A Second Order Minimum-Energy Filter on the Special Orthogonal Group

Mohammad Zamani, Jochen Trumpf, and Robert Mahony

Abstract— **This work documents a case study in the application of Mortensen's nonlinear filtering approach to invariant systems on general Lie groups. In this paper, we consider the special orthogonal group $\mathrm{SO}(3)$ of all rotation matrices. We identify the exact form of the kinematics of the minimum-energy (optimal) observer on $\mathrm{SO}(3)$ and note that it depends on the Hessian of the value function of the associated optimal control problem. We derive a second order approximation of the dynamics of the Hessian by neglecting third order terms in the expansion of the dynamics. This yields a Riccati equation that together with the optimal observer equation form a second order minimum-energy filter on $\mathrm{SO}(3)$. The proposed filter is compared to the multiplicative extended Kalman filter (MEKF), arguably the industry standard for attitude estimation, by means of simulations. Our studies indicate superior transient and asymptotic tracking performance of the proposed filter as compared to the MEKF.**

I. INTRODUCTION

Optimal filtering is a core field in the systems and control literature. Deterministic nonlinear filtering was introduced by Mortensen [1] in the late 1960s and is now referred to as *minimum-energy filtering*. In this method, the optimal filtering problem is rewritten as an optimal control problem minimizing a deterministic energy functional in the plant disturbance signals followed by a further optimization over the initial value parameter in the system. For the linear case, the minimum energy filter for the standard quadratic cost functional has been shown [2]–[4] to be equivalent to the classical Kalman filter equations [5]. An advantage of the minimum-energy filtering approach is that it does not require a stochastic interpretation of the underlying signals in the estimation problem and facilitates the application of well known theories from optimal control. For example, Krener [6] proved that under some conditions, including the uniform observability of the system, a minimum-energy estimate converges exponentially fast to the true state. In more recent work Aguiar *et al.* [7] applied the minimum-energy principle to systems with perspective outputs by embedding the nonlinear geometry in an overarching Euclidean space. The resulting estimator is 'optimal' on the matrix space $\mathbb{R}^{4\times 4}$ and arguably near optimal when the estimate is projected to the special Euclidean group $\mathrm{SE}(3)$ of rigid-body transformations. In separate work, Coote *et al.* [8] proposed a near-optimal nonlinear filter applying minimum-energy filtering directly to the geometric structure of the unit circle S^1. In follow on work, Zamani *et al.* [9] obtained a near-optimal minimum-energy filter posed directly on the Special Orthogonal Group $\mathrm{SO}(3)$. These two filters are designed directly on the underlying geometric structure of the system state space and include explicit bounds on their distance to optimality. These filters, however, do not use systematic design principles in the design of the filter, relying rather on identification of a suitable Lyapunov function for filter design and the optimality analysis.

This present paper considers the application of Mortensen's nonlinear filtering approach [1] to invariant systems on $\mathrm{SO}(3)$ with vectorial measurements. The present work extends our prior work [9] in that here we incorporate a measurement model and use Mortensen's approach [1] to provide a systematic approach to deriving the proposed estimator. We derive the exact form of an optimal observer on $\mathrm{SO}(3)$ that can be interpreted as a gradient flow with respect to a time-varying metric derived from the Hessian of the value function of the corresponding optimal control problem. This result links optimal filtering on Lie groups to the nonlinear observer work that has been actively investigated in the last five years, cf. [10], [11]. We use Mortensen's method to obtain a second order approximation to the ordinary differential equation (ODE) governing the evolution of the Hessian of the value function and following the same philosophy as Mortensen in [1] we derive a three-by-three positive definite matrix Riccati equation to compute the associated filter gain. The filter obtained has the same general structure and tuning parameters as the multiplicative extended Kalman filter (MEKF) [12], however, it has additional terms in the Riccati equation that appear in the derivation due to the geometry of the state space. The proposed filter can be implemented with essentially the same computational complexity as the MEKF. In a suit of simulations the proposed filter is compared against the MEKF. According to a recent survey on nonlinear attitude filtering methods, cf. the conclusions section in [13], "the extended Kalman filter, especially in the form known as the multiplicative extended Kalman filter, remains the method of choice for the great majority of applications." Our simulation studies indicate that the proposed filter consistently outperforms the MEKF both in transient and asymptotic responses. A key property of the proposed filter is that it is highly robust to choice of initial parameters (initial condition, initial gain estimate) and can be implemented with minimal tuning.

The remainder of the paper is organized as follows. Section II introduces the notation. The problem of minimum-

The authors are with the Research School of Engineering, Australian National University, Canberra, ACT, Australia. Mohammad.Zamani@anu.edu.au, Jochen.Trumpf@anu.edu.au and Robert.Mahony@anu.edu.au.

energy filtering on SO(3) is formally discussed in Section III. In Section IV we apply Mortensen's approach to derive the proposed filtering formulas. The obtained results are presented and formally proved in Section V. Next, Section VI includes some remarks on the implementation of the proposed filter along with a simulation study, comparing the performance of the proposed filter against the MEKF. Finally, Section VII concludes the paper.

