

# 언센티드 칼만필터 훈련 알고리즘에 의한 순환신경망의 파라미터 추정 및 비선형 채널 등화에의 응용

## Parameter Estimation of Recurrent Neural Networks Using A Unscented Kalman Filter Training Algorithm and Its Applications to Nonlinear Channel Equalization

권오신

Oh Shin Kwon

군산대학교 공과대학 전자정보공학부 교수

School of Electronic and Information Engineering

Kunsan National University

### 요약

실시간 순환형 훈련 알고리즘(RTRL)과 같이 경사법에 의해 훈련되는 순환형 뉴럴 네트워크(RNN)은 수렴속도가 매우 느린 단점을 지니고 있다. 이 알고리즘은 또한 오차 역전달 처리과정에서 결코 쉽지 않은 미분 계산을 필요로 한다. 본 논문에서는 완전하게 결합된 RNN의 훈련을 위하여 소위 언센티드 칼만필터라고 불리는 미분없는 칼만필터 훈련 알고리즘을 시스템의 상태공간 상에서 표현하였다. 미분없는 칼만필터 훈련 알고리즘은 순환형 뉴럴 네트워크 훈련시 미분 계산 없이 매우 빠른 수렴속도와 좋은 추정 성능을 보여준다. 비선형 채널 등화 실험을 통하여 미분 없는 칼만필터 훈련 알고리즘을 이용한 RNN의 성능이 향상되었음을 보였다.

### Abstract

Recurrent neural networks(RNNs) trained with gradient based such as real time recurrent learning(RTRL) has a drawback of slow convergence rate. This algorithm also needs the derivative calculation which is not trivialized in error back propagation process. In this paper a derivative free Kalman filter, so called the unscented Kalman filter(UKF), for training a fully connected RNN is presented in a state space formulation of the system. A derivative free Kalman filter learning algorithm makes the RNN have fast convergence speed and good tracking performance without the derivative computation. Through experiments of nonlinear channel equalization, performance of the RNNs with a derivative free Kalman filter learning algorithm is evaluated.

**Key Words** : Recurrent Neural Network, A Derivative Free Kalman Filter, Unscented Kalman Filter, Channel Equalization, Time-varying Channel.

### 1. 서론

최근 여러 분야에서 다양하게 이용되고 있는 신경망(Recurrent Neural Network, RNN)은 기본적으로 비선형 상태방정식으로 나타내는 동적 시스템으로 표시할 수 있다.

이러한 동특성(dynamic nature)으로 인하여 순환신경망은 시스템 모델링과 신호처리 예를 들어 예측(prediction) 및 적응제어, 시스템인식, 음성인식과 같은 다양한 분야에 성공적으로 널리 응용되어 지고 있다.[1] 또한 다른 많은 영역에서

입·출력의 순차특성의 편리함 때문에 순환신경망이 시계열 예측 및 순차변환(sequence transduction) 등과 같은 분야에 매우 흥미롭게 이용되어 지는 것을 알 수 있다.

특히 디지털 통신의 채널 등화(channel equalization)에서는 이러한 순차적인 데이터 처리의 좋은 예를 보여주며 또한 통신 리시버에서 등화기처럼 사용되는 적응필터는 파라미터가 실시간으로 학습하여 업데이트 되는 것에 대단히 쉽게 적용할 수 있다. 최근 순환신경망(RNN)은 네트워크 구조의 다양성과 훈련 알고리즘의 편리함으로 인하여 통신채널 등화에 많이 응용되어 왔다.[2],[3],[4],[5]

한편 이와는 별도로 순환신경망에 대한 구조적인 관점에서 많은 발전이 이루어져 왔는데, 전체적으로 연결되는 네트워크에서부터 부분적으로 연결하는 방법에 이르거나, 단일층의 레이어에서부터 다층 레이어의 네트워크 등의 구조가 이에 좋은 예이다. 그러나 이러한 눈부신 발전에도 불구하고

접수일자 : 2005년 9월 1일

완료일자 : 2005년 10월 7일

본 논문은 2002년도 군산대학교 교수장기해외연수경비의 지원에 의하여 연구되었음.

아직까지 해결해야 할 문제점이 많이 남아 있는데 그것은 동특성(dynamical behavior)에 대한 다양한 분석과 함께 네트워크의 동역학에 의해 유기되는 복잡한 움직임에 대처하기 위한 훈련 알고리즘 등을 개발 하는 것이다. 그래서 지금까지 네트워크의ダイ나믹한 특성에 관한 집중적인 연구와 이에 대응하기 위한 다양한 훈련 알고리즘 등이 많이 개발 되어 왔다.[6] 본 논문에서는 이러한 학습 알고리즘의 개발에 초점을 맞추었다. 지금까지 순환신경망에 적용하였던 전형적인 훈련 알고리즘은 온라인 상태에서 적용하는 실시간 순환학습(Real-Time Recurrent Learning, RTRL)방법과 오프라인에서 적용하는 역전달 학습(Back-Propagation Through Time, BP TT)방법을 들 수 있다.[7],[8]

그러나 이러한 방법은 모두 1차 미분요소 정보를 이용하는 경사법(gradient method)에 기반을 두는 데, 이러한 방법을 이용하여 학습할 경우 네트워크 상에서 파라미터의 갱신 과정 내내 영향을 끼치는 등의 문제점이 있다. 또한 오차경사(error gradient)를 구하는 과정에서 매우 복잡한 절차를 거쳐야 한다.

