MIMO Channel Prediction Using Recurrent Neural Networks

Chris Potter and Kurt Kosbar
Missouri S&T, Rolla, MO

Adam Panagos
Dynetics, Inc.
Huntsville, AL

ABSTRACT

Adaptive modulation is a communication technique capable of maximizing throughput while guaranteeing a fixed symbol error rate (SER). However, this technique requires instantaneous channel state information at the transmitter. This can be obtained by predicting channel states at the receiver and feeding them back to the trasnmitter. Existing algorithms used to predict single-input single-output (SISO) channels with recurrent neural networks (RNN) are extended to multiple-input multiple-output (MIMO) channels for use with adaptive modulation and their performance is demonstrated in several examples.

KEY WORDS

Multiple-input multiple-output (MIMO), Channel prediction, Recurrent neural networks, Online training, Adaptive modulation , Flat fading

1 Introduction

Multiple-input multiple-output (MIMO) wireless communication systems have recently received a considerable amount of attention [1, 2]. There have been numerous papers which report promising gains in capacity and diversity [3]. These results all stem from the assumption that the receiver and/or transmitter have perfect knowledge of the channel. In a typical telemetry application this is not the case and the receiver must settle for an estimate of the channel. Example channel estimation techniques include sending pilot symbols in bursts [4], periodically, [5], or in space time block codes [6].

Accurate channel estimation is often crucial for satisfactory decoding performance at the receiver. Additional performance improvements can be obtained using adaptive modulation techniques if channel state information (CSI) is available at the transmitter. If full CSI is available, the water-filling algorithm can be applied to achieve capacity [1] or pre-coding can be performed on transmitted symbols to achieve a better symbol error rate (SER) [7] . With partial CSI, the

transmitter can adapt the modulation scheme to maximize spectral efficiency given a minimum SER specification [8].

A feedback link between the receiver and transmitter is commonly used to obtain CSI at the transmitter. When a channel undergoes fast fading (for example, from doppler shifts due to transmitter/receiver motion), CSI may become outdated quickly, prohibiting the transmitter from adapting correctly. Performance can be improved in this case by sending back a *prediction* of the CSI. Since the doppler shift is significantly lower than the carrier frequency, channel estimation can be performed at baseband to accomodate a low sampling rate. This allows the incorporation of more stochastic memory into the CSI prediction [9].

Previous works have used linear prediction techniques to predict CSI. This method assumes the channel can be represented as a linear combination of past channel estimates. In many telemetry applications, however, the received signal is corrupted by non-linear effects such as distortion from non-uniform scattering. Phenomenon such as this makes non-linear prediction techniques desirable. One candidate solution to the non-linear prediction problem is to employ neural networks which contain non-linear activation functions. This was originally done in [10] for multi-layer perceptron (MLP) neural networks. This work was then extended to recurrent neural networks (RNN) which are analogous to time-varying infinite impulse response (IIR) filters for channel prediction [11]. All of these approaches are valid for single-input single-output (SISO) systems. The contribution of this work is to extend these prediction results to the MIMO case. During this process a correction to the Jacobian in [12] for the Extended Kalman Filter (EKF) algorithm is obtained.

The next section describes the input-output MIMO signal model. Section 3 provides the channel estimate/prediction model. The RNN used for channel estimation and the RNN weight update equations are provided in Sections 4 and 5 respectively. Finally, MIMO channel prediction results using the proposed RNN are presented, followed by a conclusion summarizing the work.

2 Input-Output Description

A MIMO wireless baseband flat fading communication system with N_r receive antennas and N_t transmit antennas is modeled by

$$\boldsymbol{r}(k) = \boldsymbol{H}(k)\boldsymbol{s}(k) + \boldsymbol{n}(k),$$

where $\boldsymbol{r}$ is the $N_r \times 1$ received vector, $\boldsymbol{s}$ is the $N_t \times 1$ transmitted symbol vector with each s_i belonging to constellation $\mathcal{C}$ with symbol energy E_s, and $\boldsymbol{n}$ is the white noise vector of size $N_r \times 1$ with $n_i \overset{iid}{\sim} \mathcal{CN}(0, N_o)$. The $N_r \times N_t$ channel matrix $\boldsymbol{H}(k) = \{h_{mn}(k)\}$ describes the complex channel gain between the m^{th} receive antenna and the n^{th} transmit antenna as

$$\boldsymbol{H}(k) = \boldsymbol{f}(\boldsymbol{G}(k)), \tag{1}$$

where $\boldsymbol{f} : \mathbb{C}^{N_r} \longrightarrow \mathbb{C}^{N_r}$ is a mapping defined by

$$f_j(a) = \begin{cases} a & \text{for distortionless systems} \\ \tanh(a) & \text{otherwise} \end{cases} \tag{2}$$

and $g_{mn}(k)$ are assumed to be frequency flat fading sub-channels with temporal correlation modeled by Jakes [13]. Specifically, each sub-channel autocorrelation function satisfies

