System Identification and CDMA Communication

A (partial) sample report by
Nathan A. Goodman

Abstract

This (sample) report describes theory and simulations associated with a class project on system identification and code-division multiple access (CDMA) communication. The goals of the project are to gain experience applying principles of digital signal processing, to obtain understanding of how signal correlation is used in signal processing applications, and to practice technical writing and scientific communication.

1. Introduction

Correlation is a fundamental quantity used to analyze and quantify the content of information-bearing signals. Correlation quantifies the similarity between two signals; therefore, correlation can be used to find hidden structure embedded well below noise and distortion levels. In electrical and computer engineering, for example, signal correlation can be used to detect communication symbols or radar pulses reflected by a target, to estimate channel properties, or to estimate signal parameters. In many cases, correlation is even the optimum processing operation. In fact, the well-known Fourier Transform is essentially a correlation operator.

In this report, we study and demonstrate the usefulness of signal correlation through two application examples. The first example is system identification (ID) where the cross-correlation between an unknown system's input and output signals can be used to estimate the channel's impulse response. We will examine system ID performance as a function of the input signal's structure and energy. The second application example is the detection of digital communication symbols buried in noise and multi-user interference.

II. System Identification

A. *Background*

Suppose that a signal can be measured only after it passes through an unknown linear, time-invariant (LTI) system. It is sometimes important to characterize the unknown system so that distortions introduced by the system can be removed from the measured signal. In some cases, knowledge of the system can even be used to modify the input to the system in an advantageous way. For example, if a communication channel can be characterized, the transmitted waveform can be pre-conditioned such that it arrives at the receiver in the desired shape, thus reducing the need for adaptive equalization.

The system ID technique tested in this project is based on cross-correlation between the measured output signal and a known input signal. We know that the output, $y(n)$, of a LTI system is

$$y(n) = x(n) * h(n) = \sum_{k=-\infty}^{\infty} h(k) x(n-k) \tag{1}$$

where $x(n)$ is the input signal and $h(n)$ is the system's impulse response. We also know that the cross-correlation between two sequences can be expressed as the convolution of the first sequence with the time-reversed copy of the second. Thus, the cross-correlation between the output and input signals is

$$r_{yx}(n) = y(n) * x(-n). \tag{2}$$

When (1) is substituted into (2), the result is

$$r_{yx}(n) = h(n) * x(n) * x(-n)$$
$$= h(n) * r_{xx}(n) \tag{3}$$

where $r_{xx}(n)$ is the autocorrelation of the input signal. Taking the discrete-time Fourier Transform (DTFT) of both sides of (3) yields

$$S_{yx}(f) = H(f) S_{xx}(f) = H(f) |X(f)|^2 \tag{4}$$

where $S_{yx}(f)$ is the DTFT of $r_{yx}(n)$, $H(f)$ is the system's transfer function, and $S_{xx}(f) = |X(f)|^2$ is the input signal's energy spectrum. From (4), the system transfer function is

$$H(f) = \frac{S_{yx}(f)}{|X(f)|^2}, \tag{5}$$

and $h(n)$ can be found from the inverse DTFT of (5).

In practical situations, however, it is impossible to measure a noise-free version of the system's output signal. Hence, a better model for the output signal is

$$y(n) = x(n) * h(n) + w(n) \tag{6}$$

where $w(n)$ is a noise process. Substituting (6) (rather than (1)) into (2) results in an additional noise term not accounted for in the preceding analysis. The output of the cross-correlation is then

$$r_{yx}(n) = h(n) * r_{xx}(n) + w(n) * x(-n), \tag{7}$$

and the DTFT now yields

$$S_{yx}(f) = H(f)|X(f)|^2 + W(f)X^*(f).$$ (8)

Finally, solving for $H(f)$ shows the corruption that occurs due to the presence of noise. The estimate of the system transfer function is now

$$\hat{H}(f) = \frac{S_{yx}(f)}{|X(f)|^2} = H(f) + \frac{W(f)X^*(f)}{|X(f)|^2}$$
$$= H(f) + \frac{W(f)}{X(f)}$$ (9)

B. Simulation Results

We have developed a simulation to test and quantify the performance of the system ID technique described above as a function of input signal structure and signal-to-noise ratio (SNR). We focus on a finite-impulse response (FIR) system and assume that the order (and, therefore, length) of the system is known. In practice, the system order would usually need to be estimated as well. We began with simple signals and with noise power set to zero to verify the simulation was working properly and to find the correct location of the impulse response in the final output sequence. Once this was accomplished, we studied the effects of signal structure and noise power as described below.