더욱이 1차 미분요소 정보(Jacobian matrix)를 이용하는 RTRL 과 BPTT 는 확장된 칼만필터(Extended Kalman Filters, EKF)를 이용한 학습방법과 같은 2차 미분요소의 정보(Hessian matrix)에 기반을 둔 학습 알고리즘에 비하여 상대적으로 느린 수렴속도를 지니게 된다. 순환형 EKF 학습 알고리즘의 핵심은 학습하는 기간동안 전개되고 유지되는 2차 미분요소의 정보를 기호화한 평균값(mean) 과 공분산 행렬(covariance matrix)에 있다고 할 수 있다. 그러나 EKF 를 이용한 학습방법 역시 이를 도구화 하거나 동조할 때 어려움을 지니고 있고, 파라미터를 갱신하는 과정에서 전형적인 선형 시스템인 경우로 제한 되어진다.[9] 또한 EKF는 최적의 비선형 파라미터를 추정하기 위한 1차 근사화를 제공하며, 비선형 시스템을 선형화 시키는 과정에서 미분행렬의 계산을 필요로 한다. 본 논문에서는 소위 엔센티드 칼만필터(Unsented Kalman Filter)라고 불리우는 미분없는 칼만필터(derivertive-free Kalman filter)가 순환신경망에서의 학습 알고리즘으로 이용되는 방법에 대하여 제시한다.[9],[10] 다음에서 알 수 있듯이UKF를 이용한 훈련 알고리즘은 EKF를 이용한 학습 방법의 대안이 될 수 있음을 보인다. UKF를 이용할 경우 EKF를 이용한 학습 알고리즘 전개 시 필요한Jacobians matrix 계산과정을 거치지 않기 때문에 수렴속도가 향상되고 상용화하기가 쉬워진다. 시뮬레이션을 통하여 순환신경망의 학습시 UKF 를 이용한 학습 알고리즘의 뛰어난 적응성과 비선형 채널 등화에 응용할 경우 기존의 RTRL 알고리즘을 적용하였을 때와 비교하여 새로운 방법에 대한 우수한 성능을 입증하였다. 학습 알고리즘으로서 두 종류의 칼만필터 즉, 확장 칼만필터(extended Kalman filter, EKF)와 엔센티드 칼만필터(Unscented Kalman filter, UKF)를 적용한다. EKF 알고리즘은 신경회로망의 학습에 널리 이용되며 그래디언트 기반 알고리즘에 비해 빠른 수렴성능을 보장하는 것으로 알려져 있다. 그러나 비선형 시스템의 선형 근사화 과정에서 많은 오차를 포함할 수 있는 반면에, UKF 알고리즘은 신경회로망 학습 시 Jacobian 행렬의 연산을 필요로 하지 않는 장점을 가지고 있다.

최근, EKF로 학습되는 순환신경망의 응용들은 많이 발표되었지만 시변 채널에 대한 성능 평가는 아직까지 발표되지 않았다. 더욱이, 최근에 발표된 UKF 알고리즘의 순환신경망에의 응용 또한 발표되지 않았다. 따라서 본 논문에서는 UKF 알고리즘을 순환신경망 등화기의 학습에 적용하고, 시변채

널에 대한 등화 실험을 통해서 RTRL알고리즘을 사용하였을 때와 와 성능을 비교 검토하여 제한한 등화기들의 우수성을 증명한다.

## 2. 완전 결합된 순환신경망

완전 결합된 순환신경망을 나타내는 일반식은 식(1)과 같이 나타낼 수 있다 완전 결합된 순환신경망은 그림1.에서와 같이  $q \times 1$ 개의 벡터  $x(k)$ 는 비선형 이산시간 시스템에서 네트워크의 상태를 나타내며,  $(l+1) \times 1$ 개의 벡터  $u(k)$ 는 바이어스를 포함한 네트워크의 입력,  $p \times 1$ 개의 벡터  $y(k)$ 는 네트워크의 출력을 나타낸다. 또한 네트워크 상에서 잡음을 고려하지 않은 동적 특성은 식(1),(2)와 같이 표시할 수 있다.[11]

$$\begin{aligned} x(k+1) &= \varphi(W_x(k)x(k) + W_u(k)u(k)) \\ &= \varphi(W(k)z(k)) \end{aligned} \quad (1)$$

$$y(k) = C x(k+1) \quad (2)$$

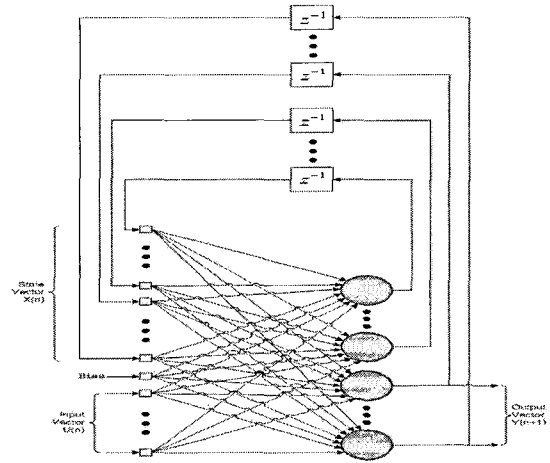


그림1. 완전 결합된 순환신경망의 구조

Fig.1 A layout of fully connected recurrent neural network.

단  $W_x(k)$ 는  $q \times q$  행렬,  $W_u(k)$ 는  $q \times (l+1)$  행렬,  $C$  는  $q \times q$  행렬,  $\varphi$ 는  $q \times q$  대각행렬을 나타낸다. 또한 두개로 분리된 웨이트 행렬은  $q \times (q+l+1)$  개의 통합 웨이트 행렬  $W(k)$ 로 나타낼 수 있다.

$$W(k) = [ W_x(k) \ W_u(k) ] \quad (3)$$

또한  $(q+1) \times 1$  개의 벡터  $z(k)$ 는 식(4)와 같이 상태벡터  $x(k)$ 와 입력벡터  $u(k)$ 로 구성된다.

$$z(k) = \begin{bmatrix} x(k) \\ u(k) \end{bmatrix} \quad (4)$$

또한 입력벡터  $u(k)$ 의 첫 번째 요소는 바이어스로 벡터  $W_u(k)$ 의 첫 번째 요소이다. 상태공간의  $q$ 차원은 결국 시스템의 차수가 되며, 그러므로 그림 1로 표시되는 상태공간모델은 1개의 입력과  $q$ 개의 출력순환 모델의 형태를 갖는다. 식 (1)은 시스템의 프로세서 방정식이며, 식(2)는 추정방정식을 나타낸다. 네트워크의 상태공간을 나타내는 프로세서 방정식(1)은 다음 형태로 표현할 수 있다.

$$x(k+1) = \begin{bmatrix} \varphi(\omega_1^T(k)z(k)) \\ \varphi(\omega_2^T(k)z(k)) \\ \vdots \\ \varphi(\omega_q^T(k)z(k)) \end{bmatrix} \quad (5)$$

단  $\varphi(\cdot)$ 는 활성화함수이며, 웨이트 벡터  $\omega_i(k)$ 는 순환신경망의  $i$ -번째 뉴론에 연결된 웨이트이며, 웨이트 행렬  $W^T(k)$ 의  $i$ -번째 열에 대응 된다.