II. Notation

The rotation group is denoted by SO(3).

$$\mathrm{SO}(3) = \{R \in \mathbb{R}^{3\times 3} \,|\, R^T R = I, \det(R) = 1\},$$

where I is the 3 by 3 identity matrix. The associated Lie algebra $\mathfrak{so}(3)$ is the set of skew-symmetric matrices,

$$\mathfrak{so}(3) = \{A \in \mathbb{R}^{3\times 3} \,|\, A = -A^T\}.$$

For $\Omega = [a,b,c]^T \in \mathbb{R}^3$, the lower index operator $(.)_\times : \mathbb{R}^3 \longrightarrow \mathfrak{so}(3)$ yields the skew-symmetric matrix

$$\Omega_x = \begin{bmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{bmatrix}.$$

Inversely, the operator vex : $\mathfrak{so}(3) \longrightarrow \mathbb{R}^3$ extracts the skew coordinates, $\mathrm{vex}(\Omega_\times) = \Omega$. The Frobenius norm of a matrix $X \in \mathbb{R}^{3\times 3}$ is given by $\|X\| := \sqrt{\langle X, X\rangle} := \sqrt{\mathrm{trace}(X^T X)}$. We define a cost $\phi_\Gamma : \mathrm{SO}(3) \longrightarrow \mathbb{R}_0^+$, by

$$\phi_\Gamma(R) := \frac{1}{2}\,\mathrm{trace}\left[(R-I)^T \Gamma (R-I)\right],$$

where $\Gamma \in \mathbb{R}^{3\times 3}$ is symmetric positive definite. Note that $\phi_\Gamma(R)$ coincides with the squared Frobenius norm of $\Gamma^{0.5}(R-I)$ and hence is nonnegative. We define the following projection operators. The symmetric projector $\mathbb{P}_s$ is defined by $\mathbb{P}_s(M) := 0.5(M + M^T)$. The skew-symmetric projector $\mathbb{P}_a$ is defined by $\mathbb{P}_a(M) := 0.5(M - M^T)$. Recall that the vector product of the two vectors $U, \Omega \in \mathbb{R}^3$ satisfies

$$(U \times \Omega) = \mathrm{vex}(2\mathbb{P}_a(\Omega U^T)). \tag{1}$$

III. Minimum-Energy Filtering on SO(3)

Consider the system on SO(3)

$$\begin{cases} \dot{R}(t) &= R(t)(A(t) + \delta(t)),\ R(0) = R_0, \\ y_i(t) &= R(t)^T a_i + \varepsilon_i(t),\ i = 1, \cdots, n\,, \end{cases} \tag{2}$$

where R is an SO(3)-valued state signal. The vectors a_i are known vector directions in the reference frame, i.e. a fixed frame considered as a reference frame. The direction measurements $\{y_i\}$ and the measurement errors $\{\varepsilon_i\}$ are measured in the body fixed frame, i.e. a frame attached to a moving rigid body with orientation R. The signals A and δ denote the measured angular velocity input and the input measurement error, respectively, and take values in $\mathfrak{so}(3)$. The unknown variables here are R_0, δ and the $\{\varepsilon_i\}$. All signals are deterministic functions of time and we assume sufficient regularity of all signals to ensure existence of unique maximal solutions of the system (2).

The principle of minimum-energy filtering adapted to our system (2) is as follows. Consider the cost

$$\begin{aligned} J(t; R_0, \delta|_{[0,\,t]}, \{\varepsilon_i|_{[0,\,t]}\}) &= \int_0^t \frac{1}{2}\left(\|\mathrm{vex}(\delta)\|^2 + \sum_i \|\varepsilon_i\|^2\right) d\tau \\ + \phi_{K_0^{-1}}(R_0) &= \int_0^t \frac{1}{2}\,\mathrm{trace}\left(\frac{1}{2}\delta^T\delta + \sum_i \varepsilon_i \varepsilon_i^T\right) d\tau \\ &+ \frac{1}{2}\,\mathrm{trace}\left[(R_0 - I)^T K_0^{-1}(R_0 - I)\right], \end{aligned} \tag{3}$$

in which $K_0 \in \mathbb{R}^{3\times 3}$ is symmetric positive definite. At each time t, given the measurements $\{y_i|_{[0,\,t]}\}$ and $A|_{[0,\,t]}$, the goal is to obtain an estimate $\hat{R}(t)$ of the true state $R(t)$ by minimizing the cost (3). The cost arguments are the unknowns (R_0, $\delta|_{[0,\,t]}$, $\{\varepsilon_i|_{[0,\,t]}\}$). Hence, equivalently, we seek the minimizing unknowns, given that they are compatible with the measurements in fulfilling the system equations (2). The cost functional (3) encodes the total energy involved in any anticipated set of unknowns and by minimizing it we achieve a minimum-energy estimate.

A set of minimizing unknowns (R_0^*, $\delta^*|_{[0,\,t]}$, $\{\varepsilon_i^*|_{[0,\,t]}\}$) replaced in the system equations (2) yields the optimal state trajectory $R^*_{[0,\,t]}$. The subscript $[0,\,t]$ indicates that the optimization takes place on the interval $[0,\,t]$. We pick the final optimal state $R^*_{[0,\,t]}(t)$ as our minimum-energy estimate at time t, $\hat{R}(t) := R^*_{[0,\,t]}(t)$. This could be a lengthy process as at any time t we need to obtain the optimal state trajectory $R^*_{[0,\,t]}$ and pick its final value as our minimum-energy estimate at that time. In practice, we will seek a recursive filter that at each time t yields the minimum-energy estimate as its state value.

