Gaussian Product Derivation of Kalman Filter

Julius S. Gyorfi
KBR, Inc. at NASA Ames Research Center
Moffett Field, CA 94035 USA
Julius.gyorfi@us.kbr.com

***Abstract*—A common form of the discrete-time Kalman filter equations in state space is derived as a product of Gaussians in measurement space. The derivation is generalized to multidimensional spaces and explicitly shows all the steps in the interest of clarity.**

I. Introduction

THE Kalman filter is as indispensable as it is inscrutable. Its numerous applications and extensions are an integral part of modern engineering practices. There is, however, a steep learning curve for those who are trying to grasp the fundamental principles and underlying theory. Although a plethora of articles have been devoted to these topics, they tend to fall into one of two categories: either highly theoretical or too oversimplified. Kalman's original paper [1] is an example of the former. Alternatively, Welch and Bishop [2] offer an excellent introductory tutorial without delving into the derivation of the equations. A middle ground is possible. Given the conditions that guarantee optimality, the filter equations can be directly derived as the product of the Gaussian noise distributions. Faragher [3] alluded to this, albeit in the context of one-dimensional state and measurement vectors. We generalize the derivation to the multidimensional case which is nontrivial due to the noncommutative nature of matrix multiplication, although our figures are one-dimensional out of necessity.

A. Stochastic Linear Systems

Kalman formulated the filter as a state space solution of the Weiner optimal filtering problem [1] which can be stated succinctly as estimating the actual states of a system from a set of noisy measurements. There are three conditions in particular that enable an optimal solution to this problem: (i) that the system in question is linear, (ii) that the process noise is normally distributed with zero mean, and (iii) that the measurement noise is normally distributed with zero mean.

The first optimality condition is satisfied by a general discrete-time linear, time-invariant system:

$$\mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} + \mathbf{w} \tag{1}$$

where $\mathbf{x}_k$ is an $n \times 1$ state vector at time k, $\mathbf{A}$ is an $n \times n$ state transition matrix, and $\mathbf{w}$ is an $n \times 1$ zero-mean, normally distributed process noise random vector

$$p(\mathbf{w}) \;\sim\; \mathcal{N}(\mathbf{0}, \mathbf{Q}) \tag{2}$$

which satisfies the second optimality condition. The covariance $\mathbf{Q}$ is an $n \times n$ matrix defined as

$$\mathbf{Q} \;\equiv\; \mathbb{E}[\mathbf{w}\mathbf{w}^T] \tag{3}$$

Because the process noise is a normally distributed random vector, the state vector $\mathbf{x}$ will also be defined by a normal distribution

$$p(\mathbf{x}_k) \;\sim\; \mathcal{N}(\bar{\mathbf{x}}_k, \mathbf{P}_k) \tag{4}$$

where $\bar{\mathbf{x}}_k$ is the expected value of the state vector

$$\bar{\mathbf{x}}_k \;\equiv\; \mathbb{E}[\mathbf{x}_k] \tag{5}$$

and $\mathbf{P}_k$ is the $n \times n$ covariance matrix for $\mathbf{x}_k$

$$\mathbf{P}_k \;\equiv\; \mathbb{E}[\mathbf{x}_k\mathbf{x}_k^T] \tag{6}$$

The state $\mathbf{x}_k$ is assumed to be unobservable but it can be indirectly measured as a linear combination of some or all of the states:

$$\mathbf{z}_k \;=\; \mathbf{H}\mathbf{x}_k + \mathbf{v} \tag{7}$$

where $\mathbf{z}_k$ is an $m \times 1$ measurement vector at time k, $\mathbf{H}$ is an $m \times n$ measurement matrix, and $\mathbf{v}$ is an $m \times 1$ zero-mean, normally distributed measurement noise random vector which satisfies the third optimality condition. The random nature of the measurement noise means the measurement vector will be defined by a normal distribution

$$p(\mathbf{v}) \;\sim\; \mathcal{N}(\mathbf{0}, \mathbf{R}) \tag{8}$$

where $\mathbf{R}$ is the measurement covariance

$$\mathbf{R} \;\equiv\; \mathbb{E}[\mathbf{v}\mathbf{v}^T] \tag{9}$$

The measurement space may have different dimensionality from the state space ($m \neq n$) and may be in different units. We will examine the behavior of the state distribution in both state and measurement spaces as we derive the Kalman filter equations. An important consideration to remember is that, while it is always possible to go from state space to measurement space via multiplication by the $\mathbf{H}$ matrix (see Section II), it is generally not possible to perform the inverse operation since $\mathbf{H}$ is not guaranteed to be square and invertible. Figure 1 shows an example state distribution in both spaces.

