Parameter Estimation in Physiological Models
Euro Summer School Lipari (Sicily-Italy)

Nonlinear Filtering and Estimation

Hien Tran

Department of Mathematics
Center for Research in Scientific Computation and
Center for Quantitative Sciences in Biomedicine
North Carolina State University

September 16-17, 2009

NC STATE UNIVERSITY

Outline

Outline

Kalman Filter

- In 1960, R.E. Kalman published his seminal paper describing an efficient recursive solution to the discrete, linear filtering problem from a series of noisy measurements.
- Since its discovery over 40 years ago, much research has gone into refining its estimation accuracy and into its extensions to *highly nonlinear models*.

References:

- R.E. Kalman, A new approach to linear filtering and prediction problems, Trans of the ASME - Journal of Basic Engineering **82** (Series D): 35-45, 1960
- F.L. Lewis, *Optimal Estimation with an Introduction to Stochastic Control Theory*, John Wiley & Sons, 1986

A Hypothetical Example

Reference: P.S. Maybeck, *Stochastic Models, Estimation, and Control, Vol. 1*, Academic Press, 1979.

Suppose you are at lost at sea during the night and trying to determine your location.

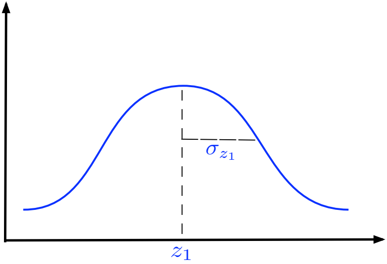

Figure: Conditional density of position based on the measured value z_1.

A Hypothetical Example

Now, a trained navigator friend takes an independent estimation of the position right after you do (so that the true position has not changed at all !).

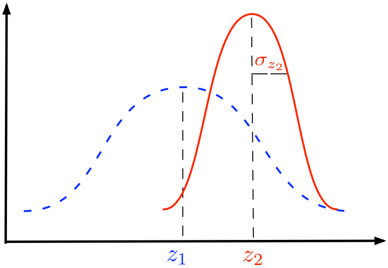

Figure: Conditional density of position based on the measured value z_2.

A Hypothetical Example

At this point, you have two measurements available to estimate your current position.

Question: How do you combine these data? (so that you have a better estimate on your position than either the first or the second estimate)

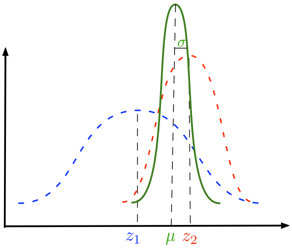

Figure: Conditional density of position based on the measured values z_1 and z_2.

Optimal Estimate

$$\mu = \frac{\sigma_{z_2}^2}{\sigma_{z_1}^2 + \sigma_{z_2}^2} z_1 + \frac{\sigma_{z_1}^2}{\sigma_{z_1}^2 + \sigma_{z_2}^2} z_2, \quad \frac{1}{\sigma^2} = \frac{1}{\sigma_{z_1}^2} + \frac{1}{\sigma_{z_2}^2}$$

- If $\sigma_{z_1} = \sigma_{z_2}$, the best estimate should be the average of the two.
- If $\sigma_{z_1} > \sigma_{z_2}$ (i.e., z_2 is a better estimate), then the formula indicates that we should weight our estimate more toward z_2.
- The variance of the optimal estimate is less than both $\sigma_{z_1}^2$ and $\sigma_{z_2}^2$.

Rewrite the optimal estimate as:

$$\begin{aligned} \mu &= z_1 + K\left[z_2 - z_1\right], \quad K = \frac{\sigma_{z_1}^2}{\sigma_{z_1}^2 + \sigma_{z_2}^2}, \\ \sigma^2 &= \sigma_{z_1}^2 - K\sigma_{z_1}^2 \end{aligned}$$

Kalman Filter: Concepts

$$\begin{aligned} \mu &= z_1 + K\left[z_2 - z_1\right], \quad K = \frac{\sigma_{z_1}^2}{\sigma_{z_1}^2 + \sigma_{z_2}^2}, \\ \sigma^2 &= \sigma_{z_1}^2 - K\sigma_{z_1}^2 \end{aligned}$$

Now, suppose that z_1 is the estimate from your model and z_2 is the measurement, *Kalman filter is a technique that combines the model estimate with measurement to derive a better estimate for the model by considering both the error in the model and the error in the data.*

The same idea can be extended to estimate the unknown parameters in the model as well as the states - *dual estimation*.