$$R_{g_{mn}g_{mn}}(\tau) = J_0(2\pi f_d T_s), \tag{3}$$

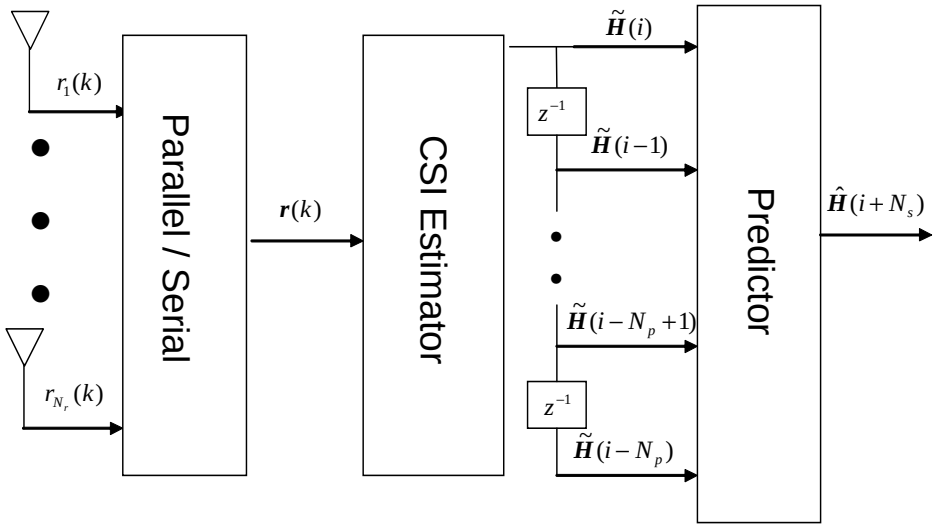

Figure 1: Block diagram of MIMO channel predictor.

where f_d is the doppler frequency and T_s is the symbol period. The non-linear hyperbolic tangent function in (2) was chosen to generalize the uniform scattering behavior of Jakes' model.

3 Prediction Model

Accurate prediction of future CSI requires knowledge of previous channel statistics. For simplicity, the estimate obtained at discrete time $i \triangleq N_e k$ is modeled as

$$\widetilde{\boldsymbol{H}}(i) = \boldsymbol{H}(i) + \widetilde{\boldsymbol{E}}(i), \tag{4}$$

where $e_{mn}(k) \overset{iid}{\sim} \mathcal{CN}(0, \sigma^2_{\boldsymbol{E}})$ describes the estimation error of $(m,n)^{th}$ sub-channel. Observing the block diagram in Figure 1, the N_p most recent estimates are used for the N_s step prediction $\widehat{\boldsymbol{H}}(i+N_s)$. Since Jakes' model assumes the channel is bandlimited to f_d Hz, to allow the predictor to accurately measure the temporal correlation the receiver must avoid spectral aliasing by performing channel estimation at least every $\frac{1}{2f_d}$ seconds.

4 Recurrent Neural Networks

The RNN with n input neurons, m hidden neurons, and r output neurons shown in Figure 2 was used for non-linear MIMO channel prediction. All outputs of the hidden layers are fed back to the input layer. The activation functions are described by the mapping $\boldsymbol{\Phi} : \mathbb{C}^m \longrightarrow \mathbb{C}^m$ with real and imaginary components defined by

$$\boldsymbol{\Phi} = \boldsymbol{\phi} + j\boldsymbol{\phi} = \tanh(\cdot) + j\tanh(\cdot).$$

Let $\boldsymbol{h}(n) = \mathrm{vec}(\boldsymbol{H}(n))$ where $\mathrm{vec}(\cdot)$ is a $N_r N_t \times 1$ vector obtained by stacking the columns of $\boldsymbol{H}$ from left to right. Using the equivalent state space representation model in [14] the input-output relationship can be extended to MIMO systems by

$$\begin{aligned} \boldsymbol{y}(k+1) &= \boldsymbol{C}\boldsymbol{x}(k+1) \\ \boldsymbol{x}(k+1) &= \boldsymbol{\Phi}(\boldsymbol{W}_x(k)\boldsymbol{x}(k) + \boldsymbol{W}_u(k)\boldsymbol{u}(k)), \end{aligned} \tag{5}$$

where $\boldsymbol{C}$ is a $r \times m$ real valued matrix and $\boldsymbol{W}_x$ and $\boldsymbol{W}_u$ are complex valued neural network weight matrices of size $q \times q$ and $q \times n$ respectively. Letting $\boldsymbol{W}(k) = [\boldsymbol{W}_x(k)\ \boldsymbol{W}_u(k)]$ and $\boldsymbol{z}(k) = [\boldsymbol{x}(k)\ \boldsymbol{u}(k)]^T$ the real (I) and imaginary (Q) components of the state vector are

$$\boldsymbol{x}(k+1) = \boldsymbol{\phi}\bigg(\boldsymbol{W}^I(k)\boldsymbol{z}^I(k) - \boldsymbol{W}^Q(k)\boldsymbol{z}^Q(k)\bigg) + j\boldsymbol{\phi}\bigg(\boldsymbol{W}^I(k)\boldsymbol{z}^Q(k) + \boldsymbol{W}^Q(k)\boldsymbol{z}^I(k)\bigg),$$

which is the desired form for deriving the RNN weight updates.

5 MIMO RNN Weight Updates

For accurate channel prediction, the neural network weights should minimize an appropriate cost function. The squared error between the neural network prediction and the most recent channel estimate is proposed as

$$J(k) = \frac{1}{2}\boldsymbol{e}^H(k+1)\boldsymbol{e}(k+1), \tag{6}$$

where

$$\boldsymbol{e}(k+1) \triangleq \widetilde{\boldsymbol{h}}(k) - \boldsymbol{C}\boldsymbol{x}(k+1). \tag{7}$$

Although this is not the actual prediction error, it provides a realistic way of measuring the performance, as opposed to [11] which uses the actual channel in (7). The real time recurrent learning (RTRL) and Extended Kalman Filter (EKF) algorithms are considered for training the RNN weights. The RTRL seeks to minimize the instantaneous value of (6) whereas the EKF minimizes the expected value of (6).

