

Cubature Filters:

New Generation of Nonlinear Filters that will Impact the Literature

Simon Haykin
McMaster, University
Hamilton, Ontario, Canada

email: haykin@mcmaster.ca
Web site: <http://soma.mcmaster.ca>



Outline of The Lecture

- 1. Introductory Remarks**
- 2. The Bayesian Filter**
- 3. Two Approaches for Approximating the Bayesian Filter**
- 4. The Cubature Kalman Filter**
- 5. Unique Properties of the Cubature Kalman Filter**
- 6. Cubature Filtering for Continuous-Discrete nonlinear Systems**
- 7. Applications of Cubature Filters**
- 8. Example I: Tracking a Manoeuvring ship**
- 9. Example II: Training a Recurrent Neural Network**
- 10. Concluding Remarks**

1. Introductory Remarks

“Optimality versus Robustness”

In many algorithmic applications, global optimality may not be practically feasible:

- (i) Naturally infeasible computability
- (ii) Curse-of-dimensionality due to large-scale nature of the problem at hand

Hence, the practical requirement of having to settle for a **sub-optimal solution** of the system design, which involves:

Trade-off between two conflicting design objectives:

- global optimality
- computational tractability and robust behaviour

Criterion for sub-optimality

DO AS BEST AS YOU CAN, AND NOT MORE

- **This statement is the essence of what the human brain does on a daily basis:**

Provide the “best” solution in the most reliable fashion for the task at hand, given limited resources.

- **Key question: How do we define “best”?**

Naturally, the answer to this question is problem-dependent.

2. The Bayesian Filter (Ho and Lee, 1964)

Problem statement:

Given a nonlinear dynamic system, estimate the *hidden state* of the system in a **recursive** manner by processing a sequence of *noisy observations* dependent on the state.

- The Bayesian filter provides a **unifying framework** for the optimal solution of this problem, at least in conceptual sense.

Unfortunately, except in a few special cases, the Bayesian filter is **not** implementable in practice -- hence, the need for some form of **approximation**.

State-space Model of a Nonlinear Dynamic System:

1. **Process (state) sub-model defined by the nonlinear equation:**

$$\mathbf{x}_{t+1} = \mathbf{a}(\mathbf{x}_t) + \omega_t \quad (1)$$

2. **Measurement sub-model defined by another nonlinear equation:**

$$\mathbf{y}_t = \mathbf{b}(\mathbf{x}_t) + \mathbf{v}_t \quad (2)$$

where t = discrete time

\mathbf{x}_t = hidden state of the system at time t

\mathbf{y}_t = observation at time t

ω_t = dynamic noise

\mathbf{v}_t = measurement noise

Prior Assumptions:

- **Nonlinear functions $a(\cdot)$ and $b(\cdot)$ are known**
- **Dynamic noise ω_t and measurement noise v_t are statistically independent Gaussian processes of zero mean and known covariance matrices.**

Two fundamental Update Equations

1. The time-update equation:

$$\underbrace{p(\mathbf{x}_t | \mathbf{Y}_{t-1})}_{\text{Predictive distribution}} = \int_{R^n} \underbrace{p(\mathbf{x}_t | \mathbf{x}_{t-1})}_{\text{Prior distribution}} \underbrace{p(\mathbf{x}_{t-1} | \mathbf{Y}_{t-1})}_{\text{Old posterior distribution}} d\mathbf{x}_{t-1} \quad (3)$$

where R^n denotes the n -dimensional state space.

The \mathbf{Y}_{t-1} denotes the past history of the observations:

$$\mathbf{Y}_{t-1} = \{\mathbf{y}_{t-1}, \dots, \mathbf{y}_1\}$$

2. The measurement-update equation:

$$\underbrace{p(\mathbf{x}_t | \mathbf{Y}_t)}_{\text{Updated posterior distribution}} = \frac{1}{Z_t} \underbrace{p(\mathbf{x}_t | \mathbf{Y}_{t-1})}_{\text{Predictive distribution}} \underbrace{l(\mathbf{y}_t | \mathbf{x}_t)}_{\text{Likelihood function}} \quad (4)$$

where Z_t is the normalizing constant defined by

$$Z_t = \int_{R^n} p(\mathbf{x}_t | \mathbf{Y}_{t-1}) l(\mathbf{y}_t | \mathbf{x}_t) d\mathbf{x}_t$$

