Probabilistic Derivation of the Basic Kalman Filter*

Aaron A. D'Souza
adsouza@usc.edu

1 Introduction

Consider a system having a state $\boldsymbol{\theta}_k$ at time step k which is observed by some process generating an observation $\mathbf{z}_k$. The state transition and observation equations can be written as follows:

$$\begin{aligned}\boldsymbol{\theta}_k &= \mathbf{f}\left(\boldsymbol{\theta}_{k-1}, \mathbf{w}_k\right) \\ \mathbf{z}_k &= \mathbf{g}\left(\boldsymbol{\theta}_k, \mathbf{v}_k\right)\end{aligned}$$

where $\mathbf{w}_k$ and $\mathbf{v}_k$ are system and observation noise respectively. Given a set of observations $\mathcal{D} = \{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_k\}$, we wish to determine $p(\boldsymbol{\theta}_k|\mathcal{D}_k)$, the distribution over the state at the current time. Using Bayes rule we can write the following:

$$\begin{aligned}p(\boldsymbol{\theta}_k|\mathcal{D}_k) &= p(\boldsymbol{\theta}_k|\mathbf{z}_k, \mathcal{D}_{k-1}) && (1) \\ &\propto p(\mathbf{z}_k|\boldsymbol{\theta}_k)p(\boldsymbol{\theta}_k|\mathcal{D}_{k-1}) && (2)\end{aligned}$$

Also, given the Markov structure of the problem, we have:

$$p(\boldsymbol{\theta}_k, \boldsymbol{\theta}_{k-1}|\mathcal{D}_{k-1}) = p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1})p(\boldsymbol{\theta}_{k-1}|\mathcal{D}_{k-1}) \quad (3)$$

From Eqs. (2) and (3) we can derive the following recursive state estimaton equations:

$$\left.p(\boldsymbol{\theta}_k|\mathcal{D}_{k-1}) = \int p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1})p(\boldsymbol{\theta}_{k-1}|\mathcal{D}_{k-1})d\boldsymbol{\theta}_{k-1}\right\} \quad \text{Prediction} \quad (4)$$

$$\left.\begin{aligned}p(\boldsymbol{\theta}_k|\mathcal{D}_k) &= c_k p(\mathbf{z}_k|\boldsymbol{\theta}_k)p(\boldsymbol{\theta}_k|\mathcal{D}_{k-1}) \\ c_k &= \int p(\mathbf{z}_k|\boldsymbol{\theta}_k)p(\boldsymbol{\theta}_k|\mathcal{D}_{k-1})\end{aligned}\right\} \quad \text{Filter update} \quad (5)$$

In general these equations are difficult to evaluate. For special cases of state transition and observation functions, and noise distributions however, these equations become exactly solvable.

*This derivation is pretty detailed, and might have a typo or two. If you find any, please let me know!

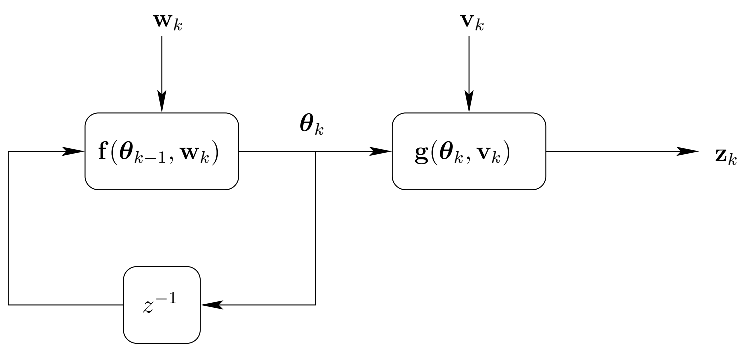

Figure 1: Generic system diagram.