B. Kalman Filter Equations

Welch and Bishop [2] present the Kalman filter as a set of five equations which we reproduce here as (10) to (14). Equations (10) and (11) comprise the *prediction* or *time update* steps in which the current state is propagated forward in time. The $\mathbf{x}_k^-$ term in (10) is referred to as the *a priori* state estimate because it has not taken into account the measurement at time k. Similarly, $\mathbf{P}_k^-$ in (11) is the *a priori* state covariance at time k.

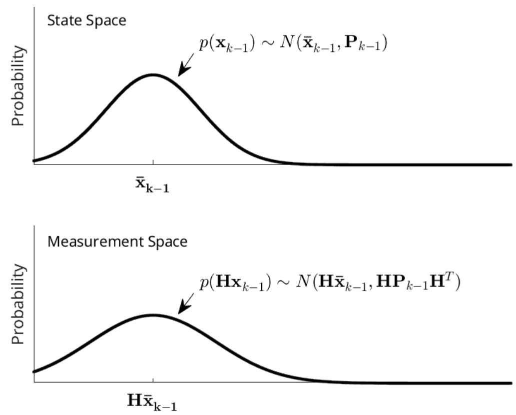

Fig. 1. State Distribution at Time $k-1$ Shown in Both State and Measurement Spaces.

$$\mathbf{x}_k^- = \mathbf{A}\mathbf{x}_{k-1} \tag{10}$$

$$\mathbf{P}_k^- = \mathbf{A}\mathbf{P}_{k-1}\mathbf{A}^T + \mathbf{Q} \tag{11}$$

$$\mathbf{K} = \mathbf{P}_k^-\mathbf{H}^T(\mathbf{H}^T\mathbf{P}_k^-\mathbf{H}^T + \mathbf{R})^{-1} \tag{12}$$

$$\mathbf{x}_k = \mathbf{x}_k^- + \mathbf{K}(\mathbf{z}_k - \mathbf{H}\mathbf{x}_k^-) \tag{13}$$

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}_k^- \tag{14}$$

Equations (12) through (14) are the *correction* or *measurement* equations. Equation (12) defines the $n \times m$ Kalman gain matrix $\mathbf{K}$ which is applied to the *residual* or *innovation* term $\mathbf{z}_k - \mathbf{H}\mathbf{x}_k^-$ in (13) to produce a correction to the *a priori* state estimate $\mathbf{x}_k^-$ which yields the *a posteriori* state estimate $\mathbf{x}_k$. The residual is the difference between the current measurement and the *a priori* state estimate, which must be premultiplied by $\mathbf{H}$ to be converted into measurement space (see Fig. 2.) The Kalman gain is also used to update the state covariance matrix in (14).

In the following sections we will derived each of the equations (10) to (14) and show that the measurement equations (12) to (14) can be calculated from the product of the process noise distribution and the measurement noise distribution. For reference, the formula for the product of Gaussians is given in the Appendix.

II. Derivation of Prediction Equations

We begin deriving (10) by applying the expectation operator $\mathbb{E}[\ ]$ to (1):

$$\mathbb{E}[\mathbf{x}_k^-] = \mathbb{E}[\mathbf{A}\mathbf{x}_{k-1} + \mathbf{w}] \tag{15}$$

Due to the linearity of the expectation operator, this can be expanded into

$$\mathbb{E}[\mathbf{x}_k^-] = \mathbf{A}\mathbb{E}[\mathbf{x}_{k-1}] + \mathbb{E}[\mathbf{w}] \tag{16}$$

Substituting (5) yields

$$\bar{\mathbf{x}}_k^- = \mathbf{A}\bar{\mathbf{x}}_{k-1} + \mathbb{E}[\mathbf{w}] \tag{17}$$

As noted above, the process noise random vector $\mathbf{w}$ is normally distributed with zero mean which means $\mathbb{E}[\mathbf{w}] = 0$.