Special Cases of the Bayesian Filter

- The celebrated **Kalman filter** is a special case of the Bayesian filter, assuming that the dynamic system is linear and both the dynamic noise and measurement noise are statistically independent Gaussian processes.
- Except for this special case and couple of other cases, exact computation of the posterior distribution $p(\mathbf{x}_t | \mathbf{Y}_t)$, defined in Eq. (3) is **not** feasible; the same remark applies to the normalizing constant.
- We therefore have to abandon optimality and be content with a **sub-optimal nonlinear filtering algorithm** that is computationally tractable.

3. Two Approaches for Approximating the Bayesian Filter

1. Direct numerical approximation of the posterior in a local sense:

- **Extended Kalman filter (simple and therefore widely used) when the nonlinearity is of a mild sort**
- **Unscented Kalman filter (heuristic in its formulation)**
- **Central-difference Kalman filter**
- **Cubature Kalman filter (New)**

2. Indirect numerical approximation of the posterior in a global sense:

Particle filters, whose

- **roots are embedded in Monte Carlo simulation; and**
- **typically, they are computationally demanding**

4. The Cubature Kalman Filter

- At the heart of the Bayesian filter, we have to compute moment integrals whose integrands are expressed in the common form

(Nonlinear function) \times (Gaussian function)

- The challenge is to numerically approximate the integral so as to completely **preserve second-order information about the state \mathbf{x}_t that is contained in the sequence of observations denoted by \mathbf{Y}_t**
- The computational tool that accommodates this requirement is the ***cubature rule***.

The Cubature Rule

- In mathematical terms, we have to compute integrals of the generic form

$$h(\mathbf{f}) = \int_{R^n} \underbrace{\mathbf{f}(\mathbf{x})}_{\text{Arbitrary nonlinear function}} \underbrace{\exp\left(-\frac{1}{2}\mathbf{x}^T \mathbf{x}\right)}_{\text{Normalized Gaussian function of zero mean and unit covariance matrix}} d\mathbf{x} \quad (5)$$

- To do the computation, a key step is to make a change of variables from the Cartesian coordinate system (in which the vector \mathbf{x} is defined) to a **spherical-radial coordinate system**:

$$\mathbf{x} = r\mathbf{z} \text{ subject to } \mathbf{z}^T \mathbf{z} = 1 \text{ and } \mathbf{x}^T \mathbf{x} = r^2 \text{ where } 0 \leq r < \infty$$

- The next step is to apply the **radial rule** using the **Gaussian quadrature**.

Reference I:

For mathematical derivation of the Cubature Kalman filter, refer to the paper:

I. Arasartnam and S. Haykin
“Cubature Kalman Filters”
IEEE Transactions on Automatic Control,
vol. 54, June 2009

5. Unique Properties of the Cubature Kalman Filter

Property 1: The cubature Kalman filter (CKF) is a *derivative-free on-line sequential-state estimator*; unlike other nonlinear filters, it relies on integration for its operation.

Property 2: Approximations of the four moment integrals involved in the derivation are all *linear* in the number of function evaluations.

Property 3: **Computational complexity** of the cubature Kalman filter grows as n^3 , where n is the dimensionality of the state space.

Property 4: The cubature Kalman filter *completely preserves second-order information about the state* that is contained in the observations.

Property 5: Regularization is built into the constitution of the cubature Kalman filter by virtue of the fact that the **prior** is known to play a role equivalent to regularization.

Property 6: The cubature Kalman filter *inherits properties of the **linear Kalman filter***, including square-root filtering for improved accuracy and reliability.

Property 7: The cubature Kalman filter is the **closest known direct approximation to the Bayesian filter**, outperforming all other nonlinear filters in a Gaussian environment:

**It eases the curse-of-dimensionality problem
but, by itself, does not overcome it.**

6. Cubature Filtering for Continuous-Discrete Nonlinear Systems

This second class of nonlinear filters deals with state-estimation problems whose state-space models are **naturally hybridized**:

(i) Continuous-time state (process) sub-model.

(ii) Discrete-time measurement sub-model.