Let $\boldsymbol{w}_i$ be the i^{th} column of $\boldsymbol{W}^T$. The weight updates for the RTRL algorithm are [12, 14]

$$\begin{aligned}
\Delta_{\boldsymbol{w}_i}(k+1) &= \gamma_g[(\boldsymbol{e}^I(k))^T\boldsymbol{C} \quad (\boldsymbol{e}^Q(k))^T\boldsymbol{C}] \begin{bmatrix} \boldsymbol{\Lambda}_i^{II}(k) + j\boldsymbol{\Lambda}_i^{IQ}(k) \\ \boldsymbol{\Lambda}_i^{QI}(k) + j\boldsymbol{\Lambda}_i^{QQ}(k) \end{bmatrix} + \gamma_m \boldsymbol{w}_i(k) \\
\boldsymbol{w}_i(k+1) &= \Delta_{\boldsymbol{w}_i}(k+1) + \boldsymbol{w}_i(k),
\end{aligned}$$

where γ_g is the learning gain, γ_m is the momentum gain, and Λ^{AB} satisfies the following recursion

$$\begin{aligned}
\begin{bmatrix} \boldsymbol{\Lambda}_i^{II}(k+1) & \boldsymbol{\Lambda}_i^{IQ}(k+1) \\ \boldsymbol{\Lambda}_i^{QI}(k+1) & \boldsymbol{\Lambda}_i^{QQ}(k+1) \end{bmatrix} &= \begin{bmatrix} \boldsymbol{\phi} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\phi} \end{bmatrix} \begin{bmatrix} \boldsymbol{W}_x^I(k) & -\boldsymbol{W}_x^Q(k) \\ \boldsymbol{W}_x^Q(k) & \boldsymbol{W}_x^I(k) \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda}_i^{II}(k) & \boldsymbol{\Lambda}_i^{IQ}(k) \\ \boldsymbol{\Lambda}_i^{QI}(k) & \boldsymbol{\Lambda}_i^{QQ}(k) \end{bmatrix} \\
&\quad + \begin{bmatrix} \boldsymbol{Z}_i^I(k) & -\boldsymbol{Z}_i^Q(k) \\ \boldsymbol{Z}_i^Q(k) & \boldsymbol{Z}_i^I(k) \end{bmatrix}
\end{aligned}$$

with the initialization

$$\begin{bmatrix} \boldsymbol{\Lambda}_i^{II}(k+1) & \boldsymbol{\Lambda}_i^{IQ}(k+1) \\ \boldsymbol{\Lambda}_i^{QI}(k+1) & \boldsymbol{\Lambda}_i^{QQ}(k+1) \end{bmatrix} = \boldsymbol{0}.$$

Before the weight updates for the EKF are given, the complex Jacobian in [12] is corrected to

$$\begin{aligned}
\boldsymbol{\Lambda}(k+1) &= \frac{\partial(\boldsymbol{x}(k+1))}{\partial(\boldsymbol{w}^I(k))} + j\frac{\partial(\boldsymbol{x}(k+1))}{\partial(\boldsymbol{w}^Q(k))} \\
&= \boldsymbol{\Lambda}_i^{II}(k) - \boldsymbol{\Lambda}_i^{QQ}(k) + j\bigg(\boldsymbol{\Lambda}_i^{QI}(k) + \boldsymbol{\Lambda}_i^{IQ}(k)\bigg).
\end{aligned}$$

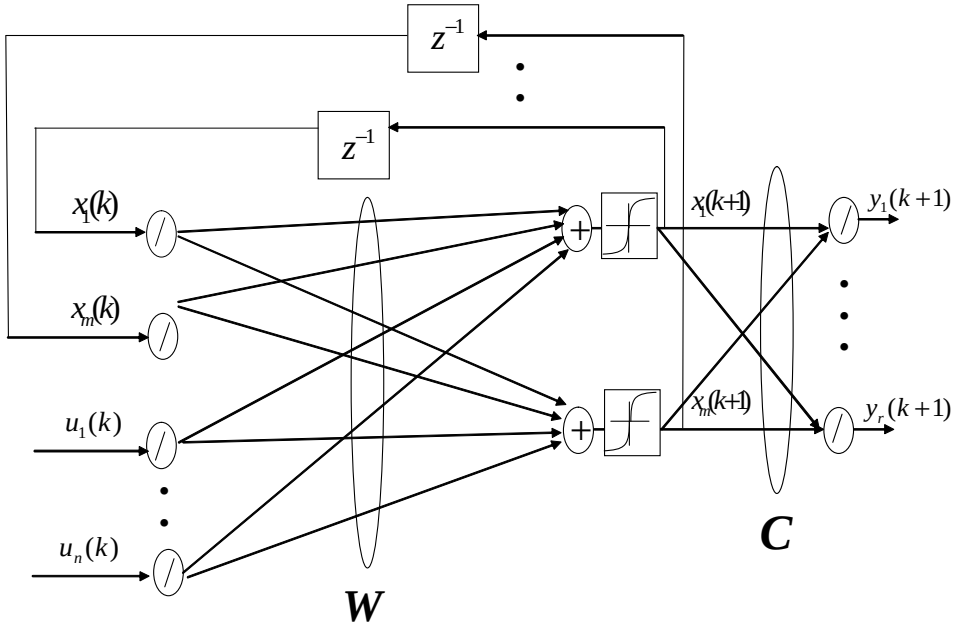

Figure 2: State space representation of a recurrent neural network.

