AN ABSTRACT OF THE THESIS OF

Qiwei Wang for the degree of Master of Science in Electrical and Computer Engineering presented on November 18, 2013.

Title: Adaptive Broadband Wireless Interference Mitigation Architecture Applying Higher Order Statistics Algorithm for Computer Platform

Abstract approved: ____

Mario E. Magaña

The broadband wireless interference in a computer platform is resulted by multiple electro-magnetic emission sources. This non-Gaussian interference is proved to be double-sided K-distributed in previous research. With the limitation of transmission power and dimension of the device, interference mitigation is an efficient way to improve received signal bit error rate (BER). When applied on the double-sided K-distributed interference in the presence of Gaussian noise, traditional interference/noise cancellation schemes are not able to produce satisfactory results. In this thesis, our target is to find an interference mitigation method with improved BER performance. By introducing a new criterion of goodness, i.e. the cross-cumulant, the new adaptive algorithm based on higher order statistics (HOS) is designed to reconstruct and to cancel the interference in a recursive fashion. It is proved to be effective on both experimental binary transmission system and the OFDM system which is widely applied in modern mobile communication devices. Compared to the previous cancellation method, the BER performance is improved considerably.

Adaptive Broadband Wireless Interference Mitigation Architecture Applying Higher Order Statistics Algorithm for Computer Platform

by

Qiwei Wang

A THESIS

submitted to

Oregon State University

in partial fulfillment of

the requirements for the

degree of

Master of Science

Presented November 18, 2013

Commencement June 2014

Master of Science thesis of Qiwei Wang presented on November 18, 2013.

APPROVED:

Major Professor, representing Electrical and Computer Engineering

Director of the School of Electrical Engineering and Computer Science

Dean of the Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Qiwei Wang, Author

ACKNOWLEDGEMENTS

To pursue a graduate degree is always full of adventures to me, with the company of great professors, my family and my friends. Here I express my most sincere appreciation to my major advisor, Dr. Mario E. Magaña, for his patience, academic excellence and support from the very beginning of my stay at Oregon State University as an international student. There is no way this thesis will be a reality without his guidance.

I also need to thank my family, particularly my parents, Hongyan and Limin, my grandparents Xuejun and Zhenglu, and my dearest sister Shuchen. I can never thank them enough for their love and support from the other side of Pacific Ocean. I would also thank David Yu and his family, who made me feel at home all the time.

During my stay at Corvallis I really feel lucky meeting my wonderful friends. Although I am not able to name them one by one, their contribution to this thesis will never been forget. Especially my gratitude goes to Weiyang Li, Kai Deng and Wenmin Wang for their friendship, help and advices during the days I dedicated to this research.

TABLE OF CONTENTS

TABLE OF CONTENTS(Continued)

PAGE

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF TABLES

Chapter 1-Introduction

In recent years, electronic devices with wireless connection become more and more widely used in people's daily life. Apart from the most commonly used laptops and cell phones, there are tablets, smart phones, digital players and even cameras that are integrated with WIFI, Bluetooth or cellular connections. Advanced wireless standards like 802.11b/g/n or 4G LTE are making all kinds of devices communicate with the speed of hundreds of megabits per second [1][2]. Users of these communication devices are always asking for a better performance. Only those products with fast transmission rate, light weight and long battery life will attract consumers, and all these characteristics require a well-engineered wireless system with most efficient design.

However, noise and interference always prevent a system to be a perfect one. Noise can be generated in any radio transmission/receiving circuits or in transmission channels. Nowadays many wireless devices transmit signals using multiple connections methods and sharing bandwidths, making it impossible to isolate each link completely, then interference becomes inevitable. Once fast rate and low error is required simultaneously, high transmitted power is needed to overcome the noise and interference. Unfortunately, sometimes transmitted power cannot be set as high as we want. In table 1.1 the Federal Communications Commission (FCC) has set a series of standards to limit the emission power of personal digital devices in order to prevent significant interference and human health issues [3], and it is predictable that with the rapid-growing usage of mobile wireless devices the regulations can get even stricter. Another problem, caused by raising the transmitted power, is that it will shorten the battery life. Higher voltage will drain the battery more quickly, resulting in greater battery size and weight.

Class A digital device at distance of 10m	
Frequency of emission (MHz)	Field strength(μV/m)
30-88	90
88-216	150
216-960	210
Above 960	300
Class B digital device at distance of 3m	
Frequency of emission (MHz)	Field strength(μV/m)
30-88	100
88-216	150
216-960	200
Above 960	500

Table 1.1 FCC emission regulations for unintended radiator

Since high transmission power is not an option for the cases stated above, a solution for noise/interference cancellation is needed to achieve the required transmission performance. Even through regulations have been made to limit the electromagnetic emission of electronic devices in table 1.1, they are in terms of relatively long distance, i.e., 3m or 10m [3]. So these regulations are made in order to make sure devices will not interfere with each other. However, as the circuits integrated in larger scale, emission sources get closer and with smaller dimension. According to FCC's regulation, the radio circuit cannot be protected from the interference which is emitted by the device itself. This thesis tries to solve the problem by finding an algorithm that cancels the interference emitted inside a computer platform.

By computer platform we mean a platform with all components that may appear in an electronic computation device like a laptop or tablet, which include crystal clocks, CPUs, RAM, hard drive, interconnect ports and so on. The electromagnetic emission produced by them can be the interference of a wireless transmitter/receiver which is also included in the same platform. All these interferences can be sorted into 2 categories: narrowband noise and broadband noise. The frequency spectrum of narrowband noise lies within the frequency span of the signal of interest, and the noise often appears as that whose peak amplitude is much greater than the signal of interest. Narrowband noise is generated by clocks of a certain frequency. The broadband noise, however, is the combination of several electromagnetic emissions, with a bandwidth that is wider than the antenna receiving bandwidth [4]. The central limit theorem is no longer applicable for this case so the broadband interference has a non-Gaussian behavior. This interference is modeled as a noise signal whose probability density function (PDF) as a double-sided K-distribution [28].

To mitigate interference, 3 main approaches are listed as follows [5] [6]:

a) Suppress the interference at the emission source

b) Decouple the transmission path and interference source

c) Mitigation at the receiving end

The first approach is the “first line of defense”. It tries to suppress the interference at source as much as possible using methods like using optimized digital pulse signal. Since radiation source components in a platform are usually designed separately, the source emission is not always carefully optimized for the platform as a whole. To make the coupling path inefficient, the most common method is to shield the receiver with a metal enclosure. However, this is not always possible because of the weight, size and cost requirement of modern wireless products. In this thesis, we use approach of c), that is, we try to find a technique that can be applied in the receiver to cancel the interference signal. This approach can be cost-effective and easy to implement in a mobile computer platform. The basic idea is to first estimate the interference and then subtract the estimated interference signal from the actual received signal. If a good estimation is made, the outcome will have a better bit error rate (BER) performance.

Here we try to find an interference mitigation mechanism for an OFDM system and implement the cancellation algorithm on a typical OFDM receiver. OFDM stands for Orthogonal Frequency Division Multiplexing, whose message symbols are separately modulated by a series of subcarriers orthogonal to each other. The advantage of this technique is that each modulated subcarrier only occupies a very small bandwidth, so it is robust with frequency selective fading and multipath fading [14]. With the help of the pilot symbol inserted in each symbol frame, the fading channel issue stated above can be solved in a wireless system. And since each subcarrier is orthogonal, a MIMO (Multiple Input Multiple Output) system can be realized with OFDM. Because of the advantages stated above, it is widely used in modern wireless communication standards including IEEE 802.11a/g/n, IEEE 802.16, DVB-T and Long Term Evolution (LTE) [5]. Since an OFDM system is somewhat sensitive to the inter-carrier interference produced by carrier frequency shift and channel noise, there are many studies on channel estimation and interference cancellation schemes. M. Chang in [15] presented a method of channel estimation based on least-squares (LS) and minimum mean-square estimation (MMSE).

This method estimates the channel impulse response using the accurate response at the pilot point, which is known to the receiver. L.Davis in [16] constructed a state-space model for the fading channel and used it for a maximum a posteriori (MAP) equalizer to estimate the Channel. In [17] the author stated a method to estimate and cancel the Gaussian noise at the receiver. It first assumes the noise after the receiver still has a zero mean but it is correlated. Then use the correlation matrix of the noise and the sampled correlation vector in the Yule-Walker equation to obtain the estimates of the coefficient of an FIR filter. After that the FIR filter reconstructs the estimate of the noise. J.S.Dhanoa in [18] presented a blind estimation of noise with minimum information by an Evolutionary Algorithm (EA). EA is an iterative algorithm that imitates biological evolution. A pool of possible estimates of the noise is randomly generated and a certain amount of the value of the estimation can be exchanged with each other during an iteration. In each generation, the one with worst match of the actual noise is eliminated and finally the iteration converges to an optimized value.

E. Alban [5] presented several estimation and mitigation methods for this particular case. For the narrow band case, a Normalized Linear Mean Square (NLMS) adaptive filter can be used to predict and cancel the interference signal and it is proved to be efficient. The BER can be improved by as much as 10dB. For the broadband case, several different methods were applied but the results were not satisfactory. Since the distribution of the interference is already known, to estimate the parameters is the most direct approach. The Method of Moment (MoM) is the one which has low complexity, but its performance depends on the sample size, while only limited number of pilot symbols in an OFDM symbol frame can be used for estimation. For the special case of K-distribution the Fractional Moments [7] method can produce an estimate with lower variance by introducing a new ratio of two moments with different order and cancel the b parameter. Another method that can achieve similar accuracy is described in [8] by Blacknell and Tough. We estimate ν by the empirical estimation of $X^2 \log|X|$, X^2 and $\log|X|$, in which X is the double sided K-distributed random variable. A Maximum Likelihood (ML) estimator can also be obtained by maximizing the log-likelihood equation, but a close form of solution cannot be obtained [8]. Based on Iskander et al.'s method to derive a close form expression for one parameter [9], a new method is developed in [6] which has a close form expression for both parameters of K-distribution by approximating the modified Bessel function of the second kind with a much simpler form. By Monte Carlo simulation the new method outperformed the other methods mentioned above when the number of samples is small. Other parameter estimation methods include the EM algorithm [10][11] and neural network methods [12][13].

In order to mitigate the broadband K-distributed interference, a method using an extended Kalman filter is presented in [5]. First a state-space model describing the K-distributed interference is required to apply the Kalman filter [19]. E. Alban derived this state-space model following the procedure introduced by Field and Tough in [20]. The vector form of a stochastic differential equation (SDE) is derived based on the fact that a double-sided K-distribution random variable can be expressed as the product of a Gaussian random variable and a square-rooted Gamma random variable [21]. These two random variables can be generated by a stochastic differential equation separately, so the product of them can be expressed by SDE using Ito's formula. After the SDE is found, the difference equation can be obtained using Euler's method. An extended Kalman filter is built based on the state-space model as described in [5][19]. This method is proved to be functional as it can lower the BER in a baseband equivalent OFDM receiver simulation. Unfortunately, its efficiency is questionable. From the results in [5], we can see that the improvement of the BER is around 5dB in the best of the cases, and the absolute value of BER is between 0.1 and 0.01, which is not acceptable in an actual OFDM system.

In this thesis, a new method using an adaptive filter is presented to improve the method stated above. An adaptive filter requires minimum knowledge of the interference signal and updates the filter coefficients according to certain "criterion of goodness". It works in an iterative fashion and converges if the step size coefficient is selected judiciously. Compared to the Kalman filter stated above, adaptive filters are more widely applied to multiple kinds of interference signals. The application of adaptive algorithms to wireless communication systems has been widely studied. In [22] M. Lee presents a design of a repeater that reproduces the channel interference in an OFDM wireless system and tries to cancel this interference at the receiver. The repeater uses an adaptive algorithm to control an FIR transversal filter that predicts the channel interference. The error of the predicted signal (which includes the interference) is compared with the actual signal seen at the receiver, which is the "criterion of goodness" to input to the adaptive algorithm. M. Kinoshita et al. applied an adaptive filter in an OFDM system in [23]. An "adaptive prefix" is added at the beginning of each OFDM symbol frame just like the cyclic prefix. For the time interval that the "adaptive prefix" is transmitted, the adaptive algorithm updates the FIR filter coefficients to predict the noise. Assuming the coefficients have already converged to their stable state, the adaptive algorithm is shut off when the adaptive prefix transmission is finished and the actual OFDM symbols arrive. An adaptive Wiener filter is used in [24] for channel estimation. In order to have the estimation of the channel frequency response, the author first obtains an accurate response of the transformed domain from the location where pilots are inserted. Then inverse transforms it to the time domain. Both Normalized Least-Mean-Square (NLMS) and Recursive Least-Square (RLS) algorithms are applied to the adaptive filter to predict the actual channel impulse response. In [25], T. Zhang et al. combine a Volterra adaptive filter and an LMS filter to cancel the inter-carrier interference (ICI) in an OFDM receiver.

A new adaptive algorithm is presented in this thesis and is applied to an OFDM receiver. An adaptive algorithm using higher order statistics are first presented in [26] by D. C. Shin and C. L. Nikias. Instead of using statistics of second order as the "criterion of goodness" like usual adaptive algorithms do (LMS, NLMS, Leaky LMS and so on), this algorithm uses a higher order statistics, i.e. the cumulant. Two advantages of this method make it a good candidate for the cancellation of wideband K-distributed interference here. First, since Gaussian noise and K-distributed interference coexist in the platform noise, it is important to minimize the effect of Gaussian noise in the canceller. Fortunately, Gaussian noise's cumulant of third order above is equal to zero [27], so it will not influence the performance of the canceller. Second, in [26] the author showed that this algorithm can be used in both narrowband and wideband case, with no prerequisite of the type of noise. That means it may work better on non-Gaussian, wideband interference compared to traditional adaptive algorithms like LMS and RLS. The only drawback of this algorithm is that it achieves better performance over traditional adaptive algorithms at the expense of more computations, but we can shorten the computation time by parallelism or efficient VLSI implementation. In this thesis we chose the fourth-order of cumulant as the “criterion of goodness” to achieve a balance of computation complexity and cancellation performance. Similar to what is done in NLMS, the cancellation is performed in a normalized fashion to achieve more stable performance for difference input signals.

The fourth-order statistics (FOS) adaptive canceller can be implemented in an OFDM receiver by assuming that a reference signal can be collected by another antenna in the platform. As a result of difference locations and properties of the reference antenna and receiving antenna, the reference signal is different than the K-distributed interference in both time and frequency domains. However, with the help of the adaptive algorithm stated above, the FOS canceller can reconstruct an estimate of the K-distributed interference and cancel it out at the output. Since the OFDM signals are modulated from QAM symbols that include in-phase and quadrature components, we need 2 FOS cancellers in the system to do the cancellation as this canceller cannot work with complex signals. Using computer simulation we can show that this canceller is very effective with or without the existence of Gaussian noise. Even though the BER performance is not very stable at low SNR, the FOS canceller outperforms the extended Kalman filter described in [5] by achieving a lower BER with multiple input SNR.

The thesis is organized as follows: in Chapter 2 we show the basic nature of noise cancelling filters and adaptive filter algorithms. Mathematical principles of the adaptive filter is introduced and analyzed. In Chapter 3 we focus on the specific problem. First, the derivation and properties of double-sided K-distributed interference is introduced. Then we present the concept of cumulant and the structure of a high order cumulent adaptive algorithm. Next in Chapter 4 the design of FOS canceller and the implement-tation on a typical OFDM system is presented. The canceller performance is verified via the computer simulation. In Chapter 5 we conclude the thesis and propose some future work that can be done using FOS-based canceller.

Chapter 2-Theory of Noise/Interference Cancellation Filters

The estimation of a random signal from another given signal is one of the most important problems in signal processing. As this technique can be applied to multiple fields including speech recognition, radar detection or image reconstruction, it has been studied since the 1940s. The cancellation of random noise or interference is also part of the application of random signal estimation, because the noise can be subtracted from the corrupted signal if a good estimate is available. A digital low-pass, high-pass or band-pass filter may work on some ideal occasions, but they are rarely the optimum choice. That is why multiple new estimation and prediction filters are designed, including digital Wiener filter and discrete Kalman filter. However, with all these specially designed estimation filters, there are still some problem that cannot be solved since all these filters need some additional requirements of the noise or interference signal. An adaptive filter may be applied on those occasions. It works in an iterative way and requires minimum knowledge of the input signal

A typical random signal estimation problem is modeled as follows: suppose there is a random process $s[n]$ that is of our interest. However, the observation of this process $x[n]$ is corrupted by another random signal $v[n]$, that is:

$$x[n] = s[n] + v[n] \tag{2.1}$$

In order to retrieve $s[n]$, a digital filter is applied to the observation-based signal $y[n]$ to produce $\hat{s}[n]$, the estimate of $s[n]$. $y[n]$ depends on the actual design of a specific filter application. To make the estimation $\hat{s}[n]$ accurate, we need to optimize a "criterion of goodness". A very straightforward idea should be the estimation error

$$e[n] = \left| s[n] - \hat{s}[n] \right| \tag{2.2}$$

The structure of the estimation procedure is illustrated in Figure 2.1. This filter structure can be used for:

a) Filtering/Smoothing: to estimate $s[n]$ given current, past or future values of $x[n]$, as stated above. If future values are used, i.e., $H(z)$ is an non-causal filter, it is called smoothing instead of filtering

b) Prediction: if $s[n] = x[n+1]$, $H(z)$ is a causal filter. When we use $H(z)$ to estimate $s[n]$ we are actually predicting $x[n]$ using its previous values.

c) Deconvolution: if $x[n] = s[n] * g[n] + v[n]$, retrieving $s[n]$ is a deconvolution operation.

