Journal de Ciencia e Ingeniería, vol.14, no.1, pp.21-26, 2022
http://jci.uniautonoma.edu.co
DOI: https://doi.org/10.46571/JCI.2022.1.2

Journal de Ciencia e Ingeniería

Corporación Universitaria Autónoma del Cauca
ISSN 2145-2628, e-ISSN 2539-066X
Educational Engineering

Gaussianity and the Kalman Filter: A Simple Yet Complicated Relationship

Gaussianidad y el filtro de Kalman: Una relación simple pero complicada

Jeffrey Uhlmann[+1] **and Simon J. Julier**[*]
[+]**Deparment of Physics and Astronomy, University of Missouri-Columbia, EEUU**
[*]**Department of Computer Science, University College London, UK**

Abstract. One of the most common misconceptions made about the Kalman filter when applied to linear systems is that it requires an assumption that all error and noise processes are Gaussian. This misconception has frequently led to the Kalman filter being dismissed in favor of complicated and/or purely heuristic approaches that are supposedly "more general" in that they can be applied to problems involving non-Gaussian noise. The fact is that the Kalman filter provides rigorous and optimal performance guarantees that do not rely on any distribution assumptions beyond mean and error covariance information. These guarantees even apply to use of the Kalman update formula when applied with nonlinear models, as long as its other required assumptions are satisfied. Here we discuss misconceptions about its generality that are often found and reinforced in the literature, especially outside the traditional fields of estimation and control.

Keywords. Educational engineering; engineering history.

Resumen. Uno de los conceptos erróneos más comunes sobre el filtro de Kalman cuando se aplica a sistemas lineales es que requiere la suposición de que todos los procesos de error y ruido son Gaussianos. Este concepto erróneo ha llevado con frecuencia a descartar el filtro de Kalman en favor de enfoques complicados y/o puramente heurísticos que supuestamente son "más generales" en el sentido de que pueden aplicarse a problemas que involucran ruido no Gaussiano. El hecho es que el filtro de Kalman proporciona garantías de rendimiento óptimas y rigurosas que no se basan en suposiciones de distribución más allá de la información de covarianza de error y media. Estas garantías incluso se aplican al uso de la fórmula de actualización de Kalman cuando se aplica con modelos no lineales, siempre que se cumplan sus otros supuestos requisitos. Aquí discutimos conceptos erróneos sobre su generalidad que a menudo se encuentran y refuerzan en la literatura, especialmente fuera de los campos tradicionales de estimación y control.

Palabras Clave. Educación en ingeniería; historia de la ingeniería.

How to Cite. J. Uhlmann and S.J. Julier, "Gaussianity and the Kalman Filter: A Simple Yet Complicated Relationship", *Jou. Cie. Ing.*, vol. 14, no. 1, pp. 21-26, 2022. doi:10.46571/JCI.2022.1.2

Received: 20/01/2022 **Revised:** 06/04/2022 **Accepted:** 22/06/2022

[1] *e-mail: uhlmannj@missouri.edu*

1. Introduction

The Kalman filter [1] is among the most versatile and widely-used tools in engineering. More than 50 years after Kalman published his original paper, his work still inspires hundreds of papers each year. Some of these papers explore new applications of the algorithm in approaches which range from industrial process control, to robotics, and even to meta-analysis of election forecasts.

- "*Since the Kalman framework requires Gaussian distributions, the model can only be constructed if ...* " [2]
- "*The Kalman filter which is used for integrated navigation requires Gaussian variables ... a multimodal un-symmetric distribution has to be approximated with a Gaussian distribution before being used in the Kalman filter.*" [3]
- "*...can be best reconciled with the KF (which requires Gaussian probability distributions) by making the assumption that ...* " [4]
- " *[The] Kalman filter requires Gaussian prior $f(x_0)$ ...* " [5]
- "*Notice that each of the distributions can be effectively approximated by a Gaussian. This is a very important result for the operation for many systems, especially the ones based on a Kalman filter since the filter explicitly requires Gaussian distributed noise on measurements for proper operation.*" [6]
- *...importance sampling ... relaxes the assumption of Gaussian observation errors required by the basic Kalman filter.* [7]
- "*However, the KF requires Gaussian initial conditions, therefore ...* "[8]
- "*Kalman filters are Bayes filters that represent posteriors with Gaussians... Kalman filter mapping relies on three basic assumptions ... Gaussian noise... the initial uncertainty must be Gaussian.*" [9]
- "*The Kalman filter is a very efficient optimal filter; however it has the precondition that the noises of the process and of the measurement are Gaussian ... when the measurement is not a Gaussian distribution, the Kalman filter cannot be used.*" [10]
- "*The Kalman filter assumes that the posterior density at every time step is Gaussian and hence exactly and completely parameterized by two parameters, its mean and covariance.*"
- "*Kalman filter cannot be used here for inference because the measurement does not involve additive Gaussian noise.*" [11]
- "*The KF requires models defined by linear Gaussian probability density functions.*" [12]
- "*The most attractive advantage of the Kalman filter lies in its optimal estimation in the sense of minimum mean squared prediction errors. However, the optimality of the Kalman filter requires two restrictive prerequisites, linear state-space models and independent Gaussian white noise for both process and measurements.*" [13]