This is applied to the EKF weight updates

$$\begin{aligned}
\boldsymbol{\Gamma}(k) &= [\boldsymbol{\Lambda}(k)\boldsymbol{P}(k)\boldsymbol{\Lambda}^H(k) + \boldsymbol{R}(k)]^{-1} \\
\boldsymbol{K}(k) &= \boldsymbol{P}(k)\boldsymbol{\Lambda}^H(k)\boldsymbol{\Gamma}(k) \\
\boldsymbol{w}_i(k+1) &= \boldsymbol{w}_i(k) + \boldsymbol{K}(k)\boldsymbol{e}(k) \\
\boldsymbol{P}(k+1) &= \boldsymbol{P}(k) - \boldsymbol{K}(k)\boldsymbol{\Lambda}(k)\boldsymbol{P}(k) + \boldsymbol{Q}(k)
\end{aligned}$$

with the following initializations

$$\begin{aligned}
\boldsymbol{R}(0) &= \mu^{-1}(\boldsymbol{I} + j\boldsymbol{I}) \\
\boldsymbol{Q}(0) &= \rho(\boldsymbol{I} + j\boldsymbol{I}) \\
\boldsymbol{P}(0) &= \epsilon^{-1}(\boldsymbol{I} + j\boldsymbol{I}).
\end{aligned}$$

A contribution of this work is the ability to predict all $N_r N_t$ sub-channels simultaneously by letting

$$\begin{aligned}
\boldsymbol{y}(k+1) &= \widehat{\boldsymbol{h}}(k+1) \\
\boldsymbol{u}(k) &= \left[\begin{array}{cccc} \widetilde{\boldsymbol{h}}(k) & \widetilde{\boldsymbol{h}}(k-1) & \cdots & \widetilde{\boldsymbol{h}}(k-N_p) \end{array} \right]^T.
\end{aligned}$$

This is more computationally attractive than predicting each sub-channel independently with a separate RNN.

6 Prediction Results

This section presents numerical examples to compare the tracking capability of the EKF and RTRL algorithms. Parameters held constant are given in Table 1. These parameters would be typical of a telemetry link between a moving aircraft and stationary base station whose line of

sight was obstructed. Consider first a one step predictor corresponding to $N_s = 1$. The MSE of the EKF and RTRL algorithms is shown in Figure 3. The EKF outperforms the RTRL algorithm and reaches an average MSE that is two orders of magnitude smaller. The comparison between the actual channel coefficients and their predictions for each MIMO sub-channel are shown in Figures 4-7. These plots verify the EKF predictions match the actual channel much better than the RTRL predictions.

Next, the two training algorithms are compared in a multi-step prediction scenario by letting $N_s = 10$. Once again the EKF performance is superior to the RTRL as seen in Figure 8. For simplicity only the real part of the sub-channel corresponding to the first transmit and receive antennas is displayed in Figures 9 and 10 to show how well the 10-step EKF predictor performs compared to the 10-step RTRL predictor.

An example comparing the EKF neural network predictor to a linear predictor using the Levinson-Durbin recursion is considered next. The fixed parameters are given in Table 2. In addition to the non-linear distortion present and time varying doppler frequency, bursts of white noise with a variance of 0.1 are present. Bursts of this nature could be used to model interference from sources such as radar transmitters. The MSE comparison in Figure 11 indicates that the EKF predictor slightly outperforms the L-D approach. The real and imaginary components of the sub-channel corresponding to the first transmit and receive antennas are displayed in Figure 12. Unlike the linear predictor, the EKF neural network predictor is able to explore the non-linear correlations present in the channel.

7 Conclusion

Existing SISO prediction algorithms were extended to the MIMO case for the EKF and RTRL algorithms. This method is computationally attractive as it requires a single RNN instead of an RNN for each MIMO sub-channel. Examples presented show the EKF algorithm outperformed the RTRL algorithm in both single and multi step prediction. When channel non-linearities are present, the neural network predictor using the EKF algorithm outperformed the Levinson-Durbin linear predictor.

Table 1: Parameters Values for channel tracking example

Parameter	Value	Description
$f_j(a)$	a	No Distortion
N_r	2	Receive Antennas
N_t	2	Transmit Antennas
N_p	5	Previous Channel Estimates
N_e	1	Samples Between Estimation
σ_e^2	.001	Estimation Error
f_d	500 Hz	Doppler Frequency
T_s	10 msec	Symbol Period
T_h	1	Prediction step
μ	.1	EKF measurement noise parameter
ρ	.1	EKF process noise parameter
ϵ	10	EKF error covariance parameter
γ_g	.001	RTRL learning gain
γ_m	.3	RTRL momentum gain

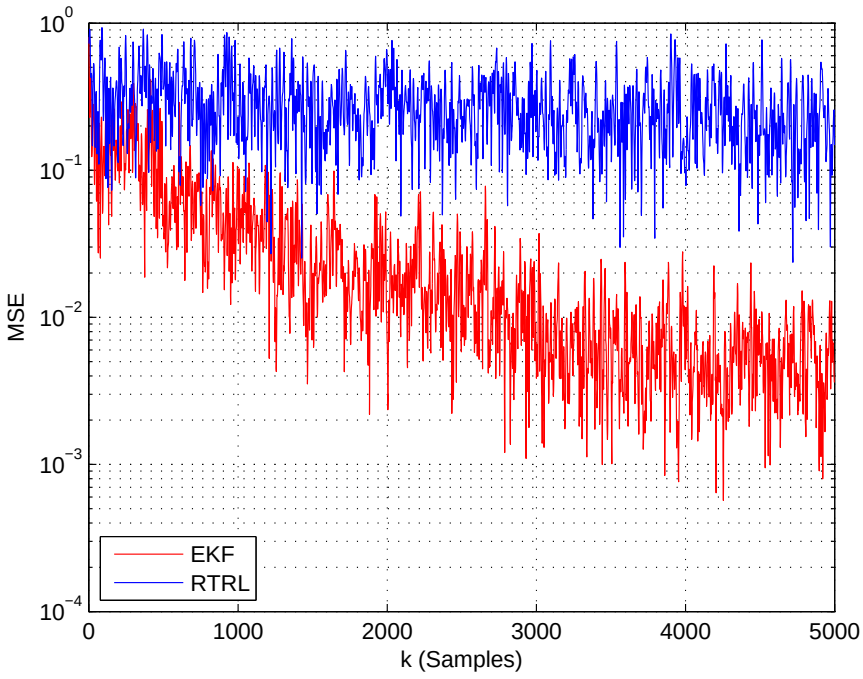

Figure 3: Mean squared error between EKF and RTRL algorithms when $N_s = 1$.

