

# 시그마포인트 칼만필터를 이용한 순환신경망 학습 및 채널 등화

## A Recurrent Neural Network Training and Equalization of Channels using Sigma-point Kalman Filter

권 오 신 \*  
Ohshin Kwon

**Abstract** - This paper presents decision feedback equalizers using a recurrent neural network trained algorithm using extended Kalman filter(EKF) and sigma-point Kalman filter(SPKF). EKF is propagated analytically through the first-order linearization of the nonlinear system. This can introduce large errors in the true posterior mean and covariance of the Gaussian random variable. The SPKF addresses this problem by using a deterministic sampling approach. The features of the proposed recurrent neural equalizer And we investigate the bit error rate(BER) between EKF and SPKF.

**Key Words** :Sigma-Point Kalman Filter, Extended Kalman Filter, Recurrent Neural Network, Channel Equalization, .

### 1. 장 서론

본 논문에서는 순환신경망 등화기의 채널에 대한 빠른 수렴속도와 우수한 추적성능을 지니는 두 종류의 훈련 알고리즘을 제안한 후 그 특성을 비교한다. 두 종류의 훈련 알고리즘으로 확장된 칼만필터(extended Kalman filter, EKF)와 시그마 포인트를 이용한 칼만필터(sigma point Kalman filter, SPKF)를 적용한다. 먼저 EKF는 신경회로망의 훈련에 자주 이용되며 빠른 수렴특성을 보인다. EKF는 신경회로망 훈련 시 Jacobian 행렬 연산이 필요하며 비선형 시스템의 선형화에 따른 모델링 오차에 의해 필터의 성능이 저하되는 단점을 지니고 있다. 이것은 EKF가 상태변수의 추정치를 명목점(nominal point)으로 사용함으로써 필터의 초기오차 또는 추정 오차를 유발 할 수 있기 때문이다. 또 다른 훈련 알고리즘으로 최근에 발표된 SPKF 알고리즘을 들 수 있는데 이 방법은 아직 순환신경망에 적용하는 방법이 그리 많이 발표되지 않았다. EKF는 때로는 필터의 준최적(sub-optimal) 성능을 갖게 하기도 하며, 초기 추정오차가 큰 경우에는 EKF 구현에 필요한 Jacobian 행렬에 심각한 문제점이 발생하므로 필터는 발산 할 수도 있는 단점을 지니고 있다. 반면에 SPKF에서는 Jacobian 행렬 계산이 필요 없으므로 큰 초기추정오차에 강인한(robust) 특성을 지니고 있다. 본 논문에서는 순환신경망의 훈련 알고리즘으로 EKF와 SPKF를 적용하고 채널에 대한 등화 실험을 통하여 두 알고리즘의 성능을 비교 검토한다.

### 2. 장 확장 칼만필터 및 훈련 알고리즘

RNE의 훈련 알고리즘은 비선형 시스템에 대한 파라미터 추정 문제로 다룰 수 있다. 비선형 시스템 방정식을 선형화하고 선형화된 시스템에 칼만필터를 적용함으로써 구현되는 확장된 칼만필터의 특징은 비선형 동적 시스템의 상태에 대한 최소변위 추

정기(minimum-variance estimator)라는 점이다. 따라서 EKF 가 신경망 훈련에 이용되는 경우 빠른 수렴 속도를 얻을 수 있다. EKF 구현에 적합한 RNE의 동특성은 다음의 비선형 이산시간 방정식과 같다.

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

$$y_d(k) = h(w(k), z(k)) + \nu(k) \quad (2)$$

윗 식에서  $z(k)$ 는 RNE의 입력들의 집합으로 다음 식(3)과 같이 표시된다.

$$z(k) = (r(k), u(k), x(k-1), x(k)) \quad (3)$$

연결강도 벡터  $w$ 는 식 (1)과 (2)에 주어진 연결강도를 이용하여 다음과 같이 재구성되었다.

$$w = [w_1^T w_2^T \dots w_p^T w_0^T]^T \quad (4)$$

EKF를 이용한 RNE 훈련은 측정 데이터를 이용하여 오차  $e(k) = y_d(k) - y(k)$ 의 평균 제곱 오차(mean-squared error, MSE)를 최소화하는 연결강도 벡터  $w$ 를 찾는 것이다. RNE를 훈련하기 위한 EKF 계산 알고리즘은 다음과 같이 주어진다.