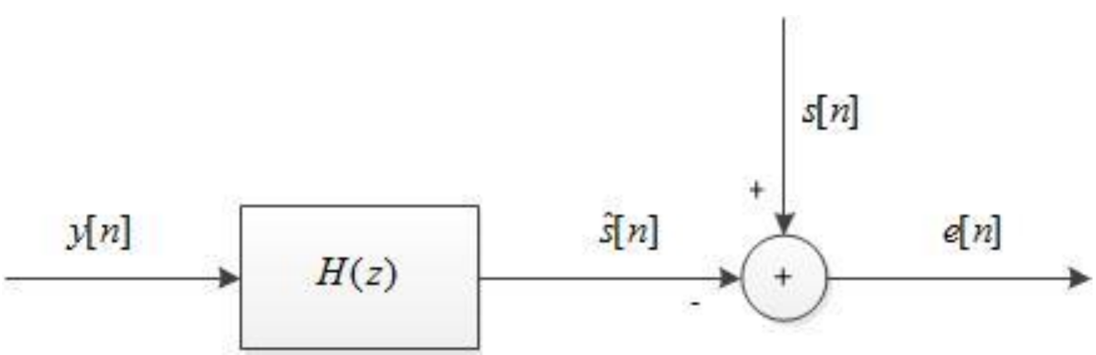


Figure 2.1 Estimation Filter Structure

Next we will introduce two widely used non-adaptive filters: Wiener filter and discrete Kalman filter. The former can be applied to estimate stationary random process, while the latter is applicable to non-stationary case as well. They are called non-adaptive because during the filtering procedure, no filter coefficient changes it value. Then adaptive algorithm will be introduced by mathematical analysis. The coefficients of an adaptive filter change through time to adapt the varying nature of input noise.

2.1 Non-adaptive filtering

2.1.1 Wiener filtering

The design of an FIR Wiener filter was first introduced by Norbert Wiener in the 1940s. This design assumed that $x[n]$ and $s[n]$ are jointly wide-sense stationary (WSS) with autocorrelations $r_x[k]$ and $r_s[k]$, respectively, and the cross-correlation $r_{sx}[k]$. $r_x[k]$, $r_s[k]$ and $r_{sx}[k]$ are assumed known. To find the optimum filter coefficients of $W(z)$, this filter tries to minimize the mean-square error (MSE). In other words, Wiener filter uses the "criterion of goodness" as follows:

$$\xi = E\{|e[n]|^2\}, \tag{2.3}$$

where $e[n]$ is defined in Eq. 2.2.

Assuming we have an N th order filter with transfer function:

$$W(z) = \sum_{n=0}^{N-1} w[n] z^{-n} \tag{2.4}$$

We need to find the value of $w(n)$ that minimizes ξ.

If $x[n]$ is the input of the filter, the output estimation of $s[n]$ is the result of the following convolution:

$$\hat{s}[n] = \sum_{l=0}^{N-1} w[l]x[n-l] \tag{2.5}$$

In order to minimize the MSE, we need set the derivatives of ξ with respect to $w^*[k]$ equal to zero, for $k = 0,1,2,...,N$, i.e.

$$\frac{\partial}{\partial w^*[k]}\xi = \frac{\partial}{\partial w^*[k]}E\{e[n]e^*[n]\} = E\{e[n]\frac{\partial}{\partial w^*[k]}e^*[n]\} = 0 \tag{2.6}$$

From Eq. 2.2 and 2.5 it follows that

$$e^*[n] = s^*[n] - \sum_{l=0}^{N-1} w^*[l]x^*[n-l] \tag{2.7}$$

Then we have

$$\frac{\partial}{\partial w^*[k]}e^*[n] = -x^*[n-k] \tag{2.8}$$

Substituting Eq. 2.8, 2.5 and 2.2 into Eq. 2.6 we have

$$E\{s[n]x^*[n-k] - \sum_{l=0}^{N-1} w[l]x[n-l]x*[n-k]\} = 0 \tag{2.9}$$

Notice that $E\{s[n]x^*[n-k]\} = r_{sx}[k]$ and $E\{x[n-l]x^*[n-k]\} = r_x[k-l]$. With the WSS assumption stated before, we now have

$$\sum_{l=0}^{N-1} w[l] r_x[k-l] = r_{sx}[k], k = 0,1,2,...,N \;; \tag{2.10}$$

This is a set of N linear equations. We can represent them in matrix form

$$\mathbf{R}_x \mathbf{w} = \mathbf{r}_{sx} \tag{2.11}$$

$\mathbf{R}_x$ is an N by N Hermitian Toeplitz matrix. The element at the (m,n) position is $r_x[m-n]$. $\mathbf{w}$ is a 1 by N column vector of filter coefficients. $\mathbf{r}_{sx}$ is a 1 by N column vector and the n th value of it is $r_{sx}[k]$. Eq. 2.11 is called Wiener-Hopf equation. By solving this equation we can have the optimized coefficients of a Wiener filter

$$\mathbf{w} = \mathbf{R}_x^{-1} \mathbf{r}_{sx} \tag{2.12}$$

And we can obtain the minimum mean-square error by substituting Eq. 2.12 into Eq. 2.3

$$\xi_{\min} = r_d(0) - \sum_{l=0}^{N-1} w[l] r_{dx}^*[l] = r_d(0) - \mathbf{r}_{sx}^H \mathbf{R}_x^{-1} \mathbf{r}_{sx} \tag{2.13}$$

The superscript H stands for Hermitian transpose.

As stated above, Wiener filtering design can be applied to multiple tasks including filtering, linear prediction and noise cancellation. The approach of noise cancellation by Wiener filter can be further developed to the method we use in this thesis for interference cancellation. Unlike the filtering problem presuming the autocorrelation of the noise is known, the noise-cancellation problem need a reference signal input, generated by a secondary receiver or sensor. The reference signal is not equal to the noise itself but has to be correlated to it. This may be the result of the difference of sensor characteristics, propagation path or any other reasons depend on the actual application. Then a noise canceller with the structure illustrated in Figure 2.2 can be designed:

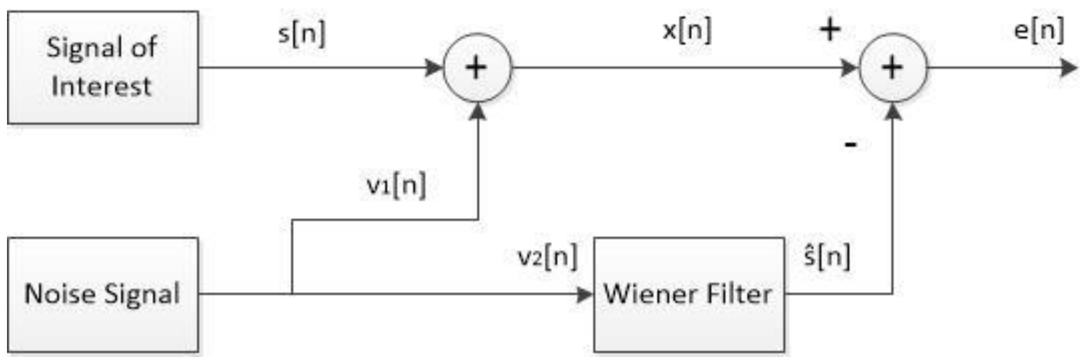


Figure 2.2: Noise canceller with Wiener filter

If we denote the actual noise as $v_1[n]$ and the reference signal as $v_2[n]$, we will have the noise-corrupted signal as

$$x[n] = s[n] + v_1[n], \tag{2.14}$$

where $s[n]$ is the signal of interests. We try to use the Wiener filter to estimate the actual noise $v_1[n]$ from the reference signal $v_2[n]$. That is, we estimate $s[n]$ by

$$\hat{s}[n] = x[n] - \hat{v}_1[n] \tag{2.15}$$

The Wiener-Hopf equation for the noise canceller can be derived directly from the Wiener-Hopf equation of the Wiener filter. By comparing the structure of the noise canceller and the Wiener filtering problem discussed before, we can use the Wiener-Hopf equation of the filter

$$\mathbf{R}_{v_2}\mathbf{w} = \mathbf{r}_{v_1v_2} \tag{2.16}$$

The k th element of $\mathbf{r}_{v_1v_2}$ is

$$r_{v_1v_2}[k] = E\{v_1[n]v_2^*[n-k]\} = E\{x[n]v_2^*[n-k]\} - E\{s[n]v_2^*[n-k]\} \tag{2.17}$$

If we assume that $s[n]$ is uncorrelated with $v_2[n]$ (which makes sense because the noise should not be statiscally related with the transmitted signal), then

$$E\{s[n]v_2^*[n-k]\} = 0 \tag{2.18}$$

And it follows that

$$r_{v_1v_2}[k] = E\{x[n]v_2^*[n-k]\} = r_{xv_2}[k] \tag{2.19}$$

So the Wiener-Hopf equation of this noise canceller is

$$\mathbf{R}_{v_2}\mathbf{w} = \mathbf{r}_{xv_2} \tag{2.20}$$

By solving this function we will be able to find the optimum filter coefficients that produce the minimum MSE estimation of the signal of interest.

The design of a Wiener noise canceller, together with the idea of using a reference signal input, is widely applied to create other more complicated noise/interference cancellation systems especially the cancellers with an adaptive algorithm, which will be presented later in this thesis.

2.1.2 Discrete Kalman filter

While a Wiener filter can produce a minimum MSE estimation by a relatively simple and stable structure, its drawback is obvious. It requires that both the signal of interest $s[n]$ and the observation $x[n]$ be jointly WSS with the knowledge of autocorrelation and cross-correlation. Unfortunately for most practical cases, WSS is not guaranteed. A discrete Kalman filter can be designed to solve the problem. Working in a recursive fashion, the discrete Kalman filter predicts the upcoming input signal using the signal already received, and updates the estimate every time a new upcoming input signal is available. The following derivation is based on the state-space model of $s[n]$, and assuming $s[n]$ is an auto-regressive (AR) process corrupted by a zero mean noise. The discussion starts with the stationary case but soon a more generalized result will be presented.

Assuming that the signal of interests is a WSS AR(p) process generated by

$$s[n] = \sum_{k=1}^{p} a[k]x[n-k] + w[n] \tag{2.21}$$

the observation is

$$x[n] = s[n] + v[n] \tag{2.22}$$

From Eqs. 2.21 and 2.22 we can construct the state-space model of the process

$$\mathbf{s}[n] = \mathbf{A}\mathbf{s}[n-1] + \mathbf{w}[n] \tag{2.23}$$

$$x[n] = \mathbf{c}^T\mathbf{s}[n] + v[n]\,, \tag{2.24}$$

where

$$\mathbf{s}[n] = \begin{pmatrix} s[n] & s[n-1] & \dots & s[n-p-1] \end{pmatrix}^T$$

$$\mathbf{A} = \begin{pmatrix} a(1) & a(2) & \dots & a(p-1) & a(p) \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

$$\mathbf{w}[n] = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \end{pmatrix}^T w[n]$$

$$\mathbf{c} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \end{pmatrix}$$

The superscript T stands for transpose.

Now, we update the estimate of $\mathbf{s}[n]$ as follows:

$$\mathbf{s}[n] = \mathbf{A}\mathbf{s}[n-1] + \mathbf{K}(x[n] - \mathbf{c}^T\mathbf{A}\hat{\mathbf{s}}[n-1]) \tag{2.25}$$

K is a 1 by p constant vector called Kalman gain vector. Similar to a Wiener filter, we need to find the value of **K** that minimizes the MSE, i.e., $E\{\|\mathbf{s}[n] - \hat{\mathbf{s}}[n]\|^2\}$.

We can generalize the result of Eq. 2.23 by substituting the constant matrix $\mathbf{A}$ with a time-varying matrix $\mathbf{A}[n-1]$, and $\mathbf{c}$ becomes a time varying q by p matrix $\mathbf{c}[n]$. Then $\mathbf{s}[n]$ will be a non-stationary process but all the derivation and applications of discrete Kalman filter are still valid. Then Eqs. 2.24 and Eq. 2.25 become

$$\mathbf{x}[n] = \mathbf{C}[n]^T \mathbf{s}[n] + \mathbf{v}[n] \tag{2.26}$$

$$\mathbf{s}[n] = \mathbf{A}[n\text{-}1]\mathbf{s}[n-1] + \mathbf{K}(\mathbf{x}[n] - \mathbf{c}[n]^T \mathbf{A}[n\text{-}1]\hat{\mathbf{s}}[n-1]) \tag{2.27}$$

If we denote $\hat{\mathbf{s}}[m|n]$ as the estimate of $\hat{\mathbf{s}}[m]$ at time m given the first n observations, then the corresponding estimation error is

$$\mathbf{e}[m|n] = \mathbf{s}[m] - \hat{\mathbf{s}}[m|n] \tag{2.28}$$

and the error covariance matrix is

$$\mathbf{P}(m|n) = \mathbf{E}\{\mathbf{e}[m|n]\mathbf{e}[m|n]^H\} \tag{2.29}$$

The Kalman gain matrix $\mathbf{K}$ can be found in two steps:

Step 1: Find $\hat{\mathbf{s}}[n|n-1]$ given $\hat{\mathbf{s}}[n-1|n-1]$ and the corresponding error covariance matrix.

Step2: Find the optimized $\hat{\mathbf{s}}[n|n]$ given the new observation $\mathbf{x}[n]$. Optimization means to find $\mathbf{K}$ that minimizes the following MSE:

$$\xi[n] = E\{\|e(n|n)\|^2\} = tr\{\mathbf{P}(n|n)\} \tag{2.30}$$

In step 1, we can make a prediction of $\mathbf{s}[n]$ from Eq. 2.23. Since the noise $\mathbf{w}[n]$ is zero mean, the estimate of $\mathbf{s}[n]$ based on the first n-1 observations is

$$\hat{\mathbf{s}}[n \mid n-1] = \mathbf{A}[n\text{-}1]\hat{\mathbf{s}}[n-1 \mid n-1] \tag{2.31}$$

It follows that the estimation error is

$$\mathbf{e}[n \mid n-1] = \mathbf{s}[n] - \hat{\mathbf{s}}[n \mid n-1] = \mathbf{A}[n-1]\mathbf{e}[n-1 \mid n-1] + \mathbf{w}[n] \tag{2.32}$$

Notice that if we assume $\hat{\mathbf{s}}[n-1 \mid n-1]$ is an unbiased estimate, $\hat{\mathbf{s}}[n \mid n-1]$ will also be unbiased since $\mathbf{w}[n]$ is zero mean.

Now, since $\mathbf{w}[n]$ and $\mathbf{e}[n-1 \mid n-1]$ are uncorrelated, the correlation matrix of $e[n \mid n-1]$ is

$$\mathbf{P}[n \mid n-1] = \mathbf{A}[n-1]\mathbf{P}[n-1 \mid n-1]\mathbf{A}[n-1]^H + \mathbf{Q}_w[n], \tag{2.33}$$

where $\mathbf{Q}_w[n]$ is the covariance matrix of the noise $\mathbf{w}[n]$.

In step two we try to find the expression of $\mathbf{P}[n \mid n]$ so a minimum value can be found. We update the estimate of $\mathbf{s}[n]$ by the following linear estimator:

$$\hat{\mathbf{s}}[n \mid n] = \mathbf{K}^{'}[n]\hat{\mathbf{s}}[n-1 \mid n-1] + \mathbf{K}[n]\mathbf{x}[n] \tag{2.33}$$

First we need to make sure that $\hat{\mathbf{s}}[n \mid n]$ is still an unbiased estimate, because this is what we assumed in step one. We need this for the next iteration from $\hat{\mathbf{s}}[n \mid n]$ to $\hat{\mathbf{s}}[n+1 \mid n+1]$.