$$\bar{\mathbf{x}}_k^- = \mathbf{A}\bar{\mathbf{x}}_{k-1} \quad \blacksquare \tag{18}$$

We also need to project the state covariance matrix forward in time. From (6), the covariance of the state vector $\mathbf{x}_k^-$ is

$$\mathbf{P}_k^- = \mathbb{E}[(\mathbf{x}_k^-)(\mathbf{x}_k^-)^T] \tag{19}$$

Substituting (1) into (19)

$$\mathbf{P}_k^- = \mathbb{E}[(\mathbf{A}\mathbf{x}_{k-1} + \mathbf{w})(\mathbf{A}\mathbf{x}_{k-1} + \mathbf{w})^T] \tag{20}$$

The transpose of a sum is equal to the sum of transposes and the transpose of a product is equal to the product of the commuted transposes:

$$\mathbf{P}_k^- = \mathbb{E}[(\mathbf{A}\mathbf{x}_{k-1} + \mathbf{w})(\mathbf{w}^T + (\mathbf{x}_{k-1})^T\mathbf{A}^T)] \tag{21}$$

Expanding and applying the linearity of the expectation operator

$$\begin{aligned}\mathbf{P}_k^- = \;& \mathbb{E}[\mathbf{A}\mathbf{x}_{k-1}\mathbf{w}^T] + \\ & \mathbb{E}[\mathbf{A}\mathbf{x}_{k-1}(\mathbf{x}_{k-1})^T\mathbf{A}^T] + \\ & \mathbb{E}[\mathbf{w}\mathbf{w}^T] + \\ & \mathbb{E}[\mathbf{w}(\mathbf{x}_{k-1})^T\mathbf{A}^T]\end{aligned} \tag{22}$$

The state vector $\mathbf{x}_{k-1}$ and the process noise $\mathbf{w}$ are assumed to be uncorrelated meaning their covariance terms $\mathbb{E}[\mathbf{A}\mathbf{x}_{k-1}\mathbf{w}^T]$ and $\mathbb{E}[\mathbf{w}(\mathbf{x}_{k-1})^T\mathbf{A}^T]$ equal zero. Removing these terms and simplifying results in

$$\mathbf{P}_k^- = \mathbf{A}\mathbb{E}[\mathbf{x}_{k-1}(\mathbf{x}_{k-1})^T]\mathbf{A}^T + \mathbb{E}[\mathbf{w}\mathbf{w}^T] \tag{23}$$

Substituting $\mathbf{P}_{k-1}$ for $\mathbb{E}[\mathbf{x}_{k-1}(\mathbf{x}_{k-1})^T]$ from (6) and $\mathbf{Q}$ for $\mathbb{E}[\mathbf{w}\mathbf{w}^T]$ from (3),

$$\mathbf{P}_k^- = \mathbf{A}\mathbf{P}_{k-1}\mathbf{A}^T + \mathbf{Q} \quad \blacksquare \tag{24}$$

Figure 2 shows the state distribution from Fig. 1 before prediction using a dashed line and the state distribution after prediction using a solid line. The predicted state distribution is normally distributed with mean $\bar{\mathbf{x}}_{k-1}^-$ and covariance $\mathbf{P}_{k-1}^-$ in state space. In measurement space, $\bar{\mathbf{x}}_k^-$ becomes $\mathbf{H}\bar{\mathbf{x}}_k^-$ via (7) and $\mathbf{P}_k^-$ becomes $\mathbf{H}\mathbf{P}_k^-\mathbf{H}^T$ from the simplification of $\mathbb{E}[(\mathbf{H}\mathbf{x})(\mathbf{H}\mathbf{x})^T]$.

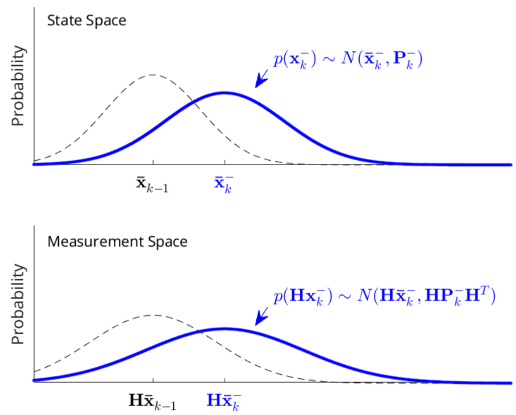

Fig. 2. Predicted State Distribution at Time k Shown in Both State and Measurement Spaces.