Technically, the cost (3) depends on the unknowns R_0, $\delta|_{[0,\,t]}$ and $\{\varepsilon_i|_{[0,\,t]}\}$, but given R_0 and $\delta|_{[0,\,t]}$, the known $\{a_i\}$ and the measurements $A|_{[0,\,t]}$ and $\{y_i|_{[0,\,t]}\}$, the $\varepsilon_i|_{[0,\,t]}$ are uniquely determined by (2). Hence we can turn this problem into an optimal control problem by substituting ε_i in (3) from (2) and treating R_0 as a constant parameter for the moment.

$$\begin{aligned} &J(t; R_0, \delta|_{[0,\,t]}) = \\ &\int_0^t \frac{1}{2}\,\mathrm{trace}\left(\frac{1}{2}\delta^T\delta + \sum_i (y_i - R^T a_i)(y_i - R^T a_i)^T\right) d\tau \\ &+ \frac{1}{2}\,\mathrm{trace}\left[(R_0 - I)^T K_0^{-1}(R_0 - I)\right]. \end{aligned} \tag{4}$$

We treat $\delta|_{[0,t]}$ as the control input and minimize (4) over δ. The initial value R_0 is treated as a fixed and known parameter when solving the optimal control problem. We will later optimize over this parameter in the derivation of the filter. The end point $R(t)$ is free but constrained by the measurements and by the dynamics (2) in the period $[0,\,t]$.

To apply the Maximum Principle we define the following pre-Hamiltonian [14].

$$\begin{aligned}\mathscr{H}(R,P,\delta,t) &:= \mathrm{trace}[\frac{1}{4}\delta(t)^T\delta(t)\\ &+\frac{1}{2}\sum_i(y_i-R^Ta_i)(y_i-R^Ta_i)^T-P^T\dot{R}],\end{aligned} \tag{5}$$

where $P\in T_R\mathrm{SO}(3)$ is the tangent space valued costate variable. Next, we minimize the Hamiltonian over δ computing the critical point of $\mathscr{H}$.

$$\begin{aligned}&\mathscr{D}_\delta\mathscr{H}\circ\Gamma=0,\ \forall\Gamma\in\mathfrak{so}(3)\\ &\quad\Rightarrow \mathrm{trace}[\frac{1}{2}\delta(t)^{*T}\Gamma-P^TR\Gamma]=0\end{aligned} \tag{6}$$

which gives $\delta^*=-2P^TR=2R^TP$. Substituting δ^* we have the optimal Hamiltonian

$$\begin{aligned}\mathscr{H}^*(R,P,t) &= \mathrm{trace}[-P^TP\\ &+\frac{1}{2}\sum_i(y_i-R^Ta_i)(y_i-R^Ta_i)^T-P^TRA].\end{aligned} \tag{7}$$

From the Maximum principle we get the Hamilton equations

$$\begin{cases}\dot{R} &= \nabla_P\mathscr{H}^*(R,P,t),\\ \dot{P} &= -\nabla_R\mathscr{H}^*(R,P,t).\end{cases} \tag{8}$$

Since $R(t)$ is free we have the boundary condition

$$P(0)=R_0\mathbb{P}_a(R_0^TK_0^{-1}(R_0-I)). \tag{9}$$

The optimal control problem boils down to solving (8) and (9) for each t. However, as was said before, we want to avoid repeating this process at each time t by obtaining a recursive solution. Next, we apply the dynamic programming principle to this problem. We define the value function

$$V(R,t):=\min_{\delta|_{[0,t]}}J(t;R_0,\delta|_{[0,t]}), \tag{10}$$

where J is the cost (4) and the minimization is subjected to the system equations (2). The Hamilton-Jacobi-Bellman equation is then [**?**]

$$\mathscr{H}^*(R,\nabla_RV(R,t),t)-\frac{\partial V(R,t)}{\partial t}=0. \tag{11}$$

From (4) the initial time boundary condition is

$$V(R_0,0)=\frac{1}{2}\mathrm{trace}\left[(R_0-I)^TK_0^{-1}(R_0-I)\right]. \tag{12}$$

Up to here we have only addressed the optimal control part of the problem (by only minimizing over δ). To complete the optimal filtering problem, we also need to optimize V over R_0. This can be equivalently posed as optimization over $R(t)$ (from (2) R_0 and $\delta|_{[0,t]}$ uniquely determine $R(t)$ and vice versa $R(t)$ and $\delta|_{[0,t]}$ uniquely determine $R(0)$). The optimum of V is characterized by the criticality condition

$$\nabla_RV(R,t)|_{R=\hat{R}(t)}=0. \tag{13}$$

Recall that the minimum-energy estimate $\hat{R}(t)$ is the minimizing argument $R^*(t)$, which yields the final condition (13). Solving Equation (13) is clearly a way to characterize $\hat{R}(t)$. However, we are interested in finding a differential equation that dynamically updates $\hat{R}(t)$.

Up to here we have introduced our optimization problem and we have shown how to approach it in the context of the Maximum Principle. In the next section we will discuss how we modify the proposed solution by Mortensen [1] to derive a second order approximation of a recursive minimum-energy filter on SO(3).