In mathematical terms, the measurement sub-model is described in the same way as in Eq. (2), but the state sub-model is different.

State (process) Sub-model

To describe the process sub-model, we have to resort to stochastic differential-equation theory, exemplified by the **Ito equation**:

$$\frac{d\mathbf{x}_t}{dt} = \mathbf{a}(\mathbf{x}, t) + \sqrt{\mathbf{Q}}\mathbf{w}_t \quad (6)$$

where \mathbf{x}_t is the unknown state at time t ,

$\mathbf{a}(\mathbf{x}, t)$ is some unknown function

\mathbf{Q} is the so-called **diffusion matrix**

\mathbf{w}_t is the standard Gaussian noise

Discretization of the Process Equation

With recursive digital computation in mind, the process sub-model would have to be **discretized in the time domain**.

The discretization can be performed using:

- (i) Numerical methods, exemplified by the **Euler method** or **Runge-Kutta method**.
- (ii) The **Ito-Taylor expansion** of order 1.5.

Cost-Reduced Square-root CKF

Having performed the discretization, we find that by proceeding in the same way as before, the resulting square-root cubature Kalman filter (SCKF) is computationally expensive.

To overcome this practical difficulty, we have developed a new nonlinear filter, named the **cost-reduced SCKF** whose distinguishing feature is summarized as follows:

The modified time-update propagates a set of cubature points without having to estimate the predicted mean and covariance at every time-step of the filtering computation.

Reference II

For mathematical derivation of the cost-reduced square-root cubature Kalman filter, refer to

**I. Arasaratnam, S. Haykin, and T. Hurd
“Cubature Filtering for Continuous-Discrete Nonlinear Systems:
Theory with an Application to Tracking”**

to be submitted for publication.

7. Practical Applications

(i) Aerospace Applications

Tracking of aircraft, satellites and guided missiles.

(ii) Training of Recurrent Neural Networks

Simply stated:

The Cubature Kalman filter and its continuous-discrete extension provide new signal-processing tools for estimating the hidden state of a nonlinear dynamic system whose nonlinearity is too difficult for the traditional use of extended Kalman filters.

8. Example I: Tracking a Manoeuvring Ship

Problem statement:

Track a ship moving in an area bounded by a shore line, assumed to be a circular disc of known radius and centered at the origin.

- **The ship's motion is modelled by a constant velocity perturbed by additive white Gaussian noise.**
- **When the ship tries to drift outside the shoreline, a gentle turning force pushes it back towards the origin.**
- **The model is interesting in that it exhibits a nonlinear behavior near the shoreline, thereby providing a good test for assessing the performance of different nonlinear filters.**

- **Dynamic State-space Model (Kushner and Budhiraja, 2000)**

$$\dot{\mathbf{x}}_t = [\dot{\xi}_t \dot{\eta}_t f_1(\mathbf{x}_t) f_2(\mathbf{x}_t)]^T + \sqrt{\mathbf{Q}_t} \beta_t \quad (7)$$

$$\begin{pmatrix} r_k \\ \theta_k \end{pmatrix} = \begin{pmatrix} \sqrt{\xi_k^2 + \eta_k^2} \\ \tan^{-1} \left(\frac{\eta_k}{\xi_k} \right) \end{pmatrix} + \mathbf{w}_k \quad (8)$$

- **where**

$$f_1(\mathbf{x}) = \begin{cases} \frac{-K\xi}{\sqrt{\xi^2 + \eta^2}}, & \sqrt{\xi^2 + \eta^2} \geq r \text{ and } \xi\dot{\xi} + \eta\dot{\eta} \geq 0; \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

$$f_2(\mathbf{x}) = \begin{cases} \frac{-K\eta}{\sqrt{\xi^2 + \eta^2}}, & \sqrt{\xi^2 + \eta^2} \geq r \text{ and } \xi\dot{\xi} + \eta\dot{\eta} \geq 0; \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

Tracking Example (continued)

- Use the Euler method with 5 steps for each measurement interval to numerically integrate Eq. (7)
- Data:
 - Radius of the disk-shape shore, $r = 5$ units
 - Gaussian process noise intensity, $Q = 0.01$
 - Gaussian measurement noise parameters, $\sigma_r = 0.01$ and $\sigma_\theta = \frac{0.5\pi}{180}$
 - Estimated initial state, $\hat{\mathbf{x}}_{0|0} = [1, 1, 1, 1]^T$ and covariance, $P_{0|0} = 10\mathbf{I}_4$ where \mathbf{I}_4 is four-dimensional identity matrix.
 - Radar scans = 1000/Monte Carlo run
 - 50 independent Monte Carlo runs

Motion of the ship.

