Cognitive Dynamic Systems

Simon Haykin
McMaster University
Hamilton, Ontario, Canada
haykin@mcmaster.ca
http://soma.mcmaster.ca

*****

1. **What is Cognition?**
2. **Cognitive Dynamic Systems Defined**
3. **Emerging Applications**
4. **Global Feedback**
5. **Why sub-optimality should be the objective of cognitive dynamic systems?**
6. **The Bayesian Filter: A powerful tool for cognitive information processing**
7. **Cubature Kalman Filter**
8. **Properties of the Cubature Kalman Filter**
9. **On-going Research Projects in My Laboratory**
10. **Concluding Remark**

Acknowledgements

1. What is Cognition?

According to the Oxford English Dictionary:

Cognition is

Knowing,
- Knowledge
- Knowledge Acquisition
- Knowledge Representation
- Contextual Knowledge
- Storage of Knowledge

Perceiving,
- Perception
- Sensing of the Environment
- Adaptation to the Environment
- Learning from the Environment

or Conceiving
- Dealing with Uncertainty:
  1. Probabilistic Reasoning
  2. Hypothesizing and Decision-making
- "The Bayesian framework"

an Act
- Control
- Approximate Dynamic Programming

etc.
- Energy Efficiency
- Robustness

The human brain has all these attributes, and there is plausible evidence for the Bayesian framework -- hence the "Bayesian brain".

2. Cognitive Dynamic Systems Defined

A Cognitive Dynamic System is a system that processes information over the course of time by performing the following functions:

- ***Sense*** **the environment;**
- ***learn*** **from the environment and adapt to its statistical variations;**
- **build a** ***predictive model*** **of prescribed aspects of the environment**

and thereby develop ***rules of behaviour*** **for the execution of prescribed tasks, in the face of environmental uncertainties,** ***efficiently and reliably in a cost-effective manner.***

3. Emerging Applications

Cognitive radio
(Candidate for 5th generation wireless communications)

Cognitive radar

Cognitive car

Cognitive genome

.
.
.

