2005:67

LICENTIATE THESIS

L

Adaptive Blind Deconvolution Using Third-Order Moments: Exploiting Asymmetry

Patrik Pääjärvi

Luleå University of Technology
Department of Computer Science and Electrical Engineering
Division of Systems and Interaction

2005:67 | ISSN: 1402-1757 | ISRN: LTU-LIC -- 05/67 -- SE

Adaptive Blind Deconvolution Using Third-Order Moments: Exploiting Asymmetry

Patrik Pääjärvi

Department of Computer Science and Electrical Engineering
Luleå University of Technology
Luleå, Sweden

Supervisor:

Professor James P. LeBlanc

Abstract

This thesis focuses on the use of third-order statistics in adaptive blind deconvolution of asymmetric impulsive signals. Traditional methods are typically based on fourth-order moments, which can discriminate signals with heavy-tailed probability functions (i.e. 'spiky' signals) from corresponding filtered versions. The work herein demonstrates that, by using third-order central moments, *asymmetry* in such signals (e.g. only positive 'spikes') can be exploited to achieve faster convergence of algorithms and increased robustness to noise. The reasons for these benefits lie mainly in the use of error functions with lower polynomial orders, which leads to simpler gradient equations, improving the convergence rate. The increased robustness to noise is due to the fact that all odd-order statistics of symmetric signals (e.g. Gaussian noise) are zero.

A previously known computationally simple, norm-constrained algorithm for gradient search is also examined. It is demonstrated that this algorithm accomplishes third-order moment maximization by gradient ascent, without the undesired effect of increasing filter norm. Norm-constrained optimization is commonly achieved using periodic normalization of the filter vector, involving costly divides and square-root operations. The investigated algorithm requires significantly fewer operations, and uses only multiplications and additions, making it well suited for implementation on fixed-point digital signal processors.

Numerical experiments, demonstrating the usefulness of the proposed methods, include blind deconvolution of sound from a diesel engine, and blind equalization of a synthetic ultra-wideband (UWB) communication channel.

Contents

Acknowledgements

I would like to thank James LeBlanc for his guidance, support and encouragement. Thanks also to my friends and colleagues at CSEE, especially to Kristina Berglund who proofread the manuscript to this thesis, and to Johan Carlson and Martin Sehlstedt for letting me use their LaTeX template. I am grateful for financial support from Gellivare Hard Rock Research, the VINNOVA ProcessIT Innovations program, and the VITAL Mål 1 EU program.

Final thanks goes to my family and to Camilla for her love and support.

Patrik Pääjärvi
Luleå November 2005

Part I

Introduction

The general topic of this thesis is *adaptive blind deconvolution* for digital real-time applications. The specific issues of focus are methods that give low computational complexity and fast convergence of algorithms when used on impulsive signals. A common thread in the three publications supplied herein is exploitation of asymmetry in impulsive signals to gain such features. An introduction to the general area follows.

1 Thesis Overview

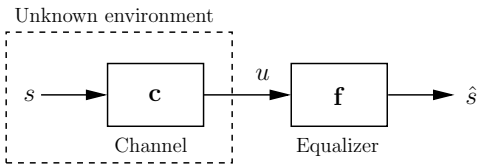

Figure 1: Block diagram of a channel-equalizer model.

Adaptive blind deconvolution can be described as reconstructing unknown, distorted signals using self-adjusting digital filters. Figure 1 shows a block-diagram model of the standard blind deconvolution problem. The signal s is a *desired* signal that needs to be transferred, measured or recorded. Assume that s cannot be directly observed: only a filtered version u is available. The object of *blind deconvolution* is to recover s from observations of u with limited, or no knowledge of either s or the system $\mathbf{c}$ through which s is filtered. In Figure 1, s is recovered from u using a *linear equalizer* $\mathbf{f}$, which acts as the *inverse system* to $\mathbf{c}$. The output signal $\hat{s}$ from the equalizer is an estimate of the desired signal.

A few notations will be defined. The signal s will be referred to as the *source signal*, $\mathbf{c}$ the *channel*, u the *channel output signal*, $\mathbf{f}$ the *equalizer*, and $\hat{s}$ the *equalizer output*

signal. A couple of examples are presented to clarify the subject.

Example: Mobile Communication

Figure 2 shows mobile communication between a cellphone situated in a car in motion and a base-station antenna. As the radio waves from the base-station propagate through the air, they get attenuated, lowering the quality of the transferred signal. However, another serious problem occurs due to the fact that the radio waves are reflected at different near- and far-field scatterers. The signal reaching the car in Figure 2 will therefore consist of a possible direct-path signal, combined with several echoed versions, each echo arriving at different times. This unavoidable effect, called *multipath spread*, causes distortion and severely limits the rate at which digital data can be transferred. Referring to Figure 1; the source signal s is here the base station signal, the channel $\mathbf{c}$ is the combined effect of the air and the scatterers (e.g. mountains and buildings), and the observed signal u is the actual radio waves that reach the car. The cellphone must use an equalizer to undo the channel effects for proper operation. Note that with s and $\mathbf{c}$ being unknown at the receiver, it is not immediately clear what equalizer parameterization should be used to achieve the desired effect.

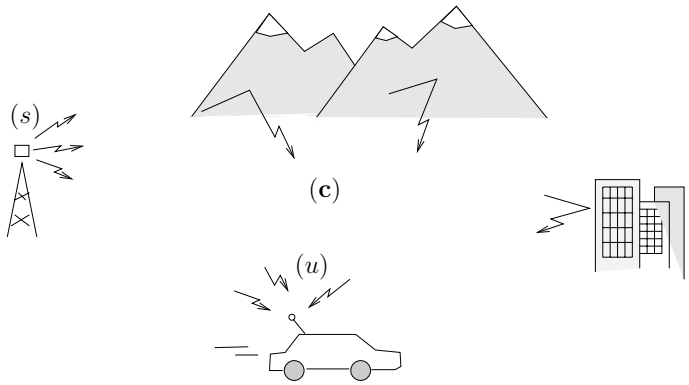

Figure 2: Mobile communication.

Example: Seismic Reflection Surveying

Figure 3 shows a simplified model of geological exploration using a seismic reflection method. Variants of this technique are used to obtain a seismic profile of an area of land, which could for example be used to detect oil or gas reserves. The principle behind the method is that different layers of earth have different densities, and a seismic wave that propagates through these layers will get reflected in the boundaries between them, due

to the change in seismic impedance. A cross-section of earth layers is shown in Figure 3. A hydraulic vibrator generates seismic waves (s) that propagate down into the ground ($\mathbf{c}$). The reflected waves (u) are recorded by a geophone situated at some distance from the vibrator. A single impulse from the vibrator will reach the geophone as a series of delayed and attenuated replicas.

If the 'channel', i.e. the combined layers of earth, can be identified, a seismic profile is achieved. Theoretically, if an equalizer can be found that, when applied to u perfectly recovers the source signal s, then all information of $\mathbf{c}$ must be contained within the equalizer. The problem of *identifying* a channel is fundamentally related to *equalizing* it.

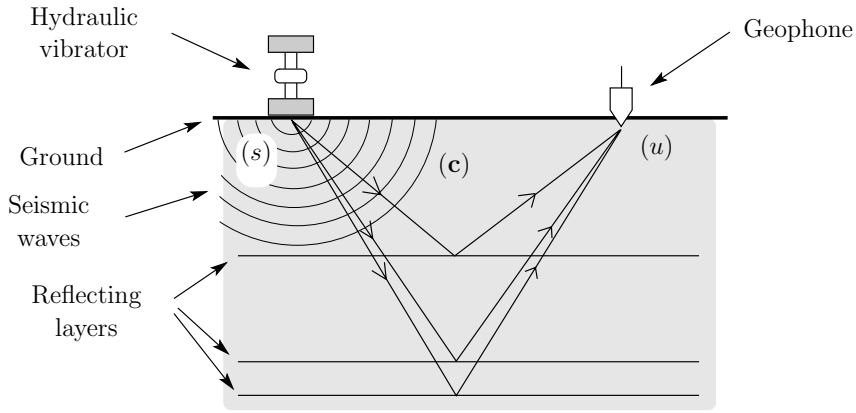

Figure 3: Seismic reflection surveying.

1.1 Time Dispersion

In both examples in Figures 2 and 3, the desired source signals suffer *time dispersion*. That is, the source signals get spread among several delayed copies of various strengths, arriving at the receiving end at different time instants: they are dispersed in time. Figure 4 shows time-amplitude plots of an impulse-shaped source signal and a corresponding time-dispersed channel output.

Dispersion in a mobile channel causes bit errors in the digital communication signal which, if not properly compensated for, can make information transfer impossible. The dispersion caused by the reflective earth layers in Figure 3 contains valuable information that needs to be retrieved. An equalizer designed to undo dispersion in a channel should, ideally, recover the source signal at its output. In most practical situations, a perfect equalizer cannot be constructed, and a sub-optimal variant that gives an estimate of s is used.

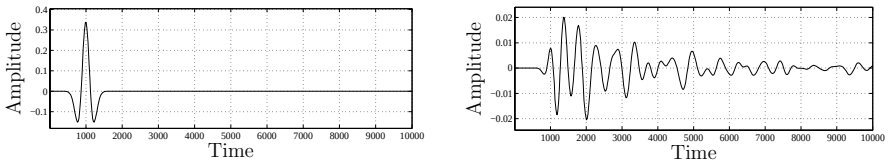

Figure 4: Example of a source signal s (left), and the time-dispersed channel output signal u (right).