$$A(k) = [R(k) + H^T(k)P(k)H(k)]^T \quad (5)$$

$$K(k) = [R(k)H(k)A(k)] \quad (6)$$

$$\hat{w}(k+1) = \hat{w}(k) + K(k)e(k) \quad (7)$$

$$P(k+1) = P(k) - K(k)H^T(k)P(k) + Q(k) \quad (8)$$

벡터  $\hat{w}(k)$ 는 시간  $k$ 에서 시스템의 상태(즉, RNE의 연결강도)의 추정치를 나타낸다. 이 추정치는 칼만 이득 행렬  $K(k)$ 와 오차  $e(k)$ 의 함수이다.  $K(k)$ 는 오차 공분산 행렬  $P(k)$ , 모든 연결강도에 대한 RNE 출력의 편미분 행렬 (즉, Jacobian 행렬)  $H(k)$  및 환산 행렬  $A(k)$ 의 함수이다. Jacobian 행렬  $H(k)$ 는 RNN에 대하여 다음 식과 같이 정의 된다.

저자 소개

\* 권오신: 군산대학교 공과대학 전자정보공학부 교수

$$H(k) = \frac{\partial h(k, \hat{w}, z)}{\partial w(k)} = [\nabla_w h_1(k) \nabla_w h_2(k) \cdots \nabla_w h_L(k)] \quad (9)$$

이때  $i$ -th 미분벡터는 다음 식(10)과 같이 정의한다.

$$\nabla_w h_i(k) = \left[ \frac{\partial h_i(k)}{\partial w_1} \quad \frac{\partial h_i(k)}{\partial w_2} \quad \cdots \quad \frac{\partial h_i(k)}{\partial w_L} \right]^T \quad (10)$$

환산행렬  $A(k)$ 는 측정잡음 공분산 행렬  $R(k)$ , Jacobian 행렬  $H(k)$  및 오차 공분산 행렬  $P(k)$ 의 함수이다. 연결강도 벡터의 추정치 갱신에 공헌하는 행렬  $P(k)$ 는 프로세스잡음 공분산 행렬  $Q(k)$ 에 의해서 증폭된다.

### 3.장 시그마 포인트 기반 칼만필터

시그마 포인트는 이전시간(prior) 확률 분포함수의 평균과 공분산 정보를 대표하고 있으며 비선형 함수를 통해 직접 시간 전파된 시그마 포인트를 사용하여 이후시간(posterior) 평균과 공분산을 계산한다. 이렇게 계산된 평균과 공분산은 확장된 칼만필터에서 선형 근사화된 식을 사용하여 시간 전파된 평균과 공분산과는 달리 작은 오차를 갖는다. UT는 비선형 변환을 통한 랜덤 변수의 통계적 특성을 계산하는 방법으로 시그마 포인트 칼만필터 알고리즘 기본을 이룬다. 우선 이산시간 비선형 함수를 다음과 같이 정의한다.

$$x_{k+1} = f(x_k) + w_k, \quad w_k \sim N(0, Q) \quad (11)$$

$$y_k = h(x_k) + v_k, \quad v_k \sim n(0, R) \quad (12)$$

상태변수와 공분산은 가우시안으로 정의된 잡음의 적용을 위해 다음과 같이 확장된다.

$$\hat{x}_0^a = E[x_0^T \quad w_0^T \quad v_0^T] = [\hat{x}_0^T \quad 0 \quad 0] \quad (13)$$

$$P_0^a = E[(x_0 - \hat{x}_0^T)(x_0 - \hat{x}_0^T)^T] = \begin{bmatrix} P_0 & 0 & 0 \\ 0 & Q & 0 \\ 0 & 0 & R \end{bmatrix} \quad (14)$$

그리고 마지막 초기화로 시그마 포인트의 가중치를 위해 다음과 같이 필터 구동 전에 미리 구성한다.