Cognitive optimization

Cognitive software

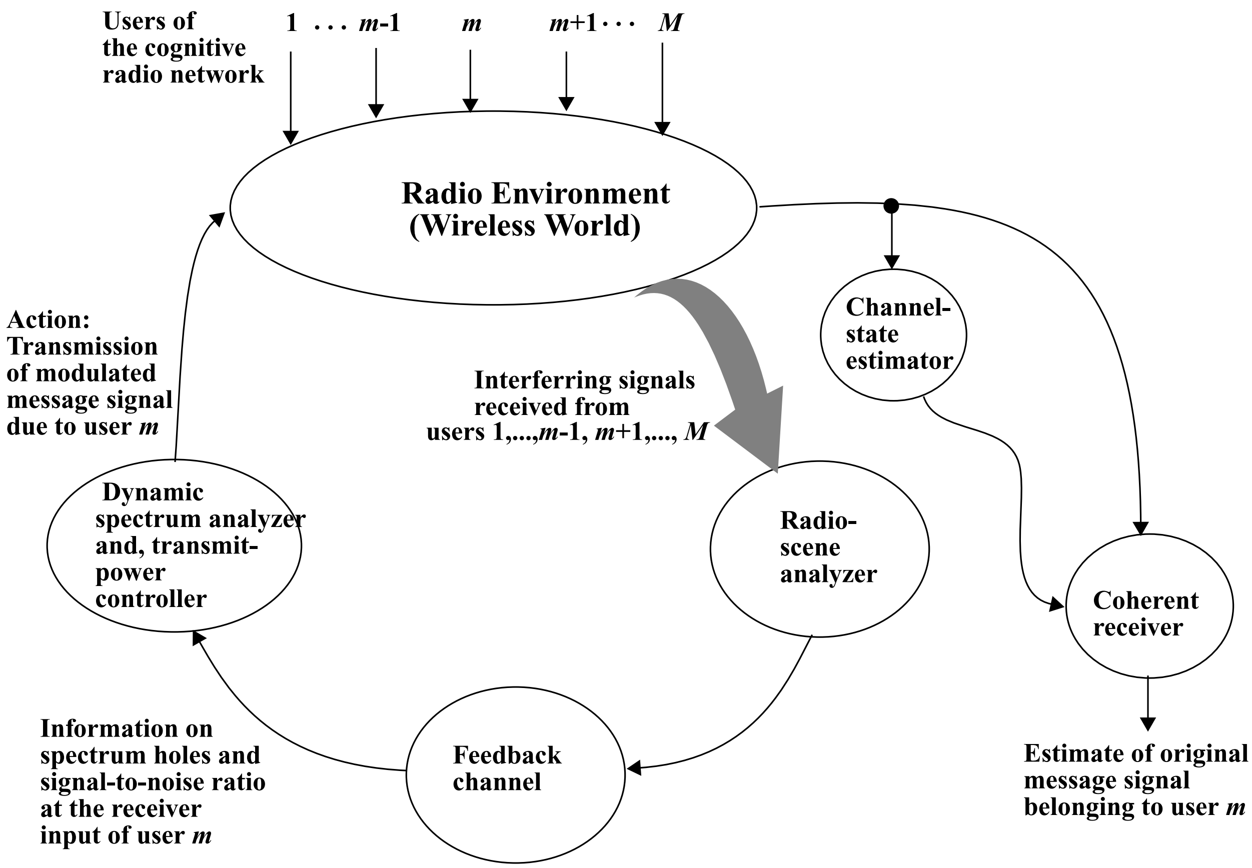
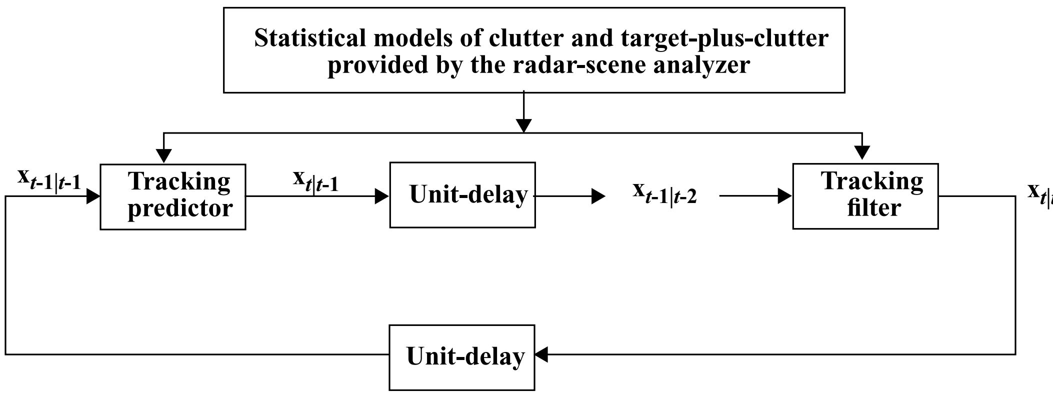


Figure 1: Cognitive signal-processing cycle for user *m* of cognitive radio network; the diagram also includes elements of the receiver of user *m*



Notations

t**:** **discrete-time**

$x_{t|t}$: filtered state vector of probabilities of targets being present in the search space at *t* given spectral measurements up to and including time *t*

The other data vectors in the diagram are similarly defined

Figure 2: Block diagram of the Bayesian direct filtering system

4. Global Feedback

"A Facilitator of Computational Intelligence"

- **The human brain is a living example of a cognitive dynamic system with global feedback in many of its parts, be that the visual system, auditory system, or motor control.**

- **Global feedback is responsible for the coordination of different constituents of a cognitive dynamic system.**

- **Global feedback is an inherent property of all cognitive dynamic systems, but global feedback by itself will *not* make a dynamic system cognitive.**

5. Why sub-optimality should be the objective of cognitive dynamic systems?

- Optimality of performance versus robustness of behaviour.