### 3. 순환신경망에 대한 RTRL 알고리즘

우선 본 절에서는 순환신경망의 학습 알고리즘 중 RTRL 알고리즘에 대하여 간략히 설명하기로 한다. RTRL 알고리즘을 설명하기 위하여 필요한 행렬을 다음과 같이 정의 한다.

- 상태벡터  $x(k)$ 를 웨이트 벡터  $\omega_i$ 에 의해 미분한 행렬.

$$\Lambda_i(k) = \frac{\partial x(k)}{\partial \omega_i(k-1)} \quad (6)$$

- 행렬  $Z(k)$ 는 벡터  $z(k)$ 를 전치하여  $i$ -번째 행으로 나타낸 것을 제외한 모든 요소를 0으로 한  $q \times (q+1)$  행렬.

$$Z_i(k) = \begin{bmatrix} 0^T \\ z^T(k) \\ 0^T \end{bmatrix} \leftarrow i\text{-th row } i=1,2,\dots,q. \quad (7)$$

- $\Phi(k)$ 는  $q \times q$  개의 대각행렬.

$$\Phi(k+1) = \text{diag} \left[ \varphi'(w_1^T(k)z(k)), \varphi'(w_2^T(k)z(k)), \dots, \varphi'(w_q^T(k)z(k)) \right] \quad (8)$$

위의 정의와 함께 뉴론  $i$ 에 대한 순환방정식  $\Lambda_i$ 은  $w_i$ 에 대한 식(5)와 체인 룰(chain rule)을 이용하여 식(9)와 같이 구할 수 있다.

$$\Lambda_i(k+1) = \Phi(k+1) [ W_x(k)\Lambda_i(k) + Z_i(k) ] \quad (9)$$

학습의 목적은 임의의 시간  $t$ 에서 에러의 제곱의 합에 의해 구해지는 비용함수(cost function)를 최소화 하는 것으로 오차  $e(k)$ 에 의해 식(10)과 같이 구한다.

$$\mathcal{J}(k) = \frac{1}{2} e^T(k) e(k) \quad (10)$$

단 측정방정식 식(2)에 의해 정의된  $p \times 1$ 개의 오차벡터  $e(k)$ 는 다음 식(11)과 같다.

$$e(k) = \mathcal{Y}(k) - y(k) \quad (11)$$

여기에서  $\mathcal{Y}(k)$ 는 원하는 출력벡터이다.

또한  $i$ -번째 뉴론의 웨이트 벡터의 변화율  $\Delta w_i(k)$ 는 식(12)와 같이 표시한다.

$$\begin{aligned} \Delta w_i(k) &= \eta \frac{\partial \mathcal{J}(k)}{\partial w_i(k)} \\ &= \eta \Lambda_i(k) e(k), \quad i=1,2,\dots,q. \end{aligned} \quad (12)$$

### 4. 파라미터 추정을 위한 언센티드 칼만필터

지난 30여년 이상 칼만필터는 제어 및 신호처리 영역에서 매우 폭넓게 필터링에 사용되어져 왔다. 특히 확장된 칼만필터(Extended Kalman Filter, EKF)는 모든 비선형 시스템을 선형화하여 사용 할 수 있는 방법으로 매우 널리 이용되어지고 있다. 그러나 EKF는 시스템을 구현하고 동조하는데 매우 까다롭고 갱신기간동안 시간 스케일 상에서 거의 선형인 시스템으로 제한되어지는 문제점을 앓고 있다.[9]

신경회로망의 파라미터 추정에 있어서, EKF는 비선형 시스템의 선형화를 통하여 최적 비선형 추정에 대한 1차 근사화를 제공한다. 이 근사화는 변환된 가우시안 랜덤변수의 사후(posterior) 평균 및 공분산을 구할 때 매우 큰 오차를 포함 할 수도 있으며, 이들은 때때로 필터 분산과 준 최적 평가 지수를 유도할 수 도 있다.[12] 그런데 Julier와 Uhlmann[10]에 의해서 제안되고 Wan 등에 의해서 더욱 발전된 언센티드 칼만필터(Unscented Kalman Filter, UKF)는 이러한 EKF 알고리즘의 단점을 해결할 수 있는 새로운 대안이 될 수 있다.[12],[13] UKF는 가우시안 분포에 대하여 프로세서 오차 및 측정오차에 대하여 Taylor 시리즈 확장에 의해 2차 근사화를 제공하므로, EKF 보다 상태 추정 및 파라미터 추정 시에 우수한 성능을 보였다.[14] 또한 UKF는 EKF 알고리즘 구현시 프로세서 방정식 및 측정방정식을 선형화 할 때 필요한 Jacobian 행렬 계산을 하지 않는데, 이점은 역행렬 계산을 피할 수 있는 장점이 있다.[9] UKF 알고리즘 구현은 언센티드 변환(Unscented Transformation, UT)으로부터 시작한다.

