Minimum-Energy Filtering on the Unit Circle

Mohammad Zamani, Jochen Trumpf, and Robert Mahony

***Abstract*— We apply Mortensen's deterministic filtering approach to derive a third order minimum-energy filter for a system defined on the unit circle. This yields the exact form of a minimum-energy filter (namely an observer plus a Riccati equation that updates the observer gain). The proposed Riccati equation is perturbed by a term depending on the third order derivative of the value function of the associated optimal control problem. The proposed filter is third order in the sense that it approximates the dynamics of the third order derivate of the value function by neglecting the fourth order derivative of the value function. Additionally, we show that the near-optimal filter proposed by Coote *et al.* in prior work can indeed be derived from a second order application of Mortensen's approach to minimum-energy filtering on the unit circle.**

I. INTRODUCTION

Optimal filtering involves estimating the state of a noisy system based on a criterion that is typically expressed as minimizing a cost function. For linear systems, the Kalman filter [1], posed in a stochastic system modeling framework, is a finite-dimensional optimal filter designed to minimize the error covariance.

Filtering theory has also been developed in a deterministic system modeling framework, where unlike in the stochastic framework, the uncertainties of a system are not modeled as stochastic processes with a priori assumptions on their statistics but rather as unknown functions of time. The deterministic filtering problem is to find a solution to the system equations that is compatible with the observations and minimizes the magnitude of the associated noise signals. Such filters are known as *minimum-energy* filters [2] since the cost used is typically the energy in the unknown noise signals. For linear systems it is known that the same Kalman filtering formulas are obtained also in a deterministic framework using a least squares cost (cf. [2]–[5]).

Minimum-energy filtering for nonlinear systems has been studied by Krener [6] who proved that under some conditions including the uniform observability of the system and the presence of a time-dependent "forgetting factor" in the cost, a minimum-energy estimate converges exponentially fast to the true state. Aguilar *et al.* [7] proposed a minimum-energy nonlinear filter for systems with perspective outputs and algebraic state constraints. They consider embedding the nonlinear geometry given by the constraints in a general Euclidean space $\mathbb{R}^n$ yielding asymptotically optimal results for systems inherently defined only on a submanifold of $\mathbb{R}^n$. In separate work, Coote *et al.* [8] designed a near-optimal minimum-energy filter that utilizes the geometric structure of the unit circle S^1. In recent work by the authors [9] a near-optimal minimum-energy filter posed directly on the Special Orthogonal Group $\mathrm{SO}(3)$ is derived. These latter two filters are designed directly on the underlying geometric structure of their application space through identification of a suitable "Lyapunov" function for the optimality analysis and include explicit bounds on their distance to optimality.

In the late 1960's Mortensen [10] proposed a systematic approach to deriving deterministic minimum-energy filters for nonlinear systems. The method was further explored by Hijab [11]. In this approach the optimal filtering problem is broken down into two steps. The first step involves applying the maximum principle of optimal control and dynamic programming to optimize an energy functional in the system "noise" signals. In the second step a further optimization takes place over the system's initial state value.

In this paper we consider applying Mortensen's minimum-energy filtering approach to design a filter on the Lie group S^1. Invariant systems such as systems defined by invariant vector fields on Lie groups appear in many applications including mechanical control systems [12]. The Lie group $S^1 \simeq SO(2)$ is important as it acts as a "training ground" for the Lie group $\mathrm{SO}(3)$ that describes the attitude of rigid bodies such as autonomous flying vehicles [9]. Using Mortensen's method we obtain the exact form of a minimum-energy filter (namely an observer on S^1 plus a Riccati equation updating the observer gain). We note that the obtained Riccati equation is perturbed by a term depending on the third order derivative of the value function of the associated optimal control problem. The proposed filter is third order in the sense that it approximates the ordinary differential equation (ODE) governing the evolution of the third order derivate of the value function by neglecting the fourth order derivative of the value function. We also point out that using Mortensen's method higher order optimal filters are straight forward to obtain and we discuss their general formulation. Although we only present derivation of the third order approximate filter the approach is straight forward to generalize to higher order approximate filters and we provide simulations for approximate filters up to the eighth order. This work improves the previous results on S^1 [8] which we prove is a second order approximation of the current filter.

The paper is organized as follows. Section II defines the system on S^1 and introduces the problem of minimum-energy filtering applied to this system. Mortensen's approach is the subject of Section III where a recursive third order approximate solution to the problem in Section II is derived. We formally presents the results and includes remarks on the numerical implementation of the proposed filter. The

The authors are with the 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.

filter derived in [8] is shown to be equal to the second order Mortensen approximate of the minimum-energy filter. A suite of simulations are included in Section IV that present the tracking performance of an eighth-order approximate of the minimum-energy filter on S^1. Moreover, a discussion around the higher derivatives of the value function and their relationship to each other is provided. Lastly Section V concludes the paper.