1.2 Adaptive Equalizers

Consider again the mobile communication in Figure 2. Note that, as the car moves, the surrounding conditions will change. Hence, the *characteristics of the channel will change over time*. Therefore, an equalizer designed to continuously undo the channel must react to these variations. An *adaptive equalizer* tracks changes in a time-varying channel, and adjusts itself to give as good an estimate of s as possible within existing constraints. Although the seismic profile will not change (rapidly) over time, finding an equalizer that recovers the source signal may be difficult to do without resorting to some *iterative* method, in which an adaptive equalizer is adjusted using some algorithm, until a sufficiently accurate model is achieved. The following section gives an overview of the subject of adaptive filters, of which adaptive equalization is one application.

2 Adaptive Filtering

Adaptive filters, i.e. filters that are self-adjusting according to some criteria, are used in many applications of signal processing. They are the subject of many textbooks, two good references are [1] and [2]. To motivate the use of adaptive filters over fixed filters, consider the following situations.

- **Changing environments:** If the parameters of the application are time-varying, an adaptive filter may be used to track such variations. The mobile channel is an example of a time-varying system. An equalizer used to continuously mitigate the dispersion caused by that channel must therefore be adaptive.

- **Unknown environments:** Deriving an equalizer is usually straightforward if a mathematical model of the channel is known. However, if the model is unknown, traditional design methods cannot be used. An adaptive equalizer may be *trained* to obtain a good inverse of a channel to which a model is not available.

- **Complex environments:** If the model of a channel is very complex, there may not be a closed-form solution to the equalizer design problem. In such cases, an equalizer may be adapted to obtain an iterative solution.

The fundamentals of adaptive filters are best explained by examples. Two common applications are shown below.

Example: System Identification

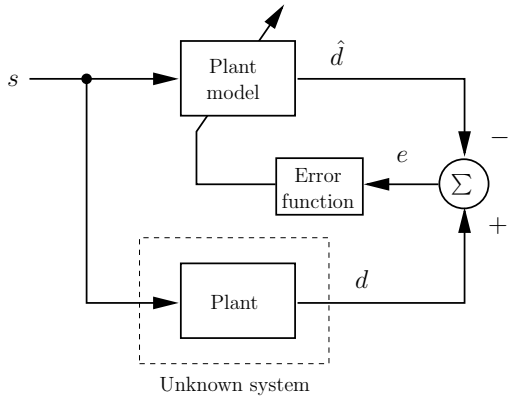

Figure 5: Adaptive system identification.

Figure 5 shows a simple model in which an adaptive filter used for identification of an unknown system. The unknown system might for example be a plant (machinery) for which an automatic control system is to be designed. For this purpose, a reasonable mathematical model of the plant is needed.

A signal s is input to both the adaptive filter and the plant. An *error signal* e is then formed as the difference between the plant output, or *reference signal*, d and the adaptive filter output $\hat{d}$. The error signal is then fed to some *error function*, which determines how the filter parameters should be adjusted to minimize e. If the filter can be adapted to make e vanish, $\hat{d}$ must equal d, and the filter becomes a model of the plant. In reality, e can never be driven to zero indefinitely, due to limitations in the adaptive filter, measurement noise, etc. Minimizing the *mean power* of e is therefore the typical approach.

This form of adaptive system identification may be employed when the unknown system has a complex structure, preventing the use of physical modeling. Another motivation for adaptive identification might be that the plant characteristics change over time. Automatic control of a time-varying plant calls for the use of continuous, or *on-line* system identification.

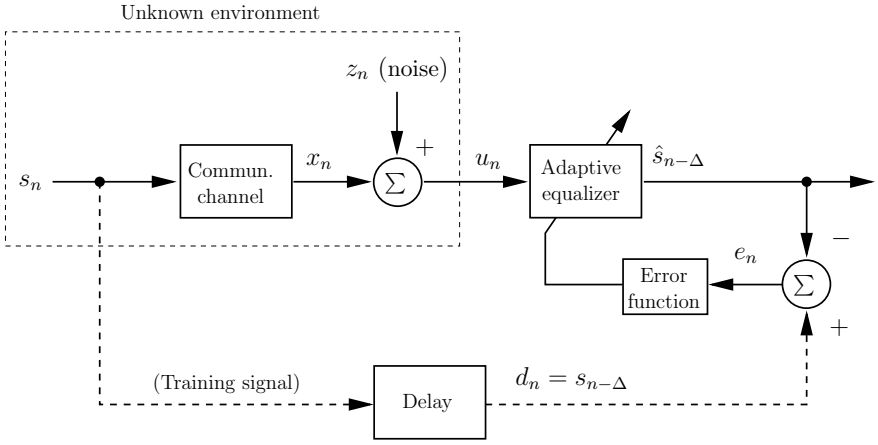

Figure 6: Adaptive channel equalization.

Example: Equalization of a Digital Communication Channel

Figure 6 shows a model of adaptive equalization of a digital communication channel. The subscript n denotes a sampling time index. The signal s_n is transferred over a channel with unknown characteristics. At the other end of the communication link, the channel output x_n plus some channel noise z_n is received as the signal u_n. To recover s_n from u_n, an adaptive equalizer is adjusted according to some function of an error signal e_n. However, unlike the system identification example, the reference signal here is the transferred signal s_n, which the equalizer output should equal ideally. The error signal e_n should therefore be the difference between what was transferred and what is actually received. Of course, e_n cannot be formed in this way. If the receiver had access to s_n, there would be no need for an equalizer in the first place.

A common method used to obtain the error signal is to send a *training signal* in s_n. The training signal is a short stream of symbols known beforehand by the receiver (e.g. stored locally in a waveform table), sent at a pre-determined time, typically at the start-up phase of communication. While the training signal is transmitted, the receiver can now form an error signal e_n, since it knows what is being sent in s_n. Adaptation of the equalizer is thus done during reception of the training signal, after which a dispersion-free communication hopefully can be initiated. The dotted line in Figure 6 symbolizes the virtual link that the training signal gives. The delay element compensates for the combined delay of the channel and equalizer.

If the communication channel in Figure 6 is time-varying, the training signal needs to be transmitted periodically, so that the equalizer may continuously track changes in the channel characteristics. Note that while this gives effective equalization, using up part of the transmissions for training reduces the useful information throughput of the link.

Given the two previous examples, a few mathematical preliminaries of adaptive digital filters are given, following approximately the development of [1].

2.1 Mean-Square Error (MSE) Minimization

The purpose of an adaptive filter is to minimize the error signal e_n in some way since, as mentioned, making e_n vanish for all n is usually not possible due to measurement noise, limited filter order and other effects. Thus, the filter must minimize some suitable *error function* of e_n. A popular choice of error function is the *mean-square error* (MSE), defined as the mean power of e_n,

$$\xi \triangleq \mathrm{E}\left\{e_n^2\right\} = \mathrm{E}\left\{(d_n - y_n)^2\right\}, \tag{1}$$

where $\mathrm{E}\{\cdot\}$ denotes expectation. The filter used is assumed to be an FIR (finite impulse response) filter of order N, represented by the real coefficient vector

$$\mathbf{f} \triangleq [f_0 \; f_1 \; \cdots \; f_N]^{\mathrm{T}}.$$

For now, $\mathbf{f}$ is assumed to be constant in time, i.e. independent of n. The filter output signal y_n is then the vector inner product between $\mathbf{f}$ and the *regressor vector* (i.e. filter state) of input samples to the filter, $\mathbf{u}_n \triangleq [u_n \; u_{n-1} \; \cdots \; u_{n-N}]^{\mathrm{T}}$,

$$y_n = \mathbf{f}^{\mathrm{T}}\mathbf{u}_n. \tag{2}$$

Inserting (2) into (1) gives

$$\begin{aligned} \xi &= \mathrm{E}\left\{d_n^2\right\} + \mathbf{f}^{\mathrm{T}}\mathrm{E}\left\{\mathbf{u}_n\mathbf{u}_n^{\mathrm{T}}\right\}\mathbf{f} - 2\mathrm{E}\left\{d_n\mathbf{u}_n^{\mathrm{T}}\right\}\mathbf{f} \\ &= \mathrm{E}\left\{d_n^2\right\} + \mathbf{f}^{\mathrm{T}}\mathbf{R}\mathbf{f} - 2\mathbf{P}^{\mathrm{T}}\mathbf{f}, \end{aligned} \tag{3}$$

with $\mathbf{R}$ and $\mathbf{P}$ defined as

$$\begin{aligned} \mathbf{R} &\triangleq \mathrm{E}\left\{\mathbf{u}_n\mathbf{u}_n^{\mathrm{T}}\right\}, \quad \text{and} \\ \mathbf{P} &\triangleq \mathrm{E}\left\{d_n\mathbf{u}_n^{\mathrm{T}}\right\} \end{aligned}$$

respectively. Under the assumption of u_n and d_n being wide-sense stationary random processes, the elements of $\mathbf{P}$ and $\mathbf{R}$ are constant.

From (3), it is seen that the mean-square error is a quadratic function of the filter vector. Taking the gradient of (3) with respect to $\mathbf{f}$ gives

$$\boldsymbol{\nabla} = 2\mathbf{R}\mathbf{f} - 2\mathbf{P}. \tag{4}$$

A vector $\mathbf{f}^*$ that minimizes ξ must be the solution to $\boldsymbol{\nabla} = \mathbf{0}$, which is

$$\mathbf{f}^* = \mathbf{R}^{-1}\mathbf{P}, \tag{5}$$

assuming the inverse of $\mathbf{R}$ exists. It should be noted that a necessary condition for $\mathbf{f}^*$ to be the unique vector that minimizes ξ is that $\mathbf{R}$ is positive definite [2].