The estimate error can be calculated as

$$\begin{aligned}\mathbf{e}[n\,|\,n] &= \mathbf{s}[n]-\mathbf{K}^{'}[n]\hat{\mathbf{s}}[n-1\,|\,n-1]-\mathbf{K}[n]\mathbf{x}[n] \\ &= [\mathbf{I}-\mathbf{K}^{'}[n]-\mathbf{K}[n]\mathbf{C}[n]]\mathbf{s}[n]+\mathbf{K}^{'}[n]\mathbf{e}[n\,|\,n-1]-\mathbf{K}[n]\mathbf{v}[n]\end{aligned} \quad (2.34)$$

We already know $E\{\mathbf{e}[n\,|\,n]\}=0$ and $E\{\mathbf{v}[n]\}=0$, so if we set

$$\mathbf{K}^{'}[n]=\mathbf{I}-\mathbf{K}[n]\mathbf{C}[n], \quad (2.35)$$

we will have $E\{\mathbf{e}[n\,|\,n]\}=0$ and $\hat{\mathbf{s}}[n\,|\,n]$ is an unbiased estimator.

Substitute Eq. 2.35 to Eq. 2.33 and 2.34 for simplicity, i.e.

$$\hat{\mathbf{s}}[n\,|\,n]=\hat{\mathbf{s}}[n\,|\,n-1]+\mathbf{K}[n](\mathbf{x}[n]-\mathbf{C}[n]\hat{\mathbf{s}}[n\,|\,n-1]) \quad (2.36)$$

$$\mathbf{e}[n\,|\,n]=[\mathbf{I}-\mathbf{K}[n]\mathbf{C}[n]]\mathbf{e}[n\,|\,n-1]-\mathbf{K}[n]\mathbf{v}[n] \quad (2.37)$$

Similar to what we do in step one, we can have the error covariance matrix given the fact that $\mathbf{v}[n]$ is uncorrelated with $\mathbf{e}[n\,|\,n]$, namely,

$$\mathbf{P}[n\,|\,n]=[\mathbf{I}-\mathbf{K}[n]\mathbf{C}[n]]\mathbf{P}[n\,|\,n-1][\mathbf{I}-\mathbf{K}[n]\mathbf{C}[n]]^{H}-\mathbf{K}[n]\mathbf{Q}_{v}[n]\mathbf{K}[n]^{H}, \quad (2.38)$$

where $\mathbf{Q}_{v}[n]$ is the covariance matrix for noise $\mathbf{v}[n]$.

Now, to find the value of $\mathbf{K}[n]$ that minimizes the MSE, we set the first order derivative of $tr\{\mathbf{P}[n\,|\,n]\}$ with respect to $\mathbf{K}[n]$ zero, and find $\mathbf{K}[n]$ as follows:

$$\mathbf{K}[n]=\mathbf{P}[n\,|\,n-1]\mathbf{C}^{H}[n](\mathbf{C}[n]\mathbf{P}[n\,|\,n-1]\mathbf{C}^{H}[n]+\mathbf{Q}_{v}[n]) \quad (2.39)$$

This completes the derivation of the Kalman gain vector, $\mathbf{K}[n]$, at time n. For time $n+1$, what needed to do is to update the estimation by taking into account the new observation $\mathbf{x}[n+1]$ and calculate the new $\mathbf{K}[n+1]$.

The only problem left is the initial condition. What we need is an unbiased estimate at time 0 without any observation. So it can be set as

$$\hat{\mathbf{s}}[0\,|\,0] = E\{\mathbf{s}[0]\} \tag{2.40}$$

Notice that in the discrete Kalman filter design, the Kalman gain matrix $\mathbf{K}[n]$ and the error covariance matrix $\mathbf{P}[n\,|\,n]$ are not functions of the actual signal of interest, $\mathbf{s}[n]$. This implies 2 things: first, we can compute the Kalman gain matrix even before the filtering procedure starts. That means a discrete Kalman filter can be very computational effective, since we do not need to go through the complex derivation for each recursive time. On the other hand, this means that if we can do some real-time updating to the filtering procedure, we may achieve even better performance.

2.2 Adaptive Filtering

Almost every estimation and prediction technique, including Wiener filtering we discussed above, requires stationary signals. Unfortunately, most of practical filtering applications involve non-stationary signals. Apart from the discrete Kalman filter described in 2.1.2, adaptive filtering is another recursive method that can be applied to non-stationary processes. Unlike the discrete Kalman filter, an adaptive filter has the same basic structure as a FIR Wiener filter but it measures the estimation error in each recursion and uses it to update the filter coefficient directly. Even for the stationary case, an adaptive filter is still advantageous at some point. In order to solve the Wiener-Hopf Eq. 2.11, we need to calculate the inverse of the autocorrelation matrix, $\mathbf{R}_x$. This may be computational costly for a filter with high order. Also, $\mathbf{R}_x$ can be almost singular, which may lead to an inaccurate calculation of $\mathbf{R}_x^{-1}$. More importantly, in most cases we have no knowledge of the ensemble average of $\mathbf{R}_x$ and $\mathbf{r}_{sx}$, so we need at least the first P measurements to complete the estimate using a time average (it is assumed the processes are ergodic). If an estimation of high accuracy is required, P needs to be large so a large time delay is introduced.

2.2.1 Adaptive filter architecture

Consider a Wiener filter for a WSS signal. The coefficient can be found by solving the Wiener-Hopf Eq. 2.11. However, if the input signal is non-stationary, we are not able to find the autocorrelation matrix $\mathbf{R}_x$ and the cross-correlation matrix $\mathbf{r}_{sx}$. So, to solve for the Wiener coefficients for a non-stationary input, we need to solve the following modified Wiener-Hopf equation instead:

$$\mathbf{R}_x[n]\mathbf{w}_n = \mathbf{r}_{sx}[n] \tag{2.41}$$

And the optimized Wiener filter coefficients become time-varying, i.e.

$$\hat{s}[n] = \sum_{k=0}^{N-1} w_n[k]x[n-k], \tag{2.42}$$

where $w_n[k]$ is the k th Wiener filter coefficient at time n. Unfortunately, it is difficult to design the time-varying filter directly since we need to optimize it at every iteration n. A better way to do this is through iteration. Instead of optimizing $w_n[k]$ for all n at one time, we only need to update it recursively by

$$\mathbf{w}_{n+1}[k] = \mathbf{w}_n[k] + \Delta\mathbf{w}_n \tag{2.43}$$

As long as $\Delta\mathbf{w}_n$ moves towards the optimum direction, we will find an optimized solution after enough iterations (assuming the convergence condition is satisfied).

The structure of a typical adaptive filter system is illustrated in Figure 2.3.

The key component of an adaptive filter system is the adaptive algorithm block. This block has two inputs, one is the error signal and the other is some reference signal. In the special case depicted in Figure 2.3, they are the observation signal $x[n]$ and the error signal $e[n] = \hat{s}[n] - s[n]$, respectively. In a later discussion we will see that these two signals can be different for different applications. Although the algorithm cannot find the optimized filter coefficients directly, it finds $\Delta\mathbf{w}_n$ that modifies the filter coefficients to have a better "criterion of goodness", according to the knowledge of the error signal and the reference signal. For the stationary case, it will converge to the solution of the Wiener-Hopf equation. For the non-stationary case, the filter can adapt to the changing statistics and change the coefficient accordingly.

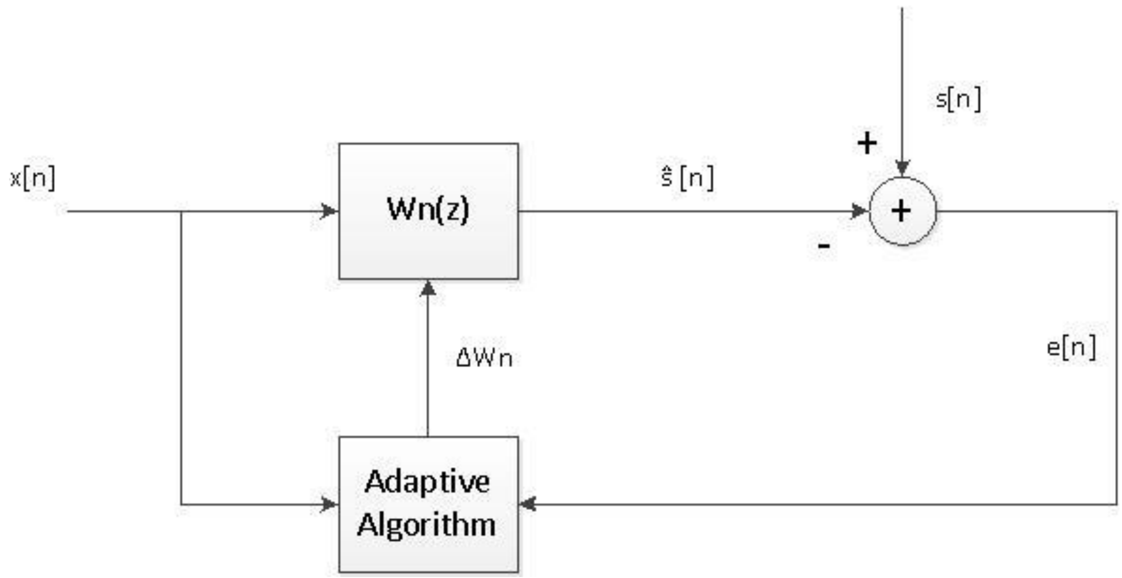


Figure 2.3 Block diagram for a typical adaptive filter system

A typical FIR adaptive filter using minimum MSE estimation has the form illustrated in Figure 2.4

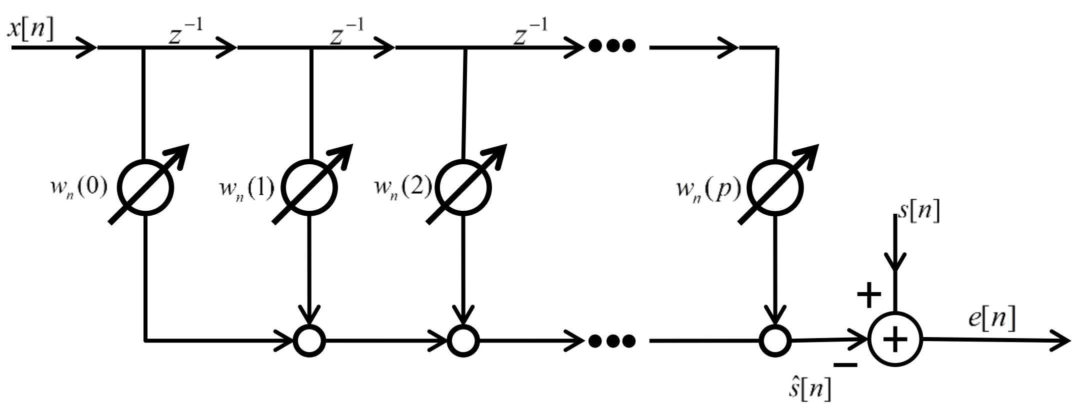


Figure 2.4 An FIR adaptive filter

2.2.2 The steepest descent method

The steepest decent method is one of the most commonly used iterative methods to find the optimized $\Delta\mathbf{w}_n$. The general idea of this method is as follows: If $\mathbf{w}_n$ is the best estimate we have at iteration time n, then at iteration time n+1, we try to move $\mathbf{w}_n$ toward the optimized direction for a short length. In a quadratic surface for instance, $\mathbf{w}_n$ is moved in the direction of the maximum descent, i.e., the direction given by the gradient of the error, and the length moved is scaled by the step-size coefficient μ. Even though this "movement" is only based on the current "location" of $\mathbf{w}_n$ and maybe is not toward the accurate optimization, it will bring $\mathbf{w}_n$ closer to the optimized point before next iteration.

For an adaptive filter of order N, the gradient of the error is

$$\nabla\xi[n] = \left(\frac{\partial\xi[n]}{\partial w[0]} \quad \frac{\partial\xi[n]}{\partial w[1]} \quad \cdots \quad \frac{\partial\xi[n]}{\partial w[N-1]} \right). \tag{2.44}$$

The gradient vector is always orthogonal to the contour of $\mathbf{w}_n$ in the current iteration. The gradient is towards the steepest ascent but what we need is its opposite direction. Thus, the way to find the $\mathbf{w}_n$ is given as follows:

$$\mathbf{w}_{n+1}[k] = \mathbf{w}_n[k] - \mu\nabla\xi[n] \tag{2.45}$$

The step-size coefficient μ can be critical for the convergence behavior of an adaptive filter. If it is too small, the filter will take longer time to converge. But a large μ will cause it not to converge at all or introduce some other unstable issues. For an adaptive filter with a stationary input, the coefficient will converge to the solution of Wiener-Hopf equation as long as the following condition is satisfied [19]:

$$0 < \mu < \frac{2}{\lambda_{\max}}, \tag{2.46}$$

where $\lambda_{\max}$ is the maximum eigenvalue of the autocorrelation matrix $\mathbf{R}_x$.

Next we will find the convergence behavior or the "learning curve" of the error $\xi[n]$ if the condition above is satisfied. We define the error vector $\mathbf{c}_n$ as

$$\mathbf{c}_n = \mathbf{w}_n - \mathbf{w}, \tag{2.47}$$

where $\mathbf{w}_n$ is the filter coefficient vector at iteration time n and $\mathbf{w}$ is the solution of Wiener-Hopf equation, i.e., the optimized coefficient vector. The eigenvalue factorization of $\mathbf{R}_x$ is given as

$$\mathbf{R}_x = \mathbf{V}\Lambda\mathbf{V}^H \tag{2.48}$$

Now if we define

$$\mathbf{u}_n = \mathbf{V}^H\mathbf{c}_n \tag{2.49}$$

It follows that [15]

$$\xi[n] = \xi_{\min} + \sum_{k=0}^{N-1} \lambda_k (1-\mu\lambda_k)^{2n} \left|u_0(k)\right|^2, \tag{2.50}$$

where $u_0(k)$ is the k th element of the rotated error vector $\mathbf{u}_n$ at time 0.

$\mathbf{u}_n$ is still the error vector but in a transformed coordinate system with respect to the eigenvectors of $\mathbf{R}_x$. From Eq. 2.50 we can see $\xi[n]$ decays to the minimum value exponentially as long as $(1-\mu\lambda_k)<1$ for all k. This condition is equivalent to the condition described by Eq.2.46. Notice that the derivation above is based on the fact that the input signals are WSS, so simply setting μ to satisfy condition 2.46 will not guarantee convergence or stability in a practical application. This result will be improved in both LMS algorithm and the FOS algorithm which will be introduced later in this thesis.

2.2.3 The LMS algorithm

2.2.3.1 Algorithm and convergence condition

A least mean-square (LMS) algorithm sets the error signal as follows:

$$e[n]=\hat{s}[n]-s[n], \tag{2.51}$$

and tries to minimize the mean-square error

$$\xi[n]=E\{|e[n]|^2\} \tag{2.52}$$

To apply the steepest descent method, we substitute $\xi[n]$ into Eq.2.45 and it follows that

$$\nabla\xi[n]=\nabla E\{|e[n]|^2\}=E\{e[n]\nabla e^*[n]\}=-E\{e[n]\mathbf{x}^*[n]\} \qquad (2.52)$$

Then Eq.2.45 becomes

$$\mathbf{w}_{n+1}[k] = \mathbf{w}_n[k] + \mu E\{e[n]\mathbf{x}^*[n]\}. \tag{2.53}$$

However, since we do not have any knowledge of the expectation $E\{e[n]\mathbf{x}^*[n]\}$, the ensemble average is needed for estimation, i.e.

$$\hat{E}\{e[n]\mathbf{x}^*[n]\} = \frac{1}{L}\sum_{l=0}^{L-1} e[n-l]\mathbf{x}^*[n-l] \tag{2.54}$$

Then, if we set the initial condition $\mathbf{w}_0 = \mathbf{0}$, the LMS algorithm can be described as

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \frac{\mu}{L}\sum_{l=0}^{L-1} e[n-l]\mathbf{x}^*[n-l] \tag{2.55}$$

In the simplest case, we set $L = 1$ and use the one-point ensemble mean to estimate the expectation $E\{e[n]\mathbf{x}^*[n]\}$. For an N th order adaptive filter, we need $N+1$ multiplications and $N+1$ additions to calculate the N coefficients. After that, calculating the output signal needs N multiplications and $N-1$ additions. So the computational complexity is $2N+1$ multiplications and $2N$ additions.