$$W_0^{(m)} = \lambda / (L + \lambda) \quad (15)$$

$$W_0^{(c)} = \lambda / (L + \lambda) + (1 - \alpha^2 + \beta) \quad (16)$$

$$W_i^{(m)} = W_i^{(c)} = 1/2(L + \lambda), \quad i = 1, \dots, 2L \quad (17)$$

여기서  $\lambda = (\alpha^2 - 1)L$ 는 조정 매개변수이며,  $\alpha$ 는 추정된 평균  $\hat{x}_0$  주위의 시그마 포인트들의 분포 범위를 의미하며 ( $1 \leq \alpha \leq 1e^{-3}$ ),  $\beta$ 는 상태변수 확률 분포의 선행 지식을 의미하며 가우시안 분포인 경우 2가 최적이다.

SPKF에서는 시그마 포인트를 설정하는 것이 중요하며 다음과 같이 설정함으로써 재귀적 필터의 구현이 시작된다.

$$x_{k-1}^a = \{ \hat{x}_{k-1}^a \pm \sqrt{(L + \lambda)P_{k-1}^a} \} \quad (18)$$

그 다음 설정된 시그마 포인트는 다음과 같이 비선형 시스템 함수를 통하여 독립적으로 시간 전파 된다.

$$X_{k|k-1}^x = f(x_{k-1}^x) + \chi_{k-1}^x \quad (19)$$

이렇게 계산된 시간 전파된 시그마 포인트를 사용하여 다음과 같이 상태변수, 공분산, 그리고 추정치의 추정치가 계산된다.

$$\hat{x}_k = \sum_{i=0}^{2L} W_i^{(m)} \chi_{i,k|k-1}^x \quad (20)$$

$$P_k = \sum_{i=0}^{2L} W_i^{(c)} [\chi_{i,k|k-1}^x - \hat{x}_k | \chi_{i,k|k-1}^x - \hat{x}_k]^T \quad (21)$$

$$\hat{y}_k = \sum_{i=0}^{2L} W_i^{(m)} [h(\chi_{i,k|k-1}^x) + \chi_{k-1}^v] \quad (22)$$

관측 값이 들어오면 다음과 같이 필터 이득 (gain)을 계산한다.

$$K_k = P_{x|y_k} P_{y_k}^{-1} \quad (23)$$

여기서 공분산  $P_{x|y_k}, P_{y_k}$ 는 다음 식과 같다.

$$P_{x|y_k} = \sum_{i=0}^{2L} W_i^{(c)} [\chi_{i,k|k-1}^x - \hat{x}_k | h(\chi_{i,k|k-1}^x) - \hat{y}_k]^T \quad (24)$$

$$P_{y_k} = \sum_{i=0}^{2L} W_i^{(c)} [h(\chi_{i,k|k-1}^x) - \hat{y}_k | h(\chi_{i,k|k-1}^x) - \hat{y}_k]^T \quad (25)$$

마지막으로 상태변수와 공분산은 다음과 같이 추정치 갱신된다.

$$\hat{x}_k = \hat{x}_k + K_k (y_k - \hat{y}_k) \quad (26)$$

$$P_k = P_k - K_k P_{y_k} K_k^T \quad (27)$$

한편 식(12)에서 표시된 비선형 이산시간 일반식은 식(2)에서 표현한 것처럼 순환신경망 등화기(RNE)의 동특성을 나타내는 식과 유사한 형태이다. 따라서 식(26)에 나타난 상태변수의 갱신은 순환신경망의 연결강도 벡터  $w$ 의 변환으로 생각할 수 있다.

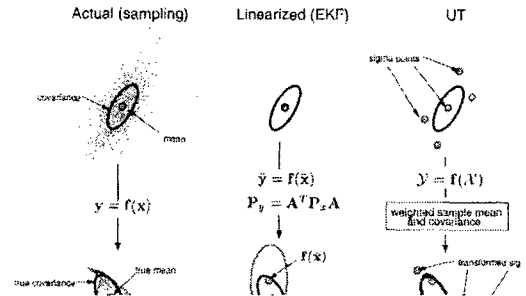


Fig. 2. Example of the sigma-point approach for mean and covariance propagation

그림2는 지금까지 설명한 확장된 칼만필터 (EKF)와 시그마 포인트 칼만필터(SPKF) 혼련 알고리즘을 적용한 경우 평균값과 공분산의 전달의 한 예를 보인 것이다.