IV. Filter Derivation Using Mortensen's Approach

In this section we modify Mortensen's program [1] in order to adapt it to the geometric setting of our problem. Recall that the relationship between a directional derivative and a gradient with respect to a Riemannian metric $\langle .,.\rangle$ is as follows.

$$\mathscr{D}_RV(R,t)\circ R\Gamma=\langle\nabla_RV(R,t),R\Gamma\rangle. \tag{14}$$

Here $\mathscr{D}_RV(R,t)\circ R\Gamma$ denotes the derivative of the function V with respect to the argument R in the tangent direction $R\Gamma$ where Γ is a Lie algebra element. The symbol $\nabla_RV(R,t)$ denotes the gradient of the function V with respect to the argument R.

Applying (14) to the final condition (13) yields

$$\{\mathscr{D}_RV(R,t)\circ R\Gamma\}|_{R=\hat{R}(t)}=0,\quad \text{for all }\Gamma\in\mathfrak{so}(3). \tag{15}$$

Next, we calculate the total time derivative of (15) using the chain rule.

$$\begin{aligned}&\frac{d}{dt}\{(\mathscr{D}_RV(R,t)\circ R\Gamma)|_{R=\hat{R}(t)}\}=\\ &\{\mathscr{D}_R^2V(R,t)\circ(\dot{\hat{R}}(t),R\Gamma)+\mathscr{D}_R(\frac{\partial V(R,t)}{\partial t})\circ R\Gamma+\\ &\mathscr{D}_RV(R,t)\circ R\mathbb{P}_a(R^T\dot{\hat{R}}\Gamma)\}_{R=\hat{R}(t)}=0,\quad \text{for all }\Gamma\in\mathfrak{so}(3).\end{aligned} \tag{16}$$

We use (11) to replace $\frac{\partial V(R,t)}{\partial t}$ in (16). This yields

$$\begin{aligned}&\frac{d}{dt}\{(\mathscr{D}_RV(R,t)\circ R\Gamma)|_{R=\hat{R}(t)}\}=\\ &\{\mathscr{D}_R^2V(R,t)\circ(\dot{\hat{R}}(t),R\Gamma)+\mathscr{D}_R\mathscr{H}^*(R,\nabla_RV(R,t),t)\circ R\Gamma+\\ &\mathscr{D}_RV(R,t)\circ R\mathbb{P}_a(R^T\dot{\hat{R}}\Gamma)\}_{R=\hat{R}(t)}=0,\quad \text{for all }\Gamma\in\mathfrak{so}(3).\end{aligned} \tag{17}$$

Using (14) and (7), the $\mathscr{H}^*$ term in (17) can be rewritten as

$$\begin{aligned}\mathscr{H}^*(R,\nabla_RV(R,t),t) &= -\mathscr{D}_RV(R,t)\circ\nabla_RV(R,t)\\ &+\frac{1}{2}\sum_i\mathrm{trace}[(y_i-R^Ta_i)(y_i-R^Ta_i)^T]-\mathscr{D}_RV(R,t)\circ RA.\end{aligned} \tag{18}$$

To simplify the resulting equation, in the following a first order derivative of H^* is derived using (18).

$$\begin{aligned}\mathscr{D}_R\mathscr{H}^*\circ R\Gamma &= -\mathscr{D}_R^2V(R,t)\circ(\nabla_RV(R,t),R\Gamma)\\ &-\mathscr{D}_RV(R,t)\circ R\mathbb{P}_a(R^T\mathscr{D}_R(\nabla_RV(R,t))\circ R\Gamma)\\ &-\sum_i\mathrm{trace}[\Gamma^T\mathbb{P}_a(R^Ta_iy_i^T)]\\ &-\mathscr{D}_R^2V(R,t)\circ(RA,R\Gamma)-\mathscr{D}_RV(R,t)\circ R\mathbb{P}_a(\Gamma A)\end{aligned} \tag{19}$$

Replacing this equation in (17) and using the final conditions (13) and (15) will cancel the first order derivatives and the gradients. Recall that the second order derivative is connected to the Hessian operator as follows.

$$\mathscr{D}_R^2 V(R,t) \circ (R\Gamma, R\Omega) = \langle \text{Hess}_R V(R,t) \circ R\Gamma, R\Omega \rangle. \quad (20)$$

This helps replacing the second derivatives with the Hessian operator. We arrive at

$$\{\langle \text{Hess}_R V(R,t) \circ \dot{\hat{R}}(t), R\Gamma \rangle - \langle \sum_i R\mathbb{P}_a(R^T a_i y_i^T), R\Gamma \rangle - \langle \text{Hess}_R V(R,t) \circ RA, R\Gamma \rangle\}_{R=\hat{R}(t)} = 0, \quad \text{for all } \Gamma \in \mathfrak{so}(3). \quad (21)$$

Finally, canceling the direction Γ and applying the inverse Hessian yields

$$\dot{\hat{R}}(t) = \hat{R}A + \text{Hess}_{\hat{R}}^{-1} V(\hat{R},t) \circ (\hat{R}\sum_i \mathbb{P}_a(\hat{R}^T a_i y_i^T)). \quad (22)$$

The observer defined by Equation (22) generates the optimal state estimate given that the Hessian of the value function $V(R,t)$ exists and is invertible. To implement the observer we will need to compute an estimate of the Hessian operator $\text{Hess}_R V(R,t)$. The goal is to obtain a finite dimensional matrix ODE playing the same role as the matrix Riccati equation in the linear deterministic filtering case [4].