Figure 1: Quotes in the literature that could be interpreted as suggesting that the Kalman filter can only be applied when errors are Gaussian-distributed.

Unfortunately, the full potential of the Kalman filter is often not appreciated and exploited due to misconceptions which have persisted since the earliest days after its emergence as a critical component in aeronautical and space applications in the 1960s. Among the most common misconceptions is that the Kalman filter can only be rigorously derived and applied to linear systems in which all the error and noise processes are *strictly* Gaussian. More specifically, it is commonly believed — and frequently stated implicitly or explicitly — that the use of a Kalman filter in the presence of non-Gaussian error processes is at the very least a sub-optimal heuristic approach that may perform well in practice if errors are approximately Gaussian but that it is mathematically non-rigorous and cannot be expected to perform well if the errors are strongly non-Gaussian. Figure 1 gives evidence of this in form of a quotes sampled from the literature across a range of fields including machine learning, estimation, systems engineering and end-user applications. (These quotes should not be taken as undermining the integrity of the sources from which they are taken but rather as examples of how someone new to the study of linear and estimation might come to believe that the Kalman filter requires Gaussianity assumptions.)

An unfortunate consequence of such misconceptions is that it is common for the Kalman filter to be dismissed from consideration for applications simply because the errors are known to be non-Gaussian. In this paper, we discuss how the Kalman filter can be derived as the optimal solution to the filtering problem under differing sets of assumptions, and we emphasize that the assumptions required for a particular derivation are not necessarily required in general for the optimality of the filter. In particular, we emphasize that linearity does not imply Gaussianity; minimizing mean-squared error does not imply Gaussianity; and that the Kalman Filter is MMSE-optimal without any assumptions of Gaussianity.

In the next section we briefly summarize the linear estimation/filtering problem and historically how the same optimal solution has been derived from two very different perspectives with very different assumptions.In particular, we emphasize that the Kalman filter can be applied rigorously – and optimally – to systems with errors from any probability distribution with finite first and second moments

2. The Estimation Problem: Kalman vs. Bayes

2.1. Estimation Problem

Consider a linear system of the form

$$\mathbf{x}_k = \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \mathbf{v}_{k-1}, \tag{1}$$

where $\mathbf{x}_k$ is the state at timestep k, $\mathbf{F}_{k-1}$ is the state transition matrix, and $\mathbf{v}_{k-1}$ is the additive process noise. We assume that this noise is independent from timestep, is zero mean, and its covariance is known. However, we do *not* assume that it is Gaussian-distributed.

The observation model for a sensor measurement of the state of the system has the same linear form:

$$\mathbf{z}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{w}_k, \tag{2}$$

where where $\mathbf{H}_k$ is a linear transformation from system to observation coordinates and $\mathbf{w}_k$ is the observation noise. Similar to $\mathbf{v}_{k-1}$, it is assumed to be zero-mean, independent, but does not have to be Gaussian.

The filter is initialized at time step 0 with an estimate $\hat{\mathbf{x}}_{0|0}$. The error in this estimate is zero mean and has a covariance $\mathbf{P}_{0|0}$. Again, this does not have to be Gaussian.