2.2 Iterative MSE Minimization: The LMS Algorithm

From Equation (5), the unique N^{th}-order filter vector $\mathbf{f}^*$ that, under the necessary condition on $\mathbf{R}$, minimizes the mean square of e_n is given. However, lacking the stationarity in u_n and d_n or other necessary conditions, there is often a need for adaptive filters which minimize the MSE in an *iterative* manner. Some more precise motivations may now be given.

- **Invertibility of R :** Existence of $\mathbf{f}^*$ requires that $\mathbf{R}$ is non-singular. In practical situations, this cannot usually be assumed. Therefore, an approximate solution to (5) must in general suffice.
- **Time-invariance of R and P :** The channel output signal x_n in Figure 6 is the convolution between the channel $\mathbf{c}$ and the source signal s_n. Even though s_n is usually a wide-sense stationary process, x_n (and hence u_n) may not be, for example if the channel is time-varying. Then, neither $\mathbf{R}$ nor $\mathbf{P}$ will be constant over time, and no unique solution for all n exists. Thus, a time-varying equalizer is required.
- **Computational complexity of calculating $\mathbf{f}^*$:** The computational cost of directly inverting $\mathbf{R}$ to obtain $\mathbf{f}^*$ may be very high. An iterative solution method is usually much more efficient.

With these motivations, some preliminaries on filter adaptation is given. The adaptive N^{th}-order FIR filter is represented at time n by the coefficient vector

$$\mathbf{f}_n \triangleq [f_{0n}\ f_{1n}\ \cdots\ f_{Nn}]^{\mathrm{T}},$$

with the subscript n indicating that the coefficients are now time-varying. Adapting $\mathbf{f}_n$ is usually done in some algorithm of the form

$$\mathbf{f}_{\text{new}} \ = \ \mathbf{f}_{\text{old}} \ + \ \text{stepsize} \ \times \ \text{gradient}, \tag{6}$$

where the 'stepsize' is a small scalar constant that determines how large or small an adaptation step should be, and the 'gradient' is a vector specifying the direction and amount of relative adjustment of the filter coefficients needed to minimize the error function. If the negative gradient of the error function is used, adaptation by *steepest descent* is obtained, i.e. the filter vector is updated in the direction in which the error function decreases most rapidly. Using the mean square of e_n as defined in (1), the gradient is obtained from (4). Note that calculating the right-hand side of (4) in practice involves estimations of $\mathbf{R}$ and $\mathbf{P}$. A more computationally efficient algorithm is obtained if the instantaneous value of the squared error signal is taken as the error function,

$$\hat{\xi}_n \triangleq e_n^2.$$

Taking the partial derivative of $\hat{\xi}_n$ with respect to $\mathbf{f}_n$ gives the *instantaneous* gradient

$$\hat{\nabla}_n \triangleq \frac{\partial \hat{\xi}_n}{\partial \mathbf{f}_n} = -2e_n\mathbf{u}_n.$$

Note that the subscript n is used to indicate that $\hat{\xi}_n$ and $\hat{\nabla}_n$ are time-varying. If the adaptive filter is iteratively updated for each n in the negative direction of $\hat{\nabla}_n$, the popular *least-mean square*, or *LMS* algorithm [3], [1] is obtained,

$$\mathbf{f}_{n+1} = \mathbf{f}_n + \mu 2e_n\mathbf{u}_n, \tag{7}$$

which has a very simple structure. To a large extent, the performance of the LMS algorithm depends on the size of the stepsize parameter μ. A short discussion on this topic is therefore justified.

2.3 Choice of Stepsize Parameter

The size of the stepsize parameter μ controls the 'length' of each update step in general algorithms of the form (6). In simple words, a large μ will rapidly move the filter vector towards the optimum setting. However, there are tradeoffs between convergence rate and performance, which is discussed shortly for the LMS algorithm.

Since the instantaneous estimate of the true gradient (4) is used in (7), the LMS algorithm does not give a true steepest-descent minimization, but instead a so called *stochastic gradient algorithm*, in which the gradient estimate is based directly on noisy samples of e_n and u_n. The resulting mean-square error ξ_n can never be smaller than $\xi_{\min}$, obtained by the optimum filter $\mathbf{f}^*$. The *misadjustment* M of the algorithm is defined as the ratio between the *excess MSE* (i.e. the difference between $\xi_{\min}$ and ξ_n) and $\xi_{\min}$,

$$M \triangleq \frac{\xi_n - \xi_{\min}}{\xi_{\min}}.$$

It can be shown that, under general settings, M is proportional to the stepsize μ. Thus, a smaller μ will in general result in a smaller misadjustment. Furthermore, if the noise level in the signals is high, a small stepsize causes the LMS algorithm to behave as a lowpass filter, averaging out the noise in the instantaneous gradient estimate.

To summarize, in general:

- **Excessively large** μ may make filter adaptation unstable, causing the coefficients to diverge.
- **Large** μ gives fast convergence but large misadjustment.
- **Small** μ gives slow convergence but small misadjustment.

A good reference on the stability and performance of adaptive filter algorithms is [4].

2.4 Extensions and Alternatives to The LMS Algorithm

Several extensions to the LMS algorithm have been developed, a handful are covered in more or less detail in [2]. A few will be mentioned here. Despite the relative simplicity of the LMS algorithm, some applications call for even lower computational cost. The *sign-error LMS*, *sign-data LMS* or *sign-sign LMS* are algorithms that exhibit LMS-like behavior, but at a substantially lower computational cost. If rate of convergence is an issue, the *recursive least squares* (RLS) algorithm is an option, though it requires relatively many computational operations. If one or several eigenvalues of the matrix $\mathbf{R}$ are zero, the LMS algorithm may exhibit unlimited growth in some filter coefficients. The *leaky LMS* algorithm is then suitable. There is usually a trade-off between low computational complexity and high performance in adaptive algorithms.

3 Blind Deconvolution

The fundamental approach in adaptive filtering is to iteratively adjust the filter coefficients so that the filter output mimics some desired signal as much as possible. However, in some situations, the desired signal *may not be known or it may not be observable*. The problem of recovering a source signal s, distorted by some unknown channel $\mathbf{c}$, is called the *blind deconvolution problem*. This section will present some fundamental aspects of adaptive blind deconvolution.

3.1 Introduction to Blind Deconvolution

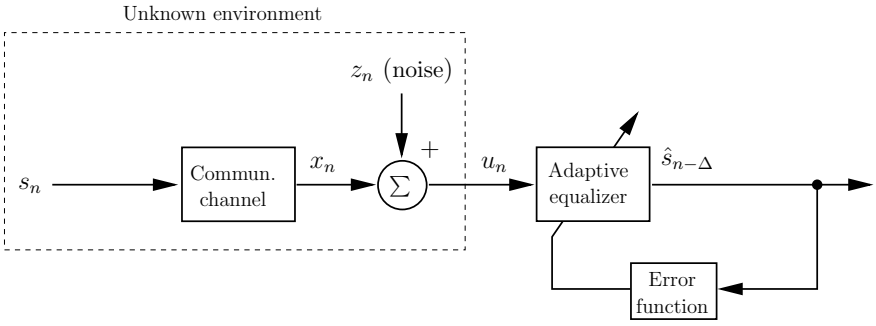

Figure 7: Blind adaptive channel equalization.

As an example of a blind deconvolution application, consider the modified channel equalization model shown in Figure 7. Comparing this model with the adaptive equalization model in Figure 6, notice that here the adaptive equalizer is adjusted based on

an error function of *the equalizer output alone*. Hence, no training signal is used, and therefore no error signal (as defined for the LMS algorithm) can be formed. The objective is still to estimate the source signal s with the equalizer output, i.e. to undo the channel.

The training signal commonly employed in digital communication is sent either once (at the start-up phase) or periodically to track time-varying channel characteristics. To maximize information throughput, it might be desirable to avoid a periodic transmission of a training sequence. This would motivate the use of blind equalization where the equalizer is adapted solely based on its own output signal. One of the earliest articles on adaptive channel equalization without training signals is [5] by Lucky from 1966.

Adapting a filter without a reference signal may seem less intuitive. Some basic ideas of how an adaptive filter may be adjusted without a training sequence follows.

3.2 Basic Concepts

To develop the basic concepts of blind deconvolution, a probabilistic view of filtering of discrete-time signals is useful. Specifically, the *probability distribution* of the samples x_n of a signal sequence $\{x_n\}$ will be of interest. First, some definitions are made.

Definition 1) The distribution D_x of x_n is *Gaussian* with mean m_x and variance σ_x^2 if

$$x_n : N(m_x, \sigma_x^2),$$

that is, every sample of $\{x_n\}$ is drawn from a normal distribution. A signal with a Gaussian distribution is said to be a *Gaussian signal*.

Definition 2) A zero-mean sequence $\{x_n\}$ is *uncorrelated* or *white* if

$$\mathrm{E}\{x_i x_j\} = \begin{cases} \sigma_x^2, & i = j \\ 0, & i \neq j \end{cases},$$

that is, if the samples in x_n are mutually independent. A white sequence with identically distributed samples is called an *independent identically distributed* (i.i.d.) sequence.

Definition 3) The *inverse* $\mathbf{c}^{-1}$ of a linear system $\mathbf{c}$ is the system for which, when applied in cascade with $\mathbf{c}$, the global output signal equals the global input signal up to a possible scale and delay.

Note: For practical reasons, Definition 3 differs from the classical notion of an ideal system inverse, which gives unit scale and no delay.

Consider the model in Figure 8, in which two systems $\mathbf{c}$ and $\mathbf{f}$ are cascaded. The signals s_n, u_n and y_n have distributions D_s, D_u and D_y respectively. The following properties apply under general assumptions [6] if $\{s_n\}$ is an i.i.d. sequence, and play fundamental roles in the theory of blind deconvolution.