The first signal studied was an impulse with power equal to 1000. Thus, the input signal was $x(n) = \sqrt{1000}\delta(n)$ where $\delta(n)$ is the standard unit impulse function for discrete-time signals. The autocorrelation sequence of this input signal is simply

$$r_{xx}(n) = \sqrt{1000}\delta(n) * \sqrt{1000}\delta(-n) = 1000\delta(n),$$ (10)

and the signal's energy spectrum is

$$|X(f)|^2 = 1000; \quad -0.5 \leq f \leq 0.5.$$ (11)

The noise power was set to one, and a 21-sample impulse response was generated using the optimum equiripple design for a bandpass filter. Since the simulated noise is different each time the simulation is executed, the estimated impulse response will also vary. One example of the estimated impulse response is shown in Fig. 1 along with the true impulse response. Figure 1 shows that the impulse response has been estimated well, but not perfectly. By comparison, if the noise power is set to $P_n = 10$, a sample result is shown in Fig. 2. The estimation error seen in Fig. 2 is clearly larger than the estimation error seen in Fig. 1. Obviously, this behavior is expected since the noise power has increased.

3
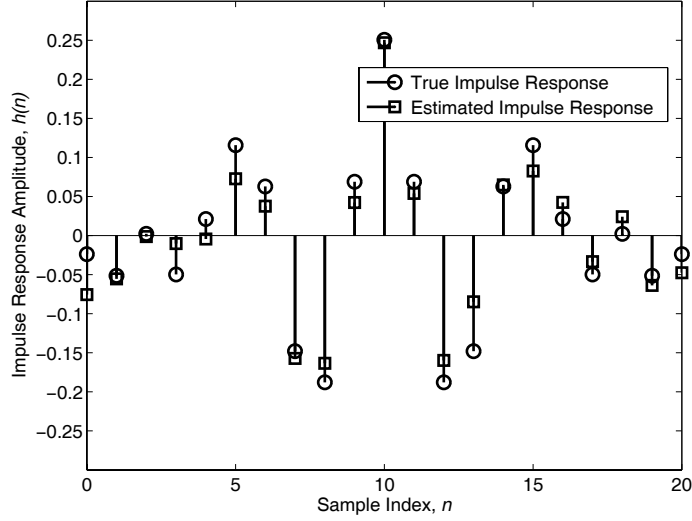
Figure 1. Sample result for the estimated system impulse response when the input signal is $x(n) = \sqrt{1000}\delta(n)$ **and the noise power is** $P_n = 1.$
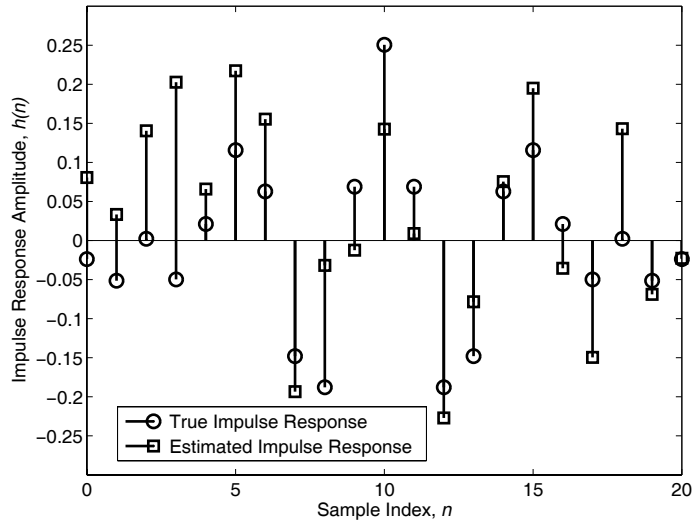