Recall that the Hessian operator is a symmetric linear map of the tangent space and can hence be represented in matrix form. There exists a symmetric matrix $K \in \mathbb{R}^{3\times 3}$ such that for all $\Omega \in \mathfrak{so}(3)$,

$$\text{Hess}_R V(R,t) \circ R\Omega = R\mathbb{P}_a(K\Omega). \quad (23)$$

This can be proved by considering the standard basis of $\mathfrak{so}(3)$, see also Proposition 3.1 in [15]. The inverse Hessian can also be represented in matrix form as follows.

$$\text{Hess}_R^{-1} V(R,t) \circ R\Omega = 2R\left((\text{trace}(K)I - K)^{-1}\text{vex}(\Omega)\right)_\times. \quad (24)$$

This immediately follows from

$$\text{vex}(KU_\times + U_\times K) = (\text{trace}(K)I - K)U, \quad (25)$$

where $K \in \mathbb{R}^{3\times 3}$ and U is the vector form of an arbitrary skew symmetric matrix $U_\times \in \mathfrak{so}(3)$. Equation (25) can be verified by considering a basis expansion. Equation (24) along with (1) converts (22) to a matrix equation.

$$\dot{\hat{R}}(t) = \hat{R}(A + \sum_i \left(Q^{-1}(y_i \times \hat{R}^T a_i)\right)_\times), \quad (26)$$

where $Q := \text{trace}(K)I - K$. The initial condition $\hat{R}(0) = I$ is obtained by evaluating (13) at $t = 0$. Equation (22) presents the dynamics of $\hat{R}(t)$ using an inverse Hessian operator that is time varying. In order to fully obtain $\hat{R}(t)$ we next need to obtain a dynamic equation for the Hessian operator.

Similar to what we did before we continue by computing the total time derivative of the second order derivative of the value function with respect to R, along the optimal trajectory $\hat{R}(t)$.

$$X := \frac{d}{dt}\{[\mathscr{D}_R^2 V(R,t) \circ (R\Gamma, R\Omega)]_{R=\hat{R}(t)}\}. \quad (27)$$

Propagating the time derivative inside the bracket yields

$$X = \{\mathscr{D}_R^2 V(R,t) \circ (R\mathbb{P}_a(R^T\dot{\hat{R}}\Gamma), R\Omega) + \mathscr{D}_R^2 V(R,t) \circ (R\mathbb{P}_a(R^T\dot{\hat{R}}\Omega), R\Gamma) + \mathscr{D}_R^3 V(R,t) \circ (\cdots) + \mathscr{D}_R^2(\frac{\partial V}{\partial t}(R,t)) \circ (R\Gamma, R\Omega)\}_{R=\hat{R}(t)}. \quad (28)$$

We neglect the third order term and from (11) we replace $\frac{\partial V(R,t)}{\partial t}$ with $\mathscr{H}^*(R, \nabla_R V(R,t), t)$. One can verify through a relatively lengthy calculations that the last term can be replaced by the following equation. Note that in deriving the following equation we neglect the third order derivatives of V and use the final condition (15) to cancel some of the terms.

$$\begin{aligned}&\mathscr{D}_R^2 \mathscr{H}^*(R, \nabla_R V(R,t), t) \circ (R\Gamma, R\Omega) \approx \\ &- 2\langle \text{Hess}_R V(R,t) \circ \text{Hess}_R V(R,t) \circ R\Gamma), R\Omega \rangle \\ &+ \langle \sum_i R\mathbb{P}_a(\Gamma\mathbb{P}_s(\hat{R}^T a_i y_i^T)), R\Omega \rangle \\ &+ \langle R\mathbb{P}_a(R^T(\text{Hess}_R V(R,t) \circ R\Gamma)A), R\Omega \rangle \\ &- \langle \text{Hess}_R V(R,t) \circ R\mathbb{P}_a(\Gamma A), R\Omega \rangle.\end{aligned} \quad (29)$$

Using (23) yields the matrix equation

$$\begin{aligned}&\mathscr{D}_R^2 \mathscr{H}^*(R, \nabla_R V(R,t), t) \circ (R\Gamma, R\Omega) \approx \\ &- 2\langle R\mathbb{P}_a(K\mathbb{P}_a(K\Gamma)), R\Omega \rangle \\ &+ \langle \sum_i R\mathbb{P}_a(\Gamma\mathbb{P}_s(\hat{R}^T a_i y_i^T)), R\Omega \rangle \\ &+ \langle R\mathbb{P}_a(\Gamma\mathbb{P}_s(KA)), R\Omega \rangle\end{aligned} \quad (30)$$

Replacing the previous equation and the observer equation (26) in (29) and using Equations (20) and (23) we get

$$\begin{aligned}X \approx \{&2\langle R\mathbb{P}_a(\Gamma\mathbb{P}_s(KA)), R\Omega \rangle + 2\langle R\mathbb{P}_a(\Gamma\mathbb{P}_s(K\beta)), R\Omega \rangle \\ &- 2\langle R\mathbb{P}_a(K\mathbb{P}_a(K\Gamma)), R\Omega \rangle \\ &+ \langle R(\mathbb{P}_a(\Gamma\mathbb{P}_s(Y^T R)), R\Omega \rangle\}_{R=\hat{R}(t)},\end{aligned} \quad (31)$$

where $\beta = \left(Q^{-1}\sum_i \text{vex}(\mathbb{P}_a(\hat{R}^T a_i y_i^T))\right)_\times$ is a skew symmetric matrix. On the other hand, X from (27) and using (23) is

$$X = \frac{d}{dt}\langle \text{Hess}_R V(R,t) \circ R\Gamma, R\Omega \rangle = \langle R\mathbb{P}_a(\dot{K}\Gamma), R\Omega \rangle. \quad (32)$$