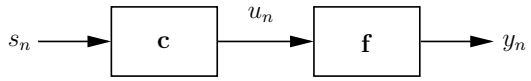

Figure 8: Two cascaded systems.

Property 1) If $D_u = D_s$ and **c** is a non-trivial system (not a simple gain), then s_n is Gaussian.

Property 2) If $D_y = D_s$ and s_n is non-Gaussian, then **f** is the inverse of **c**.

The relevance of these properties can be summarized:

- Property 1 states that if s_n is non-Gaussian, then the output signal u_n will not have the same probability distribution as s_n.
- Property 2 then means that if a filter **f** can be found for which s_n and y_n have the same distribution, then $y_n = \alpha s_{n-\Delta}$ for some scale α and some delay Δ.

One strategy for performing blind equalization for systems like the communication channel in Figure 7 would therefore be:

Approach A: Assuming s_n is a non-Gaussian i.i.d. sequence with *known* distribution, adjust the equalizer until the distribution of its output equals the distribution of s_n.

This characterization of the blind deconvolution problem was suggested by Benveniste, Goursat and Ruget in 1980 [6]. Multilevel digital communication is an example of an application where complete knowledge of the source distribution is assumed.

Another view of the problem was developed in 1981 by Donoho [7], who defined a *partial ordering*, measuring the relative 'Gaussianity' between random variables. As a consequence of the central limit theorem, under general conditions, the distribution of a non-trivial sum of independent, non-Gaussian random variables is always more close to a Gaussian, than the distributions of the individual variables. In other words, filtering a white sequence s_n with a non-trivial filter makes the output signal 'more Gaussian' than s_n. An alternative to the previous approach would hence be:

Approach B: Assuming s_n is a non-Gaussian uncorrelated sequence, adjust the equalizer until the distribution of its output is as non-Gaussian as possible.

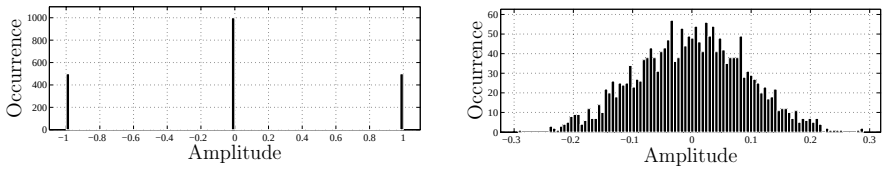

Figure 9: Histograms of a source signal s_n (left) and a corresponding time-dispersed channel output signal u_n (right).

Notice that here, unlike in Approach A, the actual distribution of s_n is not essential.

An illustration of the difference between input and output distribution of a channel is shown in Figure 9. The left plot shows the histogram of a random sequence $\{s_n\}$ of independent values of ± 1, padded with zeros. The right plot shows the histogram of $\{s_n\}$ filtered by a channel with a long impulse response. Note that the shape of the output signal histogram resembles that of a Gaussian probability density function.

In a publication preceding [6] and [7], Wiggins investigated blind deconvolution of seismic traces [8]. Knowing that the source signal had an impulsive ('spiky') appearance, Wiggins proposed a method based on adjusting the deconvolution filter to make its output as impulsive as possible. Although this idea is less formal than those proposed in [6] and [7], it may be formulated as a third blind deconvolution approach:

Approach C: Assuming s_n is a non-Gaussian signal with an *assumed* distribution $\tilde{D}_s$, adjust the equalizer to make its output signal distribution resemble $\tilde{D}_s$.

The common theme in all of the three approaches above is *characterization of the equalizer output distribution.* The error function of the blind equalization problem in Figure 7 should therefore be used to measure some feature of the equalizer output distribution (e.g. how impulsive or non-Gaussian it is). Choosing a suitable error function for a particular blind deconvolution problem can have a large impact on performance. As the main theme of this thesis is blind deconvolution problems where the source distribution is not completely known, and where computational efficiency of the methods is of importance, attention will be focused towards error functions that match these criteria.

3.3 Higher-Order Moments in Blind Deconvolution

The purpose of the error function as shown in Figure 7 is to measure some characteristic of the probability distribution of the equalizer output. Using Approach B, the error function should be a measure of how *far from Gaussian* the output distribution is. One

such measure is the *kurtosis* of y_n, defined as

$$\mathcal{K} \triangleq \frac{\mathrm{E}\{y_n^4\}}{\left(\mathrm{E}\{y_n^2\}\right)^2},$$

i.e. the normalized fourth-order moment of y_n. Kurtosis was used in [8] under the name 'Varimax norm', and was shown in [7] to give consistent discrimination between a distribution D_s and a (non-trivial) filtered version of it.

In 1980, Godard [9] proposed the use of *dispersions of order p*, defined as

$$\mathcal{D}^{(p)} \triangleq \mathrm{E}\left\{|y_n|^p - R_p\right\}^2$$

(p positive integer) for blind equalization of digital QAM (quadrature amplitude modulation) signals. $\mathcal{D}^{(p)}$ is the quadratic deviation of the p^{th}-order modulus of y_n from a dispersion constant R_p. Therefore, if minimization of $\mathcal{D}^{(p)}$ is used as criterion for a blind equalizer, its output will be driven towards a constant modulus. $\mathcal{D}^{(2)}$ is known as the widely used *constant modulus algorithm* (CMA) [10].

Both $\mathcal{K}$ and $\mathcal{D}^{(p)}$ are based on *central moments* of y_n. The p^{th}-order central moment of a zero-mean stochastic variable y_n is defined as

$$\mathcal{M}_p \triangleq \mathrm{E}\{y_n^p\}.$$

The base of CMA and kurtosis is the fourth-order central moment, $\mathcal{M}_4$. The choice of fourth-order statistics is not entirely arbitrary, some reasoning behind this is given.

- In general, error functions with polynomial structures should use as low order as possible. For example, the LMS algorithm uses the square of e_n as error function. A variant called the *least mean fourth* (LMF) algorithm, using the fourth power of e_n, has been proposed [11]. Although LMF may outperform LMS in certain situations, using higher powers in the gradient estimate may have a negative impact on the stability of the algorithm, for example if the signal-to-noise ratio is low [11].

- If the source sequence $\{s_n\}$ in Figure 7 is white, a *whitening filter* in the equalizer's place would produce an uncorrelated output. Finding the whitening filter can be done using second-order statistics. However, this only determines the *magnitude* of the unknown channel, the phase information can not be retrieved. If the channel is known to be *minimum phase*, i.e. if all of its zeros are confined within the unit circle, there exists a unique relationship between magnitude and phase. Unfortunately, most real channels are not minimum phase, and hence second-order statistics are typically insufficient for identifying $\mathbf{c}$, and higher order statistics must be employed [6], [12].

- Third-order statistics may allow identification of both magnitude and phase of general channels, but they are still less commonly used compared to fourth-order

statistics. The reason for this is that all odd-order moments of a signal with a symmetric probability density function (PDF) are zero, and many of the signals traditionally considered in blind deconvolution are symmetric (e.g. communication signals). However, it should be noted that a filtered version of an asymmetric signal tends to look more symmetric, which may lead to the false impression of the source being symmetric if a mathematical or physical model of the application is not known. As an example, consider a hammer hitting a piece of metal. The actual hammer impact, which is hard to measure, is an impulsive, asymmetric signal, but the ringings in the metal may appear symmetric if measured. The class of asymmetric signals may therefore be much larger than traditionally assumed.

Consequently, blind deconvolution methods commonly use fourth-order statistics since they impose no restrictions on channel phase characteristics or on the symmetry, or asymmetry, of the source signal PDF.

3.4 Computational Costs of Error Functions

While kurtosis has benefits such as scale invariance in its argument, the associated gradient equation is relatively complex. Godard's proposed dispersions $\mathcal{D}^{(p)}$ lead to simpler gradient expressions, and are hence popular in applications where computational efficiency is of concern. Adaptive equalizers are usually used in on-line (real-time) applications, where algorithms need to be simple. Moreover, if implemented on a digital signal processor, arithmetic operations such as divisions and square roots typically require significantly more processing power compared to multiplications and additions. The structure of $\mathcal{D}^{(p)}$ is hence attractive for implementations requiring low cost.

3.5 Higher-Order Moments and Impulsive Signals

An interesting property of higher-order moments is their ability to measure how 'heavy-tailed' the PDF of a signal is. A heavy-tailed PDF corresponds to a sparse (spiky) signal. The p^{th}-order central moment of a zero-mean stochastic variable y can be written

$$\mathrm{E}\{y^p\} = \int_{-\infty}^{\infty} y^p f(y) dy,$$

where $f(y)$ is the PDF of y. Figure 10 shows sketches of two hypothetical probability density functions along with the function y^4. The left PDF corresponds to an impulsive signal, and the right to a non-impulse signal. It is readily seen that the fourth-order moment, i.e. the integral of the product between $f(y)$ and y^4 will be larger for the impulsive signal. Hence, following Approach C for blind deconvolution, maximizing the fourth-order moment of the equalizer output would be a suitable objective if the source signal s_n is known to be sparse. Compare with Figure 9, where the heavy tails of the

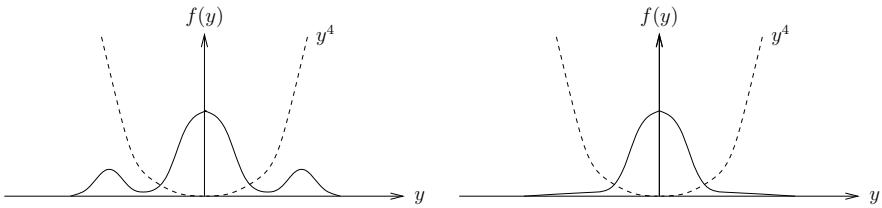

Figure 10: Hypothetical PDF's of an impulsive signal (left) and a non-impulsive signal (right) along with y^4.

