A Tutorial on Minimum Energy Filtering for Linear Systems

Jack Henderson
Australian National University

January 25, 2021

1 Introduction

This paper aims to provide the reader with an introduction to the theory of minimum energy filtering by means of a worked example on a linear system. The concept of minimum-energy filtering has been studied in the literature in many different forms over the last 60 years [1]–[5]. However, it can be difficult to find a modern and succinct explanation of minimum energy filtering and an insight into some of the underlying intuitions behind the theory.

The motivation for writing this paper comes from my own journey of learning about minimum energy filtering during the first year of my PhD, and wishing there was a better explanation of the topic. Hopefully this paper makes some progress towards that goal. It is intended to provide the reader with an understanding of the foundational concepts behind minimum-energy filtering and to provide a starting point for further study in the area.

In order to get the most out of this paper, the reader should be broadly familiar with stochastic filtering theory, including the Kalman Filter, as well as optimal control theory, vector calculus, and ordinary differential equations. The derivation presented here closely mirrors [6], but removes a lot of the complexity as we are only considering a linear system.

1.1 Problem Definition

Consider the following continuous-time linear system;

$$\dot{x} = F(t)x(t) + B(t)u(t) + w(t) \tag{1}$$

$$x(0) = x_0 \tag{2}$$

$$y(t) = H(t)x(t) + v(t) \tag{3}$$

where

$x \in \mathbb{R}^n$	System state
$F \in \mathbb{R}^{n \times n}$	System model
$u \in \mathbb{R}^p$	Control input
$B \in \mathbb{R}^{n \times p}$	Input Matrix
$w \in \mathbb{R}^n$	Model error
$y \in \mathbb{R}^m$	Sensor measurement
$H \in \mathbb{R}^{m \times n}$	Measurement model
$v \in \mathbb{R}^m$	Measurement error

We will drop the implicit dependence on t for the remainder of the example. F, B, and H are assumed to be known. It's important to note that, unlike in the derivation of the Kalman filter, we do not assume that w and v are stochastic noise processes (typically Gaussian, zero mean, white processes). Rather, we just consider w and v as deterministic, but unknown error signals. The only assumption we make is that the error signal are square integrable.

The filtering problem is to produce an estimate of the system state, $\hat{x}(t)$, which closely follows the true system state $x(t)$. An initial estimate of the system state, $\hat{x}_0$, is known, as well as the sensor measurements and control inputs, $y_{[0,t]}$ and $u_{[0,t]}$ respectively.

We could try to find an explicit solution for $\hat{x}(t)$, however this would be a function of $y_{[0,t]}$ and $u_{[0,t]}$. As t increases, this would require storing and processing an ever increasing amount of data. Instead we will aim to find a recursive implementation of the filter, which is to say we wish to find a differential equation of the form $\dot{\hat{x}}(t) = f(x(t), y(t), u(t))$. This means that, if we have an estimate of the state $\hat{x}(t)$, we can integrate (typically numerically) the differential equation forward in time to determine $\hat{x}(t^+)$. This allows us to process measurements in real time and we don't need to store or process the entire measurement history to produce the state estimate $\hat{x}(t)$.

1.2 Setting up the Minimum Energy Problem

For a chosen state, x', and time value t', we can propose a set of values for $w_{[0,t']}$ such that the system, (1), passes through the point $x(t') = x'$. Given that y is known, choosing w will also determine v. In fact, there are an infinite number of different trajectories for w that we could choose. Each choice results in a different trajectory of x, but all are plausible. Thus, we must find a way the select the 'best' of these trajectories to use as our state estimate.

We can impose the following cost functional on the error signals;

$$J_t(x(0), w_{[0,t]}) := \frac{1}{2}\|x(0) - \hat{x}_0\|^2_{K_0} + \int_0^t \frac{1}{2}\|w(\tau)\|^2_Q + \frac{1}{2}\|v(\tau)\|^2_R d\tau \tag{4}$$

weighted by symmetric matrices K_0, Q and R. The weighted norm is defined by $\|x\|_W^2 := x^\top W x$. The first term in (4) imposes a cost on how much the starting point of the trajectory, $x(0)$ deviates from the known initial estimate. The remaining terms impose a cost on the energy of the two error signals. This is the origin of the term 'minimum energy' filter.