- Global optimality of a cognitive dynamic system is not practically feasible:

  - Infeasible computability
  - Curse-of-dimensionality
  - Large-scale nature of the system

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

- Trade-off global optimality for 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"?**

6. The Bayesian Filter: A powerful tool for cognitive information processing

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

Bayesian Filter (continued)

State-space Model

1. **System (state) Model**

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

2. **Measurement model**

   $$\mathbf{y}_t = \mathbf{b}(\mathbf{x}_t) + \nu_t$$

where t = discrete time

$\mathbf{x}_t$ = state at time t

$\mathbf{y}_t$ = observation at time

ω_t = dynamic noise

ν_t = measurement noise

Assumptions:

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

Bayesian filter (continued)

Time-update equation:

$$\underbrace{p(\mathbf{x}_t|\mathbf{Y}_{t-1})}_{\text{predictive distribution}} = \int_{R^n} \underbrace{p(\mathbf{x}_t|\mathbf{x}_{t-1})}_{\text{prior distribution}} \underbrace{p(\mathbf{x}_{t-1}|\mathbf{Y}_{t-1})}_{\text{old posterior distribution}} d\mathbf{x}_{t-1}$$

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

Measurement-update equation:

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

where Z_t is the normalizing constant defined by

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

Bayesian Filter (continued)

- The celebrated Kalman filter is a special case of the Bayesian filter, assuming that the dynamic system is linear.

- Except for this special case and couple of other cases, exact computation of the predictive distribution $p(\mathbf{x}_t | \mathbf{Y}_{t\text{-}1})$ is not feasible.

- We therefore have to abandon optimality and be content with a sub-optimal nonlinear filtering algorithm that is computationally tractable.

Bayesian Filtering (continued)

Two Approaches for Approximate Nonlinear Filtering

1. **Direct numerical approximation of the posterior in a local sense:**
   - **Extended Kalman filter**
   - **Unscented Kalman filter (Julier, Ulhmann and Durrant-Whyte, 2000)**
   - **Central-difference Kalman filter (Nörgaard, Poulson, and Ravn, 2000).**
   - **Cubature Kalman filter (Arasaratnam and Haykin, 2008).**

2. **Indirect numerical approximation of the posterior in a global sense:**
   - **Particle filters (Gordon, Salmond, and Smith, 1993)**
   - **Roots embedded in Monte Carlo simulation**
   - **Computationally demanding**

Extended Kalman Filter

- **Linearize the system model around the filtered estimate $\hat{\mathbf{x}}_{t|t}$, and linearize the measurement model around the predicted estimate $\hat{\mathbf{x}}_{t|t-1}$**

- **Attributes and Limitations**

  (i) The EKF is simple to implement

  (ii) Estimation accuracy of the EKF is good for nonlinearities of a mild sort; otherwise, it is often highly suboptimal.

7. The Cubature Kalman Filter

(Arasaratnam and Haykin, IEEE Trans. Automatic Control, accepted for publication subject to revisions)

- **At the heart of the Bayesian filter, we have to compute integrals whose integrand is expressed in the form**

  (Nonlinear function) x **(Gaussian function)**

- **The challenge is to numerically approximate the integral so as to completely preserve second-order information about the state x that is contained in the sequence of observations**

- **The computational tool that accommodates this requirement is the *cubature rule* (Cools, 1997).**

Cubature Kalman Filter (continued)

The Cubature Rule

- **In mathematical terms, we have to compute an integral for the generic form**

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

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

$\mathbf{x} = r\mathbf{z}$ subject to $\mathbf{z}^T\mathbf{z} = 1$ and $\mathbf{x}^T\mathbf{x} = r^2$

where $0 \leq r < \infty$

Cubature Kalman Filter (continued)

