

# Near-Optimal Deterministic Filtering on the Unit Circle

Paul Coote, Jochen Trumpf, Robert Mahony, Jan C. Willems

**Abstract**—We present a near-optimal deterministic filter for systems that evolve on the unit circle. Unlike suboptimal filtering algorithms that rely on approximations of the system, the proposed approach preserves the non-linear system model. This leads to an explicit bound on the optimality gap in terms of the tracking error. Specifically, the optimality gap is bounded by a term that is fourth-order in the tracking error. A simulation demonstrates that the filter can track a signal on the unit circle in the presence of large disturbances. An optimal algorithm for recursive estimation of static (non-dynamic) data on the unit circle appears as a special case of the proposed filter.

## I. INTRODUCTION

There is a large body of work on filtering theory, including seminal works by Wiener [16], Kalman [9], Duncan [6], Mortensen [12] and Zakai [18] to name just a small selection. These works all address the question ‘What is a reasonable way to guess the current value of the system state?’. The classical approach is to model all unknowns — initial state, process noise and measurement noise — as random variables. Then the tools of probability theory and statistics may be used to derive the most likely value for the state vector at a given time. This *stochastic* approach has produced some very elegant theory and a wealth of success in technological applications, see e.g. [1], [7], [11] for examples and references. The paradigmatic example is the Kalman filter for linear systems that produces the maximum likelihood state vector as its own state and uses the same computational resources each time that its estimate is updated.

Unfortunately, the desirable characteristics of linear stochastic filters — primarily the property that the parametric space of Gaussian random variables is closed under transformations by linear system dynamics — does not generalise to non-linear systems. As a consequence, it is normally impossible to find a finite dimensional stochastic filter for a non-linear system: see e.g. the survey paper by Marcus [11] for some of the more subtle issues with non-linear stochastic filtering. Most applications are tackled using suboptimal algorithms, e.g. the extended Kalman filter [1], particle filter [5] or unscented filter [8], or using non-linear observers: see [13] for a relatively recent survey or [2], [4], [10] for specific applications to rigid body attitude estimation.

However, from the very beginnings of modern system theory the community has been aware of other, *non-*

*stochastic* approaches to state estimation. In particular, there exists a large body of knowledge concerning what we call ‘deterministic filtering’ (in this paper we are following the terminology used by Sontag [14] and Willems [17]). This approach has variously been named the ‘method of least squares’ [15], see also [1], or the ‘statistical method’ [7].

The deterministic approach to filtering treats measured data as *evidence* of what is really (i.e. deterministically) happening to the system. The disturbance functions are used only to resolve inconsistency concerning our dynamical model, the known input data and measurement data. No claim is made that disturbances have *a priori* known statistical properties. Rather, they are treated as unknown but deterministic signals. The idea is that we explain the observed data using a *hypothesis*, in the form of a process disturbance, measurement disturbance and initial state that reproduce the observed data. A “good” hypothesis fits the measured data (is unfalsified) and is maximally parsimonious. That is, the postulated disturbances ought to be as small as possible, while still generating the observed data. In fact, the Kalman filter can be interpreted as tacitly containing precisely such an explanation of observed data [7], [15], [14], [17].

In this paper, we investigate the use of the deterministic paradigm in the field of non-linear filtering. We present a derivation of a near-optimal filter for a deterministic non-linear system with a state evolving on the unit circle. Filtering on the unit circle is the principle behind the construction of phase-locked loops in radio receivers [1]. Moreover, this filtering task is the simplest possible case of two other, more general problems: filtering on spheres and filtering on non-linear unitary systems. For applications of systems on spheres see e.g. [3]. Our solution is near-optimal, in the sense that our filter utilises an explanation that has a small optimality gap. The optimality gap is bounded by a term that is fourth-order in the tracking error. This term is very small for typical examples, as is demonstrated in the simulation provided. Moreover, in the case of zero process noise (but unknown initial state and noisy measurements) our filter is optimal. This case is related to recursive least-squares estimation of static (non-dynamic) data, adapted to systems on the unit circle.

It is hoped that the ideas presented in this paper will lead to, for example, near-optimal filters for systems on the special orthogonal group  $SO(3)$  that describe the attitude of rigid bodies such as autonomous flying vehicles. Also, it is hoped that the case where our filter is optimal can be generalised to spheres and non-linear unitary systems — in particular to the problems of averaging on manifolds and

P. Coote, J. Trumpf and R. Mahony are with the College of Engineering and Computer Science, Australian National University, Canberra, ACT, Australia. `firstname.lastname@anu.edu.au`

J. C. Willems is with ESAT - SISTA Research Group, K.U. Leuven, Belgium. `Jan.Willems@esat.kuleuven.be`

recursive estimation on spheres.