The convergence condition given in Eq.2.46 can be further specified in an LMS adaptive filter. Unlike in the WSS input case discussed previously, $\mathbf{w}_n[k]$ is a random variable based on the estimation of $E\{e[n]\mathbf{x}^*[n]\}$, so the convergence property of it should be studied within a statistical framework. That is, instead of studying the condition that $\mathbf{w}_n[k]$ converges to a certain value, we study the expectation of $\mathbf{w}_n[k]$. For simplicity, in following derivation we assume $L = 1$. It follows Eq.2.55 that

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mu(s[n] - \mathbf{w}_n^T \mathbf{x}[n])\mathbf{x}^*[n] \tag{2.56}$$

Taking the expected value for both sides, yields

$$\begin{aligned} E\{\mathbf{w}_{n+1}\} &= E\{\mathbf{w}_n\} + \mu E\{(s[n] - \mathbf{w}_n^T \mathbf{x}[n])\mathbf{x}^*[n]\} \\ &= E\{\mathbf{w}_n\} + \mu E\{s[n]\mathbf{x}^*[n]\} - \mu E\{\mathbf{x}^*[n]\mathbf{x}^T[n]\}E\{\mathbf{w}_n\} \\ &= (\mathbf{I} - \mu\mathbf{R}_x)E\{\mathbf{w}_n\} + \mu\mathbf{r}_{sx} \end{aligned} \tag{2.57}$$

Through the same coordinate system transformation from Eq. 2.47 to 2.49, it follows that

$$E\{\mathbf{u}_n\} = (\mathbf{I} - \mu\mathbf{\Lambda})^n \mathbf{u}_0 \tag{2.58}$$

Similarly we need $\mathbf{I} - \mu\mathbf{\Lambda} < 1$ to make sure $E\{\mathbf{u}_n\}$ converges, which results in the same condition as Eq. 2.46.

However, this upper bound is not tight enough for an LMS filter and not practical to use. It only guarantees the expectation convergence but has no constraint on the variance. And still, the eigenvalue of the autocorrelation matrix is needed and we have to use multiple signal samples to estimate $\mathbf{R}_x$. This will result in time delay and additional computation complexity. One way to solve the problem is use the upper bound of maximum eigenvalue as follows:

$$\lambda_{\max} \le tr(\mathbf{R}_x) = NE\{|x[n]|^2\} \tag{2.59}$$

This will tighten the constraint of μ and also simplify the computation since $E\{|x[n]|^2\}$ can be easily estimated by time averaging. Then the convergence condition for an LMS adaptive filter becomes

$$0 < \mu < \frac{2}{NE\{|x[n]|^2\}} \tag{2.60}$$

From Eq. 2.60 we can design an improved LMS algorithm, the Normalized LMS (NLMS). If we set μ as a time variable, i.e.

$$\mu[n] = \frac{\beta}{NE\{|x[n]|^2\}} = \frac{\beta}{\|\mathbf{x}[n]\|^2}, 0 < \beta < 2 \tag{2.61}$$

and replacing μ in Eq. 2.61. Also assuming L=1, we will have the NMLS algorithm

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \beta \frac{x^*[n]}{\|\mathbf{x}[n]\|^2} e[n] \tag{2.62}$$

Comparing Eq. 2.62 to Eq. 2.55, in the LMS algorithm the actual correction made to $\mathbf{w}_n$ is proportional to $\mathbf{x}[n]$. So when $\mathbf{x}[n]$ is large, the LMS algorithm will experience noise amplification effects. NMLS solved this problem by normalizing the step-size by $\|\mathbf{x}[n]\|^2$. However, if $\|\mathbf{x}[n]\|^2 << 1$, NLMS will still experiences a similar noise problem, so we introduce a small offset to $\|\mathbf{x}[n]\|^2$ as follows:

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \beta \frac{x^*[n]}{\varepsilon + \|\mathbf{x}[n]\|^2} e[n], \tag{2.63}$$

where ε is a small but positive number.

2.2.3.2 Noise cancellation scheme

A reference input is necessary for most kinds of adaptive algorithm if they are used for noise/interference cancellation. We should notice that the filtering procedure requires the actual value of SOI to calculate the error by Eq. 2.2. However, in the specific problem of noise/interference cancellation, the actual value of SOI is unavailable to the canceller (if not, then we do not need the canceller at all). So a reference input is needed to help the adaptive algorithm finding the optimum filter coefficients. Fig.2.5 is the block diagram of a typical LMS adaptive noise canceller with reference interference. In fact, the structure illustrated in Fig. 2.5 is a modified version of Fig. 2.2, only changing the Wiener filter to an adaptive one. Similar to what we discussed in above, $v_1[n]$ and $v_2[n]$ need to be correlated to each other. If an LMS based algorithm is used in the adaptive filter, the primary input $x[n]$ itself can also be a reference input after a time delay n_0, and the output of the adaptive filter is the LMS estimate of the SOI, because minimizing $E\{|e[n]|^2\}$ is equivalent to minimizing $E\{|v_1[n]-\hat{v}_1[n]|^2\}$ [19].

In [5], E.Alban used this noise-cancellation scheme with an NMLS algorithm to cancel the narrowband noise in a computer platform. Although it is not directly applicable to the broadband case this thesis focuses on, the new adaptive algorithm introduced is developed with a similar designing idea. Based on the fundamental structure of this normalized adaptive filter, an algorithm using a different kind of "criterion of goodness" instead of MSE is designed.

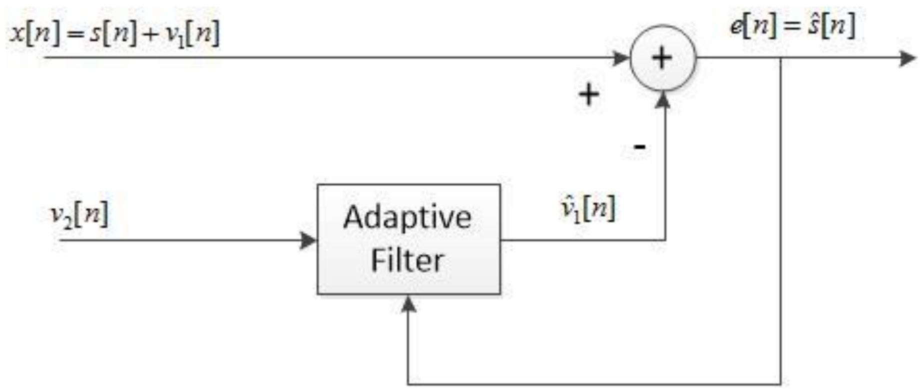


Figure 2.5 Diagram of an LMS Noise canceller with reference input

Chapter 3-A New Adaptive Algorithm Using Higher Order Statistics

E. Alban concluded in [5] that the wireless interference a computer platform suffered is K-distributed. Even though an NLMS adaptive filter is designed to eliminate the narrowband interference, it cannot be used for the broadband case. D. Shin pointed out two major difficulties of using LMS based adaptive filter in [26]: First, an LMS based adaptive filter is affected directly by uncorrelated noises at primary and reference inputs. In this particular problem, the primary and reference inputs (include the K-distributed interference) are both corrupted by uncorrelated Gaussian noise, so the performance of LMS based algorithm will be lowered. Second, the performance of LMS based algorithms highly depends on the specific problem, which means they can be very sensitive to interference signal and step size. To mitigate the broadband K-distributed interference, a more robust interference cancellation mechanism needs to be designed to meet the following requirements:

a) It is able to mitigate broadband, non-Gaussian, non-stationary interference.

b) It must work with the existence of uncorrelated Gaussian noise.

c) Its BER performance satisfies general wireless communication requirement

The LMS algorithm described in Eq. 2.51 and 2.52 minimizes the mean square error. Second order statistics is used to be the “criterion of goodness”. Here we present a new adaptive algorithm that utilizes a higher order statistics (HOS) called cumulants. The adaptive canceller designed with this algorithm will be proved to work efficiently with K-distributed interference and all 3 requirements stated above are met. We will first discuss some aspects of double sided K-distributed random process, and then come to the HOS cancellation algorithm.

3.1 K-distributed interference

3.1.1 Double-sided K-distribution

Before E. Alban modeled the computer platform interference as K-distributed, this distribution was widely used to model the clutter or reverberation of a radar or sonar system [29][30]. It can be described by the following probability density function (PDF)[7]:

$$f_X(x) = \begin{cases} \dfrac{1}{b\Gamma(\nu+1)}\left(\dfrac{x}{2b}\right)^{\nu+1} K_\nu\left(\dfrac{x}{b}\right), & x \geq 0 \\ 0, & o.w. \end{cases} \tag{3.1}$$

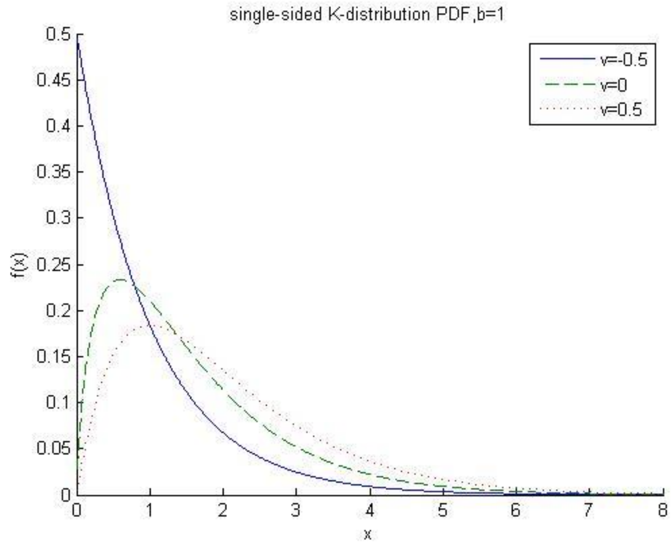
where a is a positive scalar parameter and ν is the shape parameter which satisfies $\nu > -1$. $\Gamma(\bullet)$ is the Gamma function, i.e.

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$$

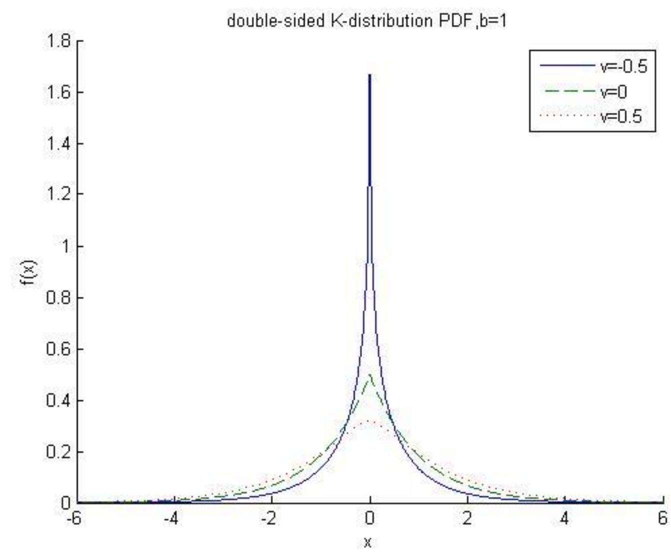
and $K_\nu(\square)$ is the modified Bessel function of order ν

$$K_\nu(x) = \int_0^\infty \exp(-x\cosh t)\cosh(\nu t)dt$$

A graph of Eq. 3.1 given $b=1$ is shown in Fig. 3.1(a). Since the PDF of a K-distributed



(a)



(b)

Figure 3.1 PDF of K-distribution

random variable (RV) is only non-zero for positive x, it is also called single-sided K-distribution. Obviously it is not appropriate to describe a random interference that can be positive or negative. We can derive the double-sided K-distribution PDF by deriving the distribution of another RV Y that $X = |Y|$, in which X is single-sided K-distributed. The double-sided K-distribution PDF is

$$f_X(x) = \frac{1}{\sqrt{\pi} b \Gamma(\nu+1)} \left(\frac{|x|}{2b} \right)^{\nu+\frac{1}{2}} K_{\nu+\frac{1}{2}} \left(\frac{|x|}{b} \right), \quad -\infty < x < +\infty, \tag{3.2}$$

in which $b > 0$ and $\nu > -1$. This distribution is used to describe the computer platform interference in [5]. A graph of Eq. 3.2 given $b = 1$ is shown in Fig 3.1(b).

3.1.2 Moments of double-sided K-distribution

The k th moments of a double-sided K-distribution RV are computed as [5]:

$$\begin{aligned} E\{x^k\} &= \int_{-\infty}^{+\infty} x^k \frac{1}{\sqrt{\pi} b \Gamma(\nu+1)} \left(\frac{|x|}{2b} \right)^{\nu+\frac{1}{2}} K_{\nu+\frac{1}{2}} \left(\frac{|x|}{b} \right) dx \\ &= \frac{2^{k-1}((-b)^k + b^k)\Gamma(\frac{1+k}{2})\Gamma(\frac{k}{2}+\nu+1)}{\sqrt{\pi}\Gamma(\nu+1)} \end{aligned} \tag{3.3}$$

Notice that when k is odd, $E\{x^k\} = 0$.

Because $E\{x\} = 0$, we only need to set $k = 2$ to find the variance:

$$\operatorname{var}(x) = E\{x^2\} = 2b^2(\nu+1) \tag{3.4}$$

The result of Eq. 3.4 is important because this is how we compute the interference signal power when computer simulation is performed. Once either b or v is fixed, the interference signal power varies with the change of the other parameter. Then the input signal-to-interference ratio (SIR) can be set to a desired value.

3.1.3 Generation of double-sided K-distributed random process

S. Kay in [21] found a simple way to generate a double-sided K-distributed random process. This random process with arbitrary power spectral density (PSD) can be generated by the product of a Gamma distribution and a Gaussian distributed random process, as follows:

$$X = \sqrt{V}U \tag{3.5}$$

If X is double-sided K-distributed with parameter b and ν, then $V \sim \Gamma(\nu+1, \frac{1}{2})$ and $U \sim N(0, b^2)$. This is how we generate a WSS double-sided K-distributed random variable in the computer simulation. Since all we need is a white broadband noise interference signal, no further manipulation after Eq. 3.5 is needed.

E. Alban in [5] also generated the process by another method. A state-space model of the process is derived from the stochastic differential functions of a double-sided K-distributed RV. This model is only used to design a Kalman filter, so the detail will not be presented in this thesis.

3.2 Higher Order Statistics (HOS) Algorithm

3.2.1 Cumulants

The high order statistics used in this algorithm are cumulants. In [31], given a set of n real value random variables $x_1, x_2, ..., x_n$, their cumulants of r th order are defined as

$$c_{k_1 k_2 \dots k_n} \triangleq (-j)^r \frac{\partial^r \ln \Phi(\omega_1, \omega_2, ..., \omega_n)}{\partial \omega_1^{k_1} \partial \omega_2^{k_2} ... \partial \omega_n^{k_n}} \Big|_{\omega_1, \omega_2, ..., \omega_n = 0}, \sum_{i=1}^{n} k_i = r \quad , \tag{3.6}$$

where $\Phi(\cdot)$ is the joint characteristic function of $x_1, x_2, ..., x_n$:

$$\Phi(\omega_1, \omega_2, ..., \omega_n) = E\{\exp(j \sum_{k=1}^{n} \omega_k x_k)\} \tag{3.7}$$

Notice that the the r th order joint moments are defined as

$$m_{k_1 k_2 \dots k_n} \triangleq (-j)^r \frac{\partial^r \Phi(\omega_1, \omega_2, ..., \omega_n)}{\partial \omega_1^{k_1} \partial \omega_2^{k_2} ... \partial \omega_n^{k_n}} \Big|_{\omega_1, \omega_2, ..., \omega_n = 0} \tag{3.8}$$

So we can express the cumulants in terms of the moments. Especially, we consider the k th moment of x_1, i.e., $k_2, k_3, ..., k_n = 0$. If we assume $E\{x_1\} = m_1 = 0$, from Eq.3.7 and 3.8 we can have

$$\begin{aligned} c_1 &= (-j) \frac{d \ln \Phi(\omega_1)}{\partial \omega_1} \Big|_{\omega_1 = 0} \\ &= (-j) \frac{1}{\Phi(\omega_1)} \cdot \frac{d\Phi(\omega_1)}{\partial \omega_1} \Big|_{\omega_1 = 0} \\ &= \frac{1}{\Phi(0)} \cdot m_1 = 0 \end{aligned} \tag{3.9}$$

Similarly,

$$c_2 = m_2 = E\{x_1^2\} = \mathrm{var}(X_1) \tag{3.10}$$

$$c_3 = m_3 = E\{x_1^3\} \tag{3.11}$$

$$c_4 = m_4 - 3c_2^2 = m_4 - 3m_2^2 \tag{3.12}$$

Next, for a random process $X(t)$, we define the correlation $c_k(\tau)$ using Eq. 3.9-3.12, i.e.

$$E\{x(t)x(t+\tau)\} = m_2(\tau) = c_2(\tau) \tag{3.13}$$

$$E\{x(t)x(t+\tau_1)x(t+\tau_2)\} = m_3(\tau_1,\tau_2) = c_3(\tau_1,\tau_2) \tag{3.14}$$

$$\begin{aligned} &E\{x(t)x(t+\tau_1)x(t+\tau_2)x(t+\tau_3)\} \\ &= m_4(\tau_1,\tau_2,\tau_3) \\ &= c_4(\tau_1,\tau_2,\tau_3) + c_2(\tau_1)c_2(\tau_3-\tau_2) \\ &\quad + c_2(\tau_2)c_2(\tau_3-\tau_1) \\ &\quad + c_2(\tau_3)c_2(\tau_2-\tau_1) \end{aligned} \tag{3.15}$$

From Eqs.3.13 and 3.15, we have

$$\begin{aligned} &c_4(\tau_1,\tau_2,\tau_3) \\ &= m_4(\tau_1,\tau_2,\tau_3) - m_2(\tau_1)m_2(\tau_3-\tau_2) - m_2(\tau_2)m_2(\tau_3-\tau_1) - m_2(\tau_3)m_2(\tau_2-\tau_1) \end{aligned} \tag{3.16}$$

In the Fourth Order Statistic (FOS) algorithm, Eq. 3.16 is used to estimate the cumulants of the inputs. [31] pointed out that third or higher order cumulants of a Gaussian process are zero. That is one of the reasons why cumulants are chosen to be the “criterion of goodness” in an algorithm mitigating non-Gaussian interference with the existence of

Gaussian noise. In other words, cumulants of third-order or above can work regardless the existence of Gaussian noise, while an LMS based algorithm cannot.