Therefore by omitting the direction $R\Omega$ from the last two equations on X we obtain the following dynamic equation for $\tilde{K}$, an approximation of K.

$$\begin{aligned}\{\mathbb{P}_a(\dot{\tilde{K}}\Gamma) = \mathbb{P}_a[&2\Gamma\mathbb{P}_s(\tilde{K}A) + 2\Gamma\mathbb{P}_s(\tilde{K}\beta) \\ &- 2\tilde{K}\mathbb{P}_a(\tilde{K}\Gamma) + \sum_i \Gamma\mathbb{P}_s(\hat{R}^T a_i y_i^T)]\}_{R=\hat{R}(t)}.\end{aligned} \quad (33)$$

Applying the Vex operator to this equation and replacing $\tilde{K}$ with $\tilde{Q} := \text{trace}(\tilde{K})I - \tilde{K}$ yields the dynamics of $\tilde{Q}$, an approximation to Q.

$$\begin{aligned}\dot{\tilde{Q}} = &\mathbb{P}_s(2\tilde{Q}A - \sum_i \tilde{Q}\left(\tilde{Q}^{-1}(y_i \times \hat{R}^T a_i)\right)_\times) - \tilde{Q}^2 \\ &+ \sum_i \text{trace}(\mathbb{P}_s(\hat{R}^T a_i y_i^T))I - \sum_i \mathbb{P}_s(\hat{R}^T a_i y_i^T),\end{aligned} \quad (34)$$

where $\tilde{Q}(0) = \text{trace}(K_0^{-1})I - K_0^{-1}$. This initial condition is obtained by differentiating Equation (12) with respect to R, using the final condition (15) and the fact that $\hat{R}(0) = I$.

V. Results

In summary, following the calculations in Section IV we have obtained the following filter. Note that for convenience we drop the *tilde* notation.

$$\dot{\hat{R}} = \hat{R}(A + \sum_i \left(Q^{-1}(y_i \times \hat{R}^T a_i)\right)_\times),\ \hat{R}(0) = I, \tag{35a}$$

$$\begin{aligned}\dot{Q} =& \mathbb{P}_s(2QA - \sum_i Q\left(Q^{-1}(y_i \times \hat{R}^T a_i)\right)_\times) - Q^2 \\ &+ \sum_i \text{trace}(\mathbb{P}_s(\hat{R}^T a_i y_i^T))I - \sum_i \mathbb{P}_s(\hat{R}^T a_i y_i^T),\end{aligned} \tag{35b}$$

where $Q(0) = \text{trace}(K_0^{-1})I - K_0^{-1}$ and K_0 is known from the cost (3). The signals A and $\{y_i\}$ are defined by the system (2). The filter in Equation (35) consists of two interconnected parts. Equation (35a) evolves on SO(3) and consists of a copy of system (2) plus an innovation term. The innovation term is a weighted distance between the (past) estimated signal and the noisy measured signal projected on the tangent space. Note that $y_i \times \hat{R}^T a_i$ encodes the rotation required to take $\hat{R}^T a_i$ to y_i in Riemannian normal coordinates. The weighting Q^{-1} is applied to the skew coordinates of this distance and the results are projected back on the tangent space. The (inverse) weighting matrix Q, dynamically generated by (35b), depends on the past estimates and the past measurements. The second equation (35b) is a time varying Riccati differential equation.

Consider the system (2) and the cost (3). Given some measurements $\{y_i|_{[0,\,t]}\}$ and their associated input $A|_{[0,\,t]}$, assume that unique solutions $\hat{R}(t)$ and $Q(t)$ to (35a) and (35b) exist on $[0,\,t]$.

Theorem 1: Assuming that $V(R,t)$ from (10) is twice differentiable and that $\text{Hess}_R V(R,t)$ is invertible, $\hat{R}(t)$ given from (35a) is optimal where $Q = \text{trace}(K)I - K$ and $\text{Hess}_R V(R,t) \circ R\Omega = R\mathbb{P}_a(K\Omega)$, for all $\Omega \in \mathfrak{so}(3)$.

Proof: From our previous calculations in Section IV it is easy to see that Equation (35a) is derived only using the optimality conditions (7) and (13). ■

Remark 1: The left invariant observer (35a) is in fact a gradient-based observer [11]. The innovation term in this equation is a gradient of the right invariant cost function $f(\hat{R}, \{y_i\}) = \sum_i |y_i - \hat{R}^T a_i|^2$ with respect to $\hat{R}$ and the right invariant metric given by $\langle A, B\rangle_Q := \text{trace}[AQB^T]$ for all $A, B \in \mathfrak{so}(3)$, where Q is a positive definite matrix.

Remark 2: Although we obtain the optimal observer (35a) we would need the true Hessian operator to be able to compute this optimal filter. The next proposition states that Equation (35b) is an approximation to the dynamics of the Hessian operator.

Proposition 1: Assuming that $V(R,t)$ is three times differentiable and that $\text{Hess}_R V(R,t)$ is invertible, $\dot{Q}$ in (35b) approximates the derivative of the true Hessian of $V(R,t)$ to the order $O(\nabla_R^3 V(R,t))$.