#### ● 언센티드 변환(UT)

언센티드 변환은 비선형 변환을 이용하여 랜덤변수의 통계적 특성을 계산하는 방법으로 UKF 알고리즘의 근간을 이룬다.[10] 비선형 변환  $y=f(x)$ 를 통하여 랜덤변수  $y$ 를 만들어 내는  $L \times 1$ 개의 랜덤변수  $x$ 를 고려하자.  $y$ 의 통계적 특성을 계산하기 위하여  $2L \times 1$  개의 시그마 벡터  $x_i$ 의 행렬  $x$ 는 다음과 같이 표현된다.

$$x_i = \begin{cases} x & i=0 \\ x + (\sqrt{(L+\lambda)P_{xx}})_i & i=1, \dots, L \\ x - (\sqrt{(L+\lambda)P_{xx}})_{i-L} & i=L+1, \dots, 2L \end{cases} \quad (13)$$

여기에서  $(\sqrt{(L+\lambda)P_{xx}})_i$ 는 square root 행렬(즉, lower-triangular Cholesky factorization)의  $i$ -번째 열이고,  $\bar{x}$ 와  $P_{xx}$ 는  $x$ 의 평균(mean)과 공분산(covariance)으로 다음과 같이 정의 된다.

$$\bar{x} = E[x] \quad (14)$$

$$P_{xx} = E[(x - \bar{x})(x - \bar{x})^T] \quad (15)$$

여기에서  $E(\cdot)$ 는 기대치 연산자이다. 또한  $\lambda$ 는  $\lambda = \alpha^2(L+x) - L$ 로 정의되는 환산계수이다. 상수  $\alpha(0.001 < \alpha < 1)$ 는  $\bar{x}$ 근처의 시그마 포인트의 퍼짐(spreading) 정도를 결정한다. 상수  $x$ 역시 환산계수이며 보통  $3-L$ 로 정해진다.  $x$ 는 근사화의 고차 모멘트를 정교하게 조정하기 위한 자유도를 제공하며, 예측오차를 줄이는데 이용된다. 랜덤변수  $x$ 가 가우시안 분포를 갖는다면 경험지식을 바탕으로  $L+x=3$ 으로 정하는 것이다. 만일  $x$ 가 다른 분포를 갖는

다면  $x$ 는 다른 값으로 정해져야 한다. 시그마 포인트  $\{x_i\}_{i=0}^{2L}$ 는 다음의 비선형 함수를 통해서 전달된다.

$$y_i = f(x_i), \quad i=0, 1, \dots, 2L \quad (16)$$

이 식은 비선형 변환벡터  $y$ 의 평균과 공분산 행렬을 추정하는데 이용되는 벡터집합을 만들어 낸다.  $y$ 의 평균과 공분산 행렬은 사후 시그마 포인트의 가중 샘플 평균 및 공분산을 이용하여 다음과 같이 근사화 된다.[12]

$$\bar{y} = \sum_{i=0}^{2L} W_i^m y_i \quad (17)$$

$$P_{yy} = \sum_{i=0}^{2L} W_i^c (y_i - \bar{y})(y_i - \bar{y})^T \quad (18)$$

단, 웨이트 계수는 식(19)와 같이 주어진다.

$$\begin{aligned} W_0^m &= \frac{\lambda}{L+\lambda} \\ W_0^c &= \frac{\lambda}{L+\lambda} + (1-\alpha^2 + \beta) \\ W_i^m &= W_i^c = \frac{1}{2(L+\lambda)}, \quad i=1, 2, \dots, 2L. \end{aligned} \quad (19)$$

위식에서 위첨자  $m$ 과  $c$ 는 각각 평균과 공분산을 의미한다.  $\beta$ 는  $x$ 의 분포에 관한 사전지식을 고려하기 위하여 이용되며, 일반적으로  $\beta=2$ 가 가우시안 분포에 대하여 최적의 선택 값이다.

### B. UKF를 이용한 온라인 파라미터 추정

순환신경망의 학습에 칼만필터를 적용하기 위하여, 네트워크의 동특성을 비선형 이산시간 시스템으로 표현하면 식 (20), (21)과 같다.

$$w(k+1) = w(k) + \omega(k) \quad (20)$$

$$y_d(k) = h(w(k), r(k), u(k), x(k+1)) + v(k) \quad (21)$$

단, 모든 벡터와 행렬은 복소수이며, 비선형함수  $h(\cdot)$ , 웨이트 벡터  $w(k)$ 는 식(22),(23)과 같이 정의된다.

$$h(w(k), z(k)) = C \varphi(w(k) z(k)) \quad (22)$$

$$w(k) = \begin{bmatrix} w_1(k) \\ w_2(k) \\ \vdots \\ w_q(k) \end{bmatrix} \quad (23)$$

여기에서  $w_i(k)(i=1, 2, \dots, q)$ 는 전환된 웨이트 행렬  $w^T(k)$ 의  $i$ -번째 열을 의미한다. 프로세서 방정식으로 알려진 식(20)은 프로세서 잡음  $\omega(k)$ 에 의해서 오염된 정적 프로세스(stationary process)로 특성화되는 RNN의 상태를 기술한다. 추정방정식으로 알려진 식(21)은 연결강도벡터  $w(k)$ 의 비선형 함수  $\varphi(\cdot)$ , 입력벡터  $u(k)$ 와 순환노드 활성화 벡터  $x(k)$ 로 구성된 벡터  $z(k)$  및 불규칙 측정잡음  $v(k)$ 의 함수로 표시되는 네트워크의 출력  $y_d(k)$ 을 나타낸다. 프로세서 잡음  $\omega(k)$ 는 전형적으로  $E[w_i w_j^T] = \delta_{ij} Q(k)$ 에 의해서 주어진 제로-평균과 백색 잡음 공분산으로 특징 지워진다. 이와 비슷하게 측정 잡음  $v(k)$  또한  $E[v_i v_j^T] = \delta_{ij} Q(k)$ 에(여기에서  $\delta$ 는 Kronecker delta 임) 의해서 주어진 제로-평균

과 백색잡음 공분산으로 특징 지워진다. 식(20),(21)로 주어진 RNN의 상태공간 모델로부터 최소화되는 MSE 비용함수는 식 (24)와 같다.