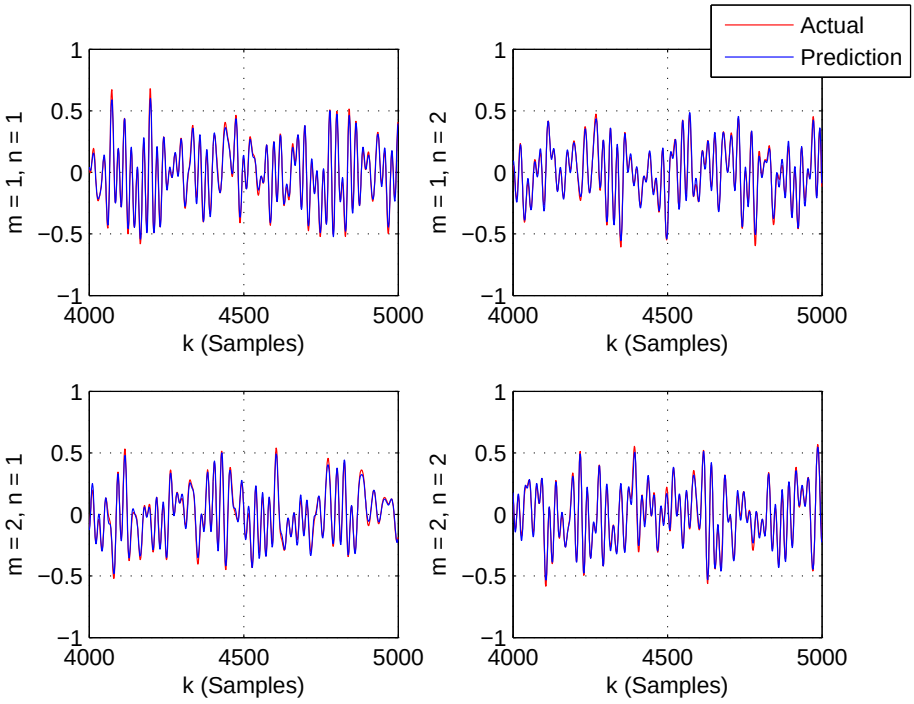

Figure 4: Real component of MIMO channel coefficients using EKF algorithm when $N_s = 1$.

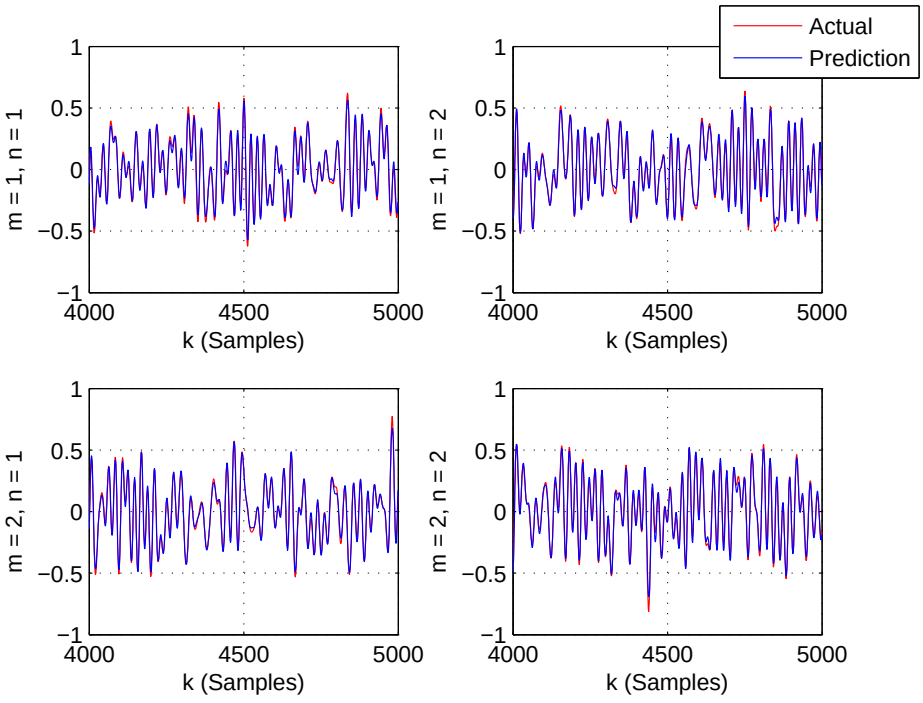

Figure 5: Imaginary component of MIMO channel coefficients using EKF algorithm when $N_s = 1$.