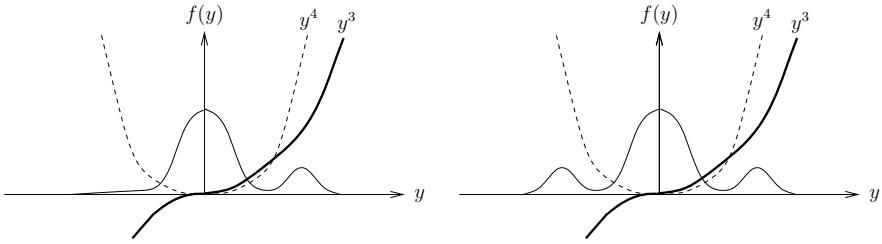

Figure 11: Hypothetical PDF's with heavy tails on one side only (left), and on both sides (right) along with y^4 and y^3.

source signal's PDF are not visually present in the output PDF due to the time dispersion caused by the channel.

As a comparison, Figure 11 shows plots of two heavy-tailed PDF's, one heavy-tailed to one side (left plot), and one heavy-tailed to both sides (right plot), along with the functions y^3 and y^4. From the left plot, it is seen that if the PDF is heavy-tailed to the positive side only, the corresponding signal will have high fourth-order moment as well as third-order moment. The absence of a negative heavy tail does not matter much, both $\mathcal{M}_4$ and $\mathcal{M}_3$ will discriminate such a signal from one without any heavy tails.

The situation is different if a signal with both positive and negative heavy tails is considered, as in the right plot in Figure 11. As mentioned earlier, the third-order moment will be zero if the PDF is symmetric around $y = 0$. This can of course be generalized to any odd-order moment.

The conclusion from this discussion is, that if the source signal is both *asymmetric and impulsive* (i.e. if it has an asymmetric, heavy-tailed PDF), both odd- and even-order moments (of order greater than two) can be used as error functions for blind deconvolution. If the signal is symmetric, however, only even-order moments will work.

4 Summary of Contributions

The three papers included in this thesis focus on adaptive blind deconvolution, or equalization, of asymmetric, impulsive signals (i.e. signals with a PDF heavy-tailed to one side). Of main interest is choosing a computationally simple error function that takes advantage of the characteristics of such signals.

Paper A - Skewness Maximization for Impulsive Sources in Blind Deconvolution

Authors: Patrik Pääjärvi and James P. LeBlanc
Reproduced from: Proceedings of the 6$^{\text{th}}$ Nordic Signal Processing Symposium, 2004 (NORSIG 2004), Espoo, Finland.

Paper A compares the use of kurtosis and *skewness* (normalized third-order moment) for adaptive blind deconvolution of asymmetric impulsive signals. It is shown that the function surface corresponding to skewness may have much fewer stationary points, compared to that of kurtosis. This should in general give faster convergence of gradient search algorithms if skewness is employed. These results are confirmed in an experiment on blind deconvolution of recorded sound from a diesel engine. A visual comparison between the function surfaces of skewness and kurtosis with respect to the coefficients of a second-order filter is made, highlighting the simpler structure of the skewness function surface.

Paper B - On-Line Adaptive Blind Deconvolution Based on Third-Order Moments

Authors: Patrik Pääjärvi and James P. LeBlanc
To appear in: IEEE Signal Processing Letters, vol. 12, no. 12, December 2005.

Paper B considers adaptive blind equalization using higher-order central moments. A comparison in terms of algorithm convergence rate and robustness to white Gaussian channel noise is made between third- and fourth-order moments. It is shown that the use of third-order moments should give faster convergence and better robustness to noise compared to fourth-order moments. Blind equalization of a synthetic ultra-wideband (UWB) channel with pulse-position modulation signaling is performed, demonstrating the faster convergence of an algorithm based on third-order moments.

Paper C - Computationally Efficient Norm-Constrained Adaptive Blind Deconvolution Using Third-Order Moments

Authors: Patrik Pääjärvi and James P. LeBlanc
Submitted to: 2006 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2006), Toulouse, France.

Paper C addresses the problem of increasing filter norm in gradient-ascent algorithms using third-order moments. The analysis of a previously known algorithm is extended using averaging analysis, to determine under what conditions on the adaptation stepsize the algorithm is stable in the filter norm. The algorithm is shown to be more computationally efficient compared to traditional methods, maintaining approximately unit filter norm over iterations. Experimental results, confirming the analysis, are supplied for adaptive blind equalization of a synthetic ultra-wideband (UWB) channel with pulse-position modulation (PPM) signaling.

5 Conclusions

The results of this work indicate that asymmetry in source signals should be exploited. The use of third-order moments allows this. Although traditional methods based on fourth-order moments may give desired results, the benefits gained from using third-order moments, which also emphasize skewness in the probability density function, can be significant. The results from Paper C indicate that third-order moment maximization might be suitable for on-line blind equalization in ultra-wideband communication, which is currently the subject of much research.

References

[1] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, New Jersey: Prentice-Hall, 1985.

[2] J. R. Treichler, C. R. Johnson, Jr., and M. G. Larimore, *Theory and Design of Adaptive Filters*. Upper Saddle River, New Jersey: Prentice-Hall, 2001.

[3] B. Widrow and M. Hoff, Jr., "Adaptive switching circuits," in *IRE WESCON Conv. Record*, vol. 4, 1960, pp. 96–104.

[4] V. Solo and X. Kong, *Adaptive Signal Processing Algorithms, Stability and Performance*. Englewood Cliffs, New Jersey: Prentice-Hall, 1995.

[5] R. W. Lucky, "Techniques for adaptive equalization of digital communication systems," *Bell System Tech. J.*, vol. 45, pp. 255–286, Feb. 1966.

[6] A. Benveniste, M. Goursat, and G. Ruget, "Robust identification of a nonminimum phase system: Blind deconvolution of a linear equalizer in data communications," *IEEE Trans. Automat. Contr.*, vol. AC-25, no. 3, pp. 385–399, June 1980.

[7] D. L. Donoho, "On minimum entropy deconvolution," in *Applied Time Series Analysis*, D. F. Findley, Ed. New York: Academic Press, 1981.

[8] R. A. Wiggins, "Minimum entropy deconvolution," *Geoexploration*, no. 16, pp. 21–35, 1978.

[9] D. N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commun.*, vol. COM-28, no. 11, pp. 1867–1875, Nov. 1980.

[10] J. R. Treichler and B. G. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, no. 2, pp. 459–472, Apr. 1983.

[11] E. Walach and B. Widrow, "The least mean fourth (LMF) adaptive algorithm and its family," *IEEE Trans. Inform. Theory*, vol. IT-30, no. 2, pp. 275–283, Mar. 1984.

[12] C. L. Nikias and J. M. Mendel, "Signal processing with higher-order spectra," *IEEE Signal Processing Mag.*, pp. 10–37, July 1993.

Part II

Paper A

Skewness Maximization for Impulsive Sources in Blind Deconvolution

Authors:
Patrik Pääjärvi and James P. LeBlanc

Reformatted version of paper originally published in:
Proceedings of the 6$^{\text{th}}$ Nordic Signal Processing Symposium, 2004 (NORSIG 2004), Espoo, Finland, pp. 304-307

Skewness Maximization for Impulsive Sources in Blind Deconvolution

Patrik Pääjärvi and James P. LeBlanc

Abstract

In blind deconvolution problems, a deconvolution filter is often determined in an iterative manner, where the filter taps are adjusted to maximize some objective function of the filter output signal. The kurtosis of the filter output is a popular choice of objective function. In this paper, we investigate some advantages of using skewness, instead of kurtosis, in situations where the source signal is impulsive, i.e. has a sparse and asymmetric distribution. The comparison is based on the error surface characteristics of skewness and kurtosis.

1 Introduction and Problem Setting

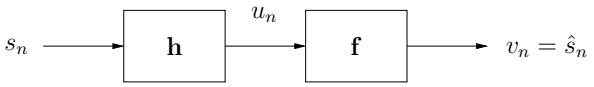

Figure 1: Block diagram of a deconvolution problem.

Fig. 1 shows a discrete-time deconvolution problem model. A source signal, s_n, whose characteristics are not completely known, is convolved with some unknown transfer function, $\mathbf{h}$. The output signal, u_n, is then applied to a deconvolution filter, $\mathbf{f}$, which, ideally, reconstructs s_n as $\hat{s}_n = v_n = u_n * \mathbf{f}$.

For geophysical applications in which the source signal has a sparse distribution (i.e. 'spiky' appearance), Wiggins [1] proposed a method called *minimum entropy deconvolution* (MED). The approach was to use the knowledge that the source signal had a sparse distribution, and try to find the deconvolution filter whose output distribution was as sparse as possible. As a measure of sparseness, Wiggins proposed the 'Varimax norm' (similar to the more commonly known *kurtosis*) as a measure of the 'spikiness' of the deconvolution filter output. The Varimax norm V for a filter output sequence v_n of M

samples is defined as

$$V = \frac{\sum_{n=0}^{M-1} v_n^4}{\left(\sum_{n=0}^{M-1} v_n^2\right)^2}. \tag{1}$$

The MED method consisted of choosing an initial filter vector $\mathbf{f}$ and then iteratively adjust the filter towards maximizing V.

Most deconvolution methods are based on some knowledge about the distribution of s_n. By using some suitable objective function $\mathcal{O}(v_n)$ of the deconvolution filter output, the filter can be adapted towards maximizing $\mathcal{O}(v_n)$.