- **We may thus express *I* as the radial integral**

  $$I = \int_0^{\infty} S(r) r^{n-1} \exp(-r^2) dr$$

  where *S*(*r*) is defined by the spherical integral

  $$S(r) = \int_{U_n} f(r\mathbf{z}) d\sigma(\mathbf{z})$$

  where σ(.) is the spherical surface measure on the region

  $$U_n = \{\mathbf{z}, \text{ subject to } \mathbf{z}^T\mathbf{z} = 1\}$$

- **Working through a fair amount of mathematical details, we finally arrive at the desired *linear approximation*:**

  $$h(f) = \int_{R^n} \mathbf{f}(\mathbf{x}) \underbrace{N(\mathbf{x}; \mathbf{0}, \mathbf{I})}_{\text{Standard Gaussian function}} d\mathbf{x}$$

  $$\approx \sum_{i=1}^{2n} \omega_i \mathbf{f}\{\xi_i\}$$

  where $\{\xi_i\}$ = cubature representations of the state vector x.

  $$\omega_i = \frac{1}{m}, \quad i = 1, 2, ..., m = 2n$$

- **The set $\left\{\xi_i, \omega_i\right\}_{i=1}^{2n}$ constitutes the cubature points used to numerically compute integrals of the form defined in Eq. (1).**

Parameter Updates of the Cubature Kalman Filter

Time update

$$\hat{\mathbf{x}}_{n|n-1} = \mathbf{E}[\mathbf{x}_n | \mathbf{Y}_{n-1}]$$

$$= \int_{R^M} \underbrace{\mathbf{a}(x_{n-1})}_{\substack{\text{Nonlinear}\\\text{state}\\\text{function}}} \underbrace{N(\mathbf{x}_{n-1}; \hat{\mathbf{x}}_{n-1|n-1}, \mathbf{P}_{n-1|n-1})}_{\text{Gaussian distribution}} d\mathbf{x}_{n-1}$$

Measurement update

$$N\Bigg(\underbrace{\begin{bmatrix}\mathbf{x}_n\\ \mathbf{y}_n\end{bmatrix}}_{\substack{\text{Joint}\\\text{variables}}}; \underbrace{\begin{bmatrix}\hat{\mathbf{x}}_{n|n-1}\\ \hat{\mathbf{y}}_{n|n-1}\end{bmatrix}}_{\substack{\text{Joint}\\\text{mean}}}, \underbrace{\begin{bmatrix}\mathbf{P}_{n|n-1} & \mathbf{P}_{xy,\, n|n-1}\\ \mathbf{P}_{yx,\, n|n-1} & \mathbf{P}_{yy,\, n|n-1}\end{bmatrix}}_{\substack{\text{Joint}\\\text{covariance matrix}}}\Bigg)$$

$$\hat{\mathbf{y}}_{n|n-1} = \int_{R^M} \underbrace{\mathbf{b}(\mathbf{x}_n)}_{\substack{\text{Nonlinear}\\\text{measurement}\\\text{function}}} \underbrace{N(\mathbf{x}_n; \hat{\mathbf{x}}_{n|n-1}, \mathbf{P}_{n|n-1})}_{\text{Gaussian function}} d\mathbf{x}_n$$

$$P_{yy,\, n|n-1} = \int_{R^M} \underbrace{\mathbf{b}(\mathbf{x}_n)\mathbf{b}^T(\mathbf{x}_n)}_{\substack{\text{Outer product}\\\text{of}\\\text{nonlinear}\\\text{measurement}\\\text{function}\\\text{with itself}}} \underbrace{N(\mathbf{x}_n; \hat{\mathbf{x}}_{n|n-1}, \mathbf{P}_{n|n-1})}_{\text{Gaussian function}} d\mathbf{x}_n - \underbrace{\hat{\mathbf{y}}_{n|n-1}\hat{\mathbf{y}}^T_{n|n-1}}_{\substack{\text{Outer product}\\\text{of the}\\\text{estimate } \hat{\mathbf{y}}_{n|n-1}\\\text{with itself}}} + \underbrace{\mathbf{Q}_{\nu,\, n}}_{\substack{\text{Covariance}\\\text{matrix}\\\text{measurement}\\\text{noise}}}$$