NC STATE UNIVERSITY

The Linear Discrete Kalman Filter

The Kalman filter addresses the general problem of estimating the state $x \in \mathbb{R}^n$ of a linear discrete-time process

$$\begin{aligned} x_{k+1} &= Ax_k + w_k, \quad w \sim \mathcal{N}(0, V) \\ y_k &= Cx_k + v_k, \quad v \sim \mathcal{N}(0, R) \end{aligned}$$

where $x_k \in \mathbb{R}^n, y_k \in \mathbb{R}^m$ and w_k, v_k are additive white gaussian noise (AWGN) processes.

Kalman Filtering Equations

Prediction Steps (Time Update):

$$\hat{x}_k^- = A\hat{x}_{k-1}$$
$$P_k^- = AP_{k-1}A^T + V$$

Correction Steps (Measurement Update):

$$K_k = P_k^- C^T \left[CP_k^- C^T + R \right]^{-1}$$
$$\hat{x}_k = \hat{x}_k^- + K_k(y_k - C\hat{x}_k^-)$$
$$P_k = \left[I - K_k C \right] P_k^-$$

- $\lim_{R \to 0} K_k = C^{-1}$
- $\lim_{P_k^- \to 0} = 0$

On-line Estimation

Filter Parameters and Tuning

- Good filter performance can be achieved by tuning the filter parameters, the *model noise and measurement noise covariances*, V and R.
- The determination of the model noise covariance V is generally more difficult.

Outline

Nonlinear Kalman Filtering

Consider a nonlinear discrete-time model and observation:

$$\begin{aligned} x_{k+1} &= f(t_k, x_k) + w_k, \quad w \sim \mathcal{N}(0, V) \\ y_k &= h(t_k, x_k) + v_k, \quad v \sim \mathcal{N}(0, R) \end{aligned}$$

where $x_k \in \mathbb{R}^n$, $y_k \in \mathbb{R}^m$, $q \in \mathbb{R}^p$ and w_k, v_k are additive white gaussian noise (AWGN) processes.

Suboptimal filters were developed to handle these situations. These filters employ

- Linearizations of the model and measurement (Extended KF)
- Approximations of the underlying distribution to a Gaussian pdf (Unscented KF)
- Monte Carlo sampling techniques (Ensemble KF, Particle Filtering)

Extended Kalman Filter

The Extended Kalman Filter (EKF) linearizes the state dynamics around the current estimate.

Prediction Steps:

$$\begin{aligned} \hat{x}_k^- &= f(t_{k-1}, \hat{x}_{k-1}) \\ P_k^- &= \nabla f(\hat{x}_k^-) P_{k-1} \nabla f^T(\hat{x}_k^-) + V \end{aligned}$$

Correction Steps:

$$\begin{aligned} K_k &= P_k \nabla h^T(\hat{x}_k^-) \left[\nabla h(\hat{x}_k^-) P_k \nabla h^T(\hat{x}_k^-) + R\right]^{-1} \\ \hat{x}_k &= \hat{x}_k^- + K_k(y_k - h_k(\hat{x}_k^-)) \\ P_k &= \left[I - K_k \nabla h(t_k, \hat{x}_k^-)\right] P_k^- \end{aligned}$$

Extended Kalman Filter

- In the EKF, the state distribution is approximated by a gaussian random variable (GRV), which is then propagated through the linearization.
- In highly nonlinear problems, the EKF tends to be very inaccurate and underestimates the true covariance of the estimated state.
- This can lead to poor performance and filter divergence.

$\Rightarrow$ Can we do better?

- **Unscented Kalman Filter** was designed to overcome these problems !

Unscented Kalman Filter

Reference: S. Haykin, *Kalman Filtering and Neural Networks*, John Wiley & Sons, Inc., 2001.