2 Linear-Gaussian state space model

The Kalman filter [1] essentially solves the prediction and filter update equations for the special case in which the system transition and observation functions are linear in the state and noise, and the noise is Gaussian. In this case since the noise is Gaussian, the integrals are analytically tractable, and the linearity of the transition and observation functions ensure that the state and observation distributions retain their Gaussian form at each step. We can write the state transition and observation equations as follows:

$$\boldsymbol{\theta}_k = \mathbf{F}\boldsymbol{\theta}_{k-1} + \boldsymbol{\Phi}\mathbf{w}_k \tag{6}$$
$$\mathbf{z}_k = \mathbf{G}\boldsymbol{\theta}_k + \boldsymbol{\Psi}\mathbf{v}_k \tag{7}$$

where we assume $\mathbf{w}_k \sim \mathcal{N}\left(\mathbf{w}_k; \mathbf{0}, \mathbf{Q}\right)$, $\mathbf{w}_k \sim \mathcal{N}\left(\mathbf{v}_k; \mathbf{0}, \mathbf{R}\right)$. We define the following statistics for the prior and posterior distributions:

$$\left.\begin{aligned} \mathbf{a}_{k|k-1} &= \langle \boldsymbol{\theta}_k | \mathcal{D}_{k-1} \rangle \\ \mathbf{P}_{k|k-1} &= \mathrm{Cov}(\boldsymbol{\theta}_k | \mathcal{D}_{k-1}) \end{aligned}\right\} \quad \text{Prior mean and covariance}$$
$$\left.\begin{aligned} \mathbf{a}_{k} &= \langle \boldsymbol{\theta}_k | \mathcal{D}_{k} \rangle \\ \mathbf{P}_{k} &= \mathrm{Cov}(\boldsymbol{\theta}_k | \mathcal{D}_{k}) \end{aligned}\right\} \quad \text{Posterior mean and covariance}$$

From Eqs. (6) and (7) we can also infer the following:

$$\boldsymbol{\theta}_k | \boldsymbol{\theta}_{k-1} \sim \mathcal{N}\left(\boldsymbol{\theta}_k; \mathbf{F}\boldsymbol{\theta}_{k-1}, \boldsymbol{\Phi}\mathbf{Q}\boldsymbol{\Phi}^T\right)$$
$$\mathbf{z}_k | \boldsymbol{\theta}_k \sim \mathcal{N}\left(\mathbf{z}_k; \mathbf{G}\boldsymbol{\theta}_k, \boldsymbol{\Psi}\mathbf{R}\boldsymbol{\Psi}^T\right)$$

2.1 Prediction

Eq. (4) allows us to create a prior distribution of the current state $\boldsymbol{\theta}_k$ with only knowledge of past observations $\mathcal{D}_{k-1}$. The reason this is a *prior* distribution is because we have yet to make an observation $\mathbf{z}_k$ in this state. Once we make the observation, we will need to integrate it with this distribution to generate a *posterior* distribution.

$$
\begin{aligned}
p(\boldsymbol{\theta}_k|\mathcal{D}_{k-1}) &= \int p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1})p(\boldsymbol{\theta}_{k-1}|\mathcal{D}_{k-1})d\boldsymbol{\theta}_{k-1} & (8)\\
&= \int \mathcal{N}\left(\boldsymbol{\theta}_k; \mathbf{F}\boldsymbol{\theta}_{k-1}, \boldsymbol{\Phi}\mathbf{Q}\boldsymbol{\Phi}^T\right)\mathcal{N}\left(\boldsymbol{\theta}_{k-1}; \mathbf{a}_{k-1}, \mathbf{P}_{k-1}\right)d\boldsymbol{\theta}_{k-1} & (9)\\
&= k\int \exp\left\{-\frac{1}{2}A\right\}d\boldsymbol{\theta}_{k-1} & (10)
\end{aligned}
$$

The term A inside the exponent can be expanded as follows:

$$
\begin{aligned}
A &= \left(\boldsymbol{\theta}_k - \mathbf{F}\boldsymbol{\theta}_{k-1}\right)^T\left(\boldsymbol{\Phi}\mathbf{Q}\boldsymbol{\Phi}^T\right)^{-1}\left(\boldsymbol{\theta}_k - \mathbf{F}\boldsymbol{\theta}_{k-1}\right) + \left(\boldsymbol{\theta}_{k-1} - \mathbf{a}_{k-1}\right)^T\mathbf{P}_{k-1}^{-1}\left(\boldsymbol{\theta}_{k-1} - \mathbf{a}_{k-1}\right)\\
&= {\boldsymbol{\theta}_{k-1}}^T\underbrace{\left(\mathbf{F}^T\left(\boldsymbol{\Phi}\mathbf{Q}\boldsymbol{\Phi}^T\right)^{-1}\mathbf{F} + \mathbf{P}_{k-1}^{-1}\right)}_{\mathbf{B}}\boldsymbol{\theta}_{k-1} - 2\boldsymbol{\theta}_{k-1}^T\underbrace{\left(\mathbf{F}^T\left(\boldsymbol{\Phi}\mathbf{Q}\boldsymbol{\Phi}^T\right)^{-1}\boldsymbol{\theta}_k + \mathbf{P}_{k-1}^{-1}\mathbf{a}_{k-1}\right)}_{\mathbf{C}}\\
&\quad + {\boldsymbol{\theta}_k}^T\left(\boldsymbol{\Phi}\mathbf{Q}\boldsymbol{\Phi}^T\right)^{-1}\boldsymbol{\theta}_k + {\mathbf{a}_{k-1}}^T\mathbf{P}_{k-1}^{-1}\mathbf{a}_{k-1}
\end{aligned}
$$