$$\mathbf{P}_{xy,\, n|n-1} = \mathbf{P}_{yx,\, n|n-1}$$

$$= \int_{R^M} \underbrace{\mathbf{x}_n \mathbf{b}^T(\mathbf{x}_n)}_{\substack{\text{Outer}\\\text{product}\\\text{of } \mathbf{x}_n \text{ with}\\\mathbf{b}(\mathbf{x}_n)}} \underbrace{N(\mathbf{x}_n; \hat{\mathbf{x}}_{n|n-1}, \mathbf{P}_{n|n-1})}_{\text{Gaussian function}} d\mathbf{x}_n - \underbrace{\hat{\mathbf{x}}_{n|n-1}\hat{\mathbf{y}}^T_{n|n-1}}_{\substack{\text{Outer product}\\\text{of the}\\\text{estimates}\\\hat{\mathbf{x}}_{n|n-1} \text{ and } \hat{\mathbf{y}}_{n|n-1}}}$$

Recursive Cycle of the Cubature Kalman Filter

- **The Kalman gain is computed as**

$$\mathbf{G}_n = \mathbf{P}_{xy,\, n|n-1}\mathbf{P}^{-1}_{yy,n|n-1}$$

where $\mathbf{P}^{-1}_{yy,n|n-1}$ **is the inverse of the covariance matrix** $\mathbf{P}^{-1}_{yy,n|n-1}$.

- **Upon receiving the new observation** $\mathbf{y}_n$**, the filtered estimate of the state** $\mathbf{x}_n$ **is computed in accordance with the predictor-corrector formula:**

$$\underbrace{\hat{\mathbf{x}}_{n|n}}_{\text{Updated estimate}} = \underbrace{\hat{\mathbf{x}}_{n|n-1}}_{\text{Old estimate}} + \underbrace{\mathbf{G}_n}_{\text{Kalman gain}}\underbrace{(\mathbf{y}_n - \hat{\mathbf{y}}_{n|n-1})}_{\text{Innovations process}}$$

- **Correspondingly, the covariance matrix of the filtered state estimation error is computed as shown by**

$$\mathbf{P}_{n|n} = \mathbf{P}_{n|n-1} - \mathbf{G}_n\mathbf{P}_{yy,\, n|n-1}\mathbf{G}_n^T$$

Updated posterior distribution

$$p(\mathbf{x}_n|\mathbf{Y}_n) = N(\mathbf{x}_n;\hat{\mathbf{x}}_{n|n}, \mathbf{P}_{n|n})$$

Properties of the Cubature Kalman Filter

Property* 1*:* The cubature Kalman filter (CKF) is a *derivative-free on-line sequential-state estimator.

***Property* 2: Approximations of the moment integrals are all *linear* in the number of adjustable parameters.**

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

***Property* 6: The cubature Kalman filter is the closest known direct approximation to the Bayesian filter, outperforming the extended Kalman filter and the central-difference Kalman filter:**

It eases the curse-of-dimensionality problem

but, by itself, does not overcome it.