3.2.2 Primary and Reference inputs

The general structure of a HOS interference canceller is illuminated in Fig. 3.2. The received signal is corrupted by both K-distributed interference and Gaussian noise:

$$x[k] = s[k] + I_K[k] + n_p[k], \tag{3.17}$$

in which $s[k]$ is the signal of interest (SOI), $I_K[k]$ is the broadband K-distributed noise and $n_p[k]$ is the Gaussian noise. The corrupted SOI is the primary input of the canceller. An important assumption before we study the HOS interference cancellation mechanism is that there is another signal input available, we use it the reference input. As stated in Chapter 2, in a noise-cancellation problem when the original SOI is not available, a reference signal is needed to aid the algorithm to find the optimum solution. Here, the HOS canceller assumes a reference input $w[k]$ is available. $w[k]$ and the interference signal, $I_K[k]$ is related as follows:

$$I_K[k] = \sum_j g[j]w[k-j] \tag{3.18}$$

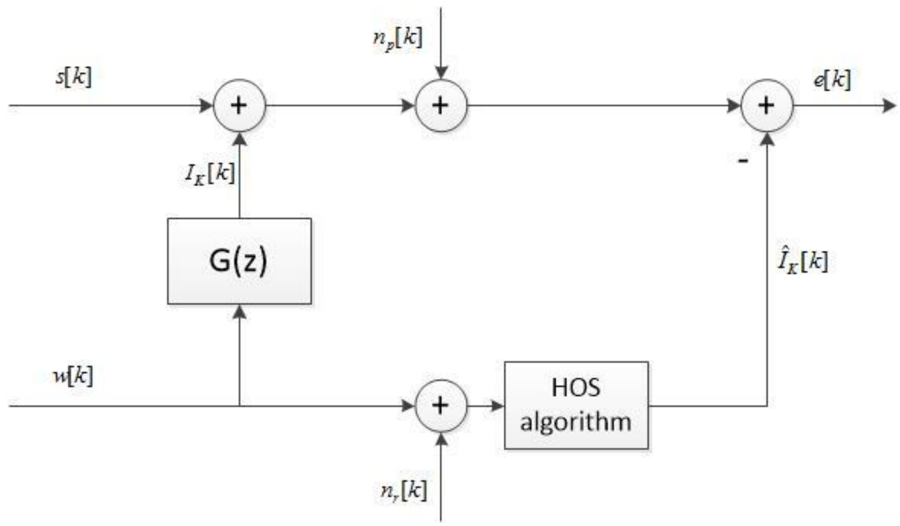


Figure 3.2 Diagram for a HOS interference canceller

That means, we can get $I_K[k]$ by passing $w[k]$ through a linear time-invariance (LTI) system $G(z)$. More specifically, $G(z)$ is a moving average (MA) filter. Thus, $I_K[k]$ and $w[k]$ are correlated.

The reference signal is also corrupted by Gaussian noise, $n_r[k]$. So the reference input of the interference canceller is

$$z[k] = w[k] + n_r[k] \tag{3.19}$$

Notice $n_r[k]$ and $n_p[k]$ are supposed to be uncorrelated so they are independently generated in computer simulation.

Then, the output of the adaptive filter in Fig. 3.2 is

$$\hat{I}_K[k] = \sum_{j=0}^{N-1} h[j] z[k-j], \tag{3.20}$$

where N is the order of the FIR adaptive filter and $h[j]$ is the j th coefficient of the filter. $\hat{I}_K[k]$ is the estimate of the input interference $I_K[n]$.

So the output of the canceller, $e[k]$, is the estimate of the SOI, namely,

$$e[k] = x[k] - \hat{I}_K[k] = \hat{s}(k) = s[k] + n_e[k] \tag{3.21}$$

$n_e[k]$ is the output noise or the estimation error, which can be expressed as

$$n_e[k] = I_K[k] - \sum_{j=0}^{N-1} h[j]w[k-j] - \sum_{j=0}^{N-1} h[j]n_r[k-j] + n_p[k] \tag{3.22}$$

3.2.3 General Scheme for the Algorithm

In [26] the general scheme for an HOS algorithm is introduced. Because the n th order cumulants of any Gaussian process is identically zero if $n \geq 3$, this general scheme applies on all high order statistics with $n \geq 3$. Let $C_{xz\ldots z}(m_1, m_2, \ldots, m_{n-1})$ denote the n th order cross-cumulants of the primary input $x[k]$ and the reference input $z[k]$, and $C_{yz\ldots z}(m_1, m_2, \ldots, m_{n-1})$ denote the n th order cross-cumulants of the adaptive filter output $y[k] = \hat{I}_K[k]$ and the reference input $z[k]$. If we use the cumulant operator $Cum[]$ introduced in [32], i.e.

$$Cum[x_1^{k_1}, x_2^{k_2}, \ldots, x_3^{k_3}] = c_{k_1 k_2 \ldots k_n} \triangleq (-j)^r \frac{\partial^r \ln \Phi(\omega_1, \omega_2, \ldots, \omega_n)}{\partial \omega_1^{k_1} \partial \omega_2^{k_2} \ldots \partial \omega_n^{k_n}} \Big|_{\omega_1, \omega_2, \ldots, \omega_n = 0}, \tag{3.23}$$

then $C_{xz\ldots z}(m_1, m_2, \ldots, m_{n-1})$ and $C_{yz\ldots z}(m_1, m_2, \ldots, m_{n-1})$ can be defined as

$$C_{xz\ldots z}(m_1,m_2,\ldots,m_{n-1}) = Cum[x[k], z[k+m_1], z[k+m_2],\ldots, z[k+m_{n-1}]] \quad (3.24)$$

$$C_{yz\ldots z}(m_1,m_2,\ldots,m_{n-1}) = Cum[y[k], z[k+m_1], z[k+m_2],\ldots, z[k+m_{n-1}]] \quad (3.25)$$

From the linear property of the cumulant operator described in [32], we have

$$\begin{aligned}
& C_{xz\ldots z}(m_1,m_2,\ldots,m_{n-1}) \\
&= Cum[s[k]+I_K[k]+n_p[k], w[k+m_1]+n_r[k+m_1], w[k+m_2]+n_r[k+m_2],\ldots, \\
& \quad w[k+m_{n-1}]+n_r[k+m_{n-1}]] \\
&= Cum[I_K[k], w[k+m_1], w[k+m_2],\ldots, w[k+m_{n-1}]] \\
& \quad + Cum[n_p[k], n_r[k+m_1], n_r[k+m_2],\ldots, n_r[k+m_{n-1}]] \\
& \quad + Cum[s[k], n_r[k+m_1], n_r[k+m_2],\ldots, n_r[k+m_{n-1}]]
\end{aligned}$$

(3.26)

Notice that n_p and n_r are Gaussian processes. So the second term of the result in Eq. 3.26 is zero. Since $E\{s[k]\}=0$, then the last term is zero. Thus,

$$\begin{aligned}
& C_{xz\ldots z}(m_1,m_2,\ldots,m_{n-1}) \\
&= Cum[I_K[k], w[k+m_1], w[k+m_2],\ldots, w[k+m_{n-1}]] \\
&= C_{Iw\ldots w}(m_1,m_2,\ldots,m_{n-1})
\end{aligned} \quad (3.27)$$

Substituting Eq. 3.18 into Eq.3.27, yields

$$\begin{aligned}
& C_{xz\ldots z}(m_1,m_2,\ldots,m_{n-1}) \\
&= Cum[\sum_j g[j]w[k-j], w[k+m_1], w[k+m_2],\ldots, w[k+m_{n-1}]] \\
&= \sum_j g[j]Cum[w[k-j], w[k+m_1], w[k+m_2],\ldots, w[k+m_{n-1}]] \\
&= \sum_j g[j]C_{w\ldots w}(j+m_1, j+m_2,\ldots, j+m_{n-1})
\end{aligned} \quad (3.28)$$

Similarly, we can express $C_{yz\ldots z}(m_1, m_2, ..., m_{n-1})$ in terms $C_{w\ldots w}(.)$, i.e.

$$C_{yz\ldots z}(m_1, m_2, ..., m_{n-1}) = \sum_{j=0}^{N-1} h[j] C_{w\ldots w}(j+m_1, j+m_2, ..., j+m_{n-1}) \tag{3.29}$$

Comparing Eq. 3.28 and 3.29, we can find that if $h[j] = g[j]$, then $C_{xz\ldots z}(m_1, m_2, ..., m_{n-1})$ and $C_{yz\ldots z}(m_1, m_2, ..., m_{n-1})$ will be identical and the adaptive filter will produce the optimum estimate of the interference $I_K[k]$. That is why we use the difference of $C_{xz\ldots z}(m_1, m_2, ..., m_{n-1})$ and $C_{yz\ldots z}(m_1, m_2, ..., m_{n-1})$ to measure the error. The new "criterion of goodness" is

$$\xi_g = \sum_{m_1} \sum_{m_2} \cdots \sum_{m_{n-1}} [C_{xz\ldots z}(m_1, m_2, ..., m_{n-1}) - C_{yz\ldots z}(m_1, m_2, ..., m_{n-1})]^2 \tag{3.30}$$

While $m_1, m_2, ..., m_{n-1}$ can be defined to include the whole $(n-1)$ dimensional space, it is not practical because of the computational complexity. What we do in reality is to choose a proper domain P that is a subset of the $(n-1)$ dimensional space. Thus, the simplified error measurement is

$$\xi = \sum_{m_1, m_2, ..., m_{n-1} \in P} [C_{xz\ldots z}(m_1, m_2, ..., m_{n-1}) - C_{yz\ldots z}(m_1, m_2, ..., m_{n-1})]^2 \tag{3.31}$$

Because

$$
\begin{aligned}
& C_{z\ldots z}(m_1, m_2, \ldots, m_{n-1}) \\
&= Cum[z[k], z[k+m_1], z[k+m_2], \ldots, z[k+m_{n-1}]] \\
&= Cum[w[k], w[k+m_1], w[k+m_2], \ldots, w[k+m_{n-1}]] \\
&\quad + Cum[n_r[k], n_r[k+m_1], n_r[k+m_2], \ldots, n_r[k+m_{n-1}]] \\
&= Cum[w[k], w[k+m_1], w[k+m_2], \ldots, w[k+m_{n-1}]]
\end{aligned}
\tag{3.32}
$$

We have

$$
C_{z\ldots z}(m_1, m_2, \ldots, m_{n-1}) = C_{w\ldots w}(m_1, m_2, \ldots, m_{n-1}) \tag{3.33}
$$

From Eq. 3.33, 3.29 and 3.31, we can express the "criterion of goodness" in terms of the cumulants of primary and reference inputs

$$
\xi = \sum_{m_1, m_2, \ldots, m_{n-1} \in P} \left[C_{xz\ldots z}(m_1, m_2, \ldots, m_{n-1}) - \sum_{j=0}^{N-1} h[j] C_{z\ldots z}(j+m_1, j+m_2, \ldots, j+m_{n-1}) \right]^2 \tag{3.34}
$$

In matrix form

$$
\xi = (\mathbf{C}_{xz\ldots z} - \mathbf{C}_{z\ldots z}\mathbf{H}_h)^T (\mathbf{C}_{xz\ldots z} - \mathbf{C}_{z\ldots z}\mathbf{H}_h) \tag{3.35}
$$

If we denote the number of points in set P as M and N is the number of taps of the FIR filter, then $\mathbf{C}_{xz\ldots z}$ is an $M \times 1$ column vector and $\mathbf{C}_{z\ldots z}$ is an $M \times N$ matrix. [26] indicated that we need $M > N$ to guarantee the reliability of the filter coefficients. $\mathbf{H}_h$ is an $N \times 1$ vector of filter coefficients

$$
\mathbf{H}_h = [h[0], h[1], \ldots, h[N-1]]^T \tag{3.36}
$$

To apply the steepest descent method, we first calculate the gradient of ξ with respect to $\mathbf{H}_h$, i.e.

$$\nabla_h(k) = \frac{\partial \xi}{\partial \mathbf{H}_h(k)} = 2(\mathbf{C}_{z\ldots z}^T \mathbf{C}_{z\ldots z} \mathbf{H}_h(k) - \mathbf{C}_{z\ldots z}^T \mathbf{C}_{xz\ldots z}) \tag{3.37}$$

Then update the filter coefficients

$$\mathbf{H}_h(k+1) = \mathbf{H}_h(k) - \mu \nabla_h(k) \tag{3.38}$$

The value of the step size μ depends on the order of the algorithm and the noise signal. The way to find a proper value of μ will be discussed later under the condition that a FOS algorithm is applied on K-distributed interference.

3.3 The FOS algorithm

According to the general scheme of HOS method, there are still 3 problems left for the specific FOS algorithm:

1) The proper selection of cross-cumulants domain. Choosing $m_i, i = 1, 2, \ldots, n-1$ by including the whole $(n-1)$ dimensional space is both computationally costly and unnecessary. So a smaller domain is needed to be decided to simplify the algorithm without the sacrifice of cancellation performance.
2) The estimation of cumulants. Both the prime and reference input signal are random processes. No statistical knowledge of them is available to the canceller. Estimations with the observations of these processes are needed.

3) A suitable step size is needed to make sure the algorithm converges.

To solve problem 1), since the fourth order statistic is used, we only need a domain such that

$$P = \{(m_1, m_2, m_3)\} \subset \Re^3 .$$

[26] provided the following domain selection when cancelling a sinusoidal interference with uniform-distributed random phase. It will be proved later to be efficient when the interference is K-distributed, where

$$\begin{array}{l} 0 \le m_1, m_2, m_3 \le L-1 \\ m_1 \ge m_2 \\ m_2 \ge m_3 \end{array} ,$$

in which the domain size L is a positive interger that is chosen to determine the number of elements in the domain. Note that the number of points M satisfies

$$M = \frac{L(L+1)(2L+1)}{6} \tag{3.39}$$

If the number of filter taps is chosen as $N = 8$, then we need at least $L = 3$ to make sure $M > N$. Thus, $M = 14$ and the points in the domain are:

$$\{(0,0,0);(1,0,0);(1,0,1);(1,1,0);\ldots;(2,2,2)\}$$

Form Eq. 3.16 we can find the way to estimate cumulants from observations over time. Assuming the input process is ergodic, then the cross-moment term in Eq. 3.16 can be estimated by a time average. This estimation also needs to be updated for each iteration when new observations are available to the canceller. If we denote the estimator of the cross-moment at the k th iteration as $\hat{R}(k, m_1, m_2, m_3)$ (because they are equivalent to covariance in this case), then the second and fourth order cross-moment estimation will be

$$\begin{gathered}\hat{R}_{xzzz}(k, m_1, m_2, m_3) = \frac{1}{k}\sum_{i=1}^{k} x[i]z[i+m_1]z[i+m_2]z[i+m_3] \\ \hat{R}_{xzzz}(k, m_1, m_2, m_3) = \frac{1}{k}\sum_{i=1}^{k} x[i]z[i+m_1]z[i+m_2]z[i+m_3]\end{gathered}, \tag{3.40}$$

$$\begin{gathered}\hat{R}_{xz}(k, \tau) = \frac{1}{k}\sum_{i=1}^{k} x[i]z[i+\tau] \\ \hat{R}_{zz}(k, \tau) = \frac{1}{k}\sum_{i=1}^{k} z[i]z[i+\tau]\end{gathered}, \tag{3.41}$$

respectively.