Figure 2. Sample result for the estimated system impulse response when the input signal is $x(n) = \sqrt{1000}\delta(n)$ **and the noise power is** $P_n = 10.$

The second signal studied is a pseudonoise random waveform of length L created using Matlab's Gaussian random number generator. The advantage of a noise waveform is that, *on average*, the autocorrelation function is again a unit impulse. In other words,

$$r_{xx}(n) = \text{E}\big[x(l)x(l-n)\big] = P_x\delta(n) \tag{12}$$

where $\text{E}[\cdot]$ is the expected value operator and P_x is the random waveform's average power. Of course, only a single realization of the random waveform can be transmitted; hence, the autocorrelation of the actual input waveform will not be a perfect impulse. Figure 3a shows a sample realization of a noise waveform with $P_x = 1000$. Figure 3b shows the autocorrelation

4
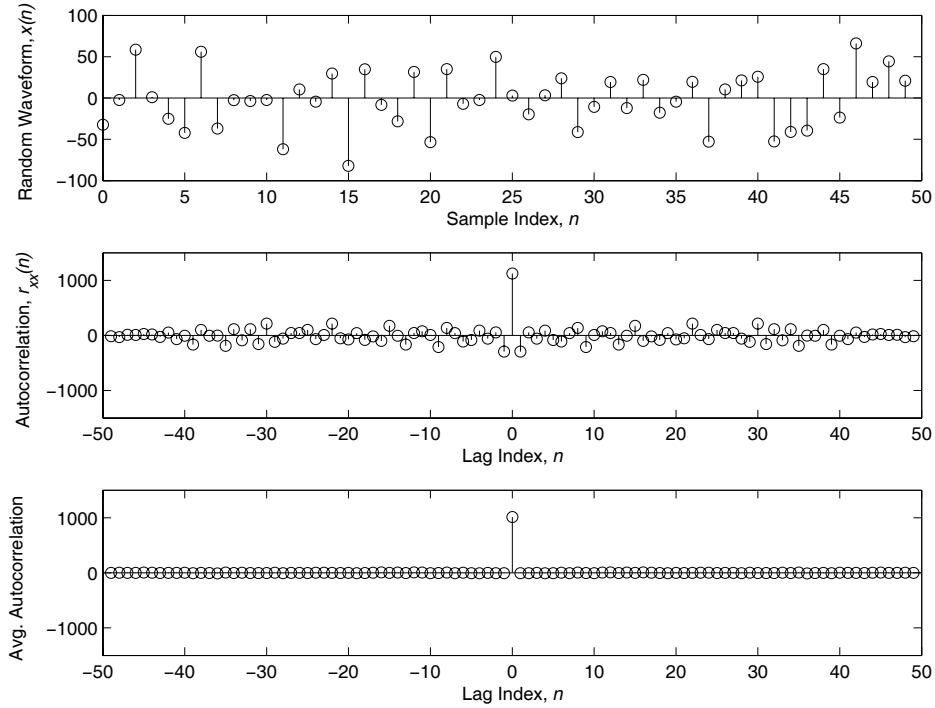
Figure 3. Top Panel: A sample realization of a pseudorandom waveform. Middle Panel: The autocorrelation of the waveform realization. Bottom Panel: Autocorrelation averaged over 500 waveform realizations.

function for the realization shown in Fig. 3a, and Fig. 3c shows the autocorrelation function that resulted from averaging the autocorrelation function over 500 realizations of the noise waveform. The average power spectrum of the random waveform is the DTFT of the average autocorrelation function. Since the average autocorrelation is the unit impulse seen in Fig. 3c, the average power spectrum is flat, which is the optimum scenario for system ID.