This paper is organised as follows. Section II introduces the deterministic filtering problem for systems on the unit circle. A near-optimal solution to that problem is given in Section III. For the case where there is no process noise (the recursive estimation problem), the optimal filter appears as a special case of our main result — this is detailed in Section IV. Section V contains a simulations of the proposed filter, and compares the results to those obtained using an extended Kalman filter. Section VI concludes.

## II. PROBLEM FORMULATION

### A. System Model

Consider a system of the form:

$$\dot{\theta} = w + \delta', \quad y = \theta + \epsilon, \quad (1)$$

where  $\theta, y, \epsilon \in S^1$  and  $\delta', w \in \mathbb{R} \triangleq T_\theta S^1$ . That is, the system evolves on the unit circle - therefore the output  $y$  and the measurement “noise”  $\epsilon$  live on the unit circle  $S^1$ . The input  $w$  is the *rate* at which  $\theta$  rotates; it is defined on the tangent space  $T_\theta S^1$  to the unit circle at  $\theta$ , as is the plant “noise”  $\delta'$ .

In real world examples,  $\delta'$  is typically “noise” in a measure of angular velocity (analogous to gyrometer noise), while  $\epsilon$  is “noise” on the state measurement (analogous to accelerometer noise).

A key concept in our paper is that  $\delta'$  and  $\epsilon$  are not modeled as noise *processes* — they are considered to be unknown deterministic functions that are inputs to the system. Hence we will not use the term “noise” again in the derivation of the filtering algorithm. Rather, we will discuss unknown *disturbance functions*. In particular, we are concerned with the possible trajectories of these deterministic disturbances, given known dynamics (1) and known input and output data.

We aim to find a state estimate  $\hat{\theta}(T)$  using known input data  $w_d$  and observed output data  $y_d$  recorded up to time  $T$  in the face of unknown disturbances  $\delta'$  and  $\epsilon$  and an unknown initial state  $\theta(0)$ .

We restrict our attention to *recursive* filters because these are required in settings where computational resources are limited, e.g. applications in robotics and embedded control systems. In these settings a small optimality gap is generally more tolerable than an untractable algorithm.

### B. The Deterministic Filtering Principle

Following the deterministic filtering paradigm [17], our filter works in the following way. It is known that  $y_d$  and  $w_d$  were the signals that *actually* occurred, therefore the disturbances and initial condition that were actually realised must generate the known data from the known dynamics (1). We formalise this requirement using the notion of an *explanation*:

*Definition 1:* An initial condition and a set of disturbance functions specified on  $t \in [0, T]$  are called an *explanation* of known (system) data on  $t \in [0, T]$  if the disturbances, known data and initial condition jointly satisfy the system equations.

We only consider hypotheses that are explanations of the observed data. However, for any  $T \geq 0$  there are infinitely many specifications of the initial condition and disturbances that are valid explanations of the observed data from the system (1).

Filter design requires a principled way to choose *good* explanations above *bad* explanations, and hence to make a *good* estimate of the value of  $\theta(T)$ . We propose that the correct principle is *parsimony* - the data must be explained using disturbance signals that are small. That is, we seek a simple, reasonable, self-consistent account of what is taking place inside our plant; a *simple* explanation being one that does not postulate large unmodeled inputs (disturbances).

This filter design paradigm can be made precise in the following way. Consider all the possible explanations of the data. Our filtering principle is to choose the explanation that minimises the cost functional:

$$J'_T = \int_0^T |Q\delta'(\tau)^2 + R(1 - \cos(\epsilon(\tau)))|d\tau + \gamma'(1 - \cos(\theta(0))),$$

for positive  $R$ ,  $Q$  and  $\gamma'$ .

*Remark 1:* The positive value of  $R$ ,  $Q$  and  $\gamma'$  allow the definition of a *simple* explanation to be scaled by the relative uncertainty in each disturbance and in the initial condition.

Given an explanation  $(\delta'_h, \epsilon_h, \theta_h(0))$  that minimises  $J'_T$ , the best estimate of the current state is found by running the system (1) with input  $w_d$ ,  $\delta'_h$  and  $\epsilon_h$  and initial condition  $\theta_h(0)$  to time  $T$ . The state  $\theta_h(T)$  is then the best estimate that can be made from the data (based on the principle of parsimony, formalised by the explicit cost functional).

Finding the filter estimate  $\theta_h(T)$  may appear to be a computationally intractable task. However, it is demonstrated below that this apparently complicated optimisation procedure does not need to be re-executed each time the data is updated. In fact, a filter need not ever execute a large optimisation algorithm. Rather, the desired result —  $\theta_h(T)$  that is consistent with a good explanation — can be found recursively.