Then, from Eq. 3.16 we can have the estimation of the fourth-order cumulants

$$\begin{aligned}&\hat{C}_{xzzz}(k, m_1, m_2, m_3) \\ &= \hat{R}_{xzzz}(k, m_1, m_2, m_3) - \hat{R}_{xz}(k, m_1)\hat{R}_{zz}(k, m_2 - m_3) \\ &\quad - \hat{R}_{xz}(k, m_2)\hat{R}_{zz}(k, m_1 - m_3) - \hat{R}_{xz}(k, m_3)\hat{R}_{zz}(k, m_1 - m_2)\end{aligned} \tag{3.42}$$

$$\begin{aligned}&\hat{C}_{zzzz}(k, m_1, m_2, m_3) \\ &= \hat{R}_{zzzz}(k, m_1, m_2, m_3) - \hat{R}_{zz}(k, m_1)\hat{R}_{zz}(k, m_2 - m_3) \\ &\quad - \hat{R}_{zz}(k, m_2)\hat{R}_{zz}(k, m_1 - m_3) - \hat{R}_{zz}(k, m_3)\hat{R}_{zz}(k, m_1 - m_2)\end{aligned} \tag{3.43}$$

To solve problem 3), a loose bound is given in [26] as follows:

$$0 < \mu < \frac{1}{tr\{\mathbf{C}_{z\ldots z}^{T}\mathbf{C}_{z\ldots z}\}} \tag{3.44}$$

Eq.3.44 is a quite loose bound. We will see in the simulation that convergence for all SIR cases cannot be guaranteed. Similar to what we did in NMLS algorithm, we can manipulate the range by introducing 2 parameters

$$0 < \mu < \frac{\mu_f}{\alpha + tr\{\mathbf{C}_{z \ldots z}^T \mathbf{C}_{z \ldots z}\}}, 0 < \mu_f < 1, \alpha > 0, \tag{3.45}$$

where μ_f is the adaptation constant parameter and α is the offset parameter that used to minimize the noise amplification effect. The proper value of μ_f is highly dependent on the input SIR and will be discussed shortly in the simulation.

3.4 Simulation for the FOS algorithm

FOS algorithm is applied on an input binary sequence corrupted by K-distributed interference. Notice that the SOI used in this simulation is only for a typical experiment but not the SOI on a computer platform as we described earlier. The ensuring results are only to verify the capability of FOS algorithm to cancel the K-distributed noise. Figure 3.3 shows the block diagram for the experimental test scheme.

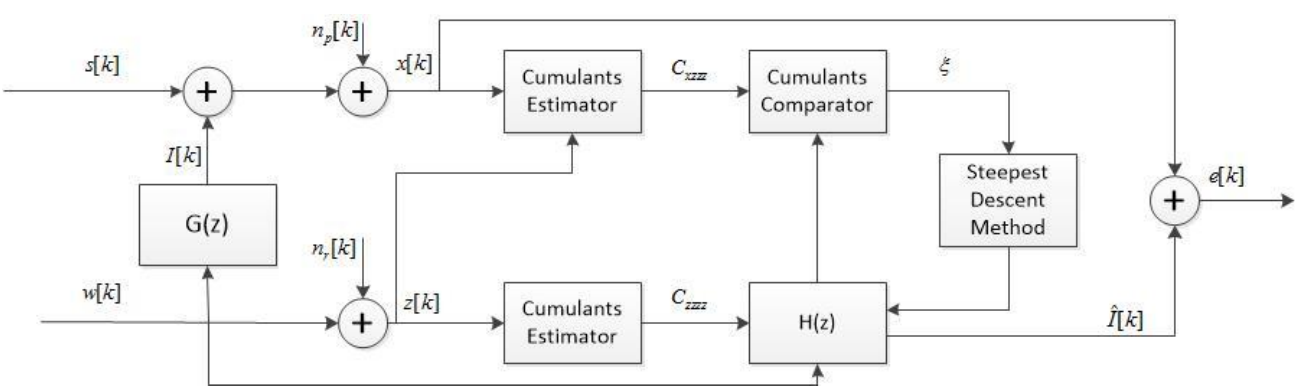


Figure 3.3 Interference simulation scheme

Simulation parameters:

Number of taps	$N = 8$
Domain size	$L = 3$
Adaptation constant parameter	$\mu_f = 0.1/0.01$
Offset parameter	$\alpha = 2$
Number of iteration	1×10^5

Table 3.1 Simulation parameters for experimental environment

μ_f is set to 2 different values to compare the convergence behavior. For the K-distributed interference, v is fixed at -0.5 and SIR are set by changing the value of b, as indicated in Eq.3.4. As the Gaussian noise is not the major noise/interference source, its power is set to be weaker than that of the K-distributed interference:

$$SNR_{Gaussian}(dB) = SIR(dB) - 5dB \tag{3.46}$$

Simulation cases:

Case 1: $\mu_f = 0.1$

Fig. 3.4-3.7 shows the comparison of the interference signal before and after the canceller, at SIR=5dB, 7.5dB, 10dB and 12.5dB, respectively.

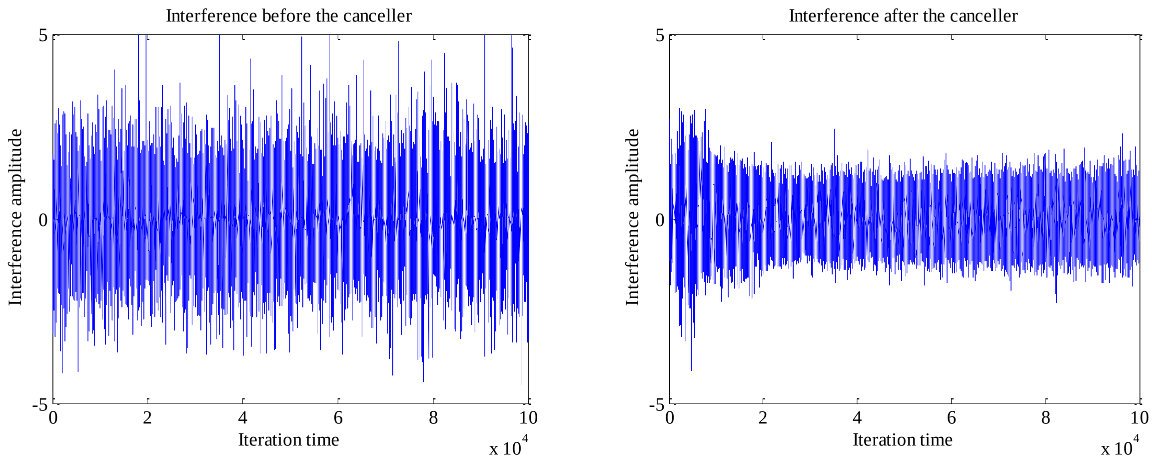


Figure 3.4 Interference signal comparison, SIR=5dB

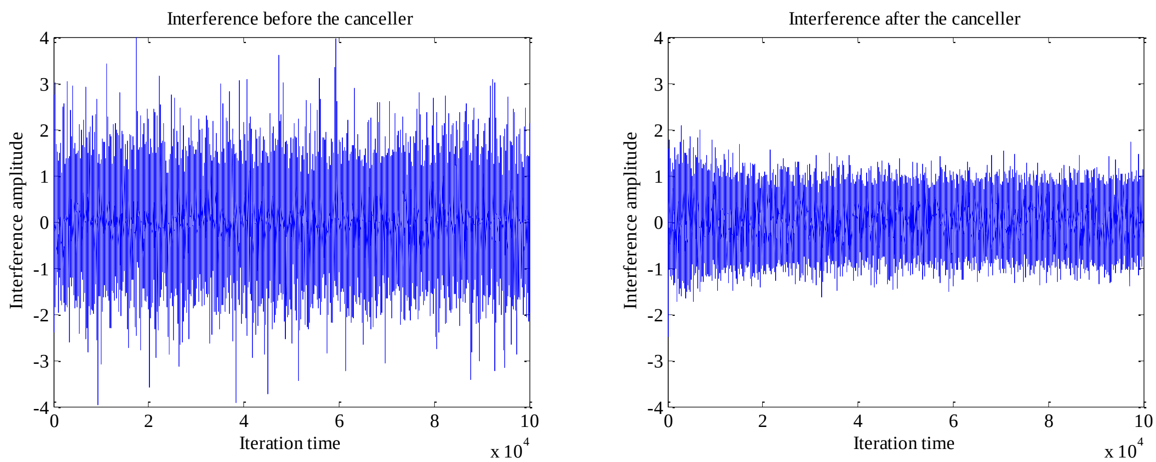


Figure 3.5 Interference signal comparison, SIR=7.5dB

Fig.3.8 shows the BER improvement after the canceller is implemented; SIR range is from 0dB to 15dB.

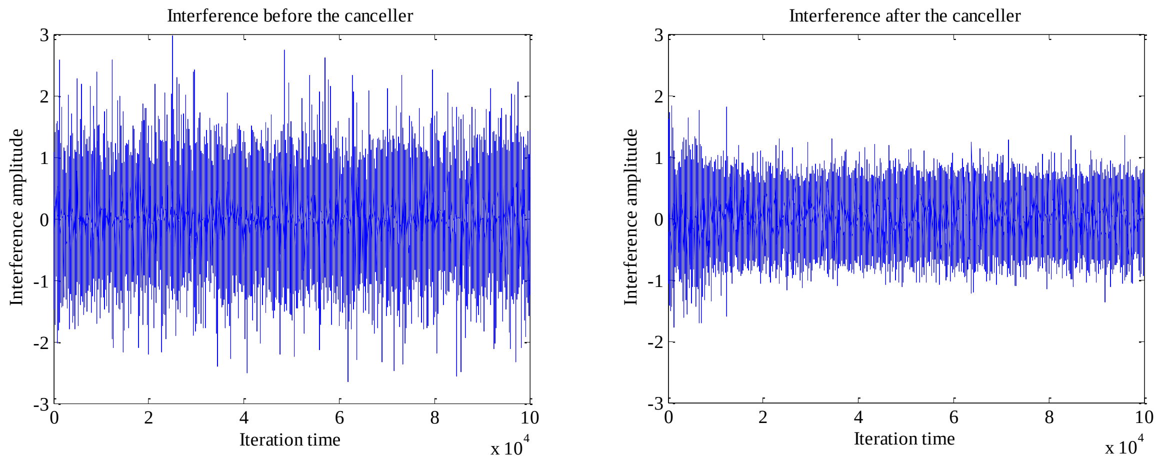


Figure 3.6 Interference signal comparison, SIR=10dB

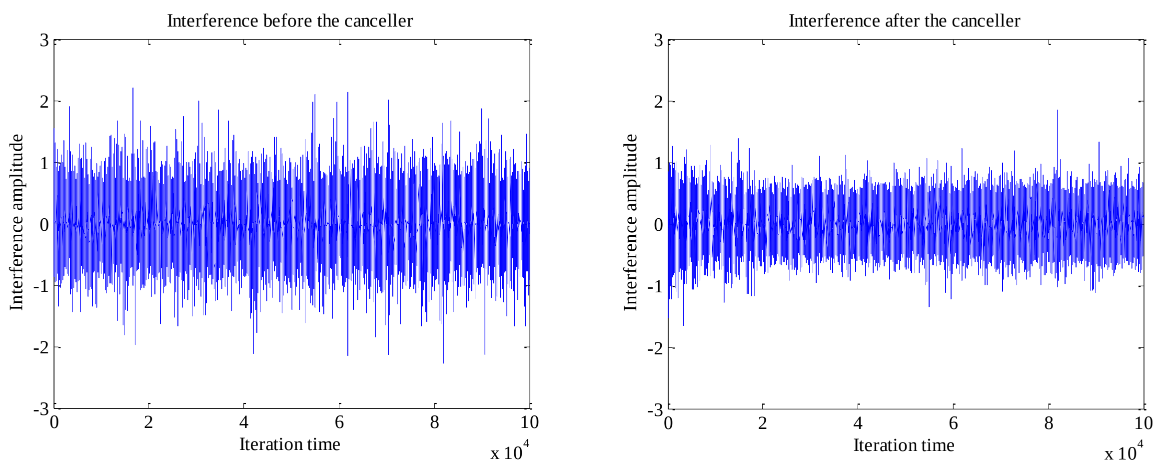


Figure 3.7 Interference signal comparison, SIR=12.5dB

From Figs 3.4-3.7 we can clearly see that the canceller works efficiently on all SIRs between 5dB and 12.5dB. The result usually converges before the first 20000 samples, which takes milliseconds in most modern wireless communication system. The BER improvement can be found in Fig 3.8, which shows 5 simulation results at each SIR from 0dB to 15dB. The best improvement can be as good as 15dB. On the other hand, Fig 3.8 shows the canceller is not working in a stable fashion on all SIRs, especially lower than

5dB. Some simulation resulted a higher BER because the algorithm is not converging. Solving this problem requires a smaller step-size coefficient. Here we choose $\mu_f = 0.01$. Since $\mu_f = 0.1$ works good when SIR is higher than 10dB, we only change the step-size for the lower SIR so the convergence time is minimized. This is called a variable step-size method and its simulation results are showed in Case 2.

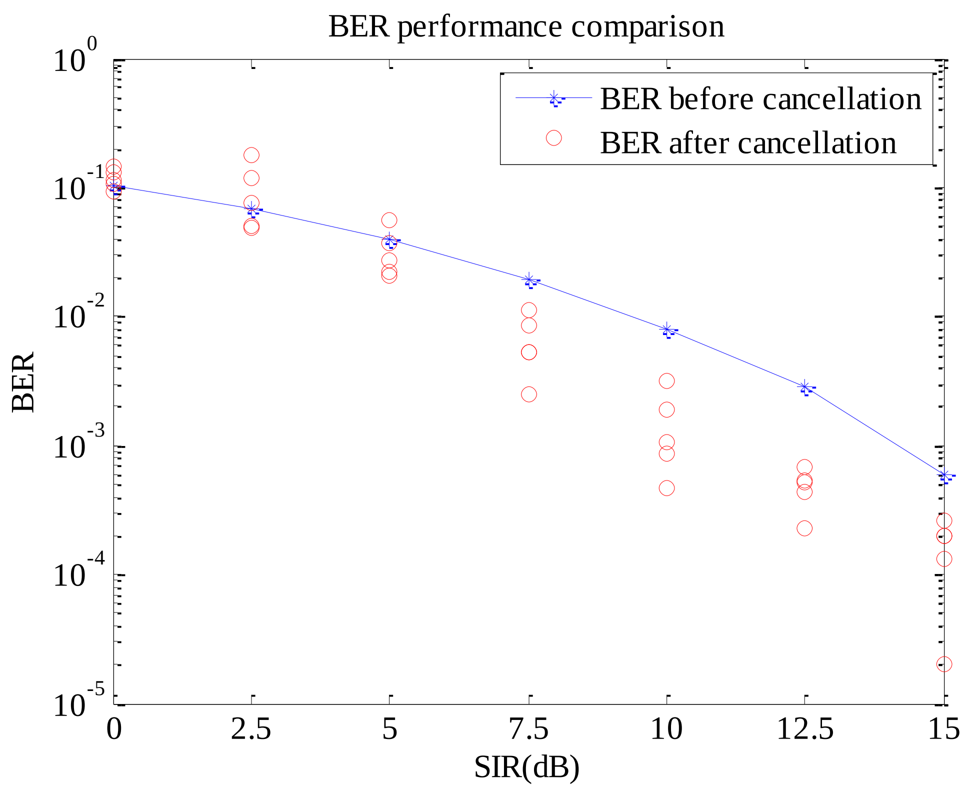


Figure 3.8 BER performance comparison with constant step-size, $\mu_f = 0.1$

Case 2: Variable step-size

Following simulation used the variable step-size scheme:

$$\mu_f = \begin{cases} 0.01, & SIR < 11 \\ 0.1, & elsewhere \end{cases} \tag{3.49}$$

Fig. 3.9 shows the BER performance. Obviously, the BER performance on low SIR is improved, which made the FOS-canceller fully applicable on this K-distributed interference cancellation problem.

