}_n \mathbf{P}_{n-1} \mathbf{F}_n^H + \mathbf{Q}_n)$. Here \mathbf{P}_{n-1} is the correlation matrix of the previous estimate. The MAP estimate is then calculated as:

$$\arg \max_{\hat{\mathbf{x}}_n} p(\hat{\mathbf{x}}_n, \mathbf{y}_n) = \arg \max_{\hat{\mathbf{x}}_n} p(\mathbf{y}_n | \hat{\mathbf{x}}_n) p(\hat{\mathbf{x}}_n) \quad (8)$$

$$= \arg \max_{\hat{\mathbf{x}}_n} e^{-(\mathbf{y}_n - \Phi_n \hat{\mathbf{x}}_n)^H \mathbf{R}_n^{-1} (\mathbf{y}_n - \Phi_n \hat{\mathbf{x}}_n)} e^{-(\hat{\mathbf{x}}_n - \mathbf{F}_n \hat{\mathbf{x}}_{n-1})^H (\mathbf{F}_n^H \mathbf{P}_{n-1} \mathbf{F}_n + \mathbf{Q}_n)^{-1} (\hat{\mathbf{x}}_n - \mathbf{F}_n \hat{\mathbf{x}}_{n-1})} \quad (9)$$

$$= \arg \min_{\hat{\mathbf{x}}_n} (\mathbf{y}_n - \Phi_n \hat{\mathbf{x}}_n)^H \mathbf{R}_n^{-1} (\mathbf{y}_n - \Phi_n \hat{\mathbf{x}}_n) + (\hat{\mathbf{x}}_n - \mathbf{F}_n \hat{\mathbf{x}}_{n-1})^H (\mathbf{F}_n^H \mathbf{P}_{n-1} \mathbf{F}_n + \mathbf{Q}_n)^{-1} (\hat{\mathbf{x}}_n - \mathbf{F}_n \hat{\mathbf{x}}_{n-1}) \quad (10)$$

This minimum value can be found analytically by setting the derivative equal to zero:

$$0 = \frac{\partial}{\partial \hat{\mathbf{x}}_n} \left((\mathbf{y}_n - \Phi_n \hat{\mathbf{x}}_n)^H \mathbf{R}_n^{-1} (\mathbf{y}_n - \Phi_n \hat{\mathbf{x}}_n) + (\hat{\mathbf{x}}_n - \mathbf{F}_n \hat{\mathbf{x}}_{n-1})^H (\mathbf{F}_n^H \mathbf{P}_{n-1} \mathbf{F}_n + \mathbf{Q}_n)^{-1} (\hat{\mathbf{x}}_n - \mathbf{F}_n \hat{\mathbf{x}}_{n-1}) \right) \quad (11)$$

$$= \frac{\partial}{\partial \hat{\mathbf{x}}_n} (\hat{\mathbf{x}}_n^H (\Phi_n^H \mathbf{R}_n^{-1} \Phi_n + (\mathbf{F}_n \mathbf{P}_{n-1} \mathbf{F}_n^H + \mathbf{Q}_n)^{-1}) \hat{\mathbf{x}}_n - \hat{\mathbf{x}}_n^H (\Phi_n^H \mathbf{R}_n^{-1} \mathbf{y}_n + (\mathbf{F}_n \mathbf{P}_{n-1} \mathbf{F}_n^H)^{-1} \mathbf{F}_n \hat{\mathbf{x}}_{n-1}) - (\mathbf{y}_n^H \mathbf{R}_n^{-1} \Phi_n + \hat{\mathbf{x}}_{n-1}^H \mathbf{F}_n^H (\mathbf{F}_n \mathbf{P}_{n-1} \mathbf{F}_n^H + \mathbf{Q}_n)^{-1}) \hat{\mathbf{x}}_n) \quad (12)$$

$$= 2(\Phi_n^H \mathbf{R}_n^{-1} \Phi_n + (\mathbf{F}_n \mathbf{P}_{n-1} \mathbf{F}_n^H + \mathbf{Q}_n)^{-1}) \hat{\mathbf{x}}_n - 2(\Phi_n^H \mathbf{R}_n^{-1} \mathbf{y}_n + (\mathbf{F}_n \mathbf{P}_{n-1} \mathbf{F}_n^H + \mathbf{Q}_n)^{-1} \mathbf{F}_n \hat{\mathbf{x}}_{n-1}) \quad (13)$$

Let

$$\hat{\mathbf{x}}_{n|n-1} = \mathbf{F}_n \hat{\mathbf{x}}_{n-1} \quad (14)$$

$$\mathbf{P}_{n|n-1} = \mathbf{F}_n \mathbf{P}_{n-1} \mathbf{F}_n^H + \mathbf{Q}_n \quad (15)$$

be the projected mean and covariance matrix, respectively:

$$\hat{\mathbf{x}}_n = \left[\Phi_n^H \mathbf{R}_n^{-1} \Phi_n + \mathbf{P}_{n|n-1}^{-1} \right]^{-1} \left[\Phi_n^H \mathbf{R}_n^{-1} \mathbf{y}_n + \mathbf{P}_{n|n-1}^{-1} \hat{\mathbf{x}}_{n|n-1} \right] \quad (16)$$