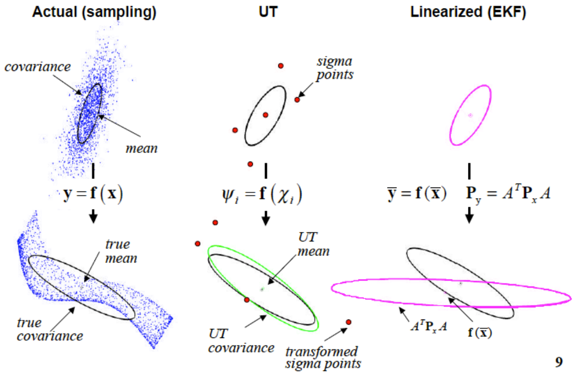

Figure: Examples of mean and covariance propagation.

Unscented Kalman Filtering

- The Unscented Kalman Filter (UKF) is built around the idea that it is easier to approximate the underlying distribution than it is to approximate the state dynamics.
- Uses a deterministic sampling approach to approximate the distribution.
- The state distribution is approximated by a GRV, but is represented by a set of sigma points, completely capturing the **true** mean and covariance of the state distribution.
- When propagated through the nonlinear system, the posterior mean and covariance are captured to second order of accuracy.
- Computational cost is equal to the EKF (of order n^3).

UKF Equations

$2n+1$ (n is the state dimension) sigma vectors are generated according to

$$\mathcal{X}_0 = \bar{x} \tag{1}$$

$$\mathcal{X}_i = \bar{x} + \left(\sqrt{(n+\lambda)P_x}\right)_i, \quad i = 1, \ldots, n \tag{2}$$

$$\mathcal{X}_i = \bar{x} - \left(\sqrt{(n+\lambda)P_x}\right)_i, \quad i = n+1, \ldots, 2n \tag{3}$$

where $\mathcal{X}_i$ denotes the i-th column of the matrix $\mathcal{X}$.

$\Rightarrow$These points are where the distribution of $\hat{x}$ are sampled. In practice, Cholesky factors are used as the matrix square root.

UKF Equations

Each sample point has an associated weight, weighting the mean estimation and the covariance estimation differently. $W^m \in \mathbb{R}^{n \times 2n+1}$ are the weights for the mean, $W^c \in \mathbb{R}^{n \times 2n+1}$ for the covariance estimate.

$$W_0^m = \lambda(n+\lambda)^{-1}$$
$$W_0^c = \lambda(n+\lambda)^{-1} + (1 - \alpha^2 + \beta)$$
$$W_i^m = W_i^c = \big(2(n+\lambda)\big)^{-1}, \quad i = 1 \ldots 2n$$

where λ, α, β are all tuning parameters.

UKF Equations

Predictions:

$$\begin{aligned}
\mathcal{X} &= \text{sigmapoints}(x_k, P_k) \\
\hat{\mathcal{X}}_i &= f(t_k, \mathcal{X}_i) \\
\hat{x}_k^- &= \sum_{i=0}^{2n} W_i^m \hat{\mathcal{X}}_i \\
\hat{P}_k^- &= V + \sum_{i=0}^{2n} W_i^c \left[\mathcal{X}_i - \hat{x}_k^-\right]\left[\mathcal{X}_i - \hat{x}_k^-\right]^T \\
\mathcal{X}_k^- &= \text{sigmapoints}(\hat{x}_k^-, \hat{P}_k^-) \\
\mathcal{Y}_k &= h(t_k, \mathcal{X}_k^-) \\
\hat{y}_k &= \sum_{i=0}^{2n} W_i^m \mathcal{Y}_i
\end{aligned}$$

where $\mathcal{X}_i$ and $\mathcal{Y}_i$ denote the i-th column.

UKF Equations

Update/Correction Equations:

$$
\begin{aligned}
P_{\bar{y}_k \bar{y}_k} &= R + \sum_{i=0}^{2n} W_i^c \left[\mathcal{Y}_i - \hat{y}_k\right]\left[\mathcal{Y}_i - \hat{y}_k\right]^T \\
P_{\bar{x}_k \bar{y}_k} &= \sum_{i=0}^{2n} W_i^c \left[\mathcal{X}_{i,k}^- - \hat{x}_k^-\right]\left[\mathcal{Y}_i - \hat{y}_k\right]^T \\
K &= P_{\bar{x}_k \bar{y}_k} P_{\bar{y}_k \bar{y}_k}^{-1} \\
\hat{x}_k &= \hat{x}_k^- + K(z_k - \hat{y}_k) \\
P_k &= P_k^- - K P_{\bar{y}_k \bar{y}_k} K^T.
\end{aligned}
$$