Donoho [2] generalized the theory behind minimum entropy deconvolution by considering a family of objective functions of a sequence v_n of length M,

$$\mathcal{O}_s^r(v_n) = \frac{\frac{1}{M}\sum_{n=0}^{M-1} |v_n|^r}{\left(\frac{1}{M}\sum_{n=0}^{M-1} |v_n|^s\right)^{r/s}}, \tag{2}$$

of which the Varimax norm is a scaled version of $\mathcal{O}_2^4(v_n)$.

Donoho noted that, as a consequence of the central limit theorem, linear combinations of identically distributed random variables become 'more Gaussian' than the individual variables. Therefore, the transfer function output signal u_n will have a distribution that is more nearly Gaussian than the distribution of s_n. Any objective function should therefore be used to reduce 'the Gaussianity' of the deconvolution filter output.

One suitable measure of Gaussianity for an MED implementation would be the *kurtosis*, K_v, of v_n,

$$K_v = E\{v_n^4\}/(E\{v_n^2\})^2, \tag{3}$$

where $E\{\cdot\}$ denotes expectation. Wiggins Varimax norm is a scaled approximation of K_v. Thus, its objective would be to find the filter whose output has a kurtosis value far from a Gaussian signal (the kurtosis value of all Gaussian distributed signals is 3).

However, for *impulsive* sources, the kurtosis may not perform well [3]. An alternative choice of objective function might be the *skewness*, S_v, of v_n, defined as

$$S_v = E\{v_n^3\}/(E\{v_n^2\})^{3/2}. \tag{4}$$

Note that skewness maximization clearly would not be suitable for deconvolution of symmetrically distributed source signals, since the skewness of any filtered version of such a signal is zero.

Next, we compare kurtosis and skewness when used as objective functions for blind deconvolution of impulsive signals. This comparison considers the error surface topologies, i.e. K_v and S_v as functions of the filter coefficients. The error surface topology will

affect the convergence of MED algorithms. In particular, the number of stationary points (i.e. points where the gradient of the error surface is zero) is an important characteristic, as an excessive number of saddle points (stationary points having a non-definite Hessian) 'stall' gradient-based filter adaption.

1.1 Notation and Definitions

To enable a comparison between skewness maximization and kurtosis maximization for blind deconvolution, we introduce notation of such gradient-based methods.

The deconvolution filter $\mathbf{f}$ used is assumed to be an FIR filter of order N, represented by the column vector

$$\mathbf{f} = [f_0 \; f_1 \; \cdots \; f_N]^{\mathrm{T}}, \tag{5}$$

where f_m denotes the m^{th} filter coefficient. The filter output at time n is given by the convolution sum

$$v_n = \sum_{k=0}^{N} f_k u_{n-k}. \tag{6}$$

A simple strategy for maximizing any objective function, $\mathcal{O}(v_n)$, is to use a gradient method wherein the filter coefficients are adapted iteratively towards increasing $\mathcal{O}(v_n)$, regarding it ultimately as a function of $\mathbf{f}$, $\mathcal{O}(\mathbf{f})$. Denote the filter vector after i iterations as $\mathbf{f}^{(i)}$, the next filter vector will be chosen as

$$\mathbf{f}^{(i+1)} = \mathbf{f}^{(i)} + \mu \nabla_{\mathcal{O}(\mathbf{f}^{(i)})}, \tag{7}$$

where

$$\nabla_{\mathcal{O}(\mathbf{f})} = \left[\frac{\partial \mathcal{O}}{\partial f_0} \; \frac{\partial \mathcal{O}}{\partial f_1} \; \cdots \; \frac{\partial \mathcal{O}}{\partial f_N} \right]^{\mathrm{T}} \tag{8}$$

is the gradient vector of $\mathcal{O}(\mathbf{f})$, and μ is some fixed or variable stepsize.

The convergence of filter adaption algorithms based on gradient ascent, such as (7), depends mainly on two factors: the choice of stepsize, μ, and the topology of the error surface $\mathcal{O}(\mathbf{f})$.

The stepsize choice is an implementation issue. It must be chosen small enough to allow convergence to a (possibly local) maximum (the stability issue), while choosing a too small stepsize incurs excessive iteration steps. The error surface topology, however, depends on the algebraic structure of the objective function used. The error surfaces of common blind deconvolution objective functions are well known to be multi-modal (i.e. to have multiple local maxima). The number of stationary points for kurtosis has been explored [4], [5], but similar results for skewness has not been found.

2 Comparison of Error Surface Topologies

An important characteristic of an error surface $\mathcal{O}(\mathbf{f})$ is the number of stationary points, i.e. the number of points where the gradient, $\nabla_{\mathcal{O}(\mathbf{f})}$, is zero. More stationary points generally means slower convergence of the gradient algorithm.

By writing out (3) as a function of the filter coefficients, we obtain

$$K_v = \frac{E\left\{\left(\sum_{k=0}^{N} f_k u_{n-k}\right)^4\right\}}{\left(E\left\{\left(\sum_{k=0}^{N} f_k u_{n-k}\right)^2\right\}\right)^2}. \tag{9}$$

Taking the gradient of (9) with respect to the m^{th} filter coefficient, f_m, and equating to zero, we obtain the following; for $m, i = 0 \ldots N$,

$$\sum_i f_i^3 \mathcal{R}_{\,m-i}^{\,0\;\;0} + 3\sum_{i\neq j} f_i^2 f_j \mathcal{R}_{\,m-i}^{\,0\;\;j-i} + \sum_{i\neq j\neq k} f_i f_j f_k \mathcal{R}_{\,m-i}^{\,j-i\;\;k-i} - \sigma_v^2 K_v \sum_i f_i \mathcal{R}_{\,m-i} = 0, \tag{10}$$

where

$$\sigma_v^2 = E\{v_n^2\} = E\left\{\left(\sum_{k=0}^{N} f_k u_{n-k}\right)^2\right\}, \tag{11}$$

and the 2^{nd} and 4^{th} moments of u_n are defined as

$$\mathcal{R}_{\,i} = E\{u_n u_{n-i}\} \qquad \mathcal{R}_{\,j\;\;k}^{\,i} = E\{u_n u_{n-i} u_{n-j} u_{n-k}\}.$$

The corresponding equation for the skewness is found similarly by writing out (4) as a function of $\mathbf{f}$, taking the gradient with respect to the m^{th} filter coefficient, f_m, and equating to zero. We obtain the following; for $m, i = 0 \ldots N$,

$$\sum_i f_i^2 \mathcal{R}_{\,m-i}^{\,0} + \sum_{i\neq j} f_i f_j \mathcal{R}_{\,m-i}^{\,j-i} - \sqrt{\sigma_v^2}\, S_v \sum_i f_i \mathcal{R}_{\,m-i} = 0, \tag{12}$$

where the 3^{rd} moment of u_n is defined as

$$\mathcal{R}_{\,j}^{\,i} = E\{u_n u_{n-i} u_{n-j}\}.$$

We note that the kurtosis-based stationary points (10) consist of a system of $N+1$ polynomial equations in $N+1$ variables $(f_0, \ldots, f_N)$. Each equation in the system has the same monomial support and a total degree of 5. The Bezout upper bound on the number of solutions (i.e. stationary points of the error surface) is then 5^{N+1}.

Similarly, the skewness system of equations consist of $N+1$ polynomials of total degree 4, yielding a Bezout upper bound on the number of stationary points of 4^{N+1}. Even for moderate filter lengths $(N+1)$, the number of possible stationary points is considerably smaller for the skewness error surface. This generally means faster convergence for gradient algorithms of the form (7).

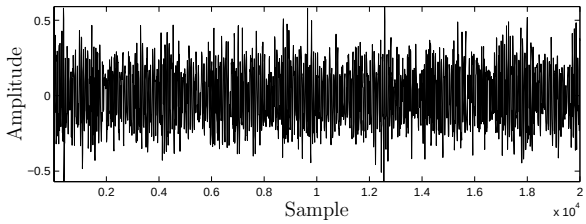

Figure 2: Measurement data from a sound recording of a diesel engine.

3 Experimental Results

3.1 MED Algorithm Comparison

To support the view in Section 2, a simulation was done in which two block-mode versions of the same MED algorithms, one using kurtosis and the other using skewness as the objective function, were applied to real measurement data.

The data, shown in Fig. 2, consisted of a sound recording of a running diesel engine. Referring to Fig. 1, the source signal s_n is the explosions from the pistons. The transfer function $\mathbf{h}$ is the engine block and housing through which the source signal propagates. The source signal is thought to be impulsive, i.e. it has a sparse and asymmetric distribution, although the measured signal appears symmetric and Gaussian, as seen in Fig. 2, after passing through the transfer function. The measurement data consists of u_n plus added noise. The deconvolution filter length was chosen to be 2000.

The MED algorithm used in the experiment was based on the filter iteration (7). Although the stepsize, μ, can be varied during iteration in several ways, a fixed stepsize was used for simplicity. Each algorithm was run 35 times, using different unit-norm initializations. Each filter was initialized with one large center tap, and the rest of the taps picked randomly from a normal distribution, with a standard deviation of 2% of the center tap magnitude. This is a reasonable approximation to the 'customary center tap initialization' of blind deconvolution folklore.

4000 iterations were performed to allow both algorithms to converge. The stepsizes for the two algorithms cannot be directly compared. In order to make a fair comparison, the stepsize for skewness was chosen just small enough to keep almost all runs stable, while the kurtosis stepsize was chosen so that about half of the runs became unstable going into convergence. In this way, the convergence rate of the kurtosis algorithm was essentially maximized for fixed stepsize.