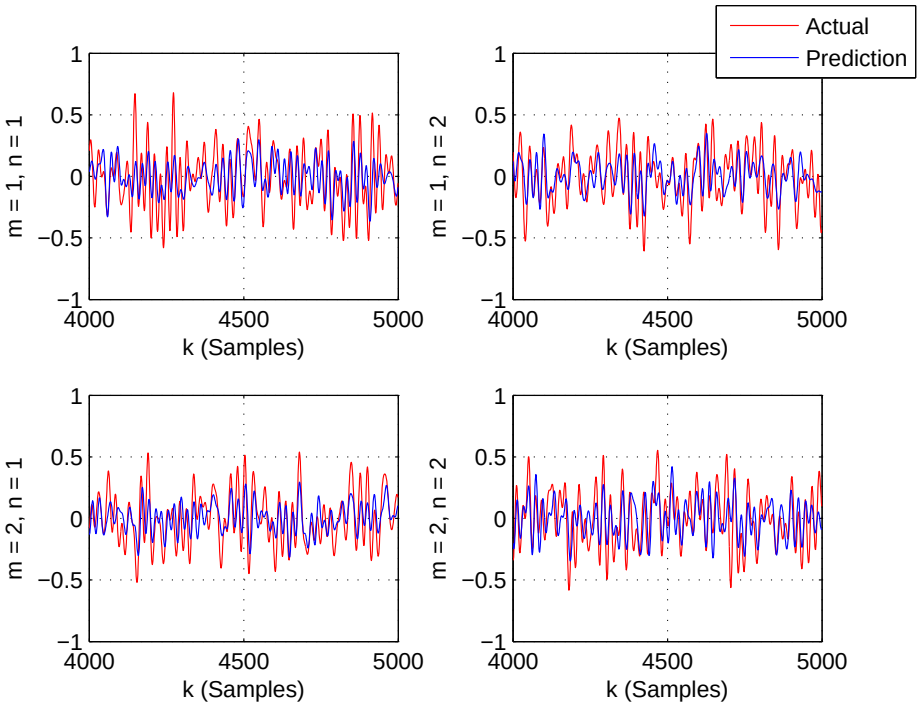

Figure 6: Real component of MIMO channel coefficients using RTRL algorithm when $N_s = 1$.

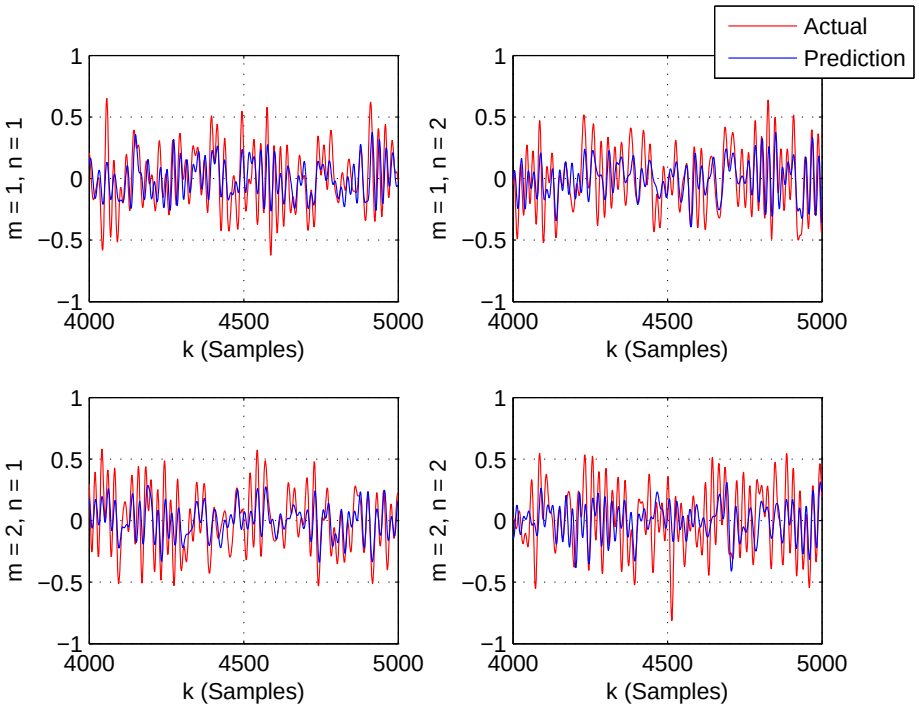

Figure 7: Imaginary component of MIMO channel coefficients using RTRL algorithm when $N_s = 1$.

Table 2: Parameters Values for non-linear channel example

Parameter	Value	Description
$f_j(a)$	$\tanh(a)$	Non-linear Distortion
N_r	2	Receive Antennas
N_t	2	Transmit Antennas
N_p	5	Previous Channel Estimates
N_e	1	Samples Between Estimation
σ_e^2	.001	Estimation Error
T_s	10 msec	Symbol Period
μ	.1	EKF measurement noise parameter
ρ	.1	EKF process noise parameter
ϵ	10	EKF error covariance parameter

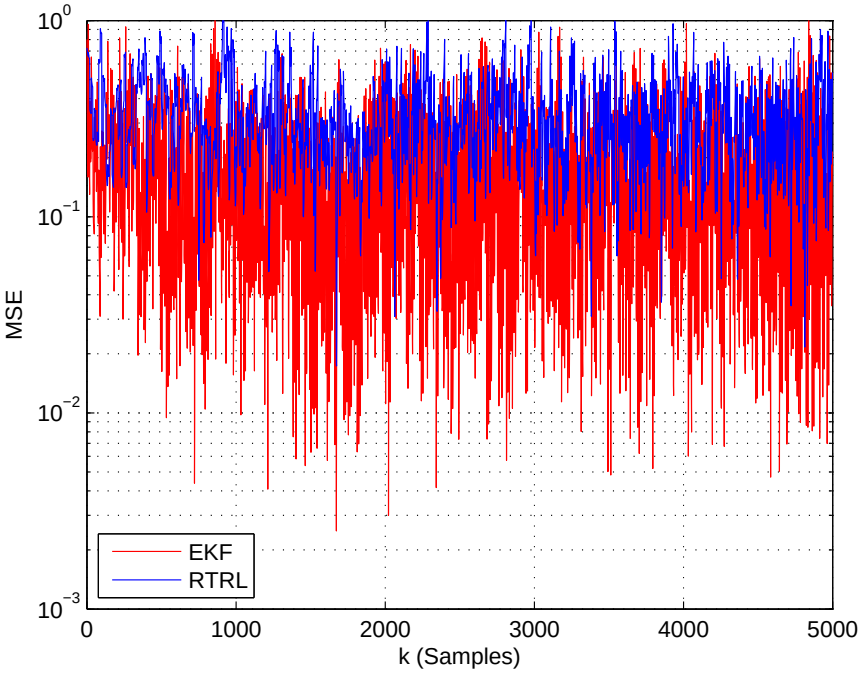

Figure 8: Mean squared error between EKF and RTRL algorithm when $N_s = 10$.