This is a quadratic in $\boldsymbol{\theta}_{k-1}$, which can be written as

$$
A = \left(\boldsymbol{\theta}_{k-1} - \boldsymbol{\mu}\right)^T\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}_{k-1} - \boldsymbol{\mu}\right) - \mathbf{C}^T\mathbf{B}\mathbf{C} + {\boldsymbol{\theta}_k}^T\left(\boldsymbol{\Phi}\mathbf{Q}\boldsymbol{\Phi}^T\right)^{-1}\boldsymbol{\theta}_k + {\mathbf{a}_{k-1}}^T\mathbf{P}_{k-1}^{-1}\mathbf{a}_{k-1}
$$

where $\boldsymbol{\Sigma} = \mathbf{B}^{-1}$ and $\boldsymbol{\mu} = \boldsymbol{\Sigma}\mathbf{C}$. Substituting this result back into the exponent of Eq. (10) we get:

$$
\begin{aligned}
p(\boldsymbol{\theta}_k|\mathcal{D}_{k-1}) &= k\int \exp\left\{-\frac{1}{2}A\right\}d\boldsymbol{\theta}_{k-1}\\
&= k\exp\left\{-\frac{1}{2}\left[{\boldsymbol{\theta}_k}^T\left(\boldsymbol{\Phi}\mathbf{Q}\boldsymbol{\Phi}^T\right)^{-1}\boldsymbol{\theta}_k + {\mathbf{a}_{k-1}}^T\mathbf{P}_{k-1}^{-1}\mathbf{a}_{k-1} - \mathbf{C}^T\mathbf{B}\mathbf{C}\right]\right\}\\
&\qquad \cdot\int \exp\left\{-\frac{1}{2}\left(\boldsymbol{\theta}_{k-1} - \boldsymbol{\mu}\right)^T\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}_{k-1} - \boldsymbol{\mu}\right)\right\}d\boldsymbol{\theta}_{k-1}\\
&= k\left(2\pi\right)^{d/2}\left|\boldsymbol{\Sigma}\right|^{1/2}\exp\left\{-\frac{1}{2}\underbrace{\left[{\boldsymbol{\theta}_k}^T\left(\boldsymbol{\Phi}\mathbf{Q}\boldsymbol{\Phi}^T\right)^{-1}\boldsymbol{\theta}_k + {\mathbf{a}_{k-1}}^T\mathbf{P}_{k-1}^{-1}\mathbf{a}_{k-1} - \mathbf{C}^T\mathbf{B}\mathbf{C}\right]}_{D}\right\}
\end{aligned}
$$

This distribution has a quadratic inside an exponential term implying that $p(\boldsymbol{\theta}_k|\mathcal{D}_{k-1})$ has a Gaussian form. We can expand the quadratic term D as follows:

$$
\begin{aligned}
D = &-\left(\mathbf{F}^T\left(\mathbf{\Phi}\mathbf{Q}\mathbf{\Phi}^T\right)^{-1}\boldsymbol{\theta}_k + \mathbf{P}_{k-1}^{-1}\mathbf{a}_{k-1}\right)^T\left(\mathbf{F}^T\left(\mathbf{\Phi}\mathbf{Q}\mathbf{\Phi}^T\right)^{-1}\mathbf{F} + \mathbf{P}_{k-1}^{-1}\right)^{-1}\left(\mathbf{F}^T\left(\mathbf{\Phi}\mathbf{Q}\mathbf{\Phi}^T\right)^{-1}\boldsymbol{\theta}_k + \mathbf{P}_{k-1}^{-1}\mathbf{a}_{k-1}\right) \\
&+ \boldsymbol{\theta}_k{}^T\left(\mathbf{\Phi}\mathbf{Q}\mathbf{\Phi}^T\right)^{-1}\boldsymbol{\theta}_k + \mathbf{a}_{k-1}{}^T\mathbf{P}_{k-1}^{-1}\mathbf{a}_{k-1}
\end{aligned}
$$