Proof: In Section IV, deriving Equation (27) was followed by using the optimal equation (35a), the final value condition (13) and neglecting the third order derivatives of the value function. Hence, the dynamics of the Hessian operator was approximated up to the second order. ■

Remark 3: Similar to our previous filter derivation in Section IV we could continue with Mortensen's approach to derive a higher order filter. However, this would require some tedious tensor algebra.

VI. Simulations

Note that rewriting the proposed filter (35) using $P := Q^{-1}$ yields the following filter that is more suitable for implementation as it avoids the inverse operation.

$$\dot{\hat{R}} = \hat{R}(A + \sum_i \left(P(y_i \times \hat{R}^T a_i)\right)_\times),\ \hat{R}(0) = I, \tag{36a}$$

$$\begin{aligned}\dot{P} =& \mathbb{P}_s(2PA - \sum_i P\left(P(y_i \times \hat{R}^T a_i)\right)_\times) + I \\ &- P(\sum_i \text{trace}(\mathbb{P}_s(\hat{R}^T a_i y_i^T))I - \sum_i \mathbb{P}_s(\hat{R}^T a_i y_i^T))P,\end{aligned} \tag{36b}$$

where $P(0) = (\text{trace}(K_0^{-1})I - K_0^{-1})^{-1}$. To achieve better implementation properties we also convert (36a) to a quaternion observer in which rotations are represented by unit quaternions. Note that if $q = [\,q_0 \quad \vec{q}\,]^T \in \mathbb{R}^4$ is a quaternion with unit norm where $\vec{q} \in \mathbb{R}^3$, then q correponds to $R(q) \in \text{SO}(3)$

$$R(q) := I_3 + 2q_0\vec{q}_\times + 2\vec{q}_\times^{\,2}. \tag{37}$$

Hence the proposed quaternion filter is given by

$$\dot{\hat{q}} = \frac{1}{2}\hat{q} \otimes (\text{vex}(A) + \sum_i P(y_i \times \hat{y}_i^T)),\ \hat{q}(0) = [\,1\,0\,0\,0\,]^T, \tag{38a}$$

$$\begin{aligned}\dot{P} =& \mathbb{P}_s(2PA - \sum_i P\left(P(y_i \times \hat{y}_i^T)\right)_\times) + I \\ &- P(\sum_i \text{trace}(\mathbb{P}_s(\hat{y}_i y_i^T))I - \sum_i \mathbb{P}_s(\hat{y}_i y_i^T))P,\end{aligned} \tag{38b}$$

where $P(0) = (\text{trace}(K_0^{-1})I - K_0^{-1})^{-1}$, the operator $\otimes$ is the quaternion multiplication and $\hat{y}_i := \hat{R}^T a_i$. It is straightforward to verify that $\hat{y}_i$ is also equal to the vector part of $\hat{q}^{-1} \otimes [\,0 \quad a\,]^T \otimes \hat{q}$, where for $q = [\,q_0 \quad \vec{q}\,]^T$ the inverse is defined as $q^{-1} := [\,q_0 \ -\vec{q}\,]^T$.

The unit quaternion MEKF is given by [12]

$$\dot{\hat{q}}_m = \frac{1}{2}\hat{q}_m \otimes (\text{vex}(A) + \sum_i P_m(y_i \times \hat{y}_i^T)),\ \hat{q}_m(0) = [1\,0\,0\,0]^T, \tag{39a}$$

$$\dot{P}_m = \mathbb{P}_s(2P_m A) + I - P_m(\sum_i \text{trace}(\mathbb{P}_s(\hat{y}_i \hat{y}_i^T))I - \sum_i \mathbb{P}_s(\hat{y}_i \hat{y}_i^T))P_m, \tag{39b}$$

where $P_m(0) = I$ and $\hat{y}_i$ is equal to the vector part of $\hat{q}_m^{-1} \otimes [\,0 \quad a\,]^T \otimes \hat{q}_m$.

Remark 4: Note that the MEKF and the proposed filter share the same observer equations (39a) and (38a). However, the proposed Riccati equation (38b) differs from the MEKF's Riccati equation (39b). The term $\mathbb{P}_s \sum_i P\left(P(y_i \times \hat{y}_i^T)\right)_\times$ in (38b) is not present in the MEKF. This term will be very small once the filter has converged and $\hat{y}_i$ is close to y_i. More importantly, note that the quadratic term in the

Riccati equation of the proposed filter is different to the MEKF by utilizing not only the information contained in the estimates $\{\hat{y}_i\}$ but also the information in the measurements $\{y_i\}$. Our simulations indicate better transient and asymptotic behaviour of the proposed filter compared to the MEKF.