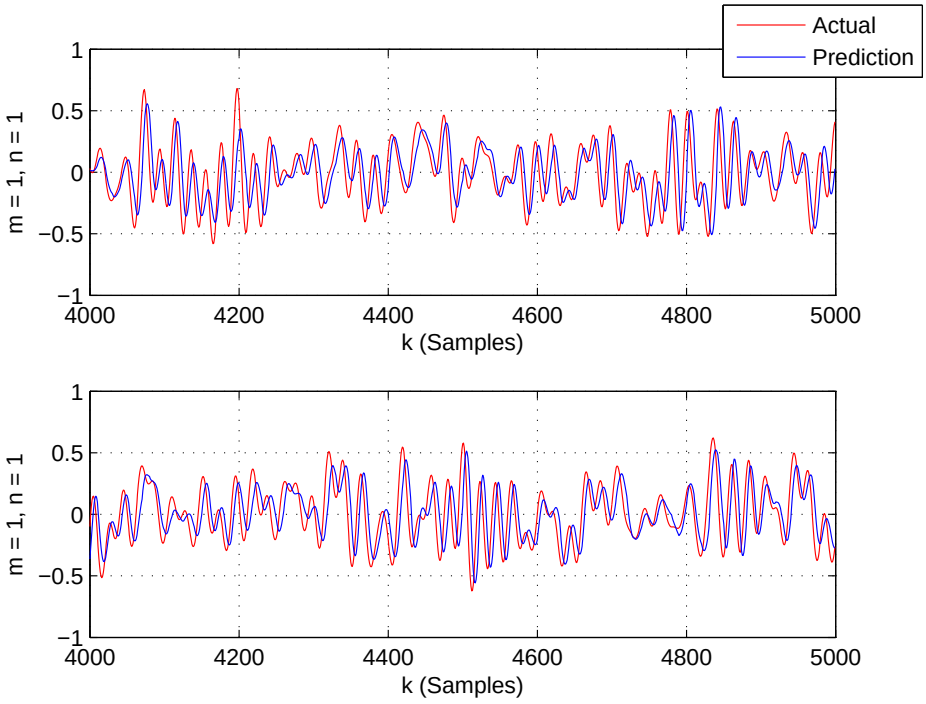

Figure 9: Real and imaginary component of a MIMO sub-channel using EKF algorithm when $N_s = 10$.

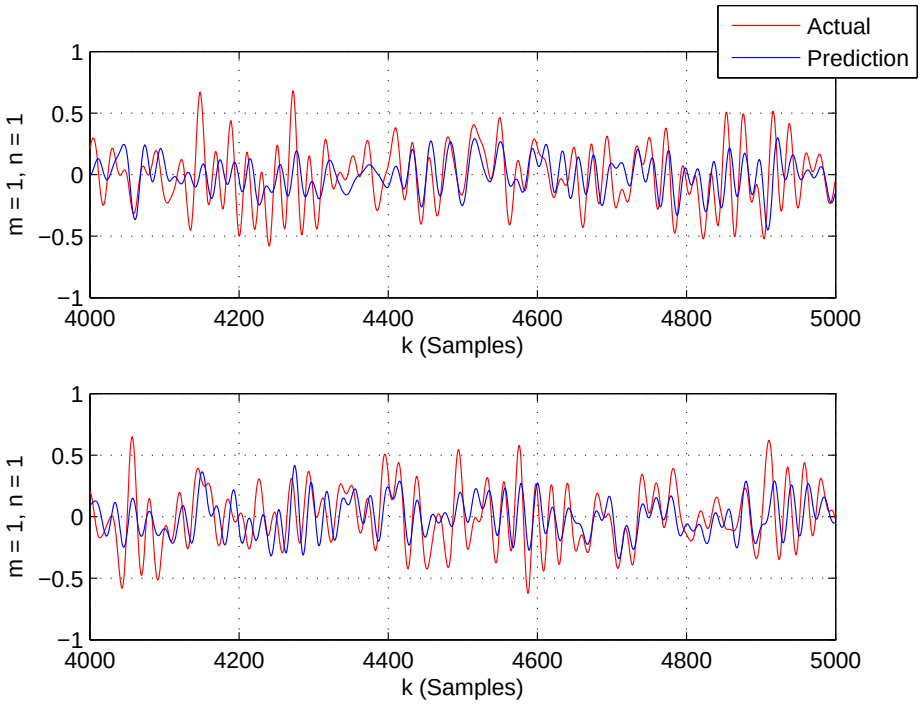

Figure 10: Real and imaginary component of a MIMO sub-channel using RTRL algorithm when $N_s = 10$.