### C. Equivalent problem formulation

The filtering problem outlined above can be considerably simplified by consideration of the system

$$\dot{\theta} = w + g\delta, \quad y = \theta + \epsilon, \quad (2)$$

and the uncertainty measure

$$J_T = \int_0^T |\delta(\tau)^2 + (1 - \cos(\epsilon(\tau)))|d\tau + \gamma(1 - \cos(\theta(0))).$$

The assignments  $\gamma = \gamma'Q^{-1}$  and  $g = \sqrt{Q}(\sqrt{R})^{-1}$  (i.e.  $\delta' = \sqrt{Q}(\sqrt{R})^{-1}\delta$ ) yield system (1) and the uncertainty measure  $J_T = Q^{-1}J'_T$ . That is,  $\min(J_T) = \min(J'_T)$ .

In the following, we consider only *explanations* of observed data  $y_d$  and  $w_d$  from system (2) and aim to choose an explanation  $(\delta_h, \epsilon_h, \theta_h(0))$  that has a low uncertainty measure  $J_T$ .

### III. MAIN RESULT

The following Theorem 1 contains the main result of the paper. It states the existence of near-optimal explanations of data and presents the associated near-optimal deterministic filter. It also provides an upper bound on the optimality gap.

*Theorem 1:* For a given time  $T$  and real-time data  $y_d : [0, T] \rightarrow S^1$  and  $w_d : [0, T] \rightarrow \mathbb{R}$  from system (2), and provided that there is a solution to (4), there exist an initial condition  $\theta_h(0)$  and signals  $\delta_h : [0, T] \rightarrow \mathbb{R}$  and  $\epsilon_h : [0, T] \rightarrow S^1$  such that:

- 1) The initial condition  $\theta_h(0)$  and disturbance functions  $\delta_h$  and  $\epsilon_h$  explain the observed data  $w_d$  and  $y_d$ ,
- 2) The recursive filter

$$\begin{aligned} \dot{\hat{\theta}} &= w_d + K \sin(y_d - \hat{\theta}), \quad \hat{\theta}(0) = 0, \\ \dot{K} &= \frac{1}{2}g^2 - \cos(y_d - \hat{\theta})K^2, \quad K(0) = \frac{1}{\gamma}, \end{aligned} \quad (3)$$

yields a value  $\hat{\theta}(T)$  that is equal to the state  $\theta_h(T)$  of system (2) generated by initial condition  $\theta_h(0)$  and inputs  $(w_d, \delta_h, \epsilon_h)$ , and,

- 3) The uncertainty measure  $J_T(\theta_h(0), \delta_h, \epsilon_h)$  is near-optimal, in the sense that

$$0 \leq J_T(\theta_h(0), \delta_h, \epsilon_h) - \min(J_T) \leq E,$$

where

$$E = \int_0^T \left| \frac{g}{K(\tau)} \sin^2\left(\frac{1}{2}(\theta(\tau) - \hat{\theta}(\tau))\right) \right|^2 d\tau.$$

*Proof:* Define the tracking error  $\Delta = \theta - \hat{\theta}$  and the innovation  $\alpha = y_d - \hat{\theta}$ , and observe that  $\dot{\Delta} = w_d + g\delta - \dot{\hat{\theta}}$ . Consider the following function:

$$\mathcal{L} = \frac{1 - \cos(\Delta)}{K}. \quad (5)$$

The time-derivative of  $\mathcal{L}$  is given by

$$\dot{\mathcal{L}} = -\frac{\dot{K}(1 - \cos(\Delta))}{K^2} + \frac{\sin(\Delta)g\delta}{K} + \frac{\sin(\Delta)(w_d - \dot{\hat{\theta}})}{K}.$$

By completing the square, one finds that

$$\begin{aligned} \dot{\mathcal{L}} &= \delta^2 - \left| \frac{g \sin(\Delta)}{2K} - \delta \right|^2 + \frac{g^2 \sin^2(\Delta)}{4K^2} \\ &\quad - \frac{(\dot{\hat{\theta}} - w_d) \sin(\Delta)}{K} - \frac{\dot{K}(1 - \cos(\Delta))}{K^2}. \end{aligned}$$

Substituting the filter dynamics (3) and (4) yields:

$$\begin{aligned} \dot{\mathcal{L}} &= |\delta|^2 + |1 - \cos(\epsilon)| - \left| \frac{1}{2} \sin(\Delta) K^{-1} g - \delta \right|^2 \\ &\quad - |1 - \cos(\alpha)| + g^2 K^{-2} \left[ \frac{\sin^2(\Delta)}{4} + \frac{\cos(\Delta)}{2} - \frac{1}{2} \right]. \end{aligned}$$