Consider terms in D that are quadratic in $\boldsymbol{\theta}_k$:

$$
\begin{aligned}
D_2 &= \boldsymbol{\theta}_k{}^T\left(\mathbf{\Phi}\mathbf{Q}\mathbf{\Phi}^T\right)^{-1}\boldsymbol{\theta}_k - \boldsymbol{\theta}_k{}^T\left(\mathbf{\Phi}\mathbf{Q}\mathbf{\Phi}^T\right)^{-1}\mathbf{F}\left(\mathbf{F}^T\left(\mathbf{\Phi}\mathbf{Q}\mathbf{\Phi}^T\right)^{-1}\mathbf{F} + \mathbf{P}_{k-1}^{-1}\right)^{-1}\mathbf{F}^T\left(\mathbf{\Phi}\mathbf{Q}\mathbf{\Phi}^T\right)^{-1}\boldsymbol{\theta}_k \\
&= \boldsymbol{\theta}_k{}^T\left[\left(\mathbf{\Phi}\mathbf{Q}\mathbf{\Phi}^T\right)^{-1} - \left(\mathbf{\Phi}\mathbf{Q}\mathbf{\Phi}^T\right)^{-1}\mathbf{F}\left(\mathbf{F}^T\left(\mathbf{\Phi}\mathbf{Q}\mathbf{\Phi}^T\right)^{-1}\mathbf{F} + \mathbf{P}_{k-1}^{-1}\right)^{-1}\mathbf{F}^T\left(\mathbf{\Phi}\mathbf{Q}\mathbf{\Phi}^T\right)^{-1}\right]\boldsymbol{\theta}_k \\
&= \boldsymbol{\theta}_k{}^T\left(\mathbf{F}\mathbf{P}_{k-1}\mathbf{F}^T + \mathbf{\Phi}\mathbf{Q}\mathbf{\Phi}^T\right)^{-1}\boldsymbol{\theta}_k
\end{aligned}
$$

Since this is the only term in D that is quadratic in $\boldsymbol{\theta}_k$, we can infer the covariance of the distribution as follows:

$$
\Rightarrow \mathbf{P}_{k|k-1} = \mathbf{F}\mathbf{P}_{k-1}\mathbf{F}^T + \mathbf{\Phi}\mathbf{Q}\mathbf{\Phi}^T
$$

Consider terms in D that are linear in $\boldsymbol{\theta}_k$

$$
\begin{aligned}
D_1 &= -2\boldsymbol{\theta}_k\left(\mathbf{\Phi}\mathbf{Q}\mathbf{\Phi}^T\right)^{-1}\mathbf{F}\left(\mathbf{F}^T\left(\mathbf{\Phi}\mathbf{Q}\mathbf{\Phi}^T\right)^{-1}\mathbf{F} + \mathbf{P}_{k-1}^{-1}\right)^{-1}\mathbf{P}_{k-1}^{-1}\mathbf{a}_{k-1} \\
&= -2\boldsymbol{\theta}_k\left(\mathbf{\Phi}\mathbf{Q}\mathbf{\Phi}^T\right)^{-1}\mathbf{F}\left[\mathbf{P}_{k-1} - \mathbf{P}_{k-1}\mathbf{F}^T\left(\mathbf{\Phi}\mathbf{Q}\mathbf{\Phi}^T + \mathbf{F}\mathbf{P}_{k-1}\mathbf{F}^T\right)^{-1}\mathbf{F}\mathbf{P}_{k-1}\right]\mathbf{P}_{k-1}^{-1}\mathbf{a}_{k-1} \\
&= -2\boldsymbol{\theta}_k\left(\mathbf{\Phi}\mathbf{Q}\mathbf{\Phi}^T\right)^{-1}\left[\mathbf{F} - \mathbf{F}\mathbf{P}_{k-1}\mathbf{F}^T\left(\mathbf{\Phi}\mathbf{Q}\mathbf{\Phi}^T + \mathbf{F}\mathbf{P}_{k-1}\mathbf{F}^T\right)^{-1}\mathbf{F}\right]\mathbf{a}_{k-1} \\
&= -2\boldsymbol{\theta}_k\left(\mathbf{\Phi}\mathbf{Q}\mathbf{\Phi}^T\right)^{-1}\left[\begin{array}{c}\mathbf{F} - \mathbf{F}\mathbf{P}_{k-1}\mathbf{F}^T\left(\mathbf{\Phi}\mathbf{Q}\mathbf{\Phi}^T + \mathbf{F}\mathbf{P}_{k-1}\mathbf{F}^T\right)^{-1}\mathbf{F} \\ -\mathbf{\Phi}\mathbf{Q}\mathbf{\Phi}^T\left(\mathbf{\Phi}\mathbf{Q}\mathbf{\Phi}^T + \mathbf{F}\mathbf{P}_{k-1}\mathbf{F}^T\right)^{-1}\mathbf{F} \\ +\mathbf{\Phi}\mathbf{Q}\mathbf{\Phi}^T\left(\mathbf{\Phi}\mathbf{Q}\mathbf{\Phi}^T + \mathbf{F}\mathbf{P}_{k-1}\mathbf{F}^T\right)^{-1}\mathbf{F}\end{array}\right]\mathbf{a}_{k-1} \\
&= -2\boldsymbol{\theta}_k\left(\mathbf{\Phi}\mathbf{Q}\mathbf{\Phi}^T + \mathbf{F}\mathbf{P}_{k-1}\mathbf{F}^T\right)^{-1}\mathbf{F}\mathbf{a}_{k-1} \\
&= -2\boldsymbol{\theta}_k\mathbf{P}_{k|k-1}^{-1}\mathbf{F}\mathbf{a}_{k-1} \\
&\Rightarrow \mathbf{a}_{k|k-1} = \mathbf{F}\mathbf{a}_{k-1}
\end{aligned}
$$