Unscented Kalman Filtering

- The UKF is a recursive implementation of the Unscented Transform (UT), which computes the statistics of a random variable that undergoes a nonlinear transformation.
- Works well on nonlinear problems.
- Similar to particle filters, only with a deterministic sampling method.
- Further numerically robust versions available in the *Square Root Filter*.

$$\mathcal{X}_i = \bar{x} \pm \left(\sqrt{(n+\lambda)P_x}\right)_i, \quad i = 1, \ldots, 2n$$

Pitfalls to Discrete Filtering

- If data are sparse, the step size taken can be large, affecting the integration accuracy.
- Dynamics that affect accuracy may be missed by a single step.
- In fixed step size integrators, there is no automatic error control.
- Discretization of the model inherently changes the model to something new.
- Discrete filters are more sensitive to amount and quality of data.

$\Rightarrow$ Solution: Continuous versions of the Kalman Filters.

Outline

Continuous Kalman Filtering

The continuous Kalman Filter is known as the Kalman-Bucy Filter.

- Continuous filters do not require an a-priori discretization of the state space dynamics.
- The state space model is augmented with a matrix Riccati equation describing the propagation of the covariance matrix. The augmented system constitutes a system of stochastic differential equations (SDEs).
- Multistep, adaptive mesh integrators can be used for state and covariance prediction, increasing accuracy and increasing information content.
- Maintain the assumption that the observations are discrete in time.

Extended Kalman-Bucy Filter

The Extended Kalman-Bucy Filter (EKBF) employs an augmented state space,

$$\begin{aligned}\dot{\hat{x}}(t) &= f(\hat{x}(t), t)\\ \dot{P}(t) &= P(t)\nabla f(\hat{x})^T + \nabla f(\hat{x})P(t) + V.\end{aligned}$$

These equations are integrated from t_k to t_{k+1}. The correction equations remain the same,

$$\begin{aligned}K_k &= P^-(t_k)\nabla h(\hat{x})^T\left[\nabla h(\hat{x})P^-(t_k)\nabla h(\hat{x})^T + R\right]^{-1}\\ P_k &= \left[I - K_k\nabla h(\hat{x})\right]P^-(t_k)\\ \hat{x}_k &= \hat{x}_k^- + K_k\left[z_k - \nabla h(\hat{x})\hat{x}_k^-\right].\end{aligned}$$

NC STATE UNIVERSITY

Extended Kalman-Bucy Filter

- The EKBF performs better than the EKF when fewer observations are available, either longitudinally or from issues arising from state observability.
- Tuning the integration tolerances will affect the tracking ability of the filter.

If the problem is too nonlinear, the EKBF will still fail. This motivates the Unscented Kalman-Bucy Filter (UKBF).

Unscented Kalman-Bucy Filter

The UKBF is a natural extension of the UKF in continuous time. The sigma points become a function of time, and are given as

$$
\begin{aligned}
\mathcal{X}(t)_0 &= \bar{x}(t) \\
\mathcal{X}(t)_i &= \bar{x}(t) + \left(\sqrt{(n+\lambda)P(t)_x}\right)_i, \quad i = 1, \ldots, n \\
\mathcal{X}(t)_i &= \bar{x}(t) - \left(\sqrt{(n+\lambda)P(t)_x}\right)_i, \quad i = n+1, \ldots, 2n
\end{aligned}
$$

Unscented Kalman-Bucy Filter

The augmented state space model is given by

$$\begin{aligned}\dot{\hat{x}}(t) &= f\big(\mathcal{X}(t), t\big)\, W^m \\ \dot{P}(t) &= \mathcal{X}(t) W f^T\big(\mathcal{X}(t), t\big) + f\big(\mathcal{X}(t), t\big)\, W \mathcal{X}^T(t) + V,\end{aligned}$$

where $\mathcal{X}(t)$ is implicitly a function of $\hat{x}(t)$ and $P(t)$ and the matrix W is given by

$$W = \big(I - [W_0^m \cdots W_{2n}^m]\big) \cdot \mathrm{diag}\big(W_c^0 \cdots W_c^{2n}\big) \cdot \big(I - [W_0^m \cdots W_{2n}^m]\big)^T.$$

The correction equations remain the same (omitted for brevity). The state space is integrated from t_k to t_{k+1}.