The term in square brackets is equal to  $-\sin^4(\Delta/2)$ . One may integrate  $\mathcal{L}$  over time  $t \in [0, T]$  according to

$$\mathcal{L}(T) = \int_0^T \dot{\mathcal{L}} d\tau + \mathcal{L}(0),$$

and substitute the filter initial conditions (3) and (4). Observe that the first two terms in the above expression for  $\dot{\mathcal{L}}$  also appear in the definition of  $J_T$ . Comparing terms shows that

$$\begin{aligned} J_T &= K^{-1}|1 - \cos(\Delta(T))| + \int_0^T |1 - \cos(\alpha(\tau))| d\tau \\ &\quad + \int_0^T \left| \frac{1}{2} \sin(\Delta(\tau)) K^{-1} g - \delta(\tau) \right|^2 d\tau \\ &\quad + \int_0^T |gK^{-1}(\sin^2(\frac{\Delta(\tau)}{2}))|^2 d\tau. \end{aligned} \quad (6)$$

Observe that  $\alpha$  is driven *solely* by the given data  $y_d$  and  $w_d$  (via the deterministic system given in (3) and (4) that has these data as inputs). This gives a lower-bound for the uncertainty measure that no filter can surpass:

$$\min(J_T) \geq \int_0^T |1 - \cos(\alpha(\tau))| d\tau. \quad (7)$$

It is unclear whether any filter can achieve  $J_T$  this small — hence, we have not derived a value for  $\min(J_T)$ , only a lower-bound. Consider a signal  $\theta_h : [0 : T] \rightarrow S^1$  which is generated by

$$\dot{\theta}_h = w_d + \frac{g}{2K} \sin(\theta_h - \hat{\theta}), \quad (8)$$

and fixed by the final condition  $\theta_h(T) - \hat{\theta}(T) = 0$ , where  $\hat{\theta}$  and  $K$  were generated by the proposed filter. Running this differential equation backwards in time uniquely defines the signal  $\theta_h$ , and in particular there exists a  $\theta_h(0)$  which returns precisely the final condition we have prescribed. Also, we can define the two signals

$$\begin{aligned} \delta_h &= \frac{1}{2K} \sin(\theta_h - \hat{\theta}), \quad \text{and} \\ \epsilon_h &= y_d - \theta_h. \end{aligned} \quad (9)$$

Now, system (8) is obviously the same as the state equation in system (2) with inputs  $w = w_d$  and  $\delta = \delta_h$ . Observe further that (9) ensures that the inputs and initial condition under consideration do indeed generate  $y_d$ . This proves the first claim of the theorem. Furthermore, the final condition for (8) proves the second claim made in the theorem. Using (6), it is apparent that

$$\begin{aligned} J_T(\theta_h(0), \delta_h, \epsilon_h) &= \int_0^T (1 - \cos \alpha(\tau)) d\tau \\ &\quad + \int_0^T |gK^{-1}(\sin^2(\frac{\Delta(\tau)}{2}))|^2 d\tau, \end{aligned}$$

from which the third claim made in the theorem follows. ■

A few comments are in order:

*Remark 2:* Statement (3) of the above theorem expresses that the optimality gap  $E$  is bounded by a term that is fourth-order in the tracking error  $\Delta$ . Most suboptimal algorithms rely on approximations, and thus do not have explicit expressions for the optimality gap. The optimality gap  $E$  and lower-bound on the minimum cost (7) can be used to assess exactly how ‘near’ to optimal the filter’s performance comes; see Section V. Statement (2) tells us that the near-optimal state estimate can be found without explicitly considering

whole explanations; a filter that does not explicitly depend on  $(\theta_h, \delta_h, \epsilon_h)$  can be used. This ensures that the proposed filtering algorithm is simple to implement. Furthermore, the filter equations (3) and (4) do not depend on  $T$ . The filter can, therefore, be used recursively, despite different compatible explanations  $(\theta_h, \delta_h, \epsilon_h)$  for different values of  $T$ . Finally, Theorem 1 requires that the gain equation (4) has a solution on  $[0, T]$ . This equation is reliably stable in practice (see Section V and Figure 4). Furthermore, with some extra effort, the fourth-order optimality gap bound can be derived with the  $g$  in (4) replaced with an arbitrary constant. This constant can be chosen to guarantee the existence of a solution to the gain equation.

The following Corollary 1 shows the form of the near-optimal deterministic filter for the original system (1).