As a comparison between the two algorithms, the kurtosis and skewness versus iteration number was recorded for each run and plotted in Fig. 3. In Fig. 4, the averages of all 35 runs are shown for both kurtosis and skewness. The two plots in Fig. 4 are normalized to the same final value, since the magnitudes of the two objective functions cannot be directly compared.

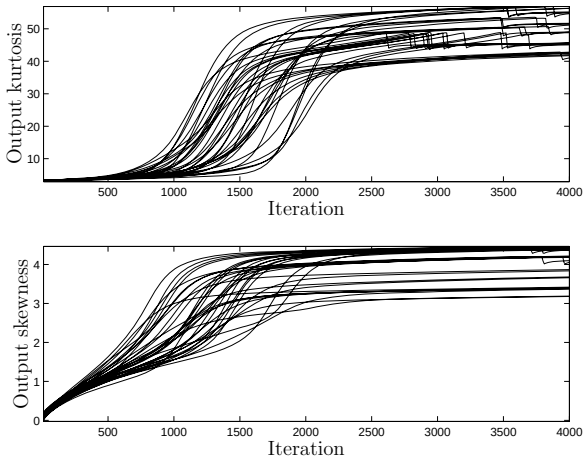

Figure 3: Kurtosis (top) and skewness (bottom) versus filter iteration.

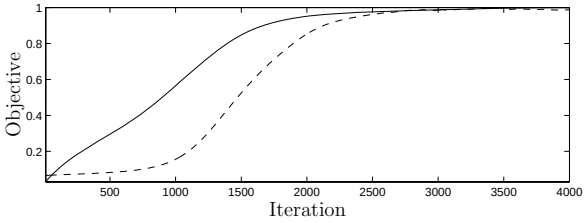

Figure 4: Kurtosis (dashed) and skewness (solid) versus filter iteration.

As seen from Fig. 4, the MED algorithm using skewness is initially steeper and reaches 50% of its final value considerably faster than the kurtosis algorithm.

The results shown in Fig. 4 provide support for the results in Section 2, namely that the error surface of skewness contains fewer stationary points, meaning less 'flat' regions at which the MED algorithm might get stalled. Fig. 5 shows the deconvolution filter outputs for one run of the kurtosis and skewness MED algorithms. Both algorithms have deconvolved the source signal and produced a sparsely distributed signal.

3.2 Error Surface Topology Comparison for a 3-Tap Filter

As an illustrative comparison, the error surfaces for skewness and kurtosis for a low-dimensional (3-tap) filter were compared visually. An impulsive signal was synthesized and filtered through a ARMA(1,1) low-pass filter. The error surfaces for skewness and kurtosis were then plotted over a set of unit-norm, three-tap deconvolution filters (i.e. the unit sphere). Figures 6 and 7 show contour plots of the error surfaces for kurtosis

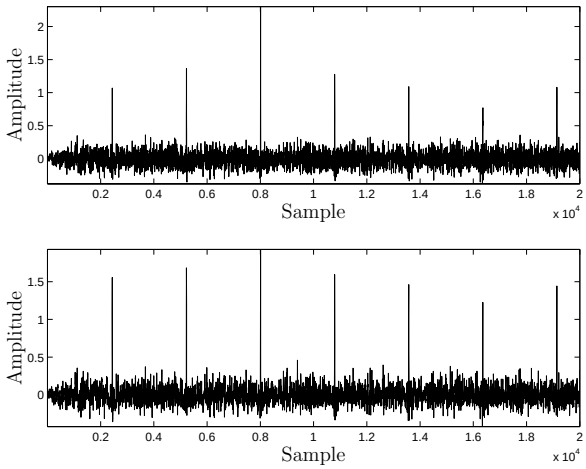

Figure 5: Kurtosis (top) and skewness (bottom) deconvolution filter outputs.

and skewness respectively. Small arrows indicate the direction of the gradient, and the stationary points are marked and classified as minima ($\times$), saddle points (s) or maxima ($\bullet$). The figures show that the error surface of kurtosis has more stationary points than the skewness error surface. As a check, it was verified that the vector fields satisfied the Euler Characteristic of the sphere [6].

4 Conclusions

The use of skewness instead of kurtosis as the objective function for minimum entropy deconvolution of impulsive sources has the benefit of an error surface with fewer saddle points, allowing better convergence behavior for simple, gradient-based methods. This has been demonstrated using both analytical and experimental results.

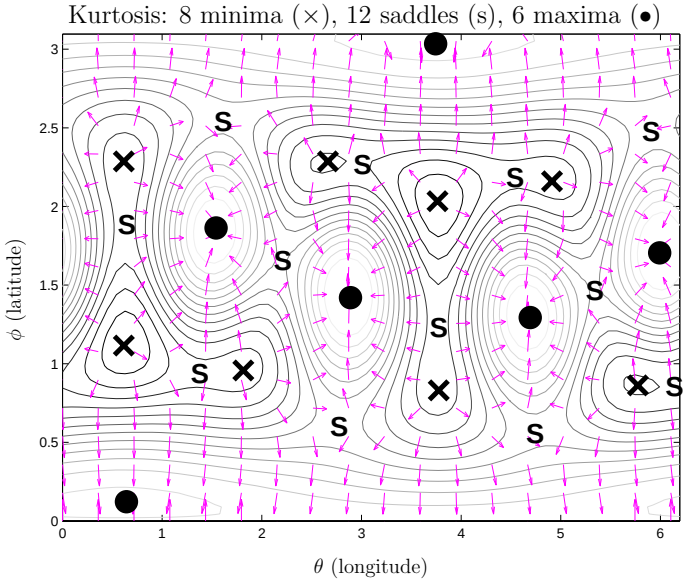

Figure 6: Kurtosis error surface.

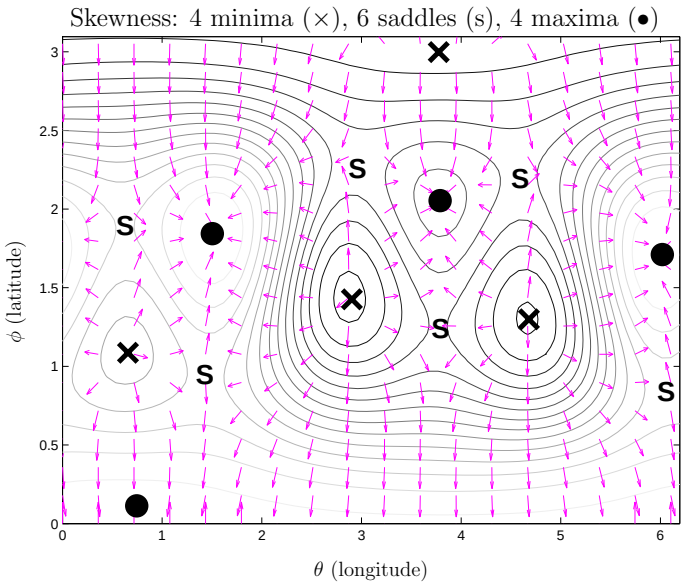

Figure 7: Skewness error surface.

References

[1] R. A. Wiggins, "Minimum entropy deconvolution," *Geoexploration*, no. 16, pp. 21–35, 1978.

[2] D. L. Donoho, "On minimum entropy deconvolution," in *Applied Time Series Analysis*, D. F. Findley, Ed. New York: Academic Press, 1981.

[3] H. Mathis and S. C. Douglas, "Bussgang blind deconvolution for impulsive signals," *Proc. IEEE*, vol. 51, no. 7, pp. 1905–1915, July 2003.

[4] Z. Ding and T. Nguyen, "Stationary points of a kurtosis maximization algorithm for blind signal separation and antenna beamforming," *IEEE Trans. Signal Processing*, vol. 48, no. 6, pp. 1587–1596, June 2000.

[5] J. P. LeBlanc, I. Fijalkow, and C. R. Johnson, Jr., "CMA fractionally spaced equalizers: Stationary points and stability under IID and temporally correlated sources," *International Journal of Adaptive Control and Signal Processing*, no. 12, 1998.

[6] V. Guillemin and A. Pollack, *Differential Topology.* Englewood Cliffs, New Jersey: Prentice-Hall, 1974.

Paper B

On-Line Adaptive Blind Deconvolution Based on Third-Order Moments

Authors:
Patrik Pääjärvi and James P. LeBlanc

Reformatted version of paper to appear in:
IEEE Signal Processing Letters, vol. 12, no. 12, Dec. 2005

On-Line Adaptive Blind Deconvolution Based on Third-Order Moments

Patrik Pääjärvi and James P. LeBlanc

Abstract

Traditional methods for on-line adaptive blind deconvolution using higher-order statistics are often based on even-order moments, due to the fact that the systems considered commonly feature symmetric source signals (i.e. signals having a symmetric probability density function). However, asymmetric source signals facilitate blind deconvolution based on odd-order moments. In this letter, we show that third-order moments give the benefits of faster convergence of algorithms and increased robustness to additive Gaussian noise. The convergence rates for two algorithms based on third- and fourth-order moments respectively are compared for a simulated Ultra-Wideband (UWB) communication channel.

1 Introduction

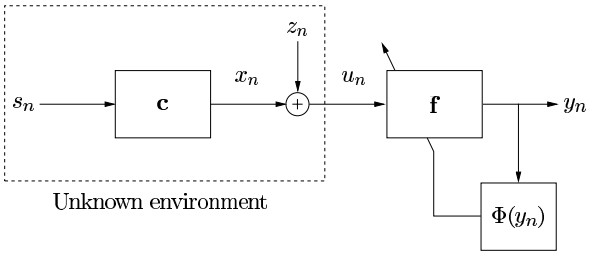

Figure 1: The signal model of a general blind deconvolution problem.