Hence, the prior distribution $p(\boldsymbol{\theta}_k|\mathcal{D}_{k-1})$ is Gaussian, with mean and variance given by:

$$
\mathbf{a}_{k|k-1} = \mathbf{F}\mathbf{a}_{k-1} \tag{11}
$$

$$
\mathbf{P}_{k|k-1} = \mathbf{F}\mathbf{P}_{k-1}\mathbf{F}^T + \mathbf{\Phi}\mathbf{Q}\mathbf{\Phi}^T \tag{12}
$$

2.2 Filter Update

$$p(\boldsymbol{\theta}_k|\mathcal{D}_k) \propto p(\mathbf{z}_k|\boldsymbol{\theta}_k)p(\boldsymbol{\theta}_k|\mathcal{D}_{k-1})$$
$$\propto \exp\left\{-\frac{1}{2}\underbrace{\left[(\mathbf{z}_k - \mathbf{G}\boldsymbol{\theta}_k)^T\left(\boldsymbol{\Psi}\mathbf{R}\boldsymbol{\Psi}^T\right)^{-1}(\mathbf{z}_k - \mathbf{G}\boldsymbol{\theta}_k) + \left(\boldsymbol{\theta}_k - \mathbf{a}_{k|k-1}\right)^T\mathbf{P}_{k|k-1}^{-1}\left(\boldsymbol{\theta}_k - \mathbf{a}_{k|k-1}\right)\right]}_{E}\right\}$$

Now:

$$E = (\mathbf{z}_k - \mathbf{G}\boldsymbol{\theta}_k)^T\left(\boldsymbol{\Psi}\mathbf{R}\boldsymbol{\Psi}^T\right)^{-1}(\mathbf{z}_k - \mathbf{G}\boldsymbol{\theta}_k) + \left(\boldsymbol{\theta}_k - \mathbf{a}_{k|k-1}\right)^T\mathbf{P}_{k|k-1}^{-1}\left(\boldsymbol{\theta}_k - \mathbf{a}_{k|k-1}\right)$$
$$= {\boldsymbol{\theta}_k}^T\left(\mathbf{G}^T\left(\boldsymbol{\Psi}\mathbf{R}\boldsymbol{\Psi}^T\right)^{-1}\mathbf{G} + \mathbf{P}_{k|k-1}^{-1}\right)\boldsymbol{\theta}_k - 2\boldsymbol{\theta}_k^T\left(\mathbf{G}^T\left(\boldsymbol{\Psi}\mathbf{R}\boldsymbol{\Psi}^T\right)^{-1}\mathbf{z}_k + \mathbf{P}_{k|k-1}^{-1}\mathbf{a}_{k|k-1}\right) + k$$