We need to find a trajectory for x which is consistent with (1) and that also minimises the cost functional. Such a trajectory will find a balance between being consistent with both the system model (a small modelling error, w) being consistent with the measurement model (a small measurement error v). We can adjust the weights, K_0, Q and R, depending on how confident we are in our initial estimate, our system model and our sensor measurements respectively.

We will approach this problem in two steps. Firstly, similar to what was discussed above, consider a specific state (x, t). Out of all the possible trajectories for $w_{[0,t]}$, there will be one which has a minimum value for the cost functional. We will introduce the value function, V, which is the value for the cost functional for the trajectory w which minimises J_t;

$$V(x,t) := \min_{w_{[0,t]}} J_t(x, w_{[0,t]}) \tag{5}$$

We can now determine the best estimate of $x(t)$ by selecting a state which minimises the value function.

$$\hat{x}(t) := \arg\min_x V(x,t) \tag{6}$$

2 Finding a Solution

Ultimately, we wish to derive a differential equation to describe how the estimate of the state, $\hat{x}$, changes over time as a function of the system input, u, and the measurement information, y. Our approach will utilise the techniques and theory from optimal control, as we can analogously consider this as an optimal control problem where v is the tracking error and w is the control input. One of the key ideas that our derivation relies on is the Hamilton-Jacobi-Bellman (HJB) Equation

One necessary condition that we will utilise is that the derivative of the value function at the minimum will be zero.

$$\left.\frac{\mathrm{d}}{\mathrm{d}x} V(x,t)\right|_{x=\hat{x}(t)} = 0 \tag{7}$$

2.1 The Optimal Hamiltonian

The Hamiltonian for this system is defined as

$$\mathcal{H}(x, \mu, w, t) := \frac{1}{2}\|w\|_Q^2 + \frac{1}{2}\|v\|_R^2 - \mu^\top (Fx + Bu + w) \tag{8}$$

where $\mu \in \mathbb{R}^n$ is the Lagrange multiplier. The terms in the Hamiltonian come from the system model (1) and the integrand of the cost functional (4). If the reader is unfamiliar with Optimal Control Theory, Kirk [7] provides a comprehensive introduction to the topic and discusses the properties of the Hamiltonian.

Based on the Pontryagin Maximum Principle, the derivative of the Hamiltonian with respect to w is zero;

$$
\begin{aligned}
0 &= \frac{\mathrm{d}}{\mathrm{d}w}\mathcal{H}(x, \mu, w, t) && (9)\\
&= \frac{\mathrm{d}}{\mathrm{d}w}\left[\frac{1}{2}\|w\|_Q^2 + \frac{1}{2}\|v\|_R^2 - \mu^\top(Fx + Bu + w)\right] && (10)
\end{aligned}
$$

We use (3) to replace v

$$
= \frac{\mathrm{d}}{\mathrm{d}w}\left[\frac{1}{2}\|w\|_Q^2 + \frac{1}{2}\|y - Hx\|_R^2 - \mu^\top(Fx + Bu + w)\right] \qquad (11)
$$

Evaluating the derivativate[1] and solving for w gives

$$
\begin{aligned}
0 &= w^\top Q - \mu^\top && (12)\\
w &= Q^{-1}\mu && (13)
\end{aligned}
$$

We can now substitute the value for w into the Hamiltonian to create the optimal Hamiltonian, $\mathcal{H}^*$

$$
\mathcal{H}^*(x, \mu, t) = \frac{1}{2}\left\|-Q^{-1}\mu\right\|_Q^2 + \frac{1}{2}\|y - Hx\|_R^2 - \mu^\top(Fx + Bu + Q^{-1}\mu) \qquad (14)
$$

Simplifying

$$
= \frac{1}{2}\|\mu\|_{Q^{-1}}^2 + \frac{1}{2}\|y - Hx\|_R^2 - \mu^\top(Fx + Bu + Q^{-1}\mu) \qquad (15)
$$

$$
\boxed{\mathcal{H}^*(x, \mu, t) = -\frac{1}{2}\|\mu\|_{Q^{-1}}^2 + \frac{1}{2}\|y - Hx\|_R^2 - \mu^\top(Fx + Bu)} \qquad (16)
$$