$$= \left[\mathbf{P}_{n|n-1} - \mathbf{P}_{n|n-1} (\mathbf{R}_n + \Phi_n \mathbf{P}_{n|n-1} \Phi_n^H)^{-1} \Phi_n \mathbf{P}_{n|n-1} \right] \left[\Phi_n^H \mathbf{R}_n^{-1} \mathbf{y}_n + \mathbf{P}_{n|n-1}^{-1} \hat{\mathbf{x}}_{n|n-1} \right] \quad (17)$$

$$= \hat{\mathbf{x}}_{n|n-1} - \mathbf{K}_n \Phi_n \hat{\mathbf{x}}_{n|n-1} + \left[\mathbf{P}_{n|n-1} \Phi_n^H \mathbf{R}_n^{-1} - \mathbf{P}_{n|n-1} \Phi_n^{-1} (\mathbf{R}_n + \Phi_n \mathbf{P}_{n|n-1} \Phi_n^H)^{-1} \Phi_n \mathbf{P}_{n|n-1} \Phi_n^H \mathbf{R}_n^{-1} \right] \mathbf{y}_n \quad (18)$$

$$= \hat{\mathbf{x}}_{n|n-1} - \mathbf{K}_n \Phi_n \hat{\mathbf{x}}_{n|n-1} + \mathbf{K}_n \left[(\Phi_n \mathbf{P}_{n|n-1} \Phi_n^H + \mathbf{R}_n) \mathbf{R}_n^{-1} - \Phi_n \mathbf{P}_{n|n-1} \Phi_n^H \mathbf{R}_n^{-1} \right] \mathbf{y}_n \quad (19)$$

$$= \hat{\mathbf{x}}_{n|n-1} - \mathbf{K}_n \Phi_n \hat{\mathbf{x}}_{n|n-1} + \mathbf{K}_n \mathbf{y}_n \quad (20)$$

$$= \hat{\mathbf{x}}_{n|n-1} + \mathbf{K}_n (\mathbf{y}_n - \Phi_n \hat{\mathbf{x}}_{n|n-1}) \quad (21)$$

Where

$$\mathbf{K}_n := \mathbf{P}_{n|n-1} \Phi_n^H \left[\mathbf{R}_n + \Phi_n \mathbf{P}_{n|n-1} \Phi_n^H \right]^{-1} \quad (22)$$

is the definition of the Kalman gain at time n . This is the exact solution that the Kalman Filter should give as a best estimate of the current state. To continue propagating the estimate to future iterations, the covariance matrix \mathbf{P}_n needs to be calculated as well. \mathbf{P}_n can then be calculated by simply finding $E[\hat{\mathbf{x}}_{n+1} \hat{\mathbf{x}}_{n+1}^H]$ using the expression derived for the estimate.

$$E[\hat{\mathbf{x}}_n \hat{\mathbf{x}}_n^H] = E[(\hat{\mathbf{x}}_{n|n-1} + \mathbf{K}_n \mathbf{y}_n - \mathbf{K}_n \Phi_n \hat{\mathbf{x}}_{n|n-1})(\hat{\mathbf{x}}_{n|n-1} + \mathbf{K}_n \mathbf{y}_n - \mathbf{K}_n \Phi_n \hat{\mathbf{x}}_{n|n-1})^H] \quad (23)$$

$$= (\mathbf{I} - \mathbf{K}_n \Phi_n) \mathbf{P}_{n|n-1} (\mathbf{I} - \mathbf{K}_n \Phi_n)^H + \mathbf{K}_n \mathbf{R}_n \mathbf{K}_n^H \quad (24)$$

$$= \mathbf{P}_{n|n-1} - \mathbf{K}_n \Phi_n \mathbf{P}_{n|n-1} - \mathbf{P}_{n|n-1} \Phi_n^H \mathbf{K}_n^H + \mathbf{K}_n (\Phi_n \mathbf{P}_{n|n-1} \Phi_n^H + \mathbf{R}_n) \mathbf{K}_n^H \quad (25)$$

$$= \mathbf{P}_{n|n-1} - \mathbf{K}_n \Phi_n \mathbf{P}_{n|n-1} - \mathbf{P}_{n|n-1} \Phi_n^H \mathbf{K}_n^H + \mathbf{P}_{n|n-1} \Phi_n^H \left[\mathbf{R}_n + \Phi_n \mathbf{P}_{n|n-1} \Phi_n^H \right]^{-1} (\Phi_n \mathbf{P}_{n|n-1} \Phi_n^H + \mathbf{R}_n) \mathbf{K}_n^H \quad (26)$$

$$= \mathbf{P}_{n|n-1} - \mathbf{K}_n \Phi_n \mathbf{P}_{n|n-1} - \mathbf{P}_{n|n-1} \Phi_n^H \mathbf{K}_n^H + \mathbf{P}_{n|n-1} \Phi_n^H \mathbf{K}_n^H \quad (27)$$

$$= \mathbf{P}_{n|n-1} - \mathbf{K}_n \Phi_n \mathbf{P}_{n|n-1} \quad (28)$$

The Equations comprising the efficient Kalman Filter update are then Equations (14), (15), (22), (21), and (28).