$$\tau(w) = e(k)^T R^{-1}(k)e(k) \quad (24)$$

이때 오차벡터  $e(k)$ 는 식(11)에서 주어졌다. 만일 측정잡음 공분산 행렬  $R(k)$ 가 상수 대각행렬이라면 임의로 정할 수 있고, 프로세서 잡음 공분산 행렬  $Q(k) = E[w(k)w^T(k)]$ 는 수렴속도와 추적성능에 영향을 미친다. 측정잡음 공분산 행렬  $R(k)$ 와 프로세서 잡음 공분산 행렬  $Q(k)$ 는 식(25),(26)과 같이 정의 한다.

$$R(k) = \mu^{-1}I \quad (25)$$

$$Q(k) = (\lambda^{-1} - 1)P(k) \quad (26)$$

여기에서  $\lambda \in (0, 1]$ 는 RLS(recursive least-squares) 알고리즘에서 forgetting 계수로 알려져 있다.[14] UKF 알고리즘은 Jacobian 행렬 계산을 할 때 미분연산을 수행하지 않고 시그마 포인트 전달을 통하여 효과적으로 평가할 수 있다. UKF 알고리즘을 이용한 RNN 연산과정은 다음과 같이 분류할 수 있다.

#### • 연결강도 벡터 및 공분산행렬의 초기화:

$$\hat{w}(0) = E[w] \quad (27)$$

$$P(0) = E[(w - \hat{w}(0))(w - \hat{w}(0))^T] \quad (28)$$

#### • 시그마 포인트 연산과정:

$$\Gamma(k) = (L + \lambda)(P(k) + Q(k)) \quad (29)$$

$$\mathcal{W}(k) = \{\hat{w}(k), \hat{w}(k) + \sqrt{\Gamma(k)}, \hat{w}(k) - \sqrt{\Gamma(k)}\} \quad (30)$$

$$D(k) = h(\mathcal{W}(k), r(k), u(k), x(k-1)) \quad (31)$$

$$\hat{\gamma}(k) = h(\hat{w}(k), r(k), u(k), x(k-1)) \quad (32)$$

#### • 추정방정식의 갱신과정:

$$P_{yy}(k) = \sum_{i=0}^{2L} W_i^c (D_i(k) - \hat{\gamma}(k))(D_i(k) - \hat{\gamma}(k))^T + R(k) \quad (33)$$

$$P_{wy}(k) = \sum_{i=0}^{2L} W_i^c (W_i(k) - \hat{w}(k)) \times (W_i(k) - \hat{w}(k))^T \quad (34)$$

$$\gamma(k) = P_{wy}(k)P_{yy}^{-1}(k) \quad (35)$$

$$\hat{w}(k+1) = \hat{w}(k) + \gamma(k)e(k) \quad (36)$$

$$P(k+1) = P(k) - \gamma(k)P_{yy}(k)\gamma^T(k) \quad (37)$$

순환신경망의 웨이트 벡터는 위에 기술한 방정식을 통하여 온라인으로 업데이트 된다.

## 5. 비선형 순환신경망 채널 등화기

### A. 통신 시스템 모델링

그림 2. 는 결정피드백 등화기(decision feedback equalize

r, DFE)를 갖는 디지털 통신시스템의 일반적인 모델을 나타낸 것으로 선형과 비선형 왜곡 상태를 모두 포함하고 있다. 정보 소스로부터 추출된 시퀀스(s(k))가 전송되고, 전송된 심볼들은 부가백색잡음(additive white Gaussian noise, AWGN)에 의해서 오염된다. 이와 같은 채널 모델은 다음과 같이 표현된다.

$$r(k) = g(\hat{r}(k)) + v(k) = g\left(\sum_{i=0}^{N-1} h_i s(k-i)\right) + v(k) \quad (38)$$

여기에서 g(·)는 비선형 왜곡을 나타내며, h는 길이가 N인 채널의 유한 임펄스 응답(finite impulse response, FIR)이다. 그리고 v(k)는 평균이 0이고 분산이  $\sigma_0^2$ 인 AWGN이다. DFE는 세 개의 변수에 의해서 결정된다. 즉 전향차수 m, 되먹임 차수 n, 그리고 결정지연 d이다. 그러므로 DFE로 입력은 전향입력  $r(k)=[r(k), r(k-1), \dots, r(k-m+1)]^T$ 과 되먹임 입력  $u(k)=[u(k-1), \dots, u(k-n)]^T$ 로 구성된다. DFE의 출력은 y(k)이고, 이 출력은 추정 심볼  $\hat{s}(k-d)$ 를 결정하기 위해서 결정장치(decision device)를 통과한다. 되먹임 차수 n은 시간 k에서 등화기의 결정에 공헌하는 전송된 심볼들이 전향차수  $m=d+1$ 에 대하여  $s(k)=[s(k), s(k-1), \dots, s(k-m-N+2)]^T$ 로 주어지므로 식(39)와 같이 된다.[15],[16]

$$n = N + m - d + 2 \quad (39)$$

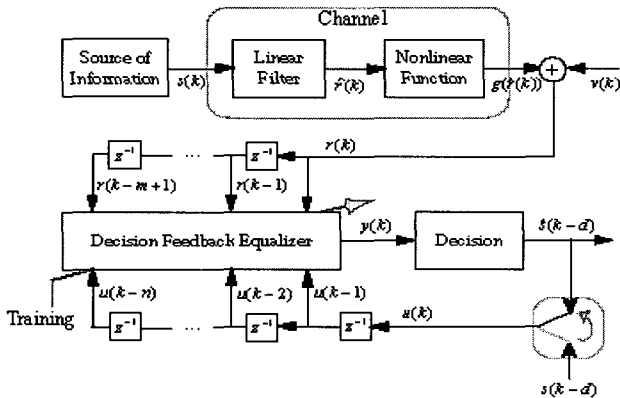


그림2. DFE로 구성된 디지털 통신 시스템  
Fig2. A digital communications system with decision feedback equalizer

다음 실험에서는 완전 결합된 RNN을 사용한 결정되먹임 순환 신경망(decision-feedback recurrent neural equalizer, DFRNE)을 DFE로 사용하였는데, 이 경우에 입력벡터 u(k)는 채널입력신호와 결정되먹임 입력 그리고 바이어스 입력으로 구성된다.