We have performed a series of simulation tests using a variety of error signals, inputs and system initial conditions. In all the simulations the two filters (39) and (38) were tested with the same initial conditions, measurements and error signals. A situation typical to our simulation tests is considered which involves a sinusoidal input A with maximum frequency of 0.1. The input measurement error signal δ in system (2) is a random process with a standard deviation of 36 degrees. Two unit reference vectors $\{a_i\}$ that are 56 degrees apart are considered. The measurement error signals $\{\varepsilon_i\}$ are assumed to be random processes with standard deviation of 45 degrees. The system (2) is initiated with a unit quaternion that has a rotation angle of 158 degrees. The simulation runs for 50 units of time with a time step of 0.01 units. Figure 1 shows that the rotation angle estimation error

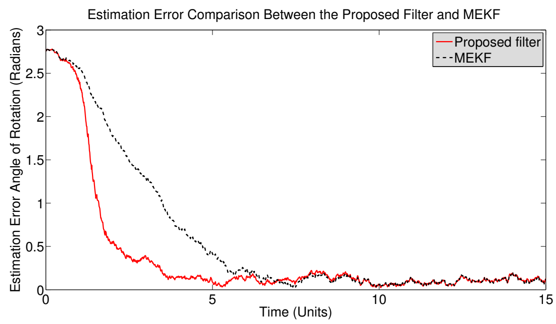

Fig. 1. **Comparing the estimation errors of the proposed filter and the MEKF. Note that this figure is zoomed to show the first 15 units of time only.**

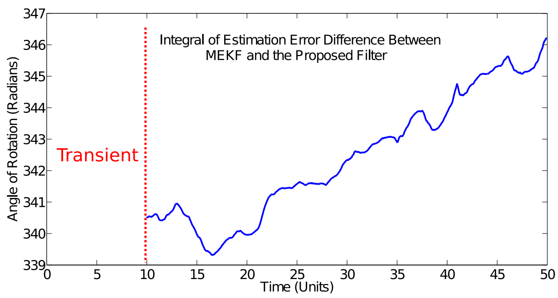

Fig. 2. **Integral of the estimation error of the MEKF minus the estimation error of the proposed filter. Note that the transient part of the figure has been removed. As can be seen the error difference is growing as time evolves which indicates that the error of the proposed filter is asymptotically smaller than the error of the MEKF.**

of the proposed filter converges to a small value more rapidly compared to the MEKF. This phenomenon is consistently observed in all simulations with different parameters which shows that the proposed filter has a more desirable transient behaviour compared to the MEKF. Figure 2 illustrates the integral of the rotation angle estimation error of the proposed filter subtracted from the rotation angle estimation error of the MEKF plotted against time. The integral is slowly growing in time which indicates the proposed filter has also a slight advantage in the asymptotic behaviour compared to the MEKF. This behaviour difference is also observed in most of our simulation tests with different initial conditions and measurement errors.

VII. Conclusions

In this paper we apply Mortensen's nonlinear filtering approach to kinematics on the Lie group SO(3). We obtain the exact form of a minimum-energy (optimal) filter for this system. We show that it involves the Hessian of the value function (of the corresponding optimal control problem) for which we derive a dynamic equation neglecting the third order derivative of the value function. By means of simulations we conclude that the proposed filter outperformances the industry standard method MEKF in several situations with different measurement and initialization errors.

VIII. Acknowledgments

This research was supported by the Australian Research Council through discovery grant DP0987411 "State Observers for Control Systems with Symmetry".

References

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

[2] E. D. Sontag, *Mathematical control theory: deterministic finite dimensional systems*. Springer, 1998.

[3] A. H. Jazwinski, *Stochastic processes and filtering theory*. Academic Press, April 1970.

[4] J. C. Willems, "Deterministic least squares filtering," *Journal of Econometrics*, vol. 118, no. 1-2, pp. 341–373, 2004.

[5] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering, Transactions of the ASME*, vol. 82, no. 1, pp. 35–45, 1960.

[6] A. Krener, "The convergence of the minimum energy estimator," in *Lecture Notes in Control and Information Sciences*, W. Kang, Ed. Springer, 2004, pp. 187–208.

[7] A. Aguiar and J. Hespanha, "Minimum-energy state estimation for systems with perspective outputs," *IEEE Transactions on Automatic Control*, vol. 51, no. 2, pp. 226–241, 2006.

[8] P. Coote, J. Trumpf, R. Mahony, and J. C. Willems, "Near-optimal deterministic filtering on the unit circle," in *IEEE Conference on Decision and Control*, 2009, pp. 5490–5495.

[9] M. Zamani, J. Trumpf, and R. Mahony, "Near-Optimal deterministic filtering on the Rotation Group," *IEEE Transactions on Automatic Control*, vol. 56, no. 6, pp. 1411–1414, 2011.

[10] S. Bonnabel, P. Martin, and P. Rouchon, "Non-linear symmetry-preserving observers on Lie groups," *IEEE Transactions on Automatic Control*, vol. 54, no. 7, pp. 1709–1713, 2009.

[11] C. Lageman, J. Trumpf, and R. Mahony, "Gradient-like observers for invariant dynamics on a Lie group," *IEEE Transactions on Automatic Control*, vol. 55, no. 2, pp. 367–377, 2010.

[12] F. Markley, "Attitude error representations for Kalman filtering," *Journal of guidance, control, and dynamics*, vol. 26, no. 2, pp. 311–317, 2003.

[13] J. L. Crassidis, F. L. Markley, and Y. Cheng, "Survey of nonlinear attitude estimation methods," *Journal of Guidance Control and Dynamics*, vol. 30, pp. 12–28, 2007.

[14] V. Jurdjevic, *Geometric control theory*. Cambridge Univ Pr, 1997.

[15] A. Bloch, P. Crouch, J. Marsden, and A. Sanyal, "Optimal control and geodesics on quadratic matrix Lie groups," *Foundations of Computational Mathematics*, vol. 8, no. 4, pp. 469–500, 2008.