Computer Experiment: Pattern Classification

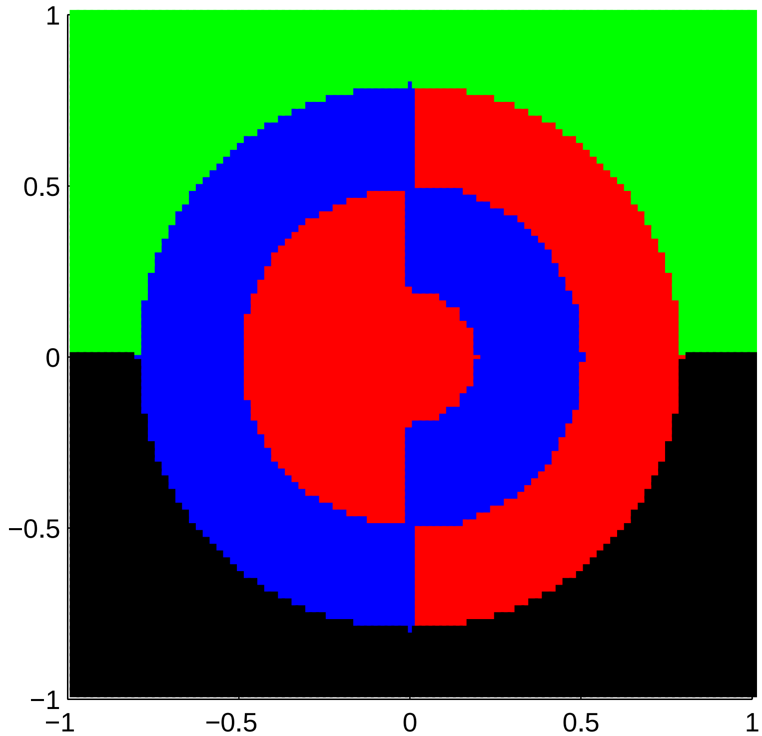


Figure 3: True classification regions

Supervised Learning

Training sample: $\{\mathbf{u}_t, \mathbf{d}_t\}$

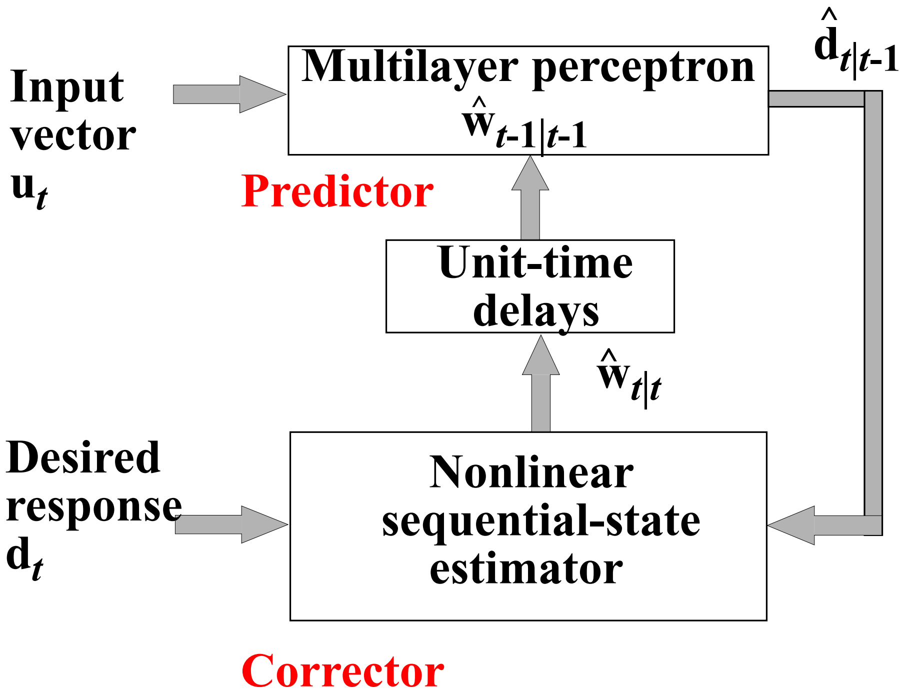


Figure 4: Block diagram of supervised learning machinery

Experimental Setup

- **Used a 2-5-5-4 FFNN with softmax output nonlinearity and the mean squared-error criterion.**