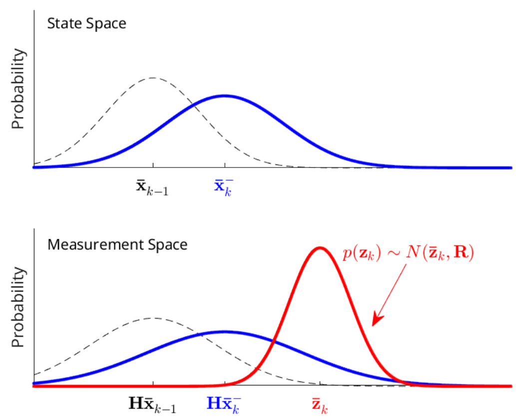

Fig. 3. Predicted State and Measurement Distributions at Time k. Note that there is no corresponding measurement distribution in state space.

III. Derivation of Correction Equations

The measurement $\mathbf{z}_k$ is a Gaussian distribution with mean $\bar{\mathbf{z}}_k$ where

$$\bar{\mathbf{z}}_k \equiv \mathbb{E}[\mathbf{z}_k] \tag{25}$$

and covariance $\mathbf{R}$ as shown in Fig. 3. Note that there is no distribution corresponding to $\mathbf{z}_k$ in state space because that would require premultiplying the measurement distribution by $\mathbf{H}^{-1}$ which is not guaranteed to exist.

The predicted state $\mathbf{H}\mathbf{x}_k^-$ and measurement $\mathbf{z}_k$ provide two different estimates of the actual state in measurement space, both normally distributed. The product of these distributions will combine the uncertainties in each to produce the best estimate of the actual state distribution at time k. Substituting the means and variances of the predicted state and measurement distributions into (40) and (41) from the Appendix will result in the new state distribution at time k with $\mu_{12} = \mathbf{H}\bar{\mathbf{x}}_k$ and $\Sigma_{12} = \mathbf{H}\mathbf{P}_k\mathbf{H}^T$ in measurement space:

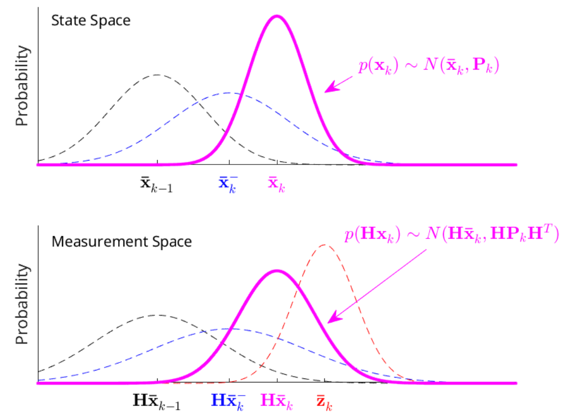

Fig. 4. State Distribution at Time k Shown in Both State and Measurement Spaces.

$$\begin{aligned}\mathbf{H}\mathbf{x}_k &= \mathbf{R}(\mathbf{H}\mathbf{P}_k^-\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{x}_k^- + \\ &\quad \mathbf{H}\mathbf{P}_k^-\mathbf{H}^T(\mathbf{H}\mathbf{P}_k^-\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{z}_k\end{aligned} \tag{26}$$

$$\mathbf{H}\mathbf{P}_k\mathbf{H}^T = \mathbf{H}\mathbf{P}_k^-\mathbf{H}^T(\mathbf{H}\mathbf{P}_k^-\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{R} \tag{27}$$

Even though $\mathbf{H}$ is not guaranteed to be invertible, it is evident that the left-hand side of both (26) and (27) depend upon $\mathbf{H}$ in a manner that enables us to determine the corresponding distribution in the state space by inspection: mean $\bar{\mathbf{x}}_k$ and covariance $\mathbf{P}_k$ in state space (see Fig. 4.) Note that we cannot perform the same transformation from measurement space to state space for the right-hand side of (26) and (27).