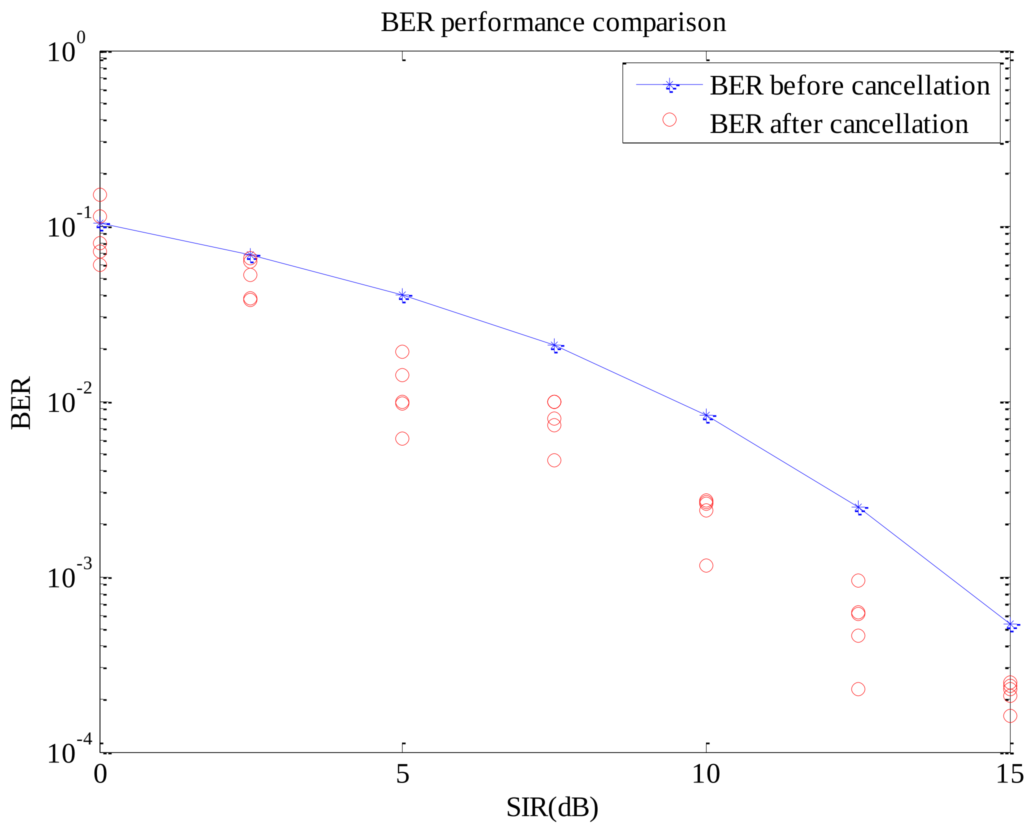


Figure 3.9 BER performance comparison with variable step-size

Chapter 4-FOS Interference Canceller in an OFDM System

4.1 OFDM System

The previous research of computer platform interference by E. Alban is based on an Orthogonal Frequency-Division Multiplexing (OFDM) system, which is widely used in modern communication standards including WiFi, WiMax and LTE. OFDM is a multicarrier modulation scheme. In an OFDM modulator, the original baseband signal is modulated by a QAM modulator first, generating a complex QAM symbol stream. A symbol stream can be split into a multiple symbol frame $X[0], X[1], ..., X[N-1]$ of length N . The QAM symbol stream is passed through a serial-to-parallel converter. It is considered as "frequency domain" signal as it contains the discrete frequency components of the OFDM modulator output signal. Then, to modulate this QAM symbol by multiple subcarriers orthogonal to each other, the inverse discrete Fourier transform (DFT) is performed, using inverse fast Fourier transform (IFFT). Thus, the QAM symbol stream is transformed to the "time domain", with multiple discrete frequency subcarriers, as follows:

$$x[n] = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} X[i] e^{j\frac{2\pi i}{N}n}, n = 0,1,...,N-1 \tag{4.1}$$

As a result, each QAM symbol $X[i]$ is modulated by multiple carriers, i.e.

$$e^{j\frac{2\pi i}{N}n}, n = 0,1,...,N-1.$$

To avoid ambiguity, we use transformed domain and modulated domain instead of "frequency domain" and "time domain", respectively. Then, a cyclic prefix of length μ is added to avoid the inter-carrier interference (ICI) in the transmission channel. Then, the output of an OFDM modulator is

$$x[N-\mu],\ldots,x[0],x[1],\ldots,x[N-1]$$

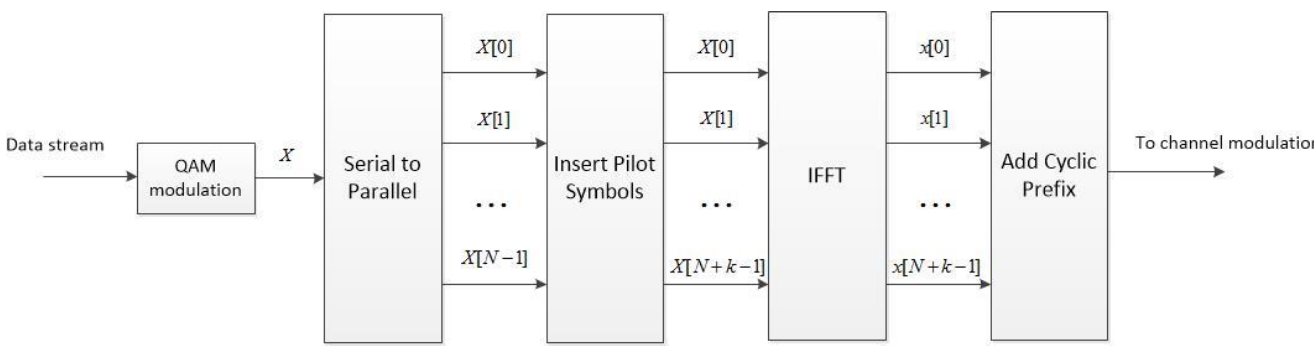


Figure 4.1(a) OFDM modulator

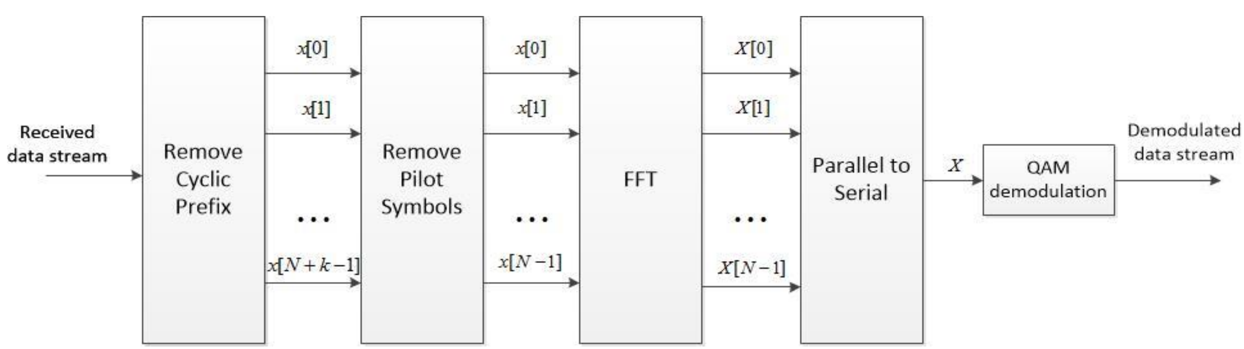


Figure 4.1(b) OFDM demodulator

At the demodulator, the cyclic prefix is first removed, and then the signal stream is passed through an FFT and a parallel-to-serial converter. After that, QAM demodulation is performed to retrieve the original signal. Fig 4.1(a) shows the structure of an OFDM modulator and Fig. 4.1(b) shows the receiver.

Notice that k pilot symbols are inserted before IFFT for channel estimation. Usually, they are added in different location of consecutive symbol frame in order to "scan" the whole bandwidth for better estimation results.

The modulation and demodulation of OFDM can be represented in matrix form by introducing the DFT matrix, i.e.

$$\mathbf{Q} = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & W_N & W_N^2 & \cdots & W_N^{N-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & W_N^{N-1} & W_N^{2(N-1)} & \cdots & W_N^{(N-1)^2} \end{bmatrix}, \tag{4.2}$$

where $W_N = e^{-j2\pi/N}$. Thus, if we denote $X[0], X[1], ..., X[N-1]$ and $x[0], x[1], ..., x[N-1]$ by two vectors $\mathbf{X}$ and $\mathbf{x}$, respectively, the IFFT of $\mathbf{X}$ can be represented as:

$$\mathbf{X} = \mathbf{Q}^{-1}\mathbf{x} \tag{4.3}$$

Notice $\mathbf{Q}$ is unitary, so we have

$$\mathbf{X} = \mathbf{Q}^{H}\mathbf{x} \tag{4.4}$$

Similarly, the FFT of $\mathbf{x}$ can be represented as

$$\mathbf{x} = \mathbf{Q}\mathbf{X} \tag{4.5}$$

Thus, the demodulation of OFDM can be written as

$$\mathbf{Y} = \mathbf{Q}\mathbf{y}, \tag{4.6}$$

where $\mathbf{y}$ is the vector which represents the received data stream, i.e.

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \boldsymbol{\nu} \tag{4.7}$$

where $\mathbf{H}$ is the channel impulse response matrix and $\boldsymbol{\nu}$ is the noise/interference.

4.2 Implementation of FOS canceller

The FOS interference canceller, although proved to be efficient on K-distributed interference, still needs some modification to work in an OFDM system. First, the received OFDM-modulated signal is complex but the FOS canceller can only work on real value signals. Second, similar to the NMLS algorithm, we need to normalize the

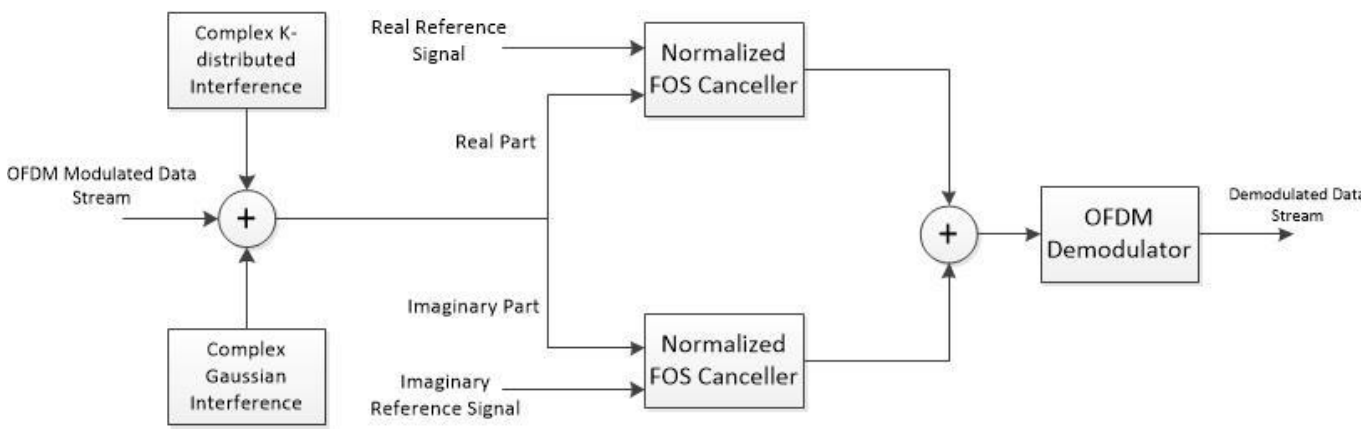


Figure 4.2 FOS canceller in an OFDM receiver

input signal to make sure the algorithm will converge. The cancellation scheme is depicted in Fig 4.2.

If the received signal is denoted as $r[n]$, the prime and reference input signal are normalized by multiplying the following scale factor:

$$a = \frac{1}{\sqrt{E_s}}, \tag{4.8}$$

where E_s is the average symbol energy.

4.3 Simulation and analysis

4.3.1 Frame composition

In order to compare our result to the performance of discrete Kalman filter designed by E. Alban in [5], the same OFDM modulation parameters are used in this simulation, i.e.

FFT size	1024
Number of data carriers	720
Number of pilot symbols	120
Prefix size	128

Table 4.1 OFDM parameters

Thus, the composition of each frame during the ODFM modulation is described as follows: After the serial to parallel conversion, the pilot symbols are inserted evenly throughout the frame. If the n th pilot symbol is denoted as P_n and the n th QAM modulated symbol is denoted as X_n, the frame can be presented as

$$[P_1, X_1, X_2, ..., X_6, P_2, X_7, ..., P_n, X_{1+6(n-1)}, ..., P_{120}, ..., X_{720}]$$

Note that in a real OFDM modulator, the location of pilot symbol should not be the same for each frame. However, since the locations of pilot symbols have nothing to do with the cancellation algorithm in this problem, the pilot symbols are set at fixed location for each frame.

The IFFT is performed right after the pilot symbol insertion. Note that the IFFT size applied here is 1024 and it is greater than the actual number of symbols in a single frame, i.e., 840. Zeros are added at the beginning and the end of each frame to make each frame contains exactly 1024 symbols, as follows:

$$[0,...,0,P_1,X_1,X_2,...,X_6,P_2,X_7,...,P_n,X_{1+6(n-1)},...,P_{120},...,X_{720},0,...0]$$

After the IFFT of length 1024, the frame is

$$[x_1,x_2...,x_{1024}]$$

where x_n denote the n th symbol in modulated domain.

Then the prefix symbols are added. The OFDM modulated frame is

$$[x_{897},...,x_{1024},x_1,x_2...,x_{1024}].$$

4.3.2 Result analysis for different step-size parameters μ_f

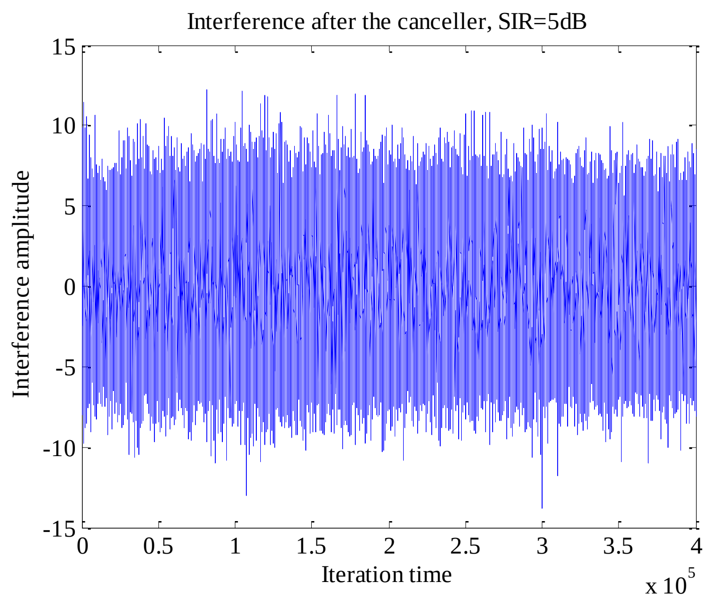
In chapter 3 the step-size problem is solved by the variable step-size approach. Even though the scheme described in Eq. 3.49 works for the simulation environment in Chapter 3, it does not apply in this case. $\mu_f = 0.01$ does not guarantee convergence for low SIR any more. In this section, the K-distribution parameters are set to

$$b = 1, v = 2.5$$

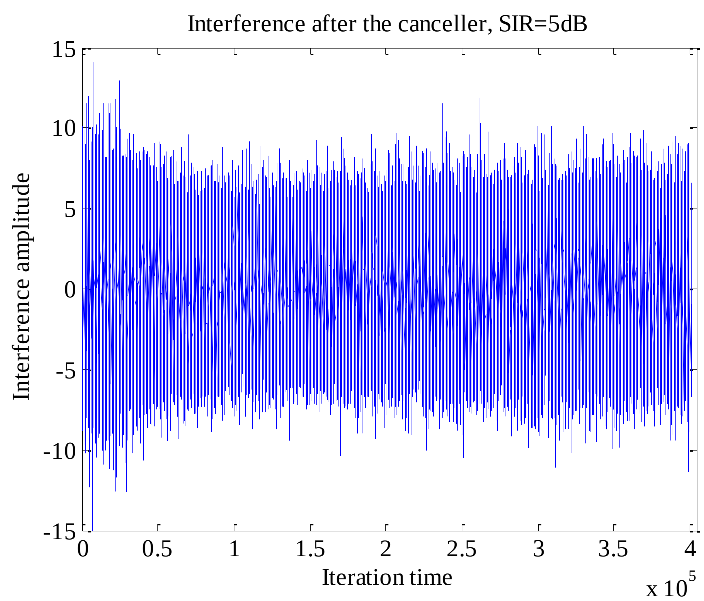
The simulation time is 5×10^5 samples for SIR=2.5, 5, 7.5 and 10dB. For SIR=12.5dB, the simulation time is set to 1.5×10^6 samples due to the low BER rate.