*Corollary 1:* For given real-time data  $y_d$  and  $w_d$  observed from system (1), and for the filtering principle outlined in Subsection II-B, a near optimal filter is given by

$$\begin{aligned} \dot{\hat{\theta}} &= w_d + K \sin(y_d - \hat{\theta}), & \hat{\theta}(0) &= 0, \\ \dot{K} &= \frac{Q}{2R} - K^2 \cos(y_d - \hat{\theta}), & K(0) &= \frac{Q}{\gamma}. \end{aligned} \quad (10)$$

The associated optimality gap is bounded by

$$E = \frac{Q}{R} \int_0^T \left| \frac{1}{K(\tau)} \sin^2\left(\frac{1}{2}(\theta(\tau) - \hat{\theta}(\tau))\right) \right|^2 d\tau.$$

*Proof:* The substitutions suggested in Subsection II-C can be applied to the filter equations (3) and (4) from Theorem 1. ■

#### IV. STATIC ESTIMATION

The non-linear filter is optimal in the case of zero input disturbance ( $\delta \equiv 0$  or  $\delta' \equiv 0$ ). The most common example of interest is the situation where the state is known to be constant ( $\dot{\theta} = 0$ ). We term this the static estimation case. We use this as our motivating example, however, the results are applicable in the more general case where there is a non-zero angular velocity  $w_d$  with no process noise.

Consider a series of disturbed measurements  $y$  of a static value on the unit circle,  $\theta$ :

$$y = \theta + \epsilon, \quad (11)$$

where  $\epsilon$  is an unknown function. For data  $y_d : [0, T] \rightarrow S^1$ , valid explanations are all  $\theta_h \in S^1$  and only those  $\epsilon_h : [0, T] \rightarrow S^1$  such that

$$y_d = \theta_h + \epsilon_h.$$

To choose good explanations over bad explanations at time  $T$ , nominate a cost functional on the initial condition and disturbance

$$J_T = \int_0^T |1 - \cos(\epsilon)| d\tau + \gamma |1 - \cos(\theta)|. \quad (12)$$

Clearly, this is a special case of Theorem 1, with  $g = 0$ . This means that the optimality gap is zero. The optimal filter is given by

$$\begin{aligned} \dot{\hat{\theta}} &= K \sin(y_d - \hat{\theta}); & \hat{\theta}(0) &= 0; \\ \dot{K} &= -K^2 \cos(y_d - \hat{\theta}); & K(0) &= \gamma^{-1}. \end{aligned} \quad (13)$$

Obviously, in the case of non-static  $\theta$  with known input data and no process disturbance ( $\delta \equiv 0$ ), one simply adds ‘+ $w_d$ ’ to the state-estimate dynamics, (13).

## V. SIMULATION

### A. System and Disturbance Setup

In order to demonstrate the properties of the filter proposed in Section III, we include a simulation of a noisy system on the unit circle. The additive noise is Gaussian — however this is merely because this is a straightforward way to generate a *disturbance function* on time  $[0, T]$ . The system simulated is of the form of (1), with noises

$$\delta' \sim \mathcal{N}\left(0, \frac{2}{5}\right); \quad \epsilon \sim \mathcal{N}\left(0, \frac{2\pi}{5}\right).$$

Note that  $\epsilon$  is defined on the unit circle, so it is technically a Von Mises distribution. A convenient intuition is that the Gaussian distribution “wraps around” the unit circle; samples from the Gaussian that have magnitude  $|\epsilon(t)| > \pi$  are shifted by  $\pm 2\pi$  in order to preserve the requirement that  $\epsilon(t) \in [-\pi, \pi]$ .

The high uncertainty in  $\epsilon$  and relatively low uncertainty in  $\delta'$  means that the constant  $g$  is quite small — that is, the optimality gap  $E$  for the non-linear filter is scaled by a small constant. This is a situation in which the proposed non-linear filter is close to optimal.

A relatively high uncertainty in the initial condition was chosen, in order to demonstrate that the proposed filter rapidly converges to a small tracking error:

$$\theta_0 \sim \mathcal{N}\left(0, \frac{\pi}{2}\right).$$

As for  $\epsilon$ , this distribution is constrained to the unit circle.

Observe that the large uncertainty in  $\epsilon$  means that measurement data  $y_d$  will occasionally appear around  $\pi$  radians above or below the true state  $\theta$ . When algorithms that apply a linear filter (and continually linearise the system around the current state estimate) see a value  $\pi$  radians away from the current estimate, they tend to take quite drastic action. In particular, the extended Kalman filter moves further than it would if it received a datum that was only e.g.  $\frac{\pi}{2}$  radians above or below the state estimate. That seems wrong in the non-linear case. Rather, there is no way to know for sure (on the unit circle) if that datum is  $\pi$  radians *above* or *below* the current estimate. Note that the proposed non-linear filter treats large innovations  $y_d(t) - \hat{\theta}(t) \approx \pi$  with scepticism; the filter corrects for the *sine* of the innovation. Occasional measurements that are  $\pi$  radians away do not cause drastic reevaluations of the state estimate.