## 4.장 시뮬레이션

### 4.1. 통신시스템모델링

그림 3에서는 선형과 비선형 왜곡 상태를 모두 포함하고 있는 결정 피드백 등화기(decision feedback equalizer, DFE)를 포함한 디지털 통신시스템을 구성하였다.

정보 소스로부터 추출된 시퀀스  $\{s(k)\}$ 가 전송되고, 전송된 심볼들은 부가백색잡음(additive white Gaussian noise, AWGN)에 의해서 오염된다. 이와 같은 채널 모델은 다음 식과 같이 표현된다.

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

여기에서  $g(\cdot)$ 는 비선형 왜곡을 나타내며  $h$ 는 길이가  $N$ 인 채널의 유한 임펄스 응답(finite impulse response, FIR)이다. 그

리고  $v(k)$ 는 평균이 0이고 분산이  $\sigma_0^2$ 인 AWGN 이다.

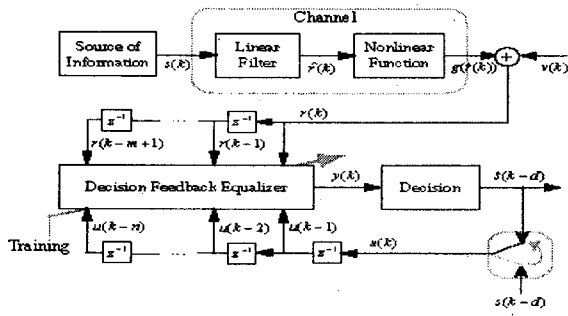


Fig3. A digital communications system with decision feedback equalizer

#### 4.2. 수렴속도에 대한 시뮬레이션

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

$$C(z) = c_0 z^{-1} + c_1 z^{-2} + c_2 z^{-3}$$

이때 채널 임펄스 응답은  $c = [c_0 \ c_1 \ c_2]^T$ 이며,  $c = [0.3482 \ 0.8704 \ 0.3482]^T$  이다. 위식은 일반적으로 채널 등화기의 모델링에 사용 되어진 식이다. 비선형 채널은 다음 식과 같이 모델링 되었는데, 여기에서 비선형성은 선형 채널의 출력에 적용되었다.

$$\begin{aligned} r(k) &= g(\hat{r}(k)) + v(k) \\ &= \hat{r}(k) + 0.2(\hat{r}(k))^2 + v(k) \end{aligned}$$

이러한 채널의 비선형 왜곡은 전송 시스템에서 기인하는 포화현상을 나타낼 수도 있다. UKF의 파라미터들은  $\alpha=0.1$ ,  $\lambda=0.99$ 이며, 결정 지연은  $d=2$ 이다. 주어진 채널에 대하여 두가지 방법(EKF, SPKF)의 시뮬레이션에 의한 수렴 특성은 그림 4.에 보였는데 MSE의 선형 스케일로 나타내었다. 이 결과는 200회의 독립적인 학습에 대한 평균 값을 구한 것인데, 매 실험마다 모든 등화기에 대하여 다른 BPSK 랜덤 시퀀스와 랜덤 초기화 연결 강도를 적용하여 14[dB]의 SNR에서 수행 하였다.

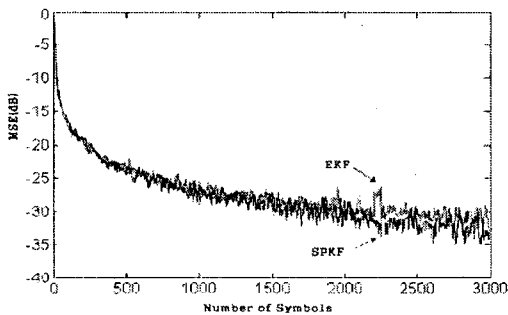


Fig. 4. Convergence properties of equalizers for channel model

그림에서 알 수 있듯이 UKF의 수렴특성은 EKF에 비하여 수렴하는 속도 면에서 그다지 빠른 특성을 보여주지는 못함을 볼 수 있다. 그러나 이 방법은 기존의 그라디언트 기반 훈련 방법에 비해서 매우 빠른 수렴 특성을 보이는 특징이 있다.