The Hamiltonian is a useful concept as it relates to the Hamilton-Jacobi-Bellman Equation

$$
\mathcal{H}^*(x, \nabla_x V(x, t), t) - \frac{\partial}{\partial t}V(x, t) = 0 \qquad (17)
$$

In the literature on the HJB equation, it is common to see a plus sign instead of the minus sign in the equation above. To paraphrase Saccon et al. [5]; the optimal control filtering problem of (6) can be thought of as a standard optimal control problem which is solved backwards in time. In this interpretation, the term

[1] Wikipedia contains a list of many useful vector calculus identities

$\|x(0) - \hat{x}_0\|^2_{K_0}$ in (4) can be considered the terminal cost, and $V(x,t)$ as the cost-to-go. This justifies the presence of the minus sign in (17) rather than the standard form of the HJB equation.

For ease of notation, we let $\mu(x,t) := \nabla_x V(x,t) = \frac{\mathrm{d}}{\mathrm{d}x} V(x,t)^\top$ which equivalently gives

$$\mathcal{H}^*(x, \mu(x,t), t) = \frac{\partial}{\partial t} V(x,t) \tag{18}$$

2.2 Derivative of the Optimal Hamiltonian

In subsequent steps of the derivation, we will need to calculate the derivative of the Optimal Hamiltonian with respect to x where μ is also a function of x. Applying the chain rule, we have

$$\frac{\mathrm{d}}{\mathrm{d}x}\mathcal{H}^*(x, \mu(x,t), t) = \frac{\partial}{\partial x}\mathcal{H}^*(x, \mu(x,t), t) + \frac{\partial}{\partial \mu}\mathcal{H}^*(x, \mu(x,t), t)\frac{\mathrm{d}}{\mathrm{d}x}\mu(x,t) \tag{19}$$

Evaluating just the first term in (19), we have

$$\frac{\partial}{\partial x}\mathcal{H}^*(x, \mu(x,t), t) = \frac{\partial}{\partial x}\left[-\frac{1}{2}\|\mu(x,t)\|^2_{Q^{-1}} + \frac{1}{2}\|y - Hx\|^2_R - \mu(x,t)^\top(Fx + Bu)\right] \tag{20}$$

$$= -(y - Hx)^\top RH - \mu(x,t)^\top F \tag{21}$$

Similarly, evaluating the second term in (19) gives

$$\frac{\partial}{\partial \mu}\mathcal{H}^*(x, \mu(x,t), t) = \frac{\partial}{\partial \mu}\left[-\frac{1}{2}\|\mu(x,t)\|^2_{Q^{-1}} + \frac{1}{2}\|y - Hx\|^2_R - \mu(x,t)^\top(Fx + Bu)\right] \tag{22}$$

$$= -\mu(x,t)^\top Q^{-1} - (Fx + Bu)^\top \tag{23}$$

Substituting (21) and (23) back into (19) gives

$$\frac{\mathrm{d}}{\mathrm{d}x}\mathcal{H}^*(x, \mu(x,t), t) =$$

$$- (y - Hx)^\top RH - \mu(x,t)^\top F - (\mu(x,t)^\top Q^{-1} + (Fx + Bu)^\top)\frac{\mathrm{d}}{\mathrm{d}x}\mu(x,t) \tag{24}$$

2.3 Solving for the State Estimate

Consider the derivative of the value function with respect to both x and t, and apply the chain rule;

$$\frac{\mathrm{d}}{\mathrm{d}x}\frac{\mathrm{d}}{\mathrm{d}t}V(x,t) = \frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{\partial}{\partial t}V(x,t) + \frac{\partial}{\partial x}V(x,t)\frac{\mathrm{d}}{\mathrm{d}t}x\right] \tag{25}$$

Substituting the first term with the HJB equation (18) and applying product rule to the second term gives

$$= \frac{\mathrm{d}}{\mathrm{d}x}\mathcal{H}^*(x, \mu(x,t), t) + \frac{\partial}{\partial x}V(x,t)\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\mathrm{d}}{\mathrm{d}t}x\right) + \left(\frac{\mathrm{d}}{\mathrm{d}t}x\right)^\top \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial}{\partial x}V(x,t)\right)^\top \tag{26}$$