**B. 수렴속도에 대한 시뮬레이션**

채널 모델 1: 비 최소 위상의 선형 채널 모델의 전달 함수를 다음 식과 같이 설정한다.

$$H(z) = 0.3482 + 0.8704z^{-1} + 0.3482z^{-2}$$

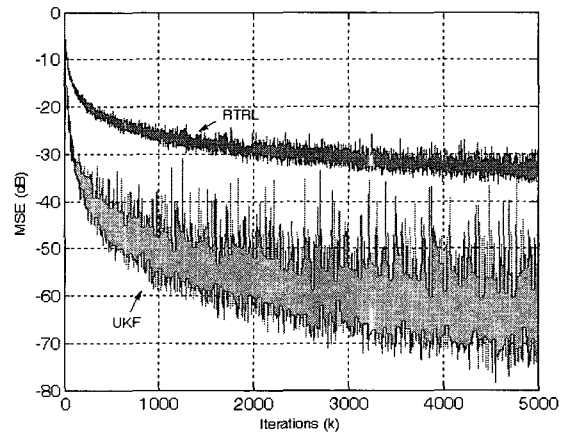
위식은 문헌 [2][3][17]에서 일반적으로 채널 등화기의 모

델링에 사용 되어진 식이다. 비선형 채널은 다음 식과 같이 모델링 되었는데, 여기에서 비선형성은 그림에서와 같이 선형 채널의 출력에 적용되었다.

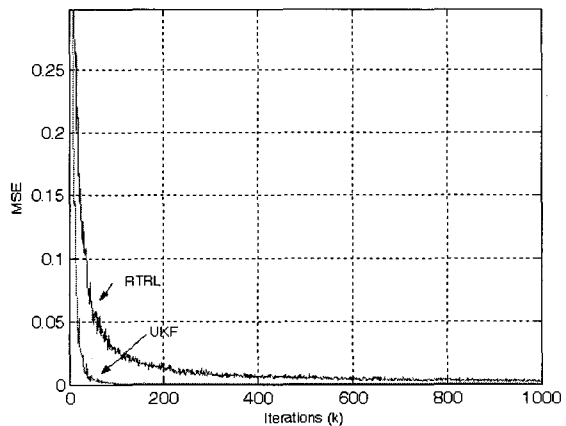
$$r(k) = \tanh(\hat{r}(k)) + v(k)$$

이러한 채널의 비선형 왜곡은 전송 시스템에서 기인하는 포화현상을 나타낼 수도 있다. RTRL의 학습 비율은 경험적으로  $\eta=0.1$ 로 선정하였으며, 이 값은 안정한 수렴을 보장하는 값이다. UKF의 파라미터 들은  $\alpha=0.1, \lambda=0.99$ 이며, 결정 지연은  $d=2$ 이다.

시뮬레이션에 의한 DFRNE의 수렴 특성은 그림 3에 보였는데 MSE의 선형 스케일과 로그 스케일로 나타내었다. 이 결과는 200회의 독립적인 학습에 대한 평균값을 구한 것이며, 매 실험마다 모든 등화기에 대하여 다른 BPSK 랜덤 시퀀스와 랜덤 초기화 연결 강도를 적용하여 14[dB]의 SNR에서 수행하였다.



(a) y-축 :Log 스케일 (10log(MSE))



(b) x-축 : 선형 스케일

그림3. 채널 모델1의 수렴특성(SNR=14dB)

Fig. 3. Convergence properties for Channel Model 1 under SNR = 14dB

그림3에서 알 수 있듯이 UKF의 수렴특성은 RTRL에 비하여 매우 빠른 속도로 수렴함을 알 수 있었다. 예를 들어 그림 3(a)에서와 같이 UKF의 MSE 값은 대략  $10^3$  개의 심볼을 학습한 후에 -50dB에 도달 하였으나, 이에 반하여 RTRL의 MSE 값은 -27dB에 도달 함을 알 수 있었으며, 또한

그림3(b)에서 알 수 있듯이, 약 100회의 학습 후에 정상상태에 도달함을 알 수 있었다. 이러한 결과를 통하여 UKF 알고리즘을 이용한 학습방법이 수렴속도와 정상상태 때의 MS E 모두 향상되었음을 확인할 수 있었다.

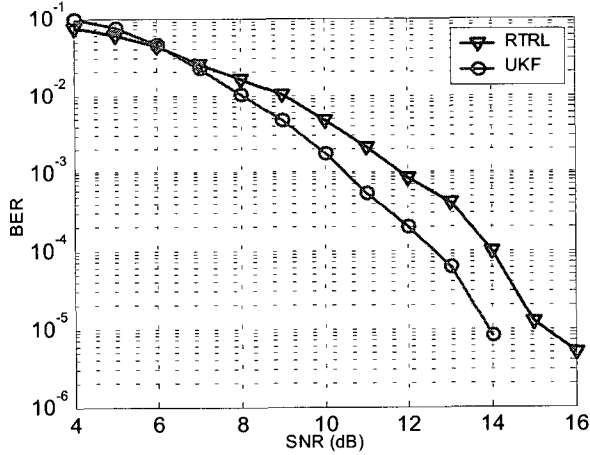


그림4. 100개의 훈련 심볼을 이용한 채널모델 1에 대한 등화기의 BER 성능

Fig.4 BER performance for Channel Mode 1 using 100 training symbols.