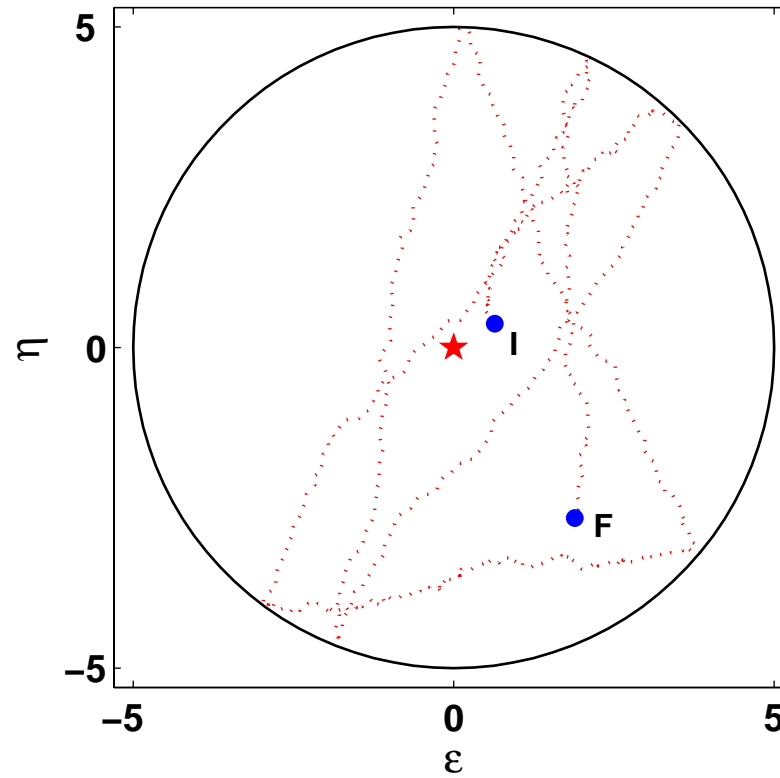
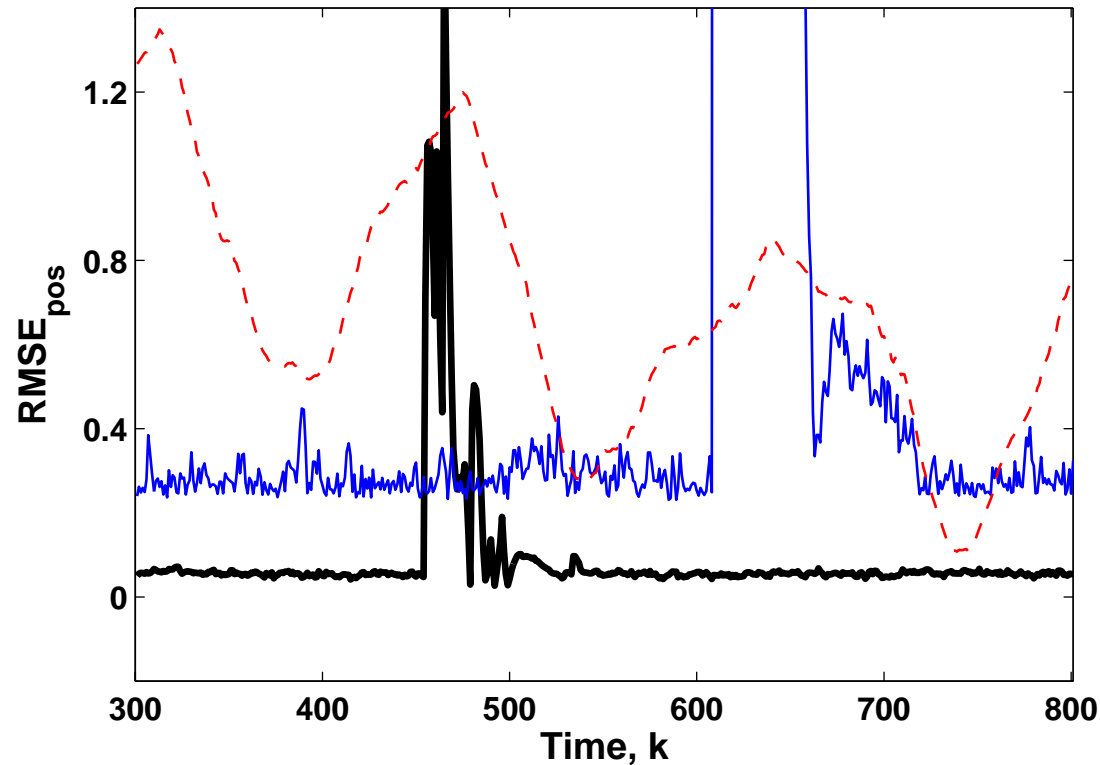