A pseudorandom input waveform has approximately the same autocorrelation function as a unit impulse. However, a single realization of the random waveform will have larger energy than the unit impulse. Hence, we expect to see improved performance over that obtained with a unit impulse of the same power. We also note that for best performance, the system ID algorithm should have knowledge of the particular waveform realization. Figure 4 shows an example of estimation performance when a pseudorandom waveform with length 50 is used and the noise power is $P_n = 10$.

The impulse response estimate in Fig. 4 is clearly improved over the estimate obtained with the unit impulse for the same noise power. In comparing Figs. 1, 2, and 4, it becomes apparent that the noise power and length of the input waveform both effect system ID performance. We hypothesize that SNR determines relative performance between input waveforms having the same power spectrum (average power spectrum in the case of pseudorandom waveform). Thus, we define SNR as
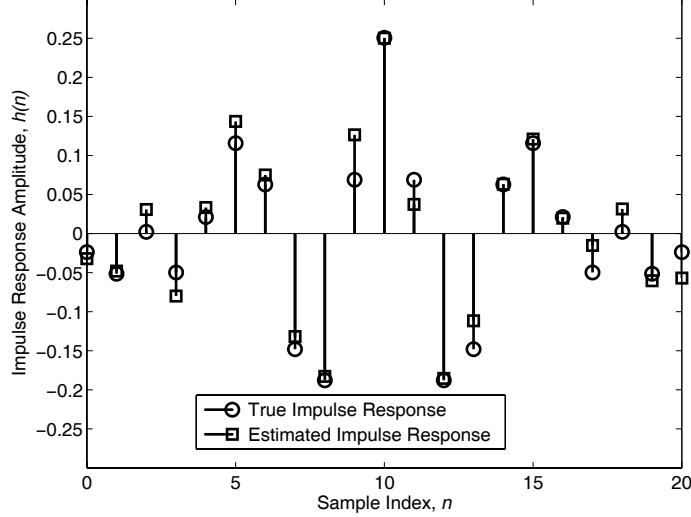
Figure 4. Sample result for the estimated system impulse response when the input signal is a pseudonoise waveform with $P_x = 1000$ and the noise power is $P_n = 10$.

$$\text{SNR} = \frac{P_x}{P_n} \tag{13}$$

where $P_x = 1000$ for the unit impulse, $P_x = \text{E}\left[x^2(n)\right] = 1000$ for the pseudorandom waveform, and P_n is the noise power. To quantify mean-squared error (MSE) performance, we repeated the system ID simulation 25000 times for the unit impulse and for pseudorandom waveforms of varying length. We also varied the SNR through control of the noise power. We then calculated the average squared-error for the 25000 trials according to

$$MSE = \frac{1}{25000} \sum_{k=1}^{25000} \left(\mathbf{h} - \hat{\mathbf{h}}_k\right)^{\text{T}} \left(\mathbf{h} - \hat{\mathbf{h}}_k\right) \tag{14}$$

where $\mathbf{h}$ is a column vector containing the values of the true impulse response, $\hat{\mathbf{h}}_k$ is a vector containing the estimated impulse response from the kth trial, and $(\cdot)^{\text{T}}$ is the transpose operator.

The results of our Monte Carlo simulations were very interesting. The squared-error values for the unit impulse were well behaved, but values for the pseudorandom waveform were sometimes extremely large. Figure 5 shows a histogram of the squared-error values for the length-50 pseudorandom waveform at an SNR of 40 dB. Despite using 100 bins for the histogram, 99.8% of the squared-error values fall into the lowest bin while the remaining values vary significantly but rarely occur. In fact, these other values occur so rarely that they cannot be seen on the same scale that is appropriate for the smallest bin. Figure 6 shows the same histogram with a reduced vertical scale.