From which we can deduce:

$$\mathbf{P}_k = \left(\mathbf{G}^T\left(\boldsymbol{\Psi}\mathbf{R}\boldsymbol{\Psi}^T\right)^{-1}\mathbf{G} + \mathbf{P}_{k|k-1}^{-1}\right)^{-1}$$
$$= \mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}\mathbf{G}^T\left(\boldsymbol{\Psi}\mathbf{R}\boldsymbol{\Psi}^T + \mathbf{G}^T\mathbf{P}_{k|k-1}\mathbf{G}\right)^{-1}\mathbf{G}\mathbf{P}_{k|k-1}$$
$$= \mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}\mathbf{G}^T\mathbf{K}^{-1}\mathbf{G}\mathbf{P}_{k|k-1} \qquad \text{where } \mathbf{K} = \boldsymbol{\Psi}\mathbf{R}\boldsymbol{\Psi}^T + \mathbf{G}^T\mathbf{P}_{k|k-1}\mathbf{G}$$

Also

$$\mathbf{a}_k = \mathbf{P}_k\left(\mathbf{G}^T\left(\boldsymbol{\Psi}\mathbf{R}\boldsymbol{\Psi}^T\right)^{-1}\mathbf{z}_k + \mathbf{P}_{k|k-1}^{-1}\mathbf{a}_{k|k-1}\right)$$
$$= \left(\mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}\mathbf{G}^T\mathbf{K}^{-1}\mathbf{G}\mathbf{P}_{k|k-1}\right)\left(\mathbf{G}^T\left(\boldsymbol{\Psi}\mathbf{R}\boldsymbol{\Psi}^T\right)^{-1}\mathbf{z}_k + \mathbf{P}_{k|k-1}^{-1}\mathbf{a}_{k|k-1}\right)$$
$$= \mathbf{a}_{k|k-1} - \mathbf{P}_{k|k-1}\mathbf{G}^T\mathbf{K}^{-1}\mathbf{G}\mathbf{a}_{k|k-1} + \mathbf{P}_{k|k-1}\mathbf{G}^T\underbrace{\left[\mathbf{I} - \mathbf{K}^{-1}\mathbf{G}\mathbf{P}_{k|k-1}\mathbf{G}^T\right]}_{F}\left(\boldsymbol{\Psi}\mathbf{R}\boldsymbol{\Psi}^T\right)^{-1}\mathbf{z}_k$$

Now:

$$F = \mathbf{I} - \mathbf{K}^{-1}\mathbf{G}\mathbf{P}_{k|k-1}\mathbf{G}^T$$
$$= \mathbf{I} - \mathbf{K}^{-1}\mathbf{G}\mathbf{P}_{k|k-1}\mathbf{G}^T - \mathbf{K}^{-1}\boldsymbol{\Psi}\mathbf{R}\boldsymbol{\Psi}^T + \mathbf{K}^{-1}\boldsymbol{\Psi}\mathbf{R}\boldsymbol{\Psi}^T$$
$$= \mathbf{K}^{-1}\boldsymbol{\Psi}\mathbf{R}\boldsymbol{\Psi}^T$$

Hence

$$\mathbf{a}_k = \mathbf{a}_{k|k-1} + \mathbf{P}_{k|k-1}\mathbf{G}^T\mathbf{K}^{-1}\left(\mathbf{z}_k - \mathbf{G}\mathbf{a}_{k|k-1}\right)$$

This allows us to characterize the posterior distribution $p(\boldsymbol{\theta}_k|\mathcal{D}_k)$ as a Gaussian with the following mean and variance:

$$\begin{aligned}
\mathbf{a}_k &= \mathbf{a}_{k|k-1} + \mathbf{P}_{k|k-1}\mathbf{G}^T\mathbf{K}^{-1}\left(\mathbf{z}_k - \mathbf{G}\mathbf{a}_{k|k-1}\right) \\
\mathbf{P}_k &= \mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}\mathbf{G}^T\mathbf{K}^{-1}\mathbf{G}\mathbf{P}_{k|k-1}
\end{aligned}$$

where

$$\mathbf{K} = \mathbf{\Psi}\mathbf{R}\mathbf{\Psi}^T + \mathbf{G}^T\mathbf{P}_{k|k-1}\mathbf{G}$$

References

[1] R. E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME — Journal of Basic Engineering*, 82:35–45, 1960.