Recall from (7) that, at the optimal trajectory, the derivative w.r.t. x of the value function is zero. Therefore,

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\mathrm{d}}{\mathrm{d}x}V(x,t)\bigg|_{x=\hat{x}(t)} = 0 \tag{27}$$

Substituting (24) into (26) and then evaluating at $x = \hat{x}(t)$ gives

$$\begin{aligned} 0 = -(y - H\hat{x})^\top RH - \mu(\hat{x},t)^\top F - \left(\mu(\hat{x},t)^\top Q^{-1} + (F\hat{x} + Bu)^\top\right)\frac{\mathrm{d}^2}{\mathrm{d}x^2}V(\hat{x},t) \\ + \mu(\hat{x},t)^\top \frac{\mathrm{d}}{\mathrm{d}x}\dot{\hat{x}} + \dot{\hat{x}}^\top \frac{\mathrm{d}^2}{\mathrm{d}x^2}V(\hat{x},t) \end{aligned} \tag{28}$$

where we have used the shorthand notation $\frac{\mathrm{d}^2}{\mathrm{d}x^2}f(x) = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\mathrm{d}}{\mathrm{d}x}f(x)\right)^\top$ and $\dot{x} = \frac{\mathrm{d}}{\mathrm{d}t}x$. From (7), we note that $\mu(\hat{x},t) = 0$, and we can simplify to

$$0 = -(y - H\hat{x})^\top RH - (F\hat{x} + Bu)^\top \frac{\mathrm{d}^2}{\mathrm{d}x^2}V(\hat{x},t) + \dot{\hat{x}}^\top \frac{\mathrm{d}^2}{\mathrm{d}x^2}V(\hat{x},t) \tag{29}$$

Solving for $\dot{\hat{x}}$ and introducing $K = \frac{\mathrm{d}^2}{\mathrm{d}x^2}V(\hat{x},t)$, we have

$$\dot{\hat{x}}^\top K = (y - H\hat{x})^\top RH + (F\hat{x} + Bu)^\top K \tag{30}$$

$$\boxed{\dot{\hat{x}} = F\hat{x} + Bu + K^{-1}H^\top R(y - H\hat{x})} \tag{31}$$

We also need to determine the initial condition of the ODE.

$$\begin{aligned} \hat{x}(0) &= \arg\min_x V(x,0) && (32) \\ &= \arg\min_x \frac{1}{2}\|x - \hat{x}_0\|^2_{\Sigma_0} && (33) \\ &= \hat{x}_0 && (34) \end{aligned}$$

2.4 Determining the Hessian of the Value Function

Note in (31) that the term K, the Hessian of the value function, is still undetermined. We can describe K in terms of an ODE and an initial condition. Recall that

$$K = \frac{\mathrm{d}^2}{\mathrm{d}x^2}V(x,t)\bigg|_{x=\hat{x}(t)} \tag{35}$$

Taking the total time derivative of K and applying the chain rule,

$$\frac{\mathrm{d}}{\mathrm{d}t}K = \frac{\partial}{\partial x}\left[\frac{\mathrm{d}^2}{\mathrm{d}x^2}V(x,t)\right]\frac{\mathrm{d}}{\mathrm{d}t}x + \frac{\partial}{\partial t}\left[\frac{\mathrm{d}^2}{\mathrm{d}x^2}V(x,t)\right] \tag{36}$$

$$= \frac{\mathrm{d}}{\mathrm{d}x}K\frac{\mathrm{d}}{\mathrm{d}t}x + \frac{\mathrm{d}^2}{\mathrm{d}x^2}\frac{\partial}{\partial t}V(x,t) \tag{37}$$

Considering just the second term

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}\frac{\partial}{\partial t}V(x,t) = \frac{\mathrm{d}^2}{\mathrm{d}x^2}\mathcal{H}(x,\mu(x,t),t) \tag{38}$$

$$= \frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{\mathrm{d}}{\mathrm{d}x}\mathcal{H}(x,\mu(x,t),t)\right]^\top \tag{39}$$