Figure 1: I - initial point, F - final point, ★ - Radar location

Performance Comparison: RMSE in position



**Figure 2: dashed red-Particle filter (PF) (1000 particles),
thin blue-Central-difference Kalman filter (CDKF)
dark black-Cubature Kalman filter (CKF)**

Performance Comparison” RMSE in velocity

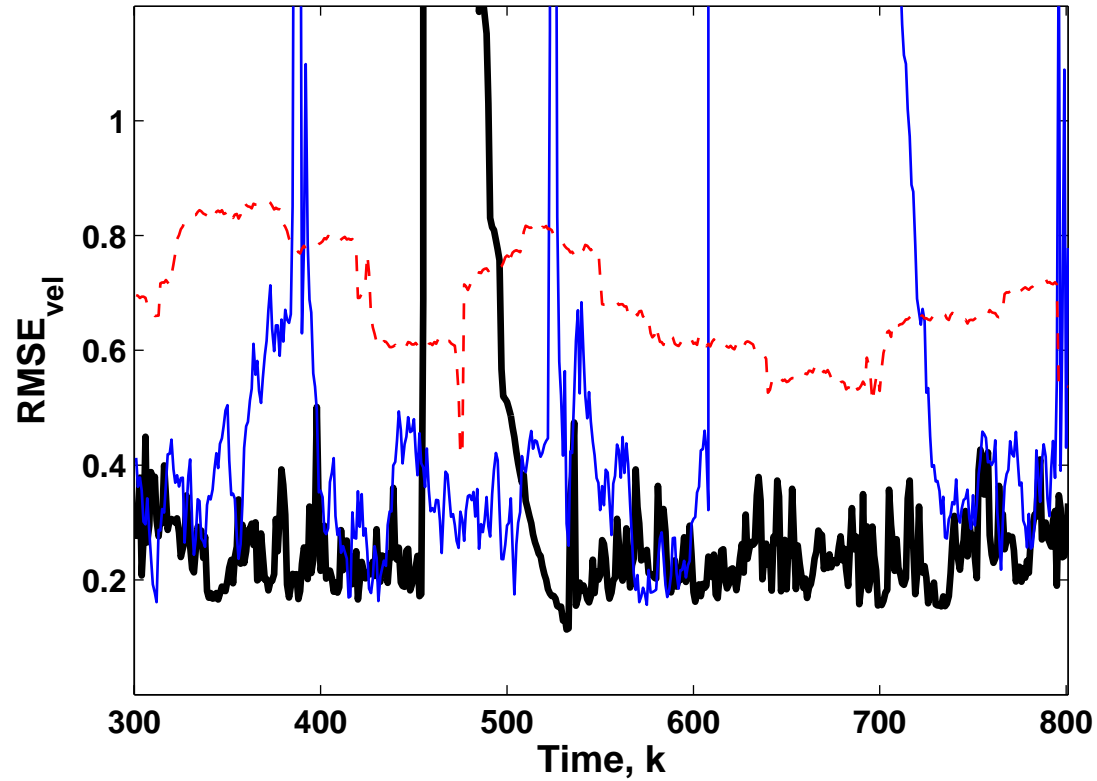


Figure 3: dashed red- PF (1000 particles), thin blue- CDKF, dark black- CKF

9. Example II: Training a Recurrent Neural Network

Mackey-Glass Attractor

$$\frac{d}{dt}x_t = bx_t + \frac{ax_{t-\Delta t}}{1 + x_{t-\Delta t}} \quad (11)$$