Given an initial condition and a sequence of measurements, the goal is to compute an estimate $\hat{\mathbf{x}}_{i|j}$ at timestep i based on all observations up to timestep j, along with its associated error covariance $\mathbf{P}_{i|j}$. The relationship between the estimate and the state is given by

$$\hat{\mathbf{x}}_{i|j} = \mathbf{x}_i + \tilde{\mathbf{x}}_{i|j}, \tag{3}$$

where $\tilde{\mathbf{x}}_{i|j}$ is the error. The mean squared error in this estimate is

$$\mathbf{P}_{i|j} \ = \ \mathrm{E}\left[\tilde{\mathbf{x}}_{i|j}\tilde{\mathbf{x}}_{i|j}^{\top}\right]. \tag{4}$$

Given that the covariance of any error distribution can never be determined *exactly* in practice, the above equality can be relaxed to the more conservative and practically-achievable requirement:

$$\mathbf{P}_{i|j} \ \geq \ \mathrm{E}\left[\tilde{\mathbf{x}}_{i|j}\tilde{\mathbf{x}}_{i|j}^{\top}\right], \tag{5}$$

where the covariance matrix $\mathbf{P}_{i|j}$ can now be interpreted as representing the best available upper bound on the expected squared error associated with $\hat{\mathbf{x}}_{i|j}$. As we discuss later, the practical question is whether an estimate which obeys this guarantee is sufficient for the problem at hand.

In Kalman's original paper he derived his now-eponymous filter from the perspective of ℓ_2-norm error minimization via othogonal projections. He also noted (his Corollary 1) that if all errors are assumed Gaussian then the system mean-and-covariance estimate can be interpreted as parameterizing a Gaussian distribution that represents the exact error distribution conditioned on the sequence of observations. In more contemporary parlance, Kalman derived a minimum-mean-squared error (MMSE) optimal filter. What is critical to note, however, is that the MMSE optimality of the filter does not in any way depend on such an assumption.

While MMSE optimality can be obtained with broad generality, i.e., with relatively weak assumptions, it does so at the expense of a strong probabilistic interpretation. For example, the Kalman filter guarantees that the expected squared error of the system estimate decreases at a certain rate, but it does not provide information necessary to answer a question about the probability that, e.g., its mean position estimate is within one meter of the true position. This motivated Ho and Lee [14] to re-derive the optimal filter from a Bayesian perspective by replacing the mean-covariance pair with the full probability density of the state, $f_{k|k}(\mathbf{x}_k|\mathbf{z}_{1:k})$. The prediction is then determined by the Chapman-Kolmogorov Equation,

$$f_{k|k-1}(\mathbf{x}_k|\mathbf{z}_{1:k-1}) = \int f(\mathbf{x}_k|\mathbf{x}')f_{k-1|k-1}(\mathbf{x}'|\mathbf{z}_{1:k})\mathrm{d}\mathbf{x}' \tag{6}$$

where $f(\mathbf{x}_k|\mathbf{x}')$ is the state transition density which encodes the process model. The update is then given from Bayes rule,

$$f_{k|k}(\mathbf{x}_k|\mathbf{z}_{1:k}) = \frac{f(\mathbf{z}_k|\mathbf{x}_k)f_{k|k-1}(\mathbf{x}_k|\mathbf{z}_{1:k-1})}{f(\mathbf{z}_{1:k})}, \tag{7}$$

where $f(\mathbf{z}_k|\mathbf{x}_k)$ is the measurement likelihood model. This incorporates the effects of the observation model.

From this Ho and Lee provided an alternate proof of Kalman's corollary that the Kalman filter is Bayes-optimal with all mean and covariance estimates interpreted as parameters for Gaussian densities corresponding to assumed-Gaussian error processes. Under these assumptions a mean and covariance estimate from the Kalman filter does not just represent the first two moments of an otherwise unknown probability distribution, it can be interpreted as the *exact* uncertainty distribution for the state. Having access to the full error distribution is clearly more informative, but it comes at the cost of assumptions that cannot be realistically satisfied in almost any nontrivial real-world application. Even beyond Gaussianity, the Bayesian derivation strictly requires *exact* and *complete* knowledge about the full statistics of all noises. The least-squares derivation, by contrast, easily accommodates conservative covariance estimates and is thus much more consistent with the practical reality that the error covariance of a sensor can never be ascertained with infinite precision.

While it is debatable whether or not the Bayesian interpretation is pedagogically more accessible or intuitive than squared-error minimization, there is no question that it appears more prominently in textbooks and introductory expositions of the Kalman filter. It should not be surprising, therefore, that lingering concerns might exist (e.g., as suggested by the quotes in Figure 1) that Gaussian-distributed noises are somehow more conducive to good filter performance based on their special role in Bayesian derivations. It would seem that the MMSE derivation should be sufficient to dispel such concerns, but its abstract mathematical guarantees apparantly cannot break the intuitive link between covariances and Gaussians that is so firmly ingrained by the Bayesian interpretation.