We can use the results from (24)

$$= \frac{\mathrm{d}}{\mathrm{d}x}\left[-(y-Hx)^\top RH - \mu(x,t)^\top F - (\mu(x,t)^\top Q^{-1} + (Fx+Bu)^\top)\frac{\mathrm{d}}{\mathrm{d}x}\mu(x,t)\right]^\top \tag{40}$$

$$= \frac{\mathrm{d}}{\mathrm{d}x}\left[-H^\top R(y-Hx) - F^\top\mu(x,t) - (\frac{\mathrm{d}}{\mathrm{d}x}\mu(x,t))^\top\left(Q^{-1}\mu(x,t) + (Fx+Bu)\right)\right] \tag{41}$$

Evaluating the derivative, we have

$$\begin{aligned} &= H^\top RH - F^\top\frac{\mathrm{d}}{\mathrm{d}x}\mu(x,t) - (\frac{\mathrm{d}}{\mathrm{d}x}\mu(x,t))^\top F - (\frac{\mathrm{d}}{\mathrm{d}x}\mu(x,t))^\top Q^{-1}\frac{\mathrm{d}}{\mathrm{d}x}\mu(x,t) \\ &\quad + \mu(x,t)^\top Q^{-1}\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\mathrm{d}}{\mathrm{d}x}\mu(x,t)\right) - (Fx+Bu)^\top\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\mathrm{d}}{\mathrm{d}x}\mu(x,t)\right) \end{aligned} \tag{42}$$

$$= H^\top RH - F^\top K - KF - KQ^{-1}K + \mu(x,t)^\top Q^{-1}\frac{\mathrm{d}}{\mathrm{d}x}K - (Fx+Bu)^\top\frac{\mathrm{d}}{\mathrm{d}x}K \tag{43}$$

Substituting back into (37) and evaluating at $x = \hat{x}(t)$ results in

$$\dot{K} = H^\top RH - F^\top K - KF - KQ^{-1}K + O(\frac{\mathrm{d}}{\mathrm{d}x}K) \tag{44}$$

where $O(\frac{\mathrm{d}}{\mathrm{d}x}K)$ represents terms containing the derivative of K, which is the third derivative of the value function. For a linear system, it is possible to show that the value function is quadratic in x, and thus the third-order derivative is identically zero. Sontag [4] shows this property in Chapter 8, specifically in Theorem 38 and Equation 8.47. Given this, the resulting ODE for K is

$$\boxed{\dot{K} = H^\top RH - F^\top K - KF - KQ^{-1}K} \tag{45}$$

The initial condition for K is straightforward to derive

$$
\begin{aligned}
K(0) &= \frac{\mathrm{d}^2}{\mathrm{d}x^2} V(x,0) && (46) \\
&= \frac{1}{2} \|x - x_0\|^2_{K_0} && (47) \\
&= K_0 && (48)
\end{aligned}
$$

We observe that, in (31), the K term is inverted. Instead of performing this inversion, we can also just define an ODE for K^{-1}.

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} K^{-1} &= -K^{-1} \frac{\mathrm{d}}{\mathrm{d}t}[K] K^{-1} && (49) \\
&= -K^{-1} \left(H^\top R H - F^\top K - K F - K Q^{-1} K \right) K^{-1} && (50)
\end{aligned}
$$

If we substitute $\Sigma = K^{-1}$, we get the familiar form of the filter;

$$
\boxed{
\begin{aligned}
\dot{\hat{x}} &= F\hat{x} + Bu + \Sigma H^\top R (y - H\hat{x}) \\
\dot{\Sigma} &= Q^{-1} + \Sigma F^\top + F\Sigma - \Sigma H^\top R H \Sigma
\end{aligned}
}
\qquad
\begin{aligned}
(51) \\
(52)
\end{aligned}
$$

3 Conclusions

3.1 Comparison with the Kalman-Bucy Filter

It should come as no surprise that the minimum-energy filter estimate has the same form as the Kalman-Bucy filter [8] which is the optimal minimum-variance estimator for the same linear system.