Remarks

If we assume both filters use the same initial conditions and covariance matrices, we observe:

- For sparse data sets, the continuous filters will outperform the discrete filters under the same filtering conditions.
- For highly nonlinear systems, the UK(B)F will outperform the EK(B)F – this is well known.
- The UK(B)F will track the unobserved states better than the EK(B)F.

Outline

Parameter Estimation

In modeling biological processes, modelers frequently wish to relate biological parameters characterizing a model, θ, to collected observations making up some data set, y. We assume that the relationship between θ and y is described by a nonlinear function G

$$G(\theta) = y$$

For example, consider a simple model for the concentration of a drug introduced in a biological system

$$\frac{dx(t)}{dt} = -ax(t) + bu(t)$$
$$y(t) = cx(t)$$

Parameter Estimation

Assuming that $x(0) = 0$, the solution, which is computed from the variation of constants formula, is given by

$$\begin{aligned} y(t) &= cb \int_0^t e^{-a(t-s)} u(s) ds \\ &\equiv G(\theta) \end{aligned}$$

where $\theta = (a, b, c)$.

Parameter Estimation

The parameter estimation problem can be solved by the Kalman Filter by writing a new state-space representation,

$$\begin{aligned} \theta_{k+1} &= \theta_k + r_k \\ y_k &= G(\theta_k) + n_k \end{aligned}$$

where r_k is the noise of the stationary parameter process and n_k is the noise of the nonlinear observation function G.

Dual Estimation

Dual estimation problems consist of estimating both the states, x_k, and the parameters, θ_k, given noisy data, y_k.

Joint Filtering

$$\begin{aligned} \dot{x} &= f(t, x, \theta) \\ \dot{\theta} &= 0 \end{aligned}$$

- Increase the number of states (large number of parameters)
- Errors propagate from the state into the parameter (which subsequently propagate back into the state)

  $\Rightarrow$ Inaccurate results or divergence of the filter

NC STATE UNIVERSITY

Dual Filter

Idea: Running two filters concurrently

- State Filter estimates the state using the current parameter estimate, $\hat{\delta}_k^-$.
- Parameter Filter estimates the parameters using the current state estimate, $\hat{x}_k^-$.
  - Do not increase the number of states for estimation.
  - Errors will not feedback into the next estimate.

Outline

A Nonlinear Spring Model

A simple nonlinear spring-mass-dashpot model:

$$\ddot{x} + \gamma \dot{x} + kx + bx^3 = 0,$$

The observed states are

$$y = (x, \dot{x})$$

Problem: Estimate parameters $\theta = (k, b, \gamma)$ using simulated data with AWGN.

Dual Estimation Results

	k	b	γ
"True"'	60	100	4
"UKF"'	59.7	96.985	3.775

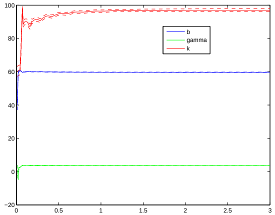

Figure: Convergence of the parameter estimation (Dual UKF).

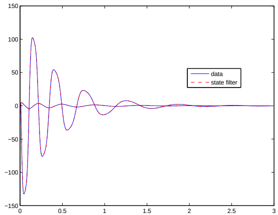

Figure: Comparison of the true states (solid) versus the dual UKF state estimation.

HIV Dynamics

An acute HIV infection with no treatment can be modeled as

$$\begin{aligned}\dot{T} &= \lambda - dT - kVT \\ \dot{T}^* &= kTV - \delta T^* \\ \dot{V} &= N\delta T^* - cV\end{aligned}$$

where T^* is infected T-cells, V is free viron particles, λ is the recruitment of uninfected T-cells, d is the per capita death rate of uninfected cells, k is the infection rate, δ is the death rate of uninfected cells, N is the number of new HIV virons and c is the clearance rate.

Collected data could be a combination of viral load (V) and healthy T-cell count (T).

HIV Model

To begin, we consider the parameter estimation problem of estimating all 6 parameters in the model, $\theta = (\lambda, d, k, \delta, N, c)$. However, the dual UKF algorithm failed to converge.