3. Conclusions

The Procrustean application of Bayes' rule to derive the Kalman filter may be suitable as a pedagogical exercise, but care must be taken to ensure that the assumptions required for the method of derivation are not confused with assumptions that are required in general for effective use of the filter. Linearity does not imply Gaussianity. Minimizing mean squared error does not imply Gaussianity. The Kalman Filter is MMSE-optimal without any assumptions of Gaussianity. In this note we have attempted to highlight this fact so that the generality and optimality of the Kalman filter can be more fully appreciated.

References

[1] Kálmán, R. E., "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering*, Vol. 82, No. 1. doi:10.1115/1.3662552.

[2] Yardim, C., Gerstoft, P., and Hodgkiss, W. S., "Atmospheric Refractivity Tracking from Radar Clutter Using Kalman and Particle Filters," in "2007 IEEE Radar Conference," Waltham, MA, USA, 2007, pp. 291–296. doi:10.1109/RADAR.2007.374230.

[3] Schön, T., Gustafsson, F., and Nordlund, P.-J., "Marginalized Particle Filters for Mixed Linear/Nonlinear State-Space Models," *IEEE Transactions on Signal Processing*, Vol. 53, No. 7, 2005, pp. 2279–2289. doi:10.1109/TSP.2005.849151.

[4] Osborn, G. W., "A Kalman Filtering Approach to the Representation of Kinematic Quantities by the Hippocampal-Entorhinal Complex," *Cognitive Neurodynamics*, Vol. 4, No. 4, 2010, pp. 315–335. doi:10.1007/s11571-010-9115-z.

[5] Suzdaleva, E., "Prior Knowledge Processing for Initial State of Kalman Filter," *International Journal of Adap*, Vol. 24, No. 3, 2008, pp. 188–202. doi:10.1002/acs.1106.

[6] Aksoy, Y. and Alatan, A. A., "Uncertainty Modeling for Efficient Visual Odometry via Inertial Ssensors on Mobile Devices Supplementary Material: Experimental Analysis of Noise on Inertial Sensors of ASUS TF201 Tablet Device," Tech. rep., Dept. of Electrical and Electronics Engineering, Middle East Technical University, 2014.

[7] Knape, J., Jonzén, N., and Sköld, M., "On Observation Distributions for State Space Models of Population Survey Data," *Journal of Animal Ecology*, Vol. 80, 2011, pp. 1269–1277. doi:10.1111/j.1365-2656.2011.01868.x.

[8] Shaferman, V. and Oshman, Y., "Cooperative Interception in a Multi-Missile Engagement," in "Proceedings of the AIAA Guidance, Navigation, and Control Conference," Chicago, IL, USA, 2009, pp. 1–23. doi:dx.doi.org/10.2514/6.2009-5783.

[9] Thrun, S., "Exploring Artificial Intelligence in the New Millennium," Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, chap. Robotic Mapping: A Survey, pp. 1–35, 2003.

[10] Rosenberg, Y. and Werman, M., "A General Filter for Measurements with any Probability Distribution," in "Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '97)," San Juan, Puerto Rico, 1997, pp. 654–659. doi:10.1109/CVPR.1997.609395.

[11] Pfister, J.-P., Dayan, P., and Lengyel, M., "Know Thy Neighbour: A Normative Theory of Synaptic Depression," in Bengio, Y., Schuurmans, D., Lafferty, J. D., Williams, C. K. I., and Culotta, A., eds., "Advances in Neural Information Processing System," Vancouver, B.C., Canada, Vol. 22, 2009, pp. 1464–1472.

[12] Buller, M. J., Tharion, W. J., Cheuvront, S. N., Montain, S. J., Kenefick, R. W., Castellani, J., Latzka, W. A., Roberts, W. S., Richter, M., Jenkins, O. C., and Hoyt, R. W., "Estimation of human core temperature from sequential heart rate observations," *Physiological Measurement*, Vol. 34, No. 7, 2013, p. 781.

[13] Bai, S., Thibault, J., and McLean, D. D., "Use of an Autoassociative Neural Network for Dynamic Data Reconciliation," *IFAC Proceedings Volumes*, Vol. 38, No. 1, 2005, pp. 65 – 70. doi:http://dx.doi.org/10.3182/20050703-6-CZ-1902.00011. 16th IFAC World Congress.

[14] Ho, Y. and Lee, R., "A Bayesian approach to problems in stochastic estimation and control," *Automatic Control, IEEE Transactions on*, Vol. 9, No. 4, 1964, pp. 333–339. doi:10.1109/TAC.1964.1105763.

Journal de Ciencia e Ingeniería

Uniautónoma DEL CAUCA

BY NC ND