A. Kalman Gain and State Update

We now consider the error between $\mathbf{H}\mathbf{x}_k$ and $\mathbf{H}\mathbf{x}_k^-$ in measurement space. Subtracting $\mathbf{H}\mathbf{x}_k^-$ from both sides of (26) yields

$$\begin{aligned}\mathbf{H}\mathbf{x}_k - \mathbf{H}\mathbf{x}_k^- &= \mathbf{R}(\mathbf{H}\mathbf{P}_k^-\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{x}_k^- + \\ &\quad \mathbf{H}\mathbf{P}_k^-\mathbf{H}^T(\mathbf{H}\mathbf{P}_k^-\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{z}_k - \\ &\quad \mathbf{H}\mathbf{x}_k^-\end{aligned} \tag{28}$$

To simplify this, we can premultiply the $\mathbf{H}\mathbf{x}_k^-$ term on the right-hand side by $(\mathbf{H}\mathbf{P}_k^-\mathbf{H}^T + \mathbf{R})(\mathbf{H}\mathbf{P}_k^-\mathbf{H}^T + \mathbf{R})^{-1}$, which is equivalent to the identity matrix

$$\begin{aligned}&\mathbf{H}\mathbf{x}_k - \mathbf{H}\mathbf{x}_k^- = \\ &\quad \mathbf{R}(\mathbf{H}\mathbf{P}_k^-\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{x}_k^- + \\ &\quad \mathbf{H}\mathbf{P}_k^-\mathbf{H}^T(\mathbf{H}\mathbf{P}_k^-\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{z}_k - \\ &\quad (\mathbf{H}\mathbf{P}_k^-\mathbf{H}^T + \mathbf{R})(\mathbf{H}\mathbf{P}_k^-\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{x}_k^-\end{aligned} \tag{29}$$

Expanding

$$\begin{aligned}&\mathbf{H}\mathbf{x}_k - \mathbf{H}\mathbf{x}_k^- = \\ &\quad \mathbf{R}(\mathbf{H}\mathbf{P}_k^-\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{x}_k^- + \\ &\quad \mathbf{H}\mathbf{P}_k^-\mathbf{H}^T(\mathbf{H}\mathbf{P}_k^-\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{z}_k - \\ &\quad \mathbf{H}\mathbf{P}_k^-\mathbf{H}^T(\mathbf{H}\mathbf{P}_k^-\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{x}_k^- - \\ &\quad \mathbf{R}(\mathbf{H}\mathbf{P}_k^-\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{x}_k^-\end{aligned} \tag{30}$$

The right-hand side of (30) can be simplified since the first and last terms cancel and the remaining terms can be combined into

$$\begin{aligned}\mathbf{H}(\mathbf{x}_k - \mathbf{x}_k^-) \; &= \\ &\mathbf{HP}_k^-\mathbf{H}^T(\mathbf{HP}_k^-\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{z}_k - \mathbf{Hx}_k^-)\end{aligned} \tag{31}$$

in measurement space and

$$\mathbf{x}_k - \mathbf{x}_k^- = \mathbf{P}_k^-\mathbf{H}^T(\mathbf{HP}_k^-\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{z}_k - \mathbf{Hx}_k^-) \tag{32}$$

in the state space. Equation (32) is equivalent to (13) with the Kalman gain $\mathbf{K}$ defined to be

$$\mathbf{K} \;\equiv\; \mathbf{P}_k^-\mathbf{H}^T(\mathbf{HP}_k^-\mathbf{H}^T + \mathbf{R})^{-1} \qquad \blacksquare \tag{33}$$

as in (12).

B. Covariance Update

All that remains to complete the derivation is the state covariance update equation (14). To derive that, we begin the same way we did before in (28) when we examined the difference between $\mathbf{Hx}_k$ and $\mathbf{Hx}_k^-$ in measurement space. This time, however, we will consider the difference between the covariances $\mathbf{HP}_k\mathbf{H}^T$ and $\mathbf{HP}_k^-\mathbf{H}^T$, substituting (27) for $\mathbf{HP}_k\mathbf{H}^T$:

$$\begin{aligned}\mathbf{HP}_k\mathbf{H}^T - \mathbf{HP}_k^-\mathbf{H}^T \; &= \\ &\mathbf{HP}_k^-\mathbf{H}^T(\mathbf{HP}_k^-\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{R} - \\ &\mathbf{HP}_k^-\mathbf{H}^T\end{aligned} \tag{34}$$