Recall that the formulation of our problem is purely deterministic, with no concept of random variables or covariance matrices. However, if we match up all of the terms in the minimum energy filter with the Kalman-Bucy filter, however we can observe the parallels to the stochastic interpretation. Σ is equivalent to the covariance of the estimate, while Q^{-1} and R^{-1} map to the process noise covariance and the measurement noise covariance respectively. This matches up well to the intuitions in the deterministic system. Σ describes the inverse of the Hessian of the value function. If the Hessian is large, it means that a small change in the estimate in any direction is going to significantly increase the cost functional, and that new measurement data is unlikely to alter the state estimate significantly. This is equivalent to the covariance of the state estimate being small, which also indicates a high degree of confidence in the current estimate.

3.2 Extending the Minimum Energy Filter

One might come to the conclusion that the minimum energy filter is just a reinterpretation or an alternative derivation of the Kalman-Bucy filter, and that there's nothing particularly different about the two approaches. However this is only true in the linear system case. Where the two approaches begin to differ is in the non-linear case.

The standard approach for dealing with non-linear systems in the stochastic framework is to linearise the system about the current state estimate, and then apply a standard linear Kalman filter, an approach known as the Extended Kalman Filter (EKF). This simple approach is often very effective for systems that are close to being linear, but when the linearisation error is high the filter behaviour is erratic and can diverge.

The derivation presented in this paper works follows through in a very similar process for non-linear systems, and does not require the same linearisation process that is used in the EKF. The key change that occurs when moving to non-linear systems is that we can no longer guarantee that the Value function is quadratic in x, which means that we cannot cancel the third-order terms out of (44). We could instead find a differential equation to describe the third order derivative, but this would contain terms of fourth-order. And solving for the fourth-order term requires the fifth-derivative and so on. Instead, one approach is to assume that these higher-order derivatives are negligible and can be discarded from the ODE anyway. If we do this, the filter no longer provides the optimal minimum-energy solution to the problem, and is called the 'second-order optimal minimum energy filter'.

Another advantage of the minimum energy filter is that it allows us to work directly with non-Euclidean state spaces such as Lie groups. For example, rather than x being an element of $\mathbb{R}^n$, we can consider state spaces such as rotation matrices, SO(3), or poses, SE(3). Zamani [9] shows an example of a minimum energy filter on SO(3), and Saccon [5] generalises the filter for arbitrary Lie Groups. These derivations require an understanding of differential geometry, but there are no fundamental differences with the derivation presented in this paper.

References

[1] R. E. Mortensen, "Maximum-likelihood recursive nonlinear filtering," *Journal of Optimization Theory and Applications*, vol. 2, no. 6, pp. 386–394, Nov. 1968. DOI: 10.1007/BF00925744.

[2] J. Willems, "Deterministic least squares filtering," *Journal of Econometrics*, vol. 118, no. 1-2, pp. 341–373, Jan. 2004. DOI: 10.1016/S0304-4076(03)00146-5.

[3] O. Hijab, "Minimum energy estimation," Ph.D. dissertation, University of California, Berkeley, 1980.

[4] E. D. Sontag, *Mathematical Control Theory*, J. E. Marsden, L. Sirovich, M. Golubitsky, and W. Jäger, Eds., ser. Texts in Applied Mathematics. New York, NY: Springer New York, 1998, vol. 6. DOI: 10.1007/978-1-4612-0577-7.

[5] A. Saccon, J. Trumpf, R. Mahony, and A. P. Aguiar, “Second-Order-Optimal Minimum-Energy Filters on Lie Groups,” *IEEE Transactions on Automatic Control*, vol. 61, no. 10, pp. 2906–2919, Oct. 2016. DOI: 10.1109/TAC.2015.2506662.

[6] M. Zamani, J. Trumpf, and R. Mahony, “Minimum-Energy Filtering for Attitude Estimation,” *IEEE Transactions on Automatic Control*, vol. 58, no. 11, pp. 2917–2921, Nov. 2013. DOI: 10.1109/TAC.2013.2259092.

[7] D. E. Kirk, *Optimal Control Theory: An Introduction.* Dover Publications, 2012.

[8] R. E. Kalman and R. S. Bucy, “New Results in Linear Filtering and Prediction Theory,” *Journal of Basic Engineering*, vol. 83, no. 1, pp. 95–108, Mar. 1961. DOI: 10.1115/1.3658902.

[9] M. Zamani, “Deterministic Attitude and Pose Filtering, an Embedded Lie Groups Approach,” Ph.D. dissertation, Australian National University, 2013.