Fig. 4.3 shows the interference signal after cancellation when SIR=7.5dB. We can clearly see that in Fig. 4.3(a) when $\mu_f = 0.01$ the result is not as good as that in Fig. 4.3(b) when $\mu_f = 0.001$. However, similar to the simulation result in Chapter 3, when the SIR is high, using a smaller μ_f may cause a very long converge time. Fig. 4.4(a) shows the interference after cancellation when SIR=12.5dB and $\mu_f = 0.001$. Notice that the even after 1.5×10^6 samples, the cancellation result does not have any obvious difference compared to the original K-distributed interference. This means that the step-size is so small that barely influences the interference signal in the limited time. In Fig. 4.4(b), step-size is set as $\mu_f = 0.01$. Both the figures and the BER show the improvement of performance. So, the variable step-size scheme is

$$\mu_f = \begin{cases} 0.001, & SIR < 10 \\ 0.01, & elsewhere \end{cases} \tag{4.9}$$

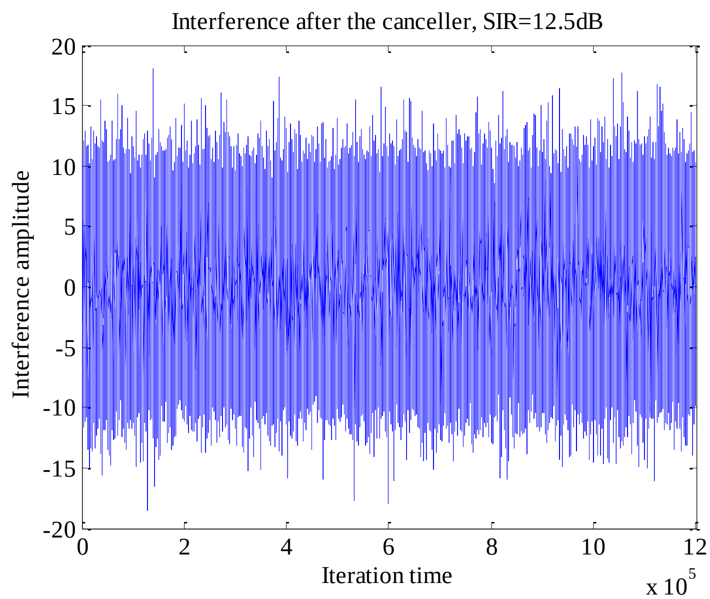


(a) $\mu_f = 0.01$

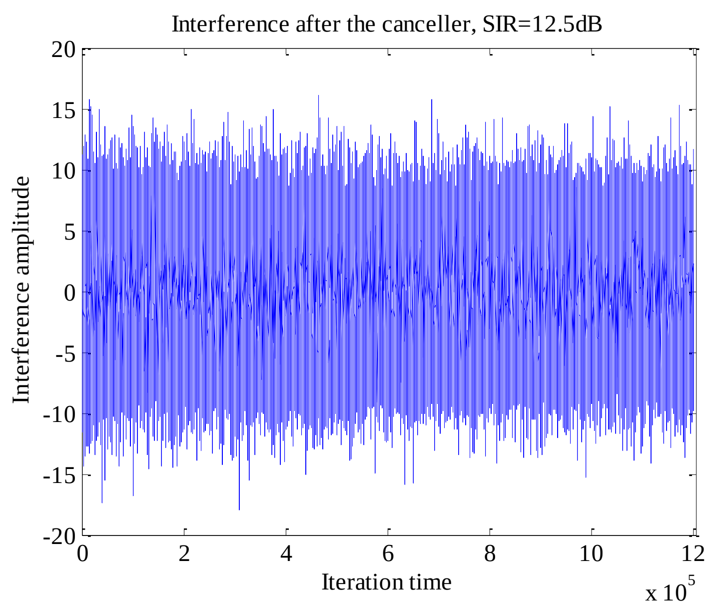


(b) $\mu_f = 0.001$

Figure 4.3 Cancellation comparison with different step-size when SIR=7.5dB



(a) $\mu_f = 0.001$



(b) $\mu_f = 0.01$

Figure 4.4 Cancellation comparison with different step-size when SIR=12.5dB

Using this scheme, the BER performance of FOS canceller when $b=1, v=2.5$ is illustrated in Fig. 4.5.

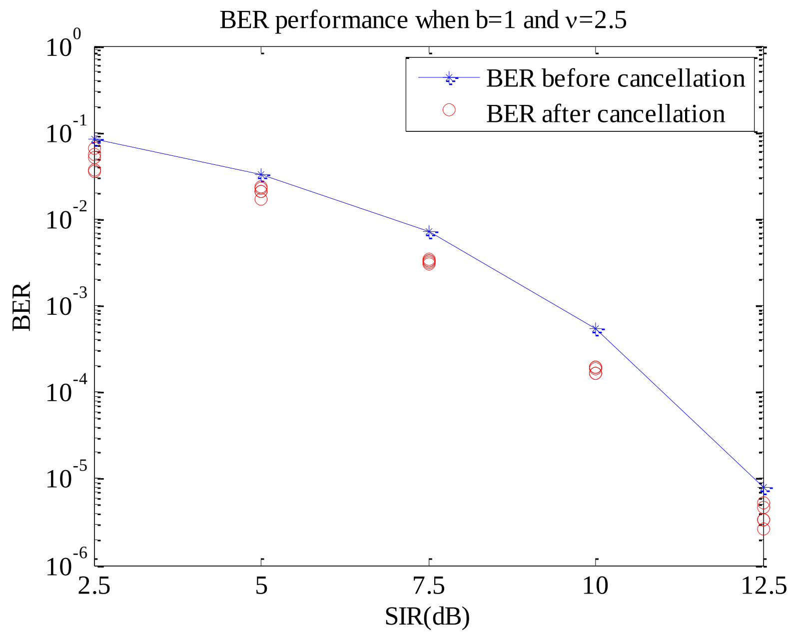


Figure 4.5 BER performance of variable step-size scheme

	SIR=5dB	SIR=12.5dB
$\mu_f = 0.001$	BER=0.0171	BER=6.67×10^{-6}
$\mu_f = 0.01$	BER=0.0227	BER=2.67×10^{-6}

Table 4.2 BER result for high and low SIR with different step-size

4.3.3 Result of different K-distribution parameters

[5] pointed out that the change of K-distribution parameters may result in different BER performance for the same cancellation algorithm. In case a) $b=2, v=1.5$ the cancellation algorithm described in [5] can improve the BER by 6db while in case b) $b=1, v=2.5$, the

improvement is less than 4dB. Now, Fig. 4.6 shows the performance of FOS canceller in both situations.

Although when the SIR is low there is no obvious difference between the two cases, it is clear that as the SIR increases, case a) still performs better than case b). However, this difference is not as high as it was in [5]. More importantly, the BER performance is much better than the result in [5], especially for small SIR. When SIR=2.5, FOS canceller and the discrete Kalman filter both can produce a BER between 0.1 and 0.01. But when SIR=10, the BER produced by FOS canceller is around 10^{-4} while the result of discrete Kalman filter is still above 0.01. This improvement is at the cost of computational complexity and the additional knowledge from the reference signal.

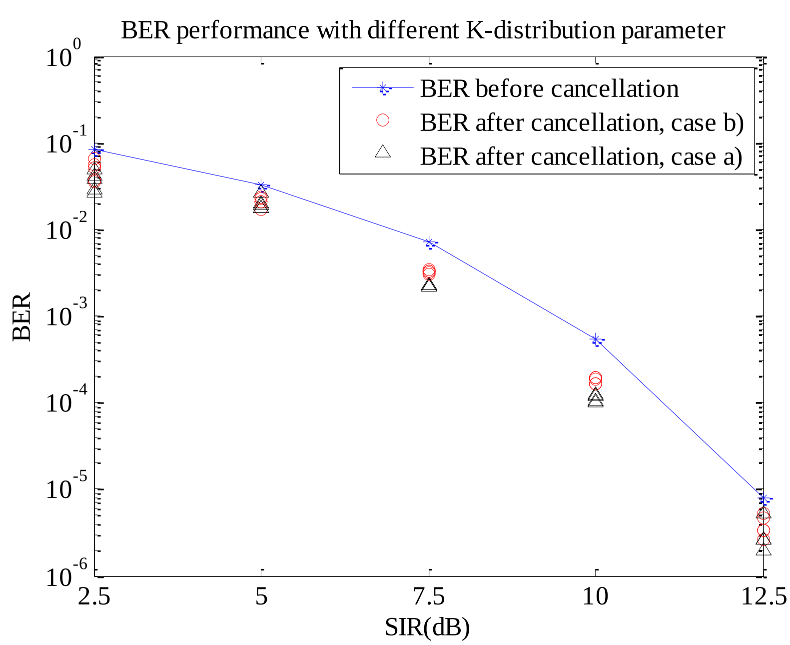


Figure 4.6 BER performance comparison for different K-distribution parameter

Chapter 5-Conclusion

This thesis applied a new noise/interference cancellation algorithm on the computer platform interference cancellation problem. Based on previous research, this new algorithm focused on the cancellation of the broadband K-distributed interference generated by multiple electronic-magnetic emission sources in a computer platform, in the presence of Gaussian noise. By introducing a new criterion of goodness instead of the regular mean square, the new adaptive algorithm is designed to reconstruct and to cancel the interference in a recursive fashion. It is proved to be effective on both experimental binary transmission system and the OFDM system which is widely applied in modern mobile communication devices. Compared to the previous cancellation method, the BER performance is improved considerably.

The broadband noise/interference in a computer platform is usually considered as Gaussian distributed. However, in E. Alban's previous work, it is proved that the interference is double-sided K-distributed. Thus, a new cancellation approach is needed for this specific interference. A discrete Kalman filter is designed but the performance is not satisfactory. It only improves the BER to around 10^{-2}, which is not enough for modern wireless communication.

Adaptive filtering is one of the efficient ways to solve the noise/interference cancellation problem. It requires minimum knowledge of the interference signal. Since the traditional LMS/NMLS are not suitable for this particular problem because of the existence of Gaussian noise, the higher order statistic algorithm is needed. Higher order cumulants are introduced as the new criterion of goodness. The new algorithm compares the higher order cumulants to improve the filter coefficients and reconstructs the interference signal from the reference signal. We verified through simulation that the fourth order statistics (FOS) algorithm is capable of mitigating the effect of K-distributed interference. A variable step-size mechanism is also introduced to ensure the convergence of the algorithm.

Furthermore, the FOS algorithm is implemented in an ODFM system and its performance is compared to that of the discrete Kalman filter. The BER performance of FOS canceller is better than the previous method. However, the large delay due to computational complexity and the availability of the reference signal are the problem left to be solved. A more efficient way to estimate the cross-cumulant from the process may simplify the computation procedure. The techniques that generate the reference signal at the receiver also need to be further studied.

Bibliography

[1] Parkvall, Stefan, and David Astely. "The Evolution of LTE towards IMT-Advanced."*Journal Of Communications* 4.3 (2009): 146-54.

[2] IEEE Standard for Information technology-- Local and metropolitan area networks-- Specific requirements-- Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications Amendment 8: IEEE 802.11 Wireless Network Management," *IEEE Std 802.11v-2011*, vol., no., pp.1,433, Feb. 9 2011

[3] F.C.Commission, "Radiated Emission Limits," code of federal regulations.Title 47(47CFR). Part 15 Subpart B,vol. 109, pp800, Oct 2012

[4] Alban, E.X.; Magana, M.E.; Skinner, H.G.; Slattery, K.P., "Statistical Modeling of the Interference Noise Generated by Computing Platforms," *Electromagnetic Compatibility, IEEE Transactions on* , vol.54, no.3, pp.574,584, June 2012

[5] E.X Alban, *Immune Radio Architecture for Platform Interference,* Oregon State University, May 2011

[6] P.R.Clayton, *Introduction to Electromagnetic Compatibility*, 2nd ed. Wiley-Interscience, January 2006

[7] Iskander, D.R.; Zoubir, A.M., "Estimation of the parameters of the K-distribution using higher order and fractional moments [radar clutter]," *Aerospace and Electronic Systems, IEEE Transactions on* , vol.35, no.4, pp.1453,1457, Oct 1999

[8] Blacknell, D.; Tough, R. J A, "Parameter estimation for the K-distribution based on [z log(z)]," *Radar, Sonar and Navigation, IEE Proceedings -* , vol.148, no.6, pp.309,312, Dec 2001

[9] Iskander, D.R.; Zoubir, A.M.; Boashash, B., "A method for estimating the parameters of the K distribution," *Signal Processing, IEEE Transactions on* , vol.47, no.4, pp.1147,1151, Apr 1999

[10] Ming Yan, "Soft decision directed MAP estimate of fast Rayleigh flat fading channels," *MILCOM 2000. 21st Century Military Communications Conference Proceedings* , vol.1, no., pp.368,372 vol.1, 2000

[11] Roberts, William J J; Furui, S., "Maximum likelihood estimation of K-distribution parameters via the expectation-maximization algorithm," *Signal Processing, IEEE Transactions on* , vol.48, no.12, pp.3303,3306, Dec 2000

[12] Mezache, A.; Sahed, M., "A method for estimating the parameters of the K-distribution using a nonlinear network based on fuzzy system and neural networks," *Signals, Circuits and Systems, 2008. SCS 2008. 2nd International Conference on* , vol., no., pp.1,6, 7-9 Nov. 2008

[13] Wachowiak, M.P.; Smolikova, R.; Zurada, J.M.; Elmaghraby, A.S., "Estimation of K distribution parameters using neural networks," *Biomedical Engineering, IEEE Transactions on* , vol.49, no.6, pp.617,620, June 2002

[14] A. Goldsmith, *Wireless Communication*, Cambridge University Press, 2005

[15] Ming-Xian Chang; Su, Yu-T, "Model-based channel estimation for OFDM signals in Rayleigh fading," *Communications, IEEE Transactions on* , vol.50, no.4, pp.540,544, Apr 2002

[16] Davis, L.M.; Collings, I.B.; Hoeher, P., "Joint MAP equalization and channel estimation for frequency-selective and frequency-flat fast-fading channels," *Communications, IEEE Transactions on* , vol.49, no.12, pp.2106,2114, Dec 2001

[17] Grami, A.; Makrakis, D.; Pasupathy, S., "Detection Using a Noise Estimation-Cancellation Technique," *Communications, IEEE Transactions on* , vol.35, no.8, pp.786,794, Aug 1987

[18] Dhanoa, J.S.; Ormondroyd, R.F.; Hughes, E.J., "Blind cancellation of wideband frequency modulated interference in a wideband acoustic communication channel," *OCEANS, 2005. Proceedings of MTS/IEEE* , vol., no., pp.1065,1069 Vol. 2, 17-23 Sept. 2005

[19] M.H.Hayes, Statistical *Digital Signal Processing and Modeling*, John Wiley & Sons Inc, 1996

[20]T. R. Field and R. J. A. Tough, “Dynamical models of weak scattering,” JOURNAL OF MATHEMATICAL PHYSICS, vol. 46, 2005

[21] Kay, S., "Representation and Generation of Non-Gaussian Wide-Sense Stationary Random Processes With Arbitrary PSDs and a Class of PDFs," *Signal Processing, IEEE Transactions on* , vol.58, no.7, pp.3448,3458, July 2010

[22] Moohong Lee; Byungjik Keum; Minjae Park; Young Serk Shim; Hwang-Soo Lee; Dae Ho Woo, "A frequency domain approach for complexity reduction in wideband radio interference cancellation repeaters," *Signal Processing, 2008. ICSP 2008. 9th International Conference on* , vol., no., pp.1971,1976, 26-29 Oct. 2008

[23] Kinoshita, M.; Sasaoka, N.; Danno, S.; Itoh, Y.; Sai, B., "A study on noise canceling system based on adaptive filter for OFDM system," *Intelligent Signal Processing and*

Communication Systems, 2005. ISPACS 2005. Proceedings of 2005 International Symposium on , vol., no., pp.697,700, 13-16 Dec. 2005

[24] Schafhuber, Dieter; Matz, G.; Hlawatsch, F., "Adaptive Wiener filters for time-varying channel estimation in wireless OFDM systems," *Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference on* , vol.4, no., pp.IV,688-91 vol.4, 6-10 April 2003

[25] Tangdi Zhang; Caixiang Wang; Xiaowen Li, "Frequency offset and channel estimation in OFDM systems by blind adaptive filter," *Mobile Congress (GMC), 2010 Global* , vol., no., pp.1,5, 18-19 Oct. 2010

[26] Dae C.Shin; Nikias, C.L., "Adaptive interference canceler for narrowband and wideband interferences using higher order statistics," *Signal Processing, IEEE Transactions on* , vol.42, no.10, pp.2715,2728, Oct 1994

[27] Nikias, C.L.; Raghuveer, Mysore R., "Bispectrum estimation: A digital signal processing framework," *Proceedings of the IEEE* , vol.75, no.7, pp.869,891, July 1987.

[28] Kay, S., "Representation and Generation of Non-Gaussian Wide-Sense Stationary Random Processes With Arbitrary PSDs and a Class of PDFs," *Signal Processing, IEEE Transactions on* , vol.58, no.7, pp.3448,3458, July 2010

[29] Watts, S., "Radar detection prediction in sea clutter using the compound K-distribution model," *Communications, Radar and Signal Processing, IEE Proceedings F* , vol.132, no.7, pp.613,620, December 1985

[30] Abraham, D.A.; Lyons, A.P., "Novel physical interpretations of K-distributed reverberation," *Oceanic Engineering, IEEE Journal of* , vol.27, no.4, pp.800,813, Oct 2002

[31] Nikias, C.L.; Raghuveer, Mysore R., "Bispectrum estimation: A digital signal processing framework," *Proceedings of the IEEE* , vol.75, no.7, pp.869,891, July 1987

[32] C.L. Nikias and A. P. Petropulu, *Higher-Order Spectra Analysis: A Nonlinear Signal Processing Framework*, PTR Prentice Hall, 1993