where t denotes continuous time,

$$a = 0.2$$

$$b = 0.1$$

$$\Delta t = 30$$

The Mackey-Glass attractor has an infinite number of degrees of freedom because we require knowledge of the initial value of x_t across a continuous-time interval.

Yet, it behaves like a strange attractor with a finite dimension.

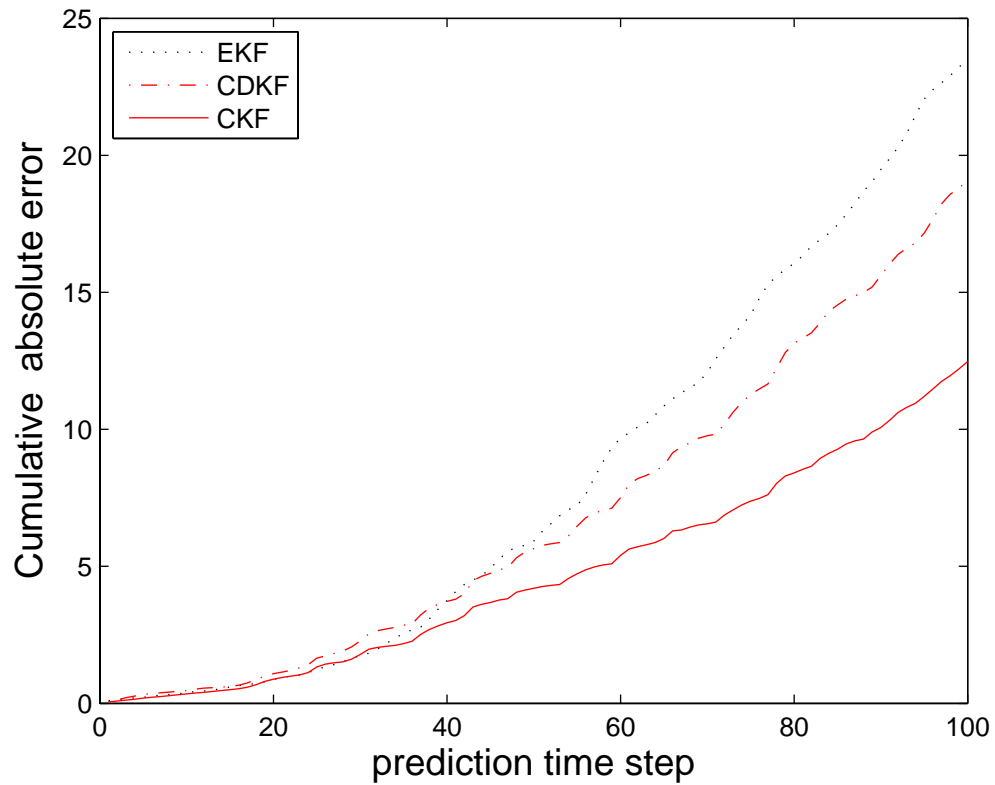


Figure 4: Ensemble-averaged cumulative absolute error curves during the autonomous prediction phase of dynamic reconstruction of the Mackey-Glass attractor

10. Concluding Remarks

The classical Kalman filter (1960) and Kalman-Bucy filter (1961) provide optimal estimates of the hidden state of linear discrete-time and continuous-time systems in Gaussian environments, respectively, being optimal in the maximum a posteriori (MAP) sense.

The two new nonlinear filters, the Cubature Kalman filter (and its square-root version) described in Reference I and the cost-reduced square-root cubature Kalman filter described in Reference II, provide optimal estimates of the hidden state of nonlinear discrete and continuous-discrete (hybrid) systems, respectively, being optimal in the sense that they are the closest direct approximations to the Bayesian filter in Gaussian environments.

**The slides for this Webiner
are available
on my website**

<http://soma.mcmaster.ca>