채널모델에 대한 비트에러율(Bit Error Rate, BER) 성능을 그림 4에 비교 하였다. 실험은 100회를 독립적으로 실행하여 평균을 구한 값이며, 매회 당 처음 100개의 심볼은 학습에 사용하였고, 나머지  $10^4$  개의 심볼은 테스트에 사용하였다. 실험에서 모든 등화기의 연결강도 벡터는 학습 심볼을 이용하여 학습과정에서 갱신되며, BER 계산 시 즉, 테스트 모드 동안에는 갱신되지 않고 학습과정에서 훈련된 연결강도 벡터를 사용한다. 그리고 전송된 심볼은 직접 결정모드일 때 계산되었다. 그림에서 알 수 있듯이 UKF 알고리즘에 의한 경우가 RTRL을 이용한 경우에 비하여  $10^{-4}$  의 BER 지점에서 약 1.3dB 가 향상되었음을 보여준다. 만일 RTRL 알고리즘을 이용하여 UKF에 의한 성능에 도달하기 위해서는 약 200개 이상의 심볼이 필요 함 을알 수 있었다.

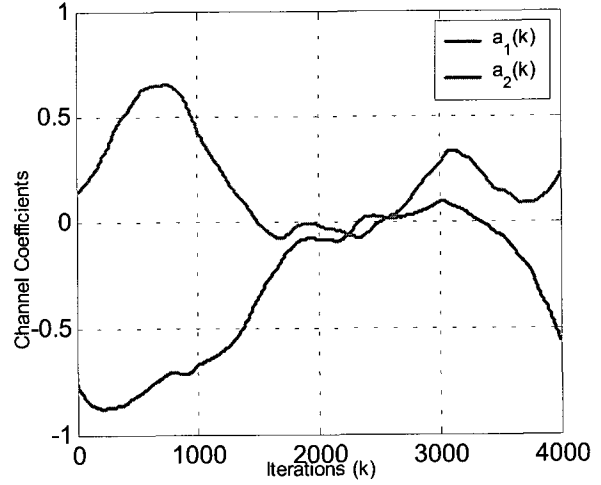
### c. 추적 성능에 대한 시뮬레이션

다음으로 DFRNE 에 대한 채널 추적 성능을 실험하였는데 왜냐하면 수렴속도는 과도상태 특성을 파악하는 것인데 비하여 채널 추적은 정상상태 특성과 연관되기 때문이다.[14] 채널 모델 2. :시변 이산시간 비선형 채널 모델의 전달함수를 다음과 같이 설정하였다.

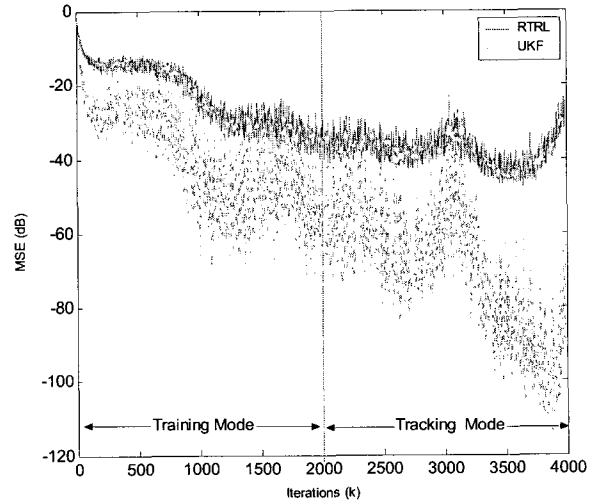
$$H_2(z) = 1.0 + a_1(k)z^{-1} + a_2(k)z^{-2}$$

이때 채널모델 1에서 설정한 비선형 왜곡을 똑같이 설정하였는데, 이 채널 모델에서 시변계수  $a_i(k)(i=1,2)$ 는 시간  $k$  에 따라 변화하는 비선형 시변 채널이다. 이러한 시변계수는 백색 가우시안 잡음 소스와 Butterworth 필터 응답의 응용에 의해서 생성된다. Butterworth 필터의 대역폭은 채널의 상대적 대역폭(페이딩률)을 결정한다. 공칭 2[kHz]채널, 2400(symbols/sec) 샘플링률 및 0.5[Hz]의 대역폭을 갖는 2차 Butterworth 필터라 가정하였다.[18] 파라미터는 채널모델 1에서 사용된 값과 동일한 값들로 설정하였다. 그림

5(a)는 페이딩률 0.5Hz 일때 시변계수  $a_1(k)$ 와  $a_2(k)$ 의 생성 예를 보여 준다.



(a) 시변계수의 변화(0.5 Hz)



(b)SNR=15dB 일때의 수렴상태

그림 5. 채널모델 2 일때의 채널 추적 정도

Fig. 5 Channel tracking capability for Channel Model 2

DFRNE에서  $k=2000$ 까지는 훈련 모드 이고,  $k=2001$ 부터는 추적 모드로 전환된다. 채널모델 1에서의 시뮬레이션과는 달리, DFRNE에서 채널의 페이딩 특성을 추적하기 위하여 추적 모드 상태에서도 그들의 웨이트 벡터를 업데이트 한다. 그림 5(b)에서 채널 추적 특성은 SNR = 15 인 상태에서 훈련모드와 추적모드를 보여준다. 예상했던 바와 같이 UKF는 RTRL에 비하여 더욱 빠른 채널 추적 성능을 보여 주고 있다. 이러한 결과는 UKF의 MSE가 RTRL의 것과 비교하여 훈련모드 에서나 추적모드 모두다 더욱 작아 짐을 알 수 있다.

### D. 연산 복잡도에 관한 비교

연산 복잡도는 등화기의 상태(q) 와 연결강도(L)갯수를 이용하여 표현할 수 있다. DFRNE에서 이용된 RTRL의 연산시간은 차수  $O(L+q)$ 에 따라 증가하고, 제한한 등화기에 이용된 UKF알고리즘은  $O(L^3)$ 에 따라 증가한다.[6],[22] U

KF 알고리즘은 RTRL 보다 연산시간이 많이 소요될 지라도 수렴속도, MSE 및 BER 성능 면에서 우수하다. 한편, 네트워크 크기가 증가함에 따라 전송된 심볼을 훈련하는데 필요한 계산비용은 약간 증가하게 된다. 하지만 제안한 UKF 알고리즘은 단지 적은 수의 뉴런을 사용할 뿐만 아니라 훈련 시에 상대적으로 적은 훈련 심볼을 필요로 한다.