We will utilize the same trick as before, this time post-multiplying $\mathbf{HP}_k^-\mathbf{H}^T$ on the right-hand side by $(\mathbf{HP}_k^-\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{HP}_k^-\mathbf{H}^T + \mathbf{R})$.

$$\begin{aligned}\mathbf{HP}_k\mathbf{H}^T - \mathbf{HP}_k^-\mathbf{H}^T \; &= \\ &\mathbf{HP}_k^-\mathbf{H}^T(\mathbf{HP}_k^-\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{R} - \\ &\mathbf{HP}_k^-\mathbf{H}^T(\mathbf{HP}_k^-\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{HP}_k^-\mathbf{H}^T + \mathbf{R})\end{aligned} \tag{35}$$

Expanding

$$\begin{aligned}\mathbf{HP}_k\mathbf{H}^T - \mathbf{HP}_k^-\mathbf{H}^T &= \\ &\mathbf{HP}_k^-\mathbf{H}^T(\mathbf{HP}_k^-\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{R} - \\ &\mathbf{HP}_k^-\mathbf{H}^T(\mathbf{HP}_k^-\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{HP}_k^-\mathbf{H}^T - \\ &\mathbf{HP}_k^-\mathbf{H}^T(\mathbf{HP}_k^-\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{R}\end{aligned} \tag{36}$$

The first and last terms on the right cancel. Substituting (33) into this then simplifying produces

$$\mathbf{HP}_k\mathbf{H}^T - \mathbf{HP}_k^-\mathbf{H}^T = -\mathbf{H}(\mathbf{KHP}_k^-)\mathbf{H}^T \tag{37}$$

Each term in (37) is premultiplied by $\mathbf{H}$ and postmultiplied by $\mathbf{H}^T$ in measurement space. The corresponding equation in state space is

$$\mathbf{P}_k - \mathbf{P}_k^- = -\mathbf{KHP}_k^- \tag{38}$$

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{KH})\mathbf{P}_k^- \qquad \blacksquare \tag{39}$$

as in (14).

IV. Conclusion

We derived a common form of the Kalman filter equations in state space using a product of Gaussians in measurement space. Our intention was to provide an alternative to those references and tutorials that are either too theoretical or too oversimplified for a neophyte to gain any insight into the principles underlying the filter.

Appendix
Product of Gaussians

Bromiley [4] derived the product of two Gaussians with scalar means and variances. We generalize their result to vector means and matrix covariances. Given Gaussians $g_1 = \mathcal{N}(\mu_{\mathbf{1}}, \mathbf{\Sigma}_1)$ and $g_2 = \mathcal{N}(\mu_{\mathbf{2}}, \mathbf{\Sigma}_2)$, where μ_1 and μ_2 are $n \times 1$ vector means and $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ are $n \times n$ covariance matrices, the product g_{12} has mean μ_{12} given by (40) and covariance $\mathbf{\Sigma}_{12}$ given by (41):

$$\mu_{\mathbf{12}} = \mathbf{\Sigma}_2(\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2)^{-1}\mu_{\mathbf{1}} + \mathbf{\Sigma}_1(\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2)^{-1}\mu_{\mathbf{2}} \tag{40}$$

$$\mathbf{\Sigma}_{12} = \mathbf{\Sigma}_1(\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2)^{-1}\mathbf{\Sigma}_2 \tag{41}$$

References

[1] Kalman, R. E., “A New Approach to Linear Filtering and Prediction Problems”, *Transactions of the ASME–Journal of Basic Engineering*, vol. 82 (Series D), pp. 35-45, 1960.

[2] Welch, G. and Bishop, G., “An Introduction to the Kalman Filter,” TR 95-041, https://www.cs.unc.edu/~welch/media/pdf/kalman_intro.pdf.

[3] Faragher, R., “Understanding the Basis of the Kalman Filter Via a Simple and Intuitive Derivation.” *IEEE Signal Processing Magazine*, pp. 128-132, Sep. 2012.

[4] Bromiley, P. A., “Products and Convolutions of Gaussian Probability Density Functions,” *Tina Memo No. 2003-003*, http://www.lucamartino.altervista.org/2003-003.pdf.