  Total number of adjustable weights: 55 plus biases

- **1000 training examples drawn randomly from the square region.**

- **DSSM:**

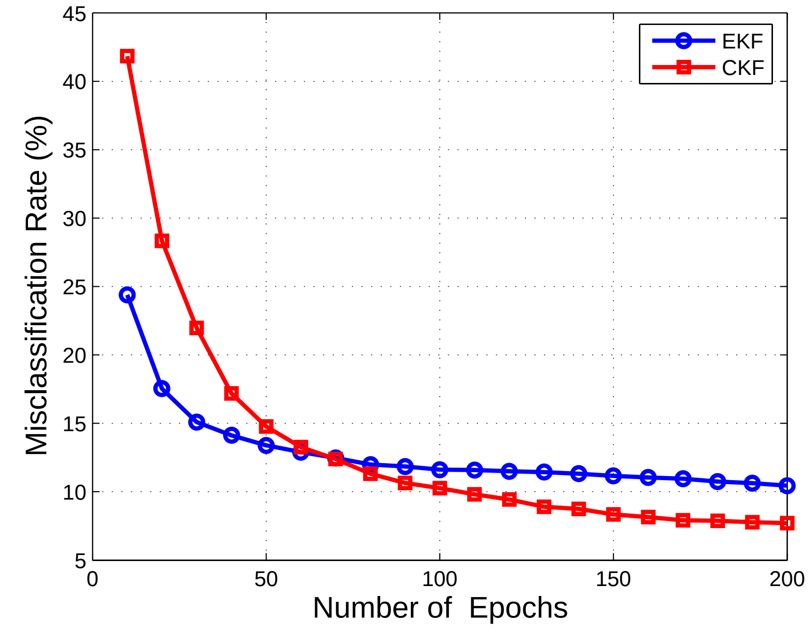
  Process equation: $\mathbf{w}_t = \mathbf{w}_{t-1} + \omega_t$

  Measurement equation: $\mathbf{y}_t = \mathbf{b}(\mathbf{w}_t, \mathbf{x}_t) + \nu_t$

- **Two training algorithms: EKF, and square-root Kalman filter (SCKF).**

- **To check robustness of filters, 10% of the training examples were mislabeled.**

Performance Comparison



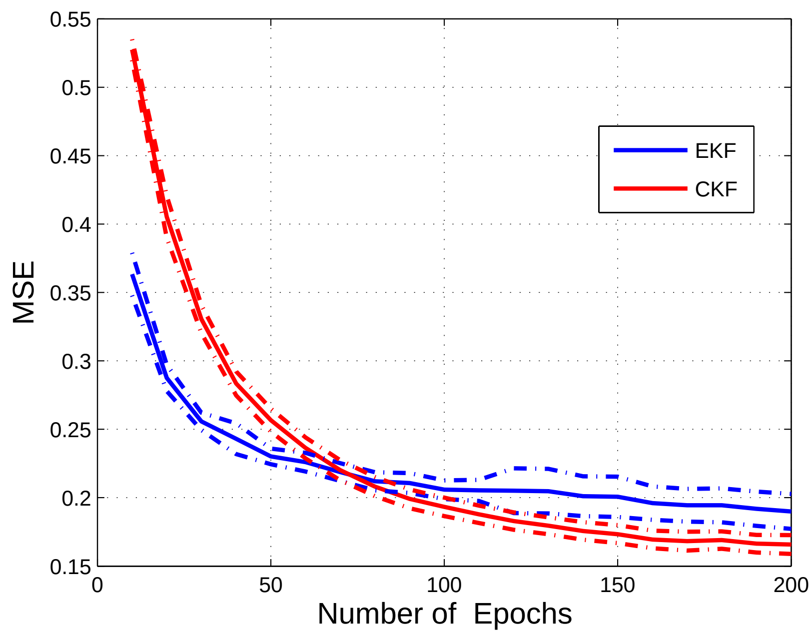


Figure 5:

9. On-going Research Projects in My Laboratory

(i) Cubature Kalman Filter:

- **Large-scale system applications involving pattern recognition and approximate dynamic programming.**

(ii) Cognitive Radio Networks:

- **Spectrum sensing**
- **Robust transmit power control**
- **Dynamic spectrum management**
- **Emergent behaviour**

(iii) Cognitive Radar Networks:

- **Sub-optimal control of inexpensive (surveillance) radar sensors, given limited computational resources**

(iv) Cocktail Party Processor:

- **Computational auditory scene analysis**

10. Concluding Remark

“Cognitive Dynamic Systems”

are

A Way of the Future

in

The 21st Century

Acknowledgements

- **Many thanks to my gifted group of outstanding Ph.D. students**

- **Deep gratitude to the Natural Sciences and Engineering Research Council (NSERC) of Canada for continued financial support**

The new Website

http://soma.mcmaster.ca

Cognitive Dynamic Systems Workshop, Niagara-on-the-Lake, May 2008, is available and slides can be downloaded from the following link

http://soma.mcmaster.ca/cds2008.php