II. Minimum-Energy Filtering on the Unit Circle

Consider the system

$$\begin{aligned}\dot{\theta}(t) &= u(t) + g\delta(t),\ \theta(0) = \theta_0,\\ y(t) &= \theta(t) + \varepsilon(t),\end{aligned} \tag{1}$$

where θ, y and ε are the state, output and output measurement error signals that evolve on the unit circle S^1. The signals u and δ denote the scalar valued input and input measurement errors that live in the tangent space $T_\theta S^1$. The scalar g is a known positive constant.

This system represents the kinematics of an object moving on a unit circle. The angle $\theta(t)$, associated to the current position of the object on the circle, is considered as the current state of the system (1). We assume we can measure the "rate" of rotation u of the system and δ denotes the unknown measurement error that we make in measuring u. We also assume we can directly measure the current state ($\theta(t)$) of the system and ε models the error that occurs in this measurement process. We will drop the explicit argument (t) for the reminder of the paper.

Remark 1: In deterministic minimum-energy filtering the signals δ and ε are considered as unknown deterministic functions of time. This is different to the stochastic setup where these signals are typically modeled as stochastic processes with some a priori assumptions on their statistical distribution.

Consider the cost

$$J(t;\theta_0,\delta|_{[0,\,t]},\varepsilon|_{[0,\,t]}) = \int_0^t \left(\frac{1}{2}\delta(\tau)^2 + (1-\cos(\varepsilon(\tau)))\right)d\tau + \frac{1}{K_0}(1-\cos(\theta(0))), \tag{2}$$

where K_0 is a positive scalar.

The problem is to design a filter that estimates the current state $\theta(t)$ of the system (1), given that a record of the past measurements $u|_{[0,\,t]}$ and $y|_{[0,\,t]}$ is available. The filtering criterion is to minimize the cost functional (2).

Note that the cost (2) depends on the unknowns ($\delta|_{[0,\,t]}$, $\varepsilon|_{[0,\,t]}$, θ_0) of the system (1) and encodes the total energy associated to them. In a sense, by minimizing (2) the goal is to find unknowns of minimum energy that together with the measurements $u|_{[0,\,t]}$ and $y|_{[0,\,t]}$ satisfy the system equations (1). Note that in general one might find infinitely many possible combinations of the unknown signals ($\delta|_{[0,\,t]}$, $\varepsilon|_{[0,\,t]}$, θ_0) that together with the measurements satisfy the system equations (1). However, the principle of minimum-energy filtering singles out a set of minimizing unknowns ($\delta^*|_{[0,\,t]}$, $\varepsilon^*|_{[0,\,t]}$, θ_0^*). Plugging them into system (1), together with the known measurements, yields the minimum-energy state trajectory $\theta^*_{[0,\,t]}$. The subscript $[0,\,t]$ indicates that the optimization takes place on the interval $[0,\,t]$. We pick the final optimal state $\theta^*_{[0,\,t]}(t)$ as our minimum-energy estimate at time t, $\hat{\theta}(t) := \theta^*_{[0,\,t]}(t)$.

At each time t, the filtering principle entails finding the full optimal state trajectory $\theta^*_{[0,\,t]}$ and picking its final value $\theta^*_{[0,\,t]}(t)$ as our minimum-energy estimate at that time. Rather than resolving this infinite dimensional optimization problem at each time instance, the aim is to design a recursive filter that updates directly the state estimate $\theta^*_{[0,\,t]}(t)$.

We begin by turning this problem into an optimal control problem. Firstly, note that the cost (2) depends on the unknowns ($\delta|_{[0,\,t]}$, $\varepsilon|_{[0,\,t]}$, θ_0), but given $\delta|_{[0,\,t]}$ and θ_0 and the known measurements $u|_{[0,\,t]}$ and $y|_{[0,\,t]}$, the signal $\varepsilon|_{[0,\,t]}$ is uniquely determined by (1). Hence, substituting ε in (2) from (1) yields the simplified cost

$$J(t;\theta_0,\delta|_{[0,\,t]}) = \int_0^t \left(\frac{1}{2}\delta^2 + (1-\cos(y-\theta))\right)d\tau + \frac{1}{K_0}(1-\cos(\theta(0))). \tag{3}$$

Now, the signal $\delta|_{[0,t]}$ is treated as the control input and (3) is minimized over δ. Initially, $\theta(0) = \theta_0$ is considered as fixed. The end point $\theta(t)$ is free but constrained by the measurements and by (1) in the period $[0,\,t]$. Note that to fully solve our original optimal filtering problem we need a further step to optimize over the initial value θ_0.