Unfortunately, the outlying squared-error values seen in Fig. 6 seemed to de-stabilize the average squared-error performance results. In other words, average squared-error plots using all 25,000 trials produced curves that were not as smooth as expected or desired. Therefore, we

6
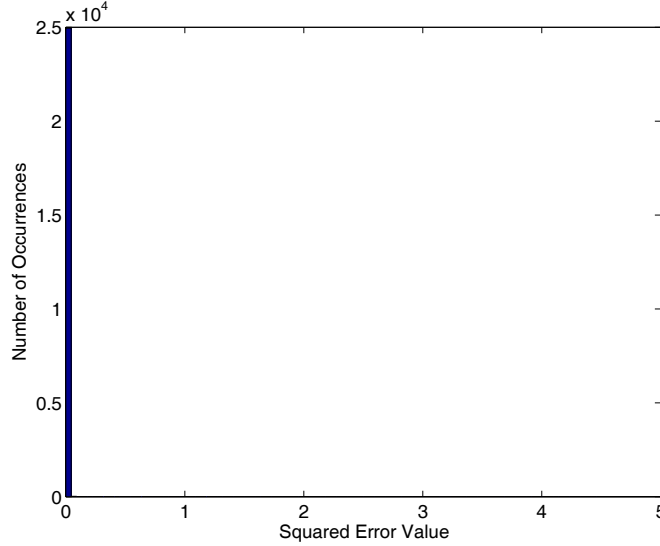
Figure 5. 100-bin histogram of squared-error values for SNR = 40 dB and a pseudorandom waveform with L = 50.
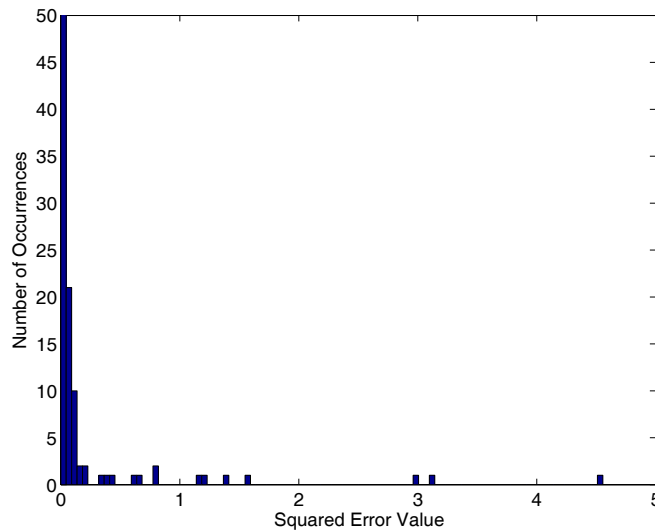

Figure 6. Vertically scaled version of Figure 5 showing rarely occurring values.

decided to ignore the outliers by using only the lowest 24,900 values in the mean calculation for a given waveform and SNR. The result of this procedure is shown in Fig. 7. In Fig. 7, it is seen that the unit impulse and the pseudorandom (labeled PN, for 'pseudonoise') waveform of length five perform nearly the same. This was unexpected since most of the pseudorandom waveform realizations have higher energy than the unit impulse. Our explanation for the unexpected result is that the autocorrelation properties of a given random waveform realization strongly affect performance. We believe that in extreme cases, poor autocorrelation properties produce the outlying results seen in Fig. 6. In less extreme cases, waveform realizations with weak autocorrelation properties still produce worse estimates than can be achieved with a unit impulse. Hence, on average and excluding the extreme outliers, the waveforms perform nearly the same.
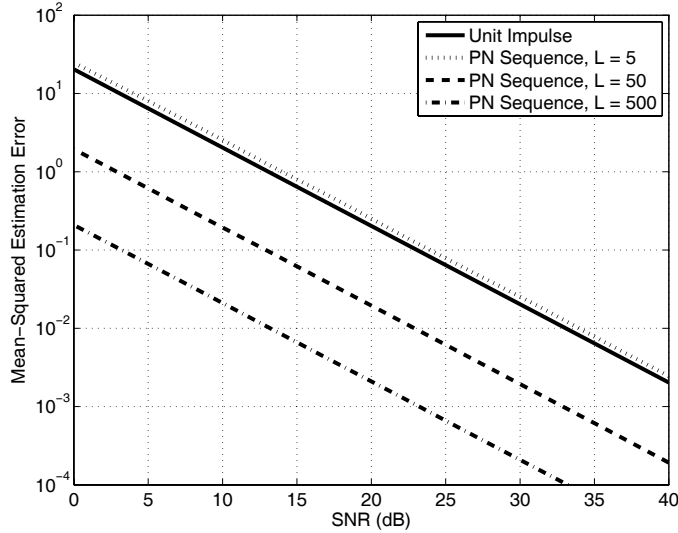
Figure 7. (Approximate) MSE performance vs. SNR and parameterized by waveform.