The non-linear filter was simulated over 250 seconds, in discrete time with a time step of 0.05 seconds. The known input  $w_d$  was the function  $w_d(t) = 2 \sin(t) + \cos(t/2)$ .

### B. Results

The trajectory taken by the system and non-linear filter is plotted in Figure 1 (upper). The filter clearly achieves good tracking.

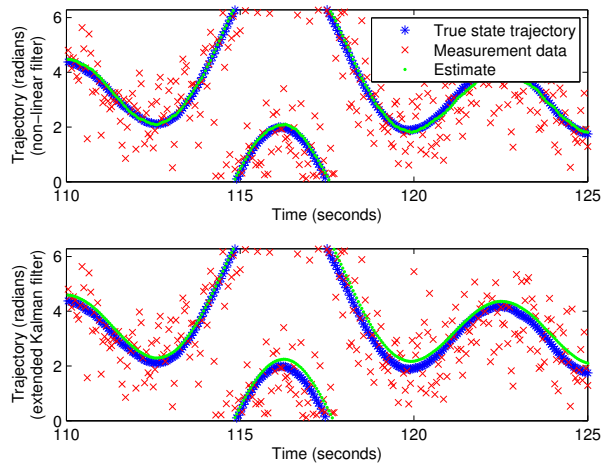


Fig. 1. The trajectory taken by the proposed non-linear filter (upper) and the extended Kalman filter (lower). Only 15 seconds are shown, for visual clarity. Clearly, both filters achieve reasonable performance.

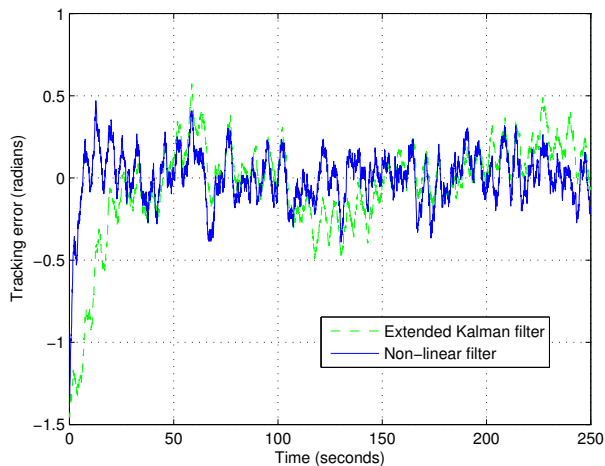


Fig. 2. Comparison of the tracking error  $\Delta$  for the proposed non-linear filter and an extended Kalman filter.

The tracking error  $\Delta(t)$  rapidly converges to approximately zero. Figure 2 shows a typical trajectory of  $\Delta$ , for a run where  $\Delta(0) \approx -1.46$ . Figure 3 (upper plot) is the corresponding histogram of the tracking error (the transient first 25 seconds of  $\Delta$  is excluded from the histogram). This leads to  $\sin^4(\frac{\Delta}{2})$  being steadily on the order of  $10^{-3}$ .

The filter gain  $K$  converges to approximately 0.5 and stays quite steady. The trajectory of  $K$  is shown in Figure 4. The fact that  $K$ ,  $Q$  and  $R$  (recall that the latter two are the uncertainties in the disturbance functions) are on the order of 1 and  $\sin^4(\frac{\Delta}{2})$  is on the order of  $10^{-3}$  yields a very small optimality gap  $E$ . Also, the large measurement noise ensures that the lower bound on the cost  $J_T$  (see (7)) is relatively high.

In fact, the proportion  $E/(\min(J_T) + E)$ , evaluated for each time  $T$ , begins at about  $10^{-1}$  and soon falls to around

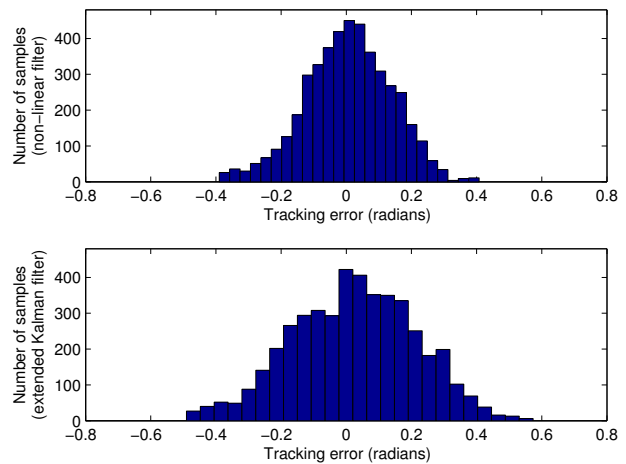


Fig. 3. Histograms of the tracking error  $\Delta$  for the proposed non-linear filter (upper) and an extended Kalman filter (lower). The first 25 seconds of tracking were disregarded because they are transient.