To apply the Maximum Principle, define the following Hamiltonian [13].

$$\mathscr{H}(\theta,p,\delta,t) := \frac{1}{2}\delta^2 + 1 - \cos(y-\theta) - p\dot{\theta}, \tag{4}$$

where $p \in T^*_\theta S^1$ is the costate variable. Minimizing the Hamiltonian over δ to compute the critical point of $\mathscr{H}$ yields

$$\nabla_\delta \mathscr{H} = 0, \Longrightarrow \delta - pg = 0, \tag{5}$$

that results in $\delta^* = gp$ where the superscripted star denotes the optimal variable. Here $\nabla_\delta \mathscr{H}$ denotes the partial derivative of $\mathscr{H}$ with respect to δ. Substituting δ^* the optimal Hamiltonian is

$$\mathscr{H}^*(\theta,p,t) = -\frac{1}{2}g^2p^2 + 1 - \cos(y-\theta) - pu. \tag{6}$$

From the Maximum principle one has the Hamilton equations

$$\begin{cases}\dot{\theta} &= \nabla_p \mathscr{H}^*(\theta,p,t),\\ \dot{p} &= -\nabla_\theta \mathscr{H}^*(\theta,p,t).\end{cases} \tag{7}$$

Since $\theta(t)$ is free one has the boundary condition

$$p(0) = \frac{\sin(\theta_0)}{K_0}. \tag{8}$$

Applying the dynamic programming principle to this problem, define the value function

$$V(\theta,t) := \min_{\delta|_{[0,\,t]}} J(t;\theta_0,\delta|_{[0,\,t]}), \tag{9}$$

where J is the cost (3) and the minimization is subjected to the system equations (1). The Hamilton-Jacobi-Bellman equation is then [13]

$$\mathscr{H}^*(\theta,\nabla_\theta V(\theta,t),t)-\nabla_t V(\theta,t)=0, \tag{10}$$

with the initial time boundary condition

$$V(\theta_0,0)=\frac{1}{K_0}(1-\cos(\theta_0)). \tag{11}$$

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 θ_0. This is equivalent to a further optimization step over $\theta(t)$ (given the measurements from (1), θ_0 and $\delta|_{[0,t]}$ uniquely determine $\theta(t)$ and vice versa $\theta(t)$ and $\delta|_{[0,t]}$ uniquely determine $\theta(0)$). Hence the optimal filtering solution $\hat{\theta}$ is characterized by the following criticality condition [10]

$$\nabla_\theta V(\theta,t)|_{\theta=\hat{\theta}(t)}=0. \tag{12}$$

Recall that the minimum-energy estimate $\hat{\theta}(t)$ is the minimizing argument $\theta^*(t)$, that yields the final condition (12). Solving Equation (12) is clearly a way to characterize $\hat{\theta}(t)$. However, the goal is to find a differential equation that dynamically updates $\hat{\theta}(t)$.

Up to here we have introduced our optimization problem and we have shown how to approach it in the context of the Maximum Principle and dynamic programming. In the next section we employ the proposed solution by Mortensen [10] to derive an approximation of a recursive minimum-energy filter on S^1.

III. Optimal Filter Derivation

In this section we apply Mortensen's filtering approach [10] to the problem introduced in Section II to obtain a third order minimum-energy filter.

The goal is to obtain a dynamic equation for the solution $\hat{\theta}$ by computing the total time derivative of (12).

$$\frac{d}{dt}\{\nabla_\theta V(\theta,t)|_{\theta=\hat{\theta}(t)}\}=0. \tag{13}$$

Applying the chain rule yields

$$\{\nabla_\theta\left(\nabla_\theta V(\theta,t)\dot{\hat{\theta}}(t)+\nabla_t V(\theta,t)\right)\}_{\theta=\hat{\theta}(t)}=0. \tag{14}$$

From (10) one can substitute $\nabla_t V(\theta,t)$ to get

$$\{\nabla^2_\theta V(\theta,t)\dot{\hat{\theta}}(t)+\nabla_\theta(\mathscr{H}^*(\theta,\nabla_\theta V(\theta,t),t))\}_{\theta=\hat{\theta}(t)}=0, \tag{15}$$