## 6. 결 론

본 논문에서는 UKF로 불리는 미분없는 Kalman Filters를 이용하여 순환 신경망에서 실시간으로 파라미터 추정하는 방법을 제시 하였으며, 이에 대한 성능의 우수함을 ISI와 비선형 왜곡이 포함되어 있는 통신채널 등화기 실험을 통하여 입증하였다. 수렴속도의 관점에서, UKF는 훈련과정에서 수렴비율이 RTRL 알고리즘을 이용할 때에 비하여 월등히 우수함을 보였다. 또한 RTRL 기법에 비하여 칼만 필터 알고리즘의 빠른 수렴 특성으로 인하여 만족스러운 성능을 얻으면서도 상대적으로 매우 적은 훈련 심볼을 필요로 하였고, BER 성능을 더욱 향상 시켰음을 알 수 있었다. 결과적으로 제안한 알고리즘은 고속채널 등화가 요구되는 디지털 통신시스템에 적합할 것으로 판단된다. 그러나 UKF가 지니고 있는 이러한 장점에도 불구하고 해결해야 할 몇 가지 문제점을 안고 있다. 그것은 계산비용의 측면에서 UKF 알고리즘은 상대적으로 매우 비싼 편으로 UKF에 대한 계산비용의 감소는 앞으로의 연구과제이며, 또한 UKF는 RTRL에 비하여 초기치 선정이 매우 민감한 편으로 UKF의 파라미터 선정에 대한 구조적인 방법은 아직도 많은 연구자들의 연구를 필요로 한다.

## 참 고 문 헌

- [1] A. F. Atiya and A. G. Parlos, "New results on recurrent network training: Unifying the algorithms and accelerating convergence," *IEEE Transactions on Neural Networks*, vol.11, pp.697-709, May 2000.
- [2] G. Kechriotis, E. Zervas, and E. S. Manolakos, "Using recurrent neural networks for adaptive communication channel equalization," *IEEE Transaction on Neural Networks* vol, 5, pp. 267-278, March 1994.
- [3] R. Parisi, E. D. Di Claudio, G. Orlandi, and B.D.Rao, "Fast adaptive digital equalization by recurrent neural networks," *IEEE Transactions on Signal Processing*, vol.45, pp. 2731- 2739, November 1997.
- [4] S. Ong, C. You, S. Choi, and D. Hong, " A decision feedback recurrent neural equalizer as an infinite impulse response filter," *IEEE Transaction on Signal Processing*, vol. 45, pp. 2851- 2858, November 1997.
- [5] H. R. Jiang and K. S. Kwak, "On modified complex recurrent neural network adaptive equalizer," *Journal of Circuits, Systems, and Computers*, vol. 11, no. 1, pp. 93-101, 2002.
- [6] B. Hammer and J. J. Steil, "Tutorial: Perspective on learning with RNNs," in *Proc. of the European Symposium on Artificial Neural Networks(ESANN)*, pp.357-369, 2002.
- [7] R. J. Williams and D. Zipser, "A learning algorithms for continually running fully recurrent neural networks," *Neural Computat -ion*, vol.1, pp.270-280, 1989.
- [8] P. J. Werbos, "Back-propagation through time: What it does and how to do it," *Proceedings of the IEEE*, vol. 78, pp.1550-1560, October 1990.
- [9] S. Julier, J. Uhlmann, and H. F. Durrant- Whyte, " A new method for the nonlinear transformation of means and covariances in filters and estimators," *IEEE Transaction on Auto matic Control*, vol. 45, pp. 477-482, March 2000.
- [10] S. J. Julier and J. K. Uhlmann, " A new extension of the Kalman filter to nonlinear systems," in *Proceeding of AeroSense: The 11th International Symposium on Aerospace /Defence Sensing, Simulation and Controls*, 1997.
- [11] S. Haykin, *Nerual Networks: a Comprehensive -ve Foundation*, 2nd Ed. Upper Saddle River, NJ: Prentice Hall, 1999.
- [12] E. A. Wan and R. van der Merwe, "The unscented Kalman filter," in *kalman filtering and neural networks*, Edited by S. Haykin. John Wiley and Sons, Inc., 2001.
- [13] E. A. Wan and R. van der Merwe, "The unscented Kalman filter for nonlinear estimation," in *Proceeding of the IEEE 2000 daptive Systems for Signal Processing, Communications and Control Symposium(AS -SPCC)*, pp. 153-158, 2000.
- [14] S. Haykin, *Adaptive Filter Theory*, 4th Ed. Upper Saddle River, NJ: Prentice Hall, 2002.
- [15] S. Chen. B. Mulgrew, and S. McLaughlin, "Adaptive Bayesian equalizer with decision feedback," *IEEE Transactions on Signal Process -ing*, vol.41, pp. 2918-2927, September 1993.
- [16] M. Solazzi, A. Uncini, E. D. Di Claudio, and R.Parisi, "Complex discriminative learning Bayesian neural equalizer," *Signal Processing*, vol. 81, pp. 2493-2502, 2001.
- [17] S. Chen, G. J. Gibson, B. Mulgrew, and S. McLaughlin, "Adaptive equalization of finite non-linear channels using multilayer perceptrons," *Signal Processing*, vol. 20, pp. 107-119, 1990.
- [18] C.Cowan and Semnani, "Time-variant equalization using a novel non-linear adaptive structure," *International Journal of Adaptive Control and Signal Processing*, vol. 12, no.2, pp. 195-206, 1998.

저 자 소 개



**권오신(Ohshin Kwon)**

1978 전북대학교 전기공학과(학사)

1981 전북대학교 전기공학과(공학석사)

1988 전북대학교 전기공학과(공학박사)

1988-2005년 현재 군산대학교 전자정보공  
학부 교수

2003-2004 University of Ottawa

방문 교수

관심분야 : 신경회로망, 적응필터, 자동 제어

Phone : 063 - 469 - 4702

Fax : 063 - 469 - 4699

E-mail : kos@kunsan.ac.kr