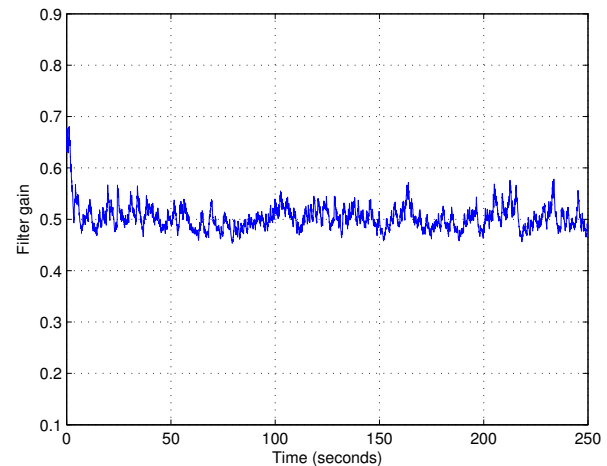


Fig. 4. The trajectory of the gain  $K$  of the non-linear filter. The gain quickly converges to the region  $K \approx 0.5$ , however it is perturbed by measurement noise (as can be seen from (10)).

$3.6 \times 10^{-4}$ , as shown in Figure 5. This means that for long times the cost attributed to sub-optimality is negligible compared to the minimum cost that *every* possible filter incurs. The claim that the filter is very near to optimal is well founded.

### C. Comparison with Extended Kalman Filter

Tuning extended Kalman filters is notoriously arcane, and we make no claim to have used every available trick and technique to tease out the best possible performance. However, we have used standard methodology and the disturbance signals have known distributions — the parameters of these distributions were used in the design of the Kalman filter.

Note that measurements  $y_d$  are taken on the unit circle, meaning that the filter has no choice but to interpret

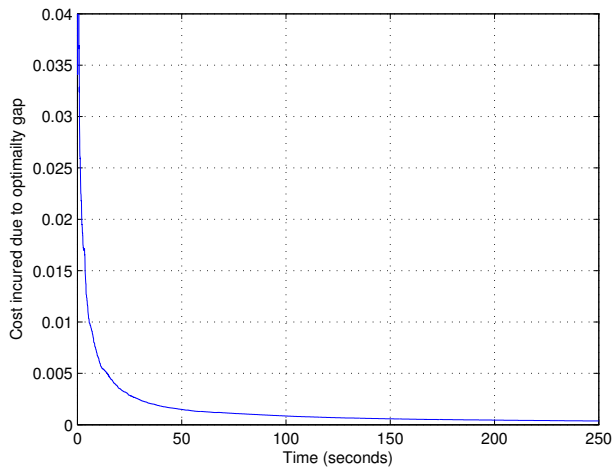


Fig. 5. The proportion of total cost incurred by each time that is due to the optimality *gap*. This is found by evaluating  $E/(\min(J_T) + E)$  (the lower bound is used for the minimum cost, see (7)). By time  $t = 250$ , this fraction is approximately  $3.6 \times 10^{-4}$ .

innovations  $0 < y_d - \hat{\theta} < \pi$  as positive and innovations  $-\pi < y_d - \hat{\theta} < 0$  as negative. This is despite the fact that the measurement noise will have some values  $|\epsilon(t)| > \pi$ . That is, the Gaussian *wraps around* the unit circle to form a Von Mises distribution, and there is no way to tell from the data which samples should be treated as having magnitude greater than  $\pi$ , and which have magnitude smaller than  $\pi$  in the opposite direction.

The trajectory taken by the system and by the extended Kalman filter is given in Figure 1 (lower). Clearly the filter achieves state tracking in this case. However, inspection of Figure 3 reveals that the tracking error tends to be a little larger than it is for the non-linear filter. The standard deviation in the tracking error is 0.133 radians for the non-linear filter; 0.189 radians for the extended Kalman filter (the transient time at the beginning of the simulation was excluded from this calculation).

Also, Figure 2 demonstrates that the transient response is not as fast as for the non-linear filter. This can be attributed to the high initial tracking error  $\Delta(0) \approx -1.46$  coupled with the high measurement error; in the first few seconds the filter is seeing lots of samples that are approximately  $\pi$  radians away from its current state estimate — some data are telling it to move one way, some are telling it to move the other way. This means that it takes extra time to settle into accurate tracking.