From the chain rule the second term is

$$\begin{aligned}\nabla_\theta(\mathscr{H}^*(\theta,\nabla_\theta V(\theta,t),t)) &= [\nabla_\theta\mathscr{H}^*(\theta,p,t)\\ &+\nabla^2_\theta V(\theta,t)\nabla_p\mathscr{H}^*(\theta,p,t)]_{p=\nabla_\theta V(\theta,t)}.\end{aligned} \tag{16}$$

Using the optimal Hamiltonian (6) yields

$$\begin{aligned}\nabla_\theta(\mathscr{H}^*(\theta,\nabla_\theta V(\theta,t),t)) &= -\sin(y-\theta)\\ &+\nabla^2_\theta V(\theta,t)(-g^2\nabla_\theta V(\theta,t)-u).\end{aligned} \tag{17}$$

Replacing (17) in (15) and assuming that $\nabla^2_\theta V(\theta,t)$ is nonzero yields the following optimal observer. Note that part of the second term in (17) is zero as $\nabla_\theta V(\theta,t)$ yields zero from the final condition (12).

$$\dot{\hat{\theta}}=u+(\nabla^2_{\hat{\theta}}V(\hat{\theta},t))^{-1}\sin(y-\hat{\theta}), \tag{18}$$

where the optimal initial guess $\hat{\theta}(0)=0$ is obtained by evaluating the final condition (12), at time zero, using the boundary condition (11).

Remark 2: In practice, some prior information about the system's initial state value $\theta(0)$ might be available. In that case the cost function (2) is modified as follows.

$$\begin{aligned}J(t;\theta_0,\delta|_{[0,t]},\varepsilon|_{[0,t]}) &= \int_0^t\left(\frac{1}{2}\delta(\tau)^2+(1-\cos(\varepsilon(\tau)))\right)d\tau\\ &+\frac{1}{K_0}(1-\cos(\theta(0)-\mu)),\end{aligned} \tag{19}$$

where μ is as a priori estimate for the initial state $\theta(0)$. This also affects the boundary condition (11) and leads to the optimal initial state estimate $\hat{\theta}(0)=\mu$.

Equation (18) is the exact kinematics of the optimal nonlinear observer. This equation consists of a copy of the system, i.e. the input u, plus a weighted innovation term $\sin(y-\hat{\theta})$. The innovation term is the difference between the output measurement y and the estimate $\hat{\theta}$ projected onto the tangent space using the sin function. Note that in this equation the inverse Hessian of the value function $(\nabla^2_{\hat{\theta}}V(\hat{\theta},t))^{-1}$ acts as a time varying gain on the innovation term $\sin(y-\hat{\theta})$.

In order to implement (18) one requires the knowledge of the Hessian $\nabla^2_{\hat{\theta}}V(\hat{\theta},t)$ of the value function. Mortensen's approach provides an algorithmic approach to derive a dynamical system that can compute $\nabla^2_{\hat{\theta}}V(\hat{\theta},t)$ on-line. From now on we will denote $P:=\nabla^2_{\hat{\theta}}V(\hat{\theta},t)$ and hence the observer (18) takes the form

$$\dot{\hat{\theta}}=u+\frac{\sin(y-\hat{\theta})}{P},\ \hat{\theta}(0)=0. \tag{20}$$

In order to obtain the kinematics of P we proceed similarly to our previous calculations by taking the total time derivative of the Hessian and by applying (12) and (10).

$$\dot{P}=\frac{d}{dt}\{\nabla^2_\theta V(\theta,t)|_{\theta=\hat{\theta}(t)}\}. \tag{21}$$

Applying the chain rule yields

$$\dot{P}=\{\nabla^3_\theta V(\theta,t)\dot{\hat{\theta}}(t)+\nabla^2_\theta\mathscr{H}^*(\theta,\nabla_\theta V(\theta,t),t)\}_{\theta=\hat{\theta}(t)}. \tag{22}$$

Note that we have used (10) to replace $\nabla_t V(\theta,t)$. Next, the second term in (22) is obtained by differentiating (17).

$$\begin{aligned}\nabla^2_\theta\mathscr{H}^*(\theta,\nabla_\theta V(\theta,t),t) &= \cos(y-\theta)\\ &+\nabla^3_\theta V(\theta,t)(-g^2\nabla_\theta V(\theta,t)-u)-g^2(\nabla^2_\theta V(\theta,t))^2.\end{aligned} \tag{23}$$