(a)

(b)
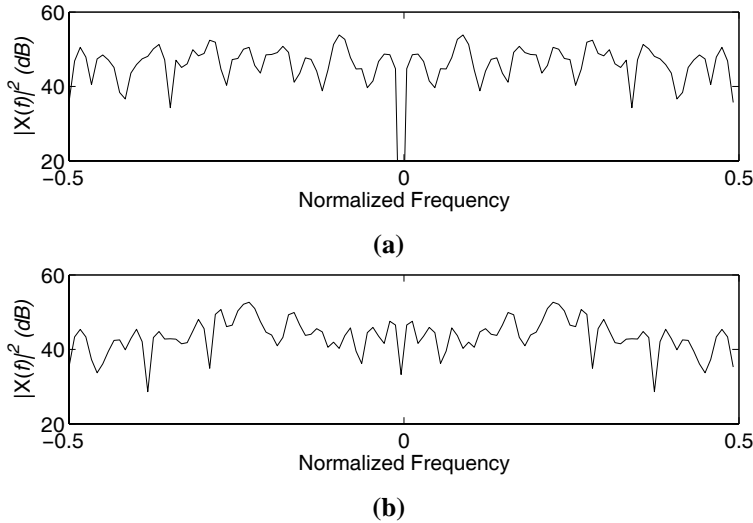
Figure 8. The power spectrum of a poorly performing pseudorandom waveform in (a), and a satisfactorily performing pseudorandom waveform in (b).

Another interesting aspect seen in Fig. 7 is the shift between curves associated with pseudorandom waveforms with different lengths. The lengths of the waveforms differ by factors of 10 when going from $L = 5$, to $L = 50$, to $L = 500$. Likewise, the performance curves are shifted by approximately $10\log_{10}(10) = 10$ dB relative to each other. Since the average energy in the pseudorandom waveforms is $\bar{E}_x = LP_x$, it appears that the relative shift in performance is directly proportional to the pseudorandom waveform's average energy.

Finally, we test our explanation that correlation properties of an individual waveform strongly affect squared-error performance. To test our hypothesis, we generated noise waveforms until we found two waveforms that performed quite differently. The power spectra of these two waveforms are shown in Figs. 8a and 8b. Figure 8a, which shows the spectrum of a

poorly performing waveform, has a noticeable null located around $f = 0$ cycles/sample. Since the estimated transfer function is obtained by dividing $S_{yx}(f)$ by the waveform's power spectrum, very low values in the power spectrum correspond to dividing by a small number. Furthermore, in the presence of noise, division by a small number can produce unstable results. Further investigation confirmed that poorly performing waveforms consistently had a deep null in their spectrum while waveforms that performed well consistently did not have a deep null.

C. Conclusions

We have successfully implemented a system identification algorithm based on cross-correlation between the input and output sequences of the system. We have analyzed performance in terms of waveform structure and SNR. This performance analysis yielded a surprising result in that the $L = 5$ pseudorandom waveform performed approximately the same as the unit impulse despite having higher average energy. This behavior was investigated further, and we discovered that pseudorandom waveforms with a deep null in their spectrum produced unstable estimates of the system's impulse response. This instability creates a much wider performance range for pseudorandom waveforms than for the unit impulse waveform. To faithfully reflect the performance of pseudorandom waveforms, we have excluded the simulation trials resulting in large estimation errors deemed to be "outliers". This approach resulted in stable performance results that we feel are representative of the true relative performance between different waveforms.

III. CDMA Communication

This sections background, results, and conclusions…

IV. Final Summary and Conclusions

Potentially a section giving the conclusions from the entire project, but no need to rehash conclusions from individual sections of the project.