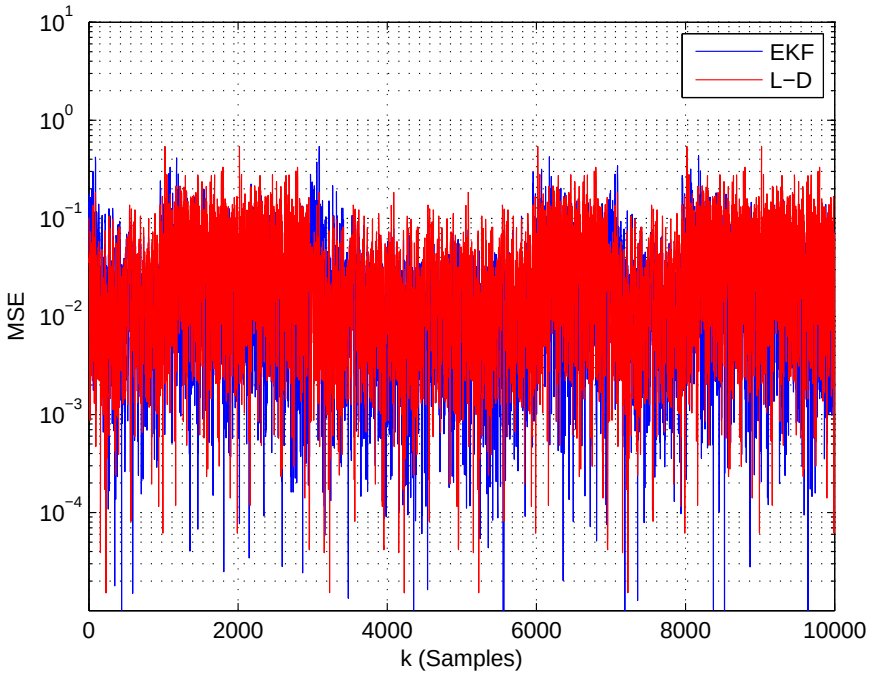

Figure 11: Mean squared error between EKF and L-D algorithm for a non-linear channel.

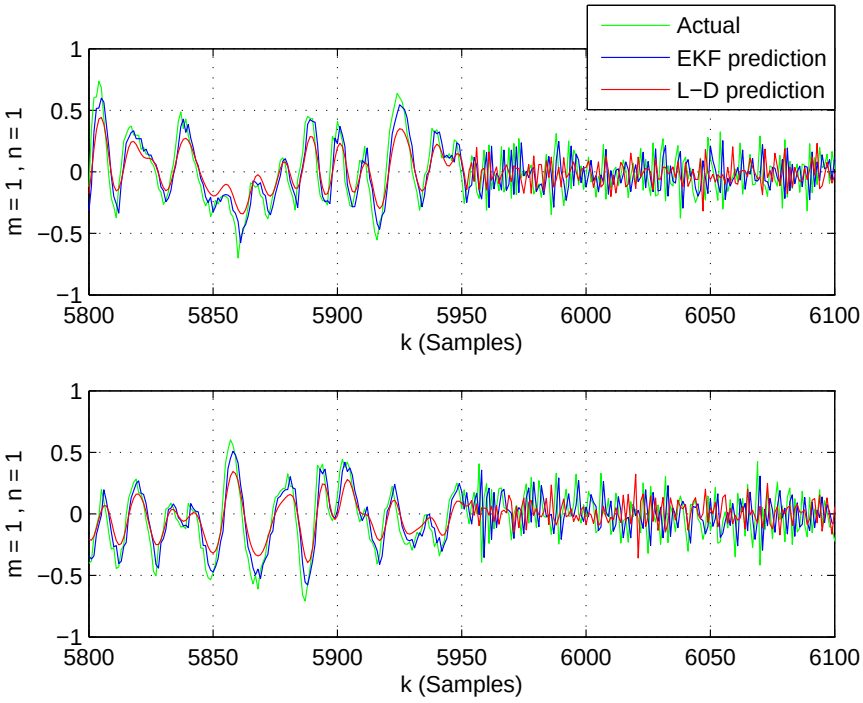

Figure 12: Real and imaginary components of a MIMO non-linear sub-channel.

REFERENCES

[1] E. Telatar, "Capacity of multi-antenna gaussian channels," *AT&T Bell Laboratories Internal Tech. Memo.*, June 1995.

[2] G. Foschini and M. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[3] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge, U.K.: Cambridge University Press, 2003.

[4] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links," *IEEE Trans. Inform. Theory*, vol. 49, pp. 951–963, Apr. 2003.

[5] S. Zhou and G. B. Giannakis, "How accurate channel prediction needs to be for transmit beamforming with adaptive modulation over rayleigh mimo channels?" *IEEE Trans. Wireless Commun.*, vol. 3, pp. 1285–1294, 2004.

[6] S. Alamouti, "A simple transmit diverstiy technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1451–1458, Aug. 1998.

[7] L. Collin *et al.*, "Optimal minimum distance based precoder for mimo spatial multiplexing systems," *IEEE Trans. Signal Processing*, vol. 52, no. 3, pp. 617–627, Mar. 2004.

[8] Z. Zhou *et al.*, "Mimo systems with adaptive modulation," *IEEE Trans. Veh. Technol.*, vol. 54, no. 5, 2005.

[9] A. Duel-Hallen, S. Hu, and H. Hallen, "Long range prediction of fading signals: enabling adaptive transmission for mobile radio channels," *IEEE Signal Processing Mag.*, vol. 17, pp. 62–75, May 2000.

[10] B. Visweswaran and T. Kiran, "Channel prediction based power control in w-cdma systems," *First Int. Conf. on 3G Mobile Comm. Techn.*, pp. 41–45, Mar. 2000.

[11] W. Liu, L.-L. Yang, and L. Hanzo, "Recurrent neural network based narrowband channel prediction," *Veh. Tech. Conf.*, vol. 5, pp. 2173–2177, May 2006.

[12] J. Choi, M. Bouchard, and T. Yeap, "Decision feedback recurrent neural equalization with fast convergence rate," *IEEE Trans. Neural Networks*, vol. 16, no. 3.

[13] A. Jazwinski, *Stochastic Process and Filtering Theory*. Acadamic Press, 1993.

[14] P. Coelho, "A new state space model for a complex rtrl neural network," *Proc IJCNN*, vol. 3, pp. 439–441, 2001.