Replacing this equation and (20) in (22) and using the final condition (12) yields

$$\dot{P}=\nabla^3_{\hat{\theta}}V(\hat{\theta},t)\frac{\sin(y-\hat{\theta})}{P}-g^2P^2+\cos(y-\hat{\theta}), \tag{24}$$

where the initial Hessian $P(0)=\frac{1}{K_0}$ is obtained by calculating the Hessian of the initial condition (11) evaluated at $\hat{\theta}(0)=0$. We denote by S the third order derivative of the value function $S:=\nabla^3_{\hat{\theta}}V(\hat{\theta},t)$. Hence the resulting Riccati equation is

$$\dot{P}=\cos(y-\hat{\theta})-g^2P^2+S\frac{\sin(y-\hat{\theta})}{P},\ P(0)=\frac{1}{K_0}. \quad (25)$$

Equation (25) encodes the exact scalar Riccati equation that updates the observer gain P in (20). Note that this Riccati equation is perturbed by a weighted correction term $\frac{\sin(y-\hat{\theta})}{P}$. In order to fully realize this equation one needs to know the weighting S that is again a time-varying signal.

Now in order to obtain the dynamics of the third order term S compute the following derivative along the optimal trajectory $\hat{\theta}$.

$$\dot{S}=\frac{d}{dt}\{\nabla^3_{\theta}V(\theta,t)|_{\theta=\hat{\theta}(t)}\}. \quad (26)$$

Applying the chain rule yields

$$\dot{S}=\{\nabla^4_{\theta}V(\theta,t)\dot{\hat{\theta}}(t)+\nabla^3_{\theta}\mathscr{H}^*(\theta,\nabla_{\theta}V(\theta,t),t)\}_{\theta=\hat{\theta}(t)}. \quad (27)$$

Note that (10) is used to replace $\nabla_t V(\theta,t)$. Next, the second term in (27) is obtained by differentiating (23).

$$\begin{aligned}&\nabla^3_{\theta}\mathscr{H}^*(\theta,\nabla_{\theta}V(\theta,t),t)=\sin(y-\theta)\\&+\nabla^4_{\theta}V(\theta,t)(-g^2\nabla_{\theta}V(\theta,t)-u)-g^2\nabla^3_{\theta}V(\theta,t)\nabla^2_{\theta}V(\theta,t)\\&-g^2\nabla^3_{\theta}V(\theta,t)\nabla^2_{\theta}V(\theta,t)-g^2\nabla^2_{\theta}V(\theta,t)\nabla^3_{\theta}V(\theta,t).\end{aligned} \quad (28)$$

Applying the final condition (12) yields

$$\begin{aligned}&\nabla^3_{\theta}\mathscr{H}^*(\theta,\nabla_{\theta}V(\theta,t),t)=\sin(y-\theta)-\nabla^4_{\theta}V(\theta,t).u\\&-3g^2\nabla^3_{\theta}V(\theta,t)\nabla^2_{\theta}V(\theta,t)\end{aligned} \quad (29)$$

Replacing (29) in (27) yields

$$\dot{S}=\sin(y-\hat{\theta})-3g^2PS+\nabla^4_{\hat{\theta}}V(\hat{\theta},t)\frac{\sin(y-\hat{\theta})}{P},\ S(0)=0, \quad (30)$$

where the initial condition $S(0)$ is obtained by evaluation the third derivative of the initial value function (11) along $\hat{\theta}$.

Note that (30) is the exact form of an optimal first order differential equation that updates the scaling of the correction term $\frac{\sin(y-\hat{\theta})}{P}$ in the Riccati equation (25). In the same way S was used as a variable to compute $\nabla^3_{\hat{\theta}}V(\hat{\theta},t)$, we could continue to introduce variables for the unknown higher order derivatives of the value function and derive ordinary differential equations for their evolution by taking their time derivative. In order to exactly realize the optimal filter, then at some point, knowledge of the nth derivative $\nabla^n_{\hat{\theta}}V(\hat{\theta},t)$ is required. Although, in general this may never be possible, there are certain situations where the structure of a system may yield knowledge of this term. For example, for linear systems with a quadratic cost functional it is known that $\nabla^3_{\hat{\theta}}V(\hat{\theta},t)=0$ and the second order filter becomes optimal. The following theorem is stated for the third order filter case but can easily be reformulated to an arbitrary order filter.

Theorem 1: Consider the system (1) and the cost (2). Given some output measurements $y|_{[0,\,t]}$ and their associated input measurements $u|_{[0,\,t]}$, assume that unique solutions $P(t)$ to (25) and $S(t)$ to (30) exist on $[0,\ t]$. Assuming that the value function (9) is four times differentiable and that the Hessian $\nabla^2_{\hat{\theta}}V(\hat{\theta},t)$ is nonzero, the three equations (20), (25) and (30) yield a minimum-energy (optimal) filter on the system defined on S^1 (1) given that $\nabla^4_{\hat{\theta}}V(\hat{\theta},t)$ is known.