한편 채널모델에 대한 비트에러율(Bit Error Rate, BER) 성능을 그림 5.에 비교한 바와 같이 EKF와 SPKF 훈련 알고리즘을 이용 하였을 때를 비교하면 두드러진 차이가 있음을 알 수 있다. 그림에서 알 수 있듯이 SPKF는 EKF를 이용한 훈련방법에 비하여 견실함을 관측할 수 있었다. 그림에서 알 수 있듯이 UKF 알고리즘

에 의한 경우가 EKF를 이용한 경우에 비하여 0-4 의 BER 지점에서 약 1.5dB 정도가 향상되었음을 보여준다. 특히 SPKF 에 의한 훈련 방법을 이용한 등화기의 경우 16[dB] 이상인 경우 전송된 심볼은 에러율이 거의 없이 완전하게 복원하고 있다.

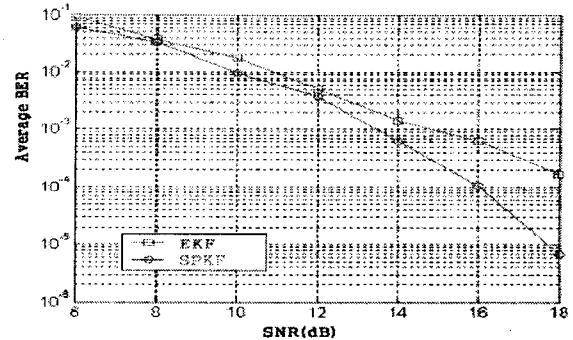


Fig.5. BER performance of the equalizer for channel model

### 5장. 결론

본 논문에서는 칼만필터를 이용하여 순환신경망을 훈련 시키는 두가지 (EKF, SPKF) 알고리즘에 대하여 비교 하였다. 제한한 훈련 알고리즘에 의한 결과는 채널에 대한 등화 실험을 통하여 상호간에 비교 되었다. 우선 EKF 에 의한 훈련 알고리즘은 비선형 추정 시스템에 대한 1차 근사화를 제공하는데 이 과정에서 매우 큰 초기 오차를 포함할 수 있어 시스템이 수렴하지 않고 발산할 수 있는 문제점이 있는 반면, SPKF에 의한 알고리즘은 같은 시스템에 대하여 2차 근사화 까지 포함하기 때문에 이러한 단점을 해결할 수 있다. 또한 두 가지 알고리즘 모두 수렴속도 면에서는 큰 차이를 보이지는 않았지만 비트 에러율 성능 면에서는 SPKF가 EKF에 비하여 매우 견실 함을 알 수 있었다. 결론적으로 통신 시스템에서 등화기는 SPKF를 이용한 알고리즘이 보다 적합한 것으로 평가된다.

### 참고문헌

- [1] S. Qureshi, "Adaptive equalization," Proceedings of the IEEE, vol. 73, no. 9, pp. 1349-1387, 1985
- [2] S. Chen, B. Mulgrew, and S. McLaughlin, "Adaptive Bayesian equalizer with decision feedback," IEEE Transactions on Signal Processing, vol. 41, pp. 2918-2927, September 1993.
- [3] A. Zerguine, A. Shafi, and M. Bettayeb, "Multilayer perception-based DFE with lattice structure," IEEE Transactions on Neural Networks, vol. 12, pp. 532-545, May 2001.
- [4] G. Kechriotis, E. Zervas, and E. S. Manolakos, "Using recurrent neural networks for adaptive communication channel equalizations," IEEE Transactions on Neural Networks, vol. 5, pp. 267-278, March 1994.
- [5] S. Haykin, "Neural Networks: a Comprehensive Foundation, 2nd Ed. Upper Saddle River, NJ: Prentice Hall, 1999.
- [6] J. Elman, "Finding structure in time," Cognitive Science, vol. 14, pp. 179-211, 1990.
- [7] S. Haykin, Adaptive Filter Theory, 4th Ed. Upper Saddle River, NJ: Prentice Hall, 2002.
- [8] 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.