Question: What's happened?

$$
\begin{aligned}
y(t) &= cb \int_0^t e^{-a(t-s)} u(s) ds \\
&\equiv G(\theta)
\end{aligned}
$$

"A priori" local analyses:

- Sensitivity
- Identifiability (Subset Selection)

NC STATE UNIVERSITY

Sensitivity Analysis

Consider a mathematical model

$$\frac{dx}{dt}(t) = f(t, x(t), \theta)$$

with observation process

$$y(t) = h(t, x(t), \theta)$$

The sensitivity of outputs with respect to parameters is defined by

$$\frac{dy_i}{d\theta_j}$$

Using the chain rule for differentiation,

$$\frac{dy}{d\theta} = \frac{\partial h}{\partial x}\frac{dx}{d\theta} + \frac{\partial h}{\partial \theta}$$

where

$$\frac{d}{dt}\frac{dx}{d\theta} = \frac{\partial f}{\partial x}\frac{dx}{d\theta} + \frac{\partial f}{\partial \theta}$$

Subset Selection

Reference: M. Fink, A. Attarian and H. Tran, Subset selection for parameter estimation in an HIV model, Proc. in Applied Mathematics and Mechanics **7**, Issue 1, (2008) 1121501-1121502.

Consider the linear least squares problem,

$$\min_{x \in \mathbb{R}^m} \|Ax - b\|_2^2$$

If $A \in \mathbb{R}^{p \times m}$ is nearly rank deficient, then this problem is very ill-conditioned. A standard technique is to compute an SVD of A and then set to zero all singular values below a certain threshold. A good threshold value to use is the numerical rank of a matrix

$$\mathrm{rank}(A, \epsilon) = \max \left\{ i \middle| \frac{|\sigma_i|}{|\sigma_1|} > \epsilon \|A\| m \right\}$$

Subset Selection in Parameter Estimation

Denote by $y(\theta)$ the model output as a function of parameter θ. We can approximate the change in the output for a change in parameter from θ to $\hat{\theta}$ as

$$y(\hat{\theta}) - y(\theta) \approx \frac{dy}{d\theta}(\hat{\theta} - \theta) + \mathcal{O}\left((\hat{\theta} - \theta)^2\right)$$

In the context of the linear least squares problem

$$\min_{\theta \in \mathbb{R}^m} \|\frac{dy}{d\theta}\Delta\theta - \Delta y\|_2^2$$

and if the matrix $A = \frac{dy}{d\theta}$ has numerical rank $k < m$, it makes sense to minimize the residual over a subspace of dimension k by modifying k parameters while keeping $m - k$ parameters constant. To determine which components of θ to modify, we look for a maximally independent set of columns of A.

Subset Selection Algorithm

SVD followed by QR with Column Pivoting:

- Compute an SVD of $A = U\Sigma V^T$ and determine a numerical rank estimate k.
- Let $V = [V_k, V_{m-k}]$, where V_k is the first k columns of V.
- Perform a QR factorization with pivoting on V_k^T to obtain

$$V_k^T P = QR$$

- Choose as the subset of components of θ the first k components of $P^T\theta$.

For the 3-dimensional HIV model, sensitivity and subset selection reveal that only 3 parameters $\theta = (\lambda, k, \delta)$ of the 6 parameters $(\lambda, d, k, \delta, N, c)$ are most identifiable and sensitive (locally).

Dual Estimation Results

	λ	k	δ
“True'"	10	8×10^{-4}	0.7
“UKF"	9.5	8.2×10^{-4}	0.701

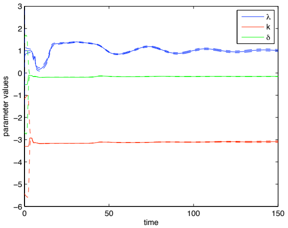

Figure: Convergence of the parameter estimation (Dual UKF).

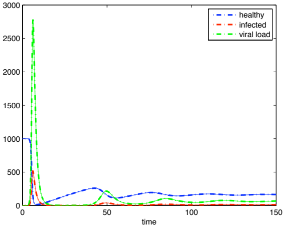

Figure: Comparison of the true states (solid) versus the dual UKF state estimation.

Acknowledgments

NC State University

- A. Attarian, H.T. Banks, J. David, S. Hu, Z. Kenz, B. Matzuka

University of Graz

- M. Bachar, J. Batzel, F. Kappel

University of Oxford

- M. Fink