Proof: Following the maximum principle and Mortensen's approach in Sections II and III and given that the assumptions in Theorem 1 hold Equations (20), (25) and (30) are obtained only using the optimality conditions (6), (10) and (12). ■

In practice, the filter equations can be truncated at any desired order by neglecting the contribution of the unknown higher order derivative of the value function. In this paper, we will assume that the fourth order derivative of the value function $\nabla_4V(\theta,t)$ is small enough to be neglected and use the following approximation to Equation (30).

$$\dot{\tilde{S}}=\sin(y-\theta)-3g^2P\tilde{S},\ \tilde{S}(0)=0, \quad (31)$$

where $\tilde{S}\approx S$.

In summary, we obtain the following third order approximate filter. Note that for convenience we drop the *tilde* notation.

$$\dot{\hat{\theta}}=u+\frac{\sin(y-\hat{\theta})}{P},\ \hat{\theta}(0)=0, \quad (32a)$$

$$\dot{P}=\cos(y-\hat{\theta})-g^2P^2+S\frac{\sin(y-\hat{\theta})}{P},\ P(0)=\frac{1}{K_0}, \quad (32b)$$

$$\dot{S}=\sin(y-\hat{\theta})-3g^2PS,\ S(0)=0, \quad (32c)$$

where the signals u and y are the measurements available from the system (1), K_0 is a scalar given in the definition of the cost (2) and g is a scalar given from system (1). The filter (32) consists of three interconnected equations that together yield a third order minimum-energy estimate $\hat{\theta}$.

Remark 3: In [8] it was shown that by using a scalar coefficient g in the system model (1) the problem of minimizing (2) subject to the system (1) is equivalent to minimizing a more general cost (33) subject to a system without the coefficient g.

$$\begin{aligned}\tilde{J}(t;\theta_0,\delta|_{[0,\,t]},\varepsilon|_{[0,\,t]})=&\int_0^t\left(\frac{1}{2}Q\delta^2+R(1-\cos(\varepsilon))\right)d\tau\\&+\frac{1}{\tilde{K}_0}(1-\cos(\theta(0))),\end{aligned} \quad (33)$$

where Q, R and $\tilde{K}_0$ are positive gains that allow us to scale the energy of the unknowns ($\delta|_{[0,\,t]}$, $\varepsilon|_{[0,\,t]}$, θ_0) according to the a priori expectations that we might have on these unknown signals in a specific application.

Proposition 1: Assuming that the value function (9) is four times differentiable and that the Hessian $\nabla^2_{\hat{\theta}}V(\hat{\theta},t)$ is nonzero, S in (32c) approximates the true kinematics of the third order derivative of the value function $\nabla^3_{\hat{\theta}}V(\hat{\theta},t)$ to the order $O(\nabla^4_{\hat{\theta}}V(\hat{\theta},t))$.

Proof: In Section III, the matrix S was derived using the optimality conditions (6), (10) and (12) and the optimal filter equations (32a) and (32b). However, the fourth order derivative of the value function was neglected and the kinematics of S was approximated up to the third order derivatives of the value function. ■

Remark 4: Using the inverse of P denoted by $K := \frac{1}{P}$ can potentially improve the numerical implementation of the filter (32) by reducing the number of inverse operations needed. Applying this idea to (32) yields

$$\dot{\hat{\theta}} = u + K\sin(y-\hat{\theta}),\ \hat{\theta}(0) = 0, \tag{34a}$$

$$\dot{K} = -K^2\cos(y-\hat{\theta}) + g^2 - SK^3\sin(y-\hat{\theta}),\ K(0) = K_0, \tag{34b}$$

$$\dot{S} = \sin(y-\hat{\theta}) - 3g^2\frac{S}{K},\ S(0) = 0. \tag{34c}$$

Remark 5: Assuming that the third order derivative of the value function $S = \nabla^3_{\hat{\theta}} V(\hat{\theta}, t)$ is negligible yields the second order filter equations

$$\dot{\hat{\theta}} = u + K\sin(y-\hat{\theta}),\ \hat{\theta}(0) = 0, \tag{35a}$$

$$\dot{K} = -K^2\cos(y-\hat{\theta}) + g^2,\ K(0) = K_0. \tag{35b}$$

These equations coincide with the near-optimal filter recently derived by Coote *et al.* [8]. It was shown in that work that the distance to optimality of the second order filter (35) is bounded by a term that is fourth order in the estimation error $\theta - \hat{\theta}$. Also in [8] the second order filter (35) shows comparable results to an extended Kalman filter (EKF) with the advantage of needing less tuning. The current work extends these result by taking into account higher order terms and also by providing a systematic approach to obtaining higher order approximations.