## VI. CONCLUSIONS AND FUTURE WORKS

We have presented a deterministic, near-optimal filter for systems that evolve on the unit circle. An optimal filter for recursive estimation of static (non-dynamic) data on the unit circle appears as a special case of our result. A simulation in the case of high measurement noise has demonstrated that the non-linear filter does not diverge and is very close to optimal. Future work includes developing a better theoretical

understanding of the optimality gap, and deriving analogous filters for general unitary and spherical systems.

## VII. ACKNOWLEDGMENTS

This work was supported by the Australian Research Council through the ARC Discovery Project DP0987411 "State Observers for Control Systems with Symmetry".

The SISTA research program of the K.U. Leuven is supported by the Research Council KUL: GOA AMBioRICS, CoE EF/05/006 Optimization in Engineering (OPTEC), IOF-SCORES4CHEM; by the Flemish Government: FWO: projects G.0452.04 (new quantum algorithms), G.0499.04 (Statistics), G.0211.05 (Nonlinear), G.0226.06 (cooperative systems and optimization), G.0321.06 (Tensors), G.0302.07 (SVM/Kernel), G. 0320.08 (convex MPC), G.0558.08 (Robust MHE), G.0557.08 (Glycemia2), G.0588.09 (Brain-machine) research communities (ICCoS, ANMMM, MLDM); G.0377.09 (Mechatronics MPC) and by IWT: McKnow-E, Eureka-Flite+, SBO LeCoPro, SBO Climaqs; by the Belgian Federal Science Policy Office: IUAP P6/04 (DYSCO, Dynamical systems, control and optimization, 2007-2011); by the EU: ERNSI; FP7-HD-MPC (INFSO-ICT-223854); and by several contact research projects.

## REFERENCES

- [1] B. Anderson and J. Moore, *Optimal Filtering*, Prentice Hall, Englewood Cliffs, NJ, 1979.
- [2] S. Bonnabel, P. Martin and P. Rouchon, Symmetry-Preserving Observers, *IEEE Transactions on Automatic Control*, Vol. 53, Issue 11, 2008, pp 2514-2526.
- [3] R. W. Brockett, Lie Theory and Control Systems Defined on Spheres, *SIAM J. Appl. Math.*, Vol. 25, No. 2, 1973, pp 213-225.
- [4] J. L. Crassidis, F. L. Markley, and Y. Cheng, Nonlinear attitude filtering methods, *Journal of Guidance, Control, and Dynamics*, Vol. 30, No. 1, 2007, pp 12-28.
- [5] A. Doucet, N. de Freitas and N. Gordon, eds. *Sequential Monte Carlo methods in practice*, Springer, New York, 2001.
- [6] T.E. Duncan, *Probability densities for diffusion processes with applications to nonlinear filtering theory and diffusion theory*, PhD thesis, Stanford University, 1967.
- [7] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press Inc., New York, 1970.
- [8] S.J. Julier and J.K. Uhlmann, Reduced sigma point filters for the propagation of means and covariances through nonlinear transformations, *Proceedings of the American Control Conference*, Vol. 2, 2002, pp 887-892.
- [9] R. E. Kalman, A New Approach to Linear Filtering and Prediction Problems, *Transactions of the ASME-Journal of Basic Engineering*, Vol. 82, Series D, 1960, pp 35-45.
- [10] R. Mahony, T. Hamel and J.M. Pflimlin, Nonlinear Complementary Filters on the Special Orthogonal Group, *IEEE Transactions on Automatic Control*, Vol. 53, Issue 5, 2008, pp 1203-1218.
- [11] S. Marcus, Algebraic and Geometric Methods in Nonlinear Filtering, *SIAM J. Control and Optimization*, Vol. 22, No. 6, 1984, pp 817-844.
- [12] R.E. Mortensen, *Optimal control of continuous-time stochastic systems*, PhD thesis, UC Berkeley, 1966.
- [13] H. Nijmeijer and T. I. Fossen, *New Directions in Nonlinear Observer Design*, Springer, New York, 1999.
- [14] E. D. Sontag, *Mathematical Control Theory: Deterministic Finite Dimensional Systems*, Second Edition, Springer, New York, 1998.
- [15] P. Swerling, Modern state estimation methods from the viewpoint of the method of least squares, *IEEE Transactions on Automatic Control*, Vol. 16, No. 6, 1971, pp 707-719.
- [16] N. Wiener, The Linear Filter for a Single Time Series, Chapter III from *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, The M.I.T. Press, Cambridge, MA, 1949, pp 81-103.
- [17] J.C. Willems, Deterministic Least Squares Filtering, *Journal of Econometrics*, Vol. 118, 2004, pp 341-373.
- [18] M. Zakai, On the optimal filtering of diffusion processes, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, Vol. 11, 1969, pp 230-243.