IV. SIMULATIONS

In this section we provide simulation results for approximate filters up to eighth order derived using the concepts of Sections III.

Firstly, the system (1) is initialized to $\theta_0 = 15°$. The signal δ is a zero mean random variable with standard deviation of 12 degrees per time units and the signal ε is a zero mean random variable with standard deviation of 9 degrees. The sinusoidal input $u(t) = 2\sin(5t)$ and the constant $g = 1$ are utilized.

Figure 1 shows the tracking performance of the eighth-order filter in the presence of noise and a 15° initialization error. As can be seen the filter achieves tracking of the true state after a short transition period. Figure 2 shows the time evolution of the filter gains during this simulation. As can be seen, after a short transition period, the odd-order derivatives (gains) converge to zero and the even-order derivatives converge to small values. Note that the higher order derivatives converge to smaller values.

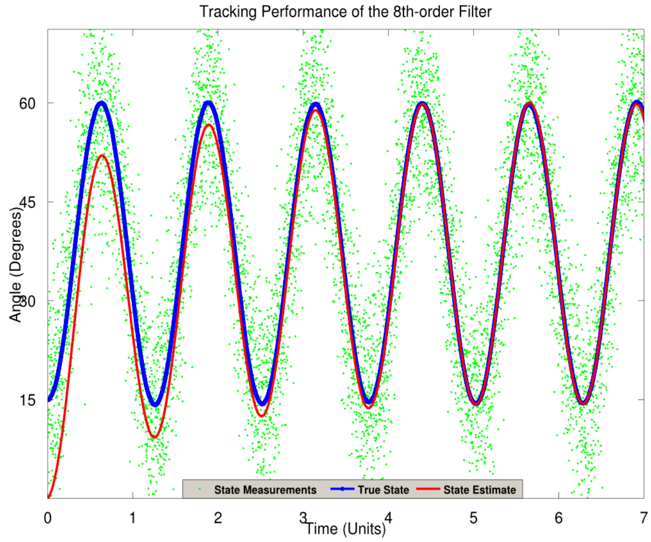

Fig. 1. Note that the figure is zoomed to the range $(0°, 75°)$. Some of the measurement data that have values less than zero degrees wrap around the circle and are hence not shown.

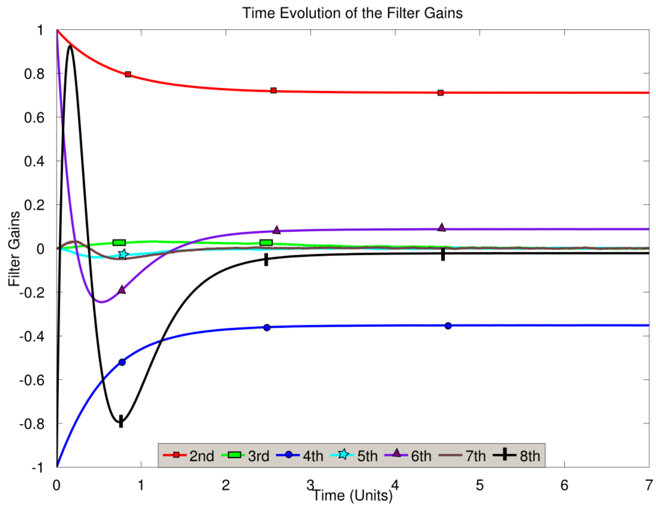

Fig. 2. Second to eighth order derivatives of the value function (9) are plotted against time. Note that the eighth order derivative is derived by neglecting the ninth order derivative.

Moreover, we have calculated Taylor series approximations of the value function (9) using the derivatives we computed. Note that the approximations are obtained at the simulation end time when the filter has converged. Also note that the expansion is around the final estimate $\hat{\theta}(t)$ centered at zero. Figure 3 shows approximations obtained using Taylor expansions of second up to seventh order. It is noted that approximations of the value function, after the overall convergence of the filter, resemble a (second order) parabola and that the higher order terms only seem to influence the global shape of the value function away from the optimal state value. Note that odd-order approximate value functions almost identically fall on the lower even-order approximates

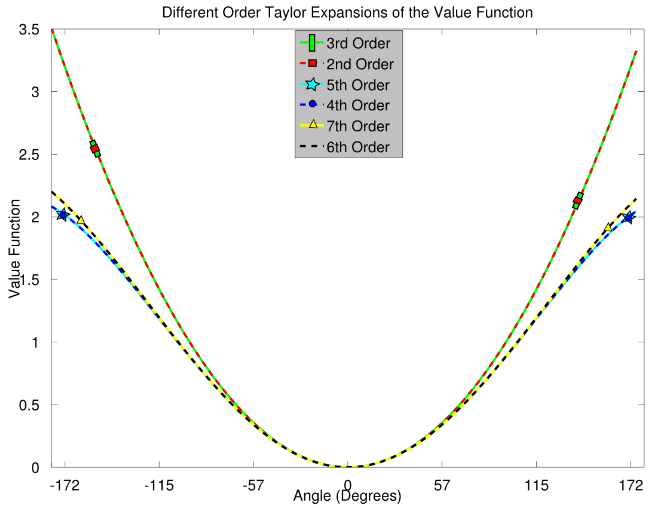

Fig. 3. Second to seventh order Taylor expansions of the value function (9) are plotted. Note that the expansion is around the final state estimate $\hat{\theta}(t)$ centered at zero.

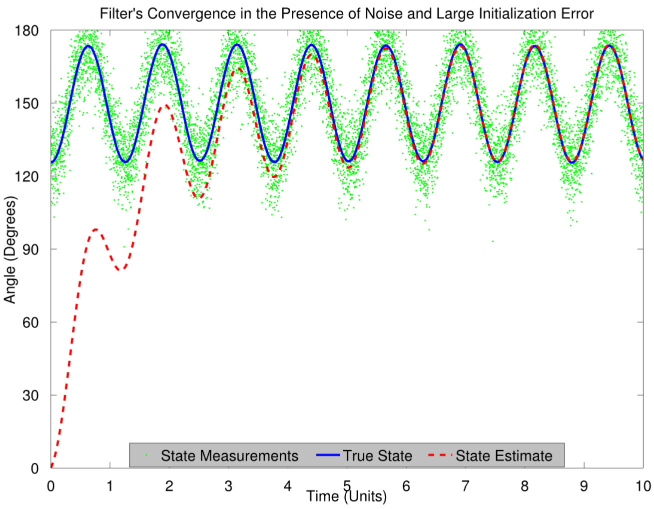

Fig. 4. The tracking performance of the eighth-order filter in the presence of a large initialization error.

of the value function. This is due to the odd-order derivatives converging to zero which was seen in Figure 2 before.

Finally, we include a simulation result in which a large initialization error (more than 120°) is considered. The same noise and input signals are used here. Figure 4 shows that despite the large initialization error the eighth-order filter converges to the true state trajectory after a short transition period. Note that the transition period is longer compared to the situation with 15° of initialization error.

V. CONCLUSIONS

We have applied Mortensen's minimum-energy filtering approach to a system defined on the Lie group S^1. We noted that this method yields a systematic program to derive higher order approximations of a minimum-energy (optimal) filter for a nonlinear system. We have shown that in general a nonlinear filter consists of an observer that includes a copy of the system plus a weighted innovation term. The observer's weighting proved to be the inverse Hessian of the value function of the associated optimal control problem. The dynamics of the Hessian is a Riccati equation perturbed by a third order derivative of the value function weighting the same innovation term. The kinematics of all higher order derivatives of the value function are first order ODEs that are in turn perturbed by a one step higher order derivative of the value function weighting the innovation term. The proposed filter on S^1 is third order in the sense that it approximates the kinematics of the third order derivative of the value function by neglecting the fourth order derivative of the value function. We provided simulation results showing the performance of an eighth-order filter derived using Mortensen's approach for the system on S^1. Simulation results indicate excellent convergence of the filter in the presence of large disturbances and initialization errors. Mortensen's approach to nonlinear optimal filtering is potentially extend-able to higher dimensional Lie groups using the general geometric theory of optimal control on Lie groups.

VI. 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. 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.

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

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

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

[5] A. Krener, "Minimum covariance, minimax and minimum energy linear estimators," *Stochastic Control Theory and Stochastic Differential systems*, pp. 490–495, 1979.

[6] ——, "The convergence of the minimum energy estimator," *New Trends in Nonlinear Dynamics and Control and their Applications*, pp. 187–208, 2004.

[7] A. P. Aguiar and J. P. Hespanha, "Minimum-energy state estimation for systems with perspective outputs and state constraints," in *Decision and Control, 2003. Proceedings. 42nd IEEE Conference on*, vol. 5. IEEE, 2003, pp. 4908–4913.

[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, to appear*.

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

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

[12] F. Bullo and A. D. Lewis, *Geometric control of mechanical systems*. Springer, 2004.

[13] M. Athans and P. Falb, *Optimal control: an introduction to the theory and its applications*. McGraw-Hill New York, 1966.