Adaptive blind deconvolution is used for equalization or identification of unknown systems when only the output of the system can be observed. Fig. 1 shows a discrete-time signal model of a general blind deconvolution problem (the subscript n denotes a time index). The object is to find the deconvolution filter $\mathbf{f}$ that approximately inverts the system $\mathbf{c}$ with limited or no knowledge of either $\mathbf{c}$ nor the source signal s_n. The system output x_n plus an additive disturbance z_n gives the observed signal u_n. The unknown, possibly time-varying system $\mathbf{c}$ may be either linear or non-linear with minimum-,

maximum- or mixed phase.

1.1 Minimum Entropy Deconvolution

In general, due to filtration through $\mathbf{c}$, the probability distribution of the system output x_n will be closer to a Gaussian than that of s_n. This is a consequence of the central limit theorem and allows for deconvolution based on measuring the 'Gaussianity' of the deconvolution filter output y_n. *Minimum Entropy Deconvolution* (MED) methods are based on using a *score function* $\Phi(y_n)$ to measure the Gaussianity (or entropy) of y_n. The probability distribution of y_n is then driven as far away from a Gaussian distribution as possible, by adapting the coefficients of $\mathbf{f}$. In general, $\mathbf{f}$ is a FIR (finite impulse response) filter and all signals are assumed to have zero mean.

To measure the Gaussianity of a signal, score functions based on higher-order central moments are commonly used. Such functions can typically be approximated by simple polynomial functions of y_n, making them specially suitable for on-line (real-time) applications, where computational efficiency is often of concern. Wiggins [1] proposed the use of the *Kurtosis* (normalized fourth-order moment) of y_n as a score function for MED. Donoho [2] generalized the theory behind MED, and considered various types of score functions, including central moments of order greater than two.

Godard [3] suggested *dispersions* of y_n as score functions for blind equalization of communication channels. The dispersion of order p (p integer > 0) is based on even-order moments of y_n, and is defined as

$$D^{(p)} = E\left\{\left(|y_n|^p - R_p\right)^2\right\}, \tag{1}$$

where R_p is a positive constant and $E\{\cdot\}$ denotes expectation. Choosing $p = 2$ leads to the popular Constant Modulus Algorithm (CMA) [4], which is based on fourth-order moments, similar to Wiggins original idea.

1.2 Symmetric and Asymmetric Source Signals

Traditional uses of blind deconvolution include linear equalization of communication channels, deconvolution of seismic traces and dereverberation of acoustic signals. Such applications are often assumed to feature *symmetric* source signals, i.e. zero-mean signals with a probability density function (PDF) that is symmetric around zero. Since all odd-order moments of symmetric signals are zero, most research focus in the field of blind deconvolution has hence been directed towards even-order moments. Although symmetric source signals dominate the field of applications for MED, *asymmetric* source signals, i.e. zero-mean signals with asymmetric PDF's (and thus with non-zero third central moment) occur in a wide range of acoustic, biomedical and mechanical signals (for example, pulse oximetry signals or hammer impacts). Asymmetry is also a feature of Impulse Radio signals [5], a proposed signaling format for Ultra-Wideband (UWB) radio [6], [5], [7].

In previous work [8], we noted that asymmetry in the source signal can be exploited by using a score function based on *third-order moments*, instead of the common fourth-order moments. The benefit of a lower-order moment is mainly a simpler score function surface (regarding $\Phi(y_n)$ as a function of the deconvolution filter coefficients). This will, in general, give faster convergence of common gradient search algorithms.

In the work presented in this paper, we compare two simple on-line score functions based on third- and fourth-order moments respectively. Since symmetric source signals have zero odd-order moments, we restrict our focus to asymmetric sources. We demonstrate that an on-line gradient search algorithm based on third-order moments should in general benefit from faster convergence and increased robustness to additive Gaussian noise, compared to algorithms based on fourth-order moments. The experimental results are obtained from simulations of an indoor Ultra-Wideband channel with Impulse Radio signaling.

2 Notation and Model Description

Referring to the discrete-time signal model in Fig. 1, we define s_n and $\mathbf{c}$ as the unknown source signal and unknown channel respectively. The sum of the channel output x_n and the disturbance z_n is the observed signal u_n, which is the input to the deconvolution filter $\mathbf{f}$. For simplicity, we will from here on refer to $\Phi(y_n)$ as an *objective* function of y_n, and the objective of the deconvolution problem is to find the filter $\mathbf{f}$ that maximizes $\Phi(y_n)$. In typical on-line situations, this is done iteratively through a gradient search algorithm.

The adaptive filter $\mathbf{f}$ is assumed to be FIR (finite impulse response) of order N. The filter after r iterations is represented by the coefficient vector

$$\mathbf{f}^{(r)} = \begin{bmatrix} f_0^{(r)} & f_1^{(r)} & \dots & f_N^{(r)} \end{bmatrix}^{\mathrm{T}}. \tag{2}$$

Using adaption by gradient ascent, $\mathbf{f}$ is recursively updated in the direction of maximizing the objective function. The filter update rule becomes

$$\mathbf{f}^{(r+1)} = \mathbf{f}^{(r)} + \mu \nabla_{\Phi(\mathbf{f}^{(r)})}, \tag{3}$$

where μ is a positive stepsize of adaption and $\nabla_{\Phi(\mathbf{f}^{(r)})}$ is the gradient of Φ with respect to $\mathbf{f}^{(r)}$,

$$\nabla_{\Phi(\mathbf{f}^{(r)})} = \begin{bmatrix} \dfrac{\partial \Phi}{\partial f_0^{(r)}} & \dfrac{\partial \Phi}{\partial f_1^{(r)}} & \cdots & \dfrac{\partial \Phi}{\partial f_N^{(r)}} \end{bmatrix}^{\mathrm{T}}. \tag{4}$$

Filter iteration can be performed either on a sample-by-sample basis (general applications), or on a symbol-by-symbol basis (digital communication applications). If the stepsize μ in (3) is small, $\mathbf{f}$ can be regarded as approximately constant in time, allowing us to drop the superscript $^{(r)}$. We then define the filter output at sampling instant n as

$$y_n = \mathbf{u}_n^{\mathrm{T}}\mathbf{f} = \mathbf{x}_n^{\mathrm{T}}\mathbf{f} + \mathbf{z}_n^{\mathrm{T}}\mathbf{f} = d_n + v_n, \tag{5}$$

with d_n being the 'filtered signal', v_n the 'filtered noise' and the signal vectors defined as $\mathbf{u}_n = [u_n\ u_{n-1}\ \ldots\ u_{n-N}]^{\mathrm{T}}$, $\mathbf{x}_n = [x_n\ x_{n-1}\ \ldots\ x_{n-N}]^{\mathrm{T}}$ and $\mathbf{z}_n = [z_n\ z_{n-1}\ \ldots\ z_{n-N}]^{\mathrm{T}}$. The two objective functions we will compare are simply the third- and fourth-order central moments respectively of the filter output y_n;

$$\Phi^{(3)}(y_n) \triangleq E\left\{y_n^3\right\}, \tag{6}$$

$$\Phi^{(4)}(y_n) \triangleq E\left\{y_n^4\right\}. \tag{7}$$

The corresponding gradients with respect to $\mathbf{f}$ are

$$\nabla_{\Phi^{(3)}} \propto E\left\{y_n^2\,\mathbf{u}_n\right\}, \tag{8}$$

$$\nabla_{\Phi^{(4)}} \propto E\left\{y_n^3\,\mathbf{u}_n\right\}. \tag{9}$$

In on-line applications, where computational power is often limited, it is customary to use an instantaneous estimate of the gradient in the filter update (3). This can be obtained from the two objective functions (8) and (9) by simply dropping the expectation operators.

From here on, we will make the following assumptions:

A1) All signals are real and zero-mean.

A2) s_n is a non-Gaussian and asymmetric signal.

A3) The disturbance z_n is a zero-mean, i.i.d. Gaussian noise process, independent of x_n, with variance σ_z^2.

A4) The stepsize parameter μ in (3) is small, so that the filter vector $\mathbf{f}^{(r)}$ can be regarded as approximately constant in time when compared to the signals, i.e. $\mathbf{f}^{(r)} = \mathbf{f}$.

A5) $\mathbf{f}$ is kept at constant (unit) norm during adaption, i.e. $\| \mathbf{f} \|^2 = \sum_i f_i^2 = 1$.

Assumption A4 is customary in adaptive filtering theory and simplifies the averaging analysis in the next section. Assumption A5 is necessary since increasing the norm of any filter $\mathbf{f}$ increases both objectives (6) and (7), while leaving the Gaussianity of the filter output unchanged.

3 Comparative Performance Analysis of 3^{RD}- and 4^{TH}-Order Objective Functions

3.1 Objective Function Surface Topology

If the objective function Φ is regarded as a function of the deconvolution filter coefficients, adaption according to (3) can be thought of as traversing a multidimensional function

surface $\Phi(\mathbf{f})$ towards any local maximum points (under the constraint of unit filter norm). The set of maximum points is a subset of the points on the function surface with zero gradient $\nabla_{\Phi(\mathbf{f})}$, with the other set members being minimum points or saddle points. Maximizing $\Phi(\mathbf{f})$ is therefore equivalent to finding a subset of solutions to

$$\nabla_{\Phi(\mathbf{f})} = \mathbf{0}. \tag{10}$$

For a filter of order N, (10) leads to a system of $N+1$ non-linear polynomial equations in the $N+1$ unknowns $\{f_0, \ldots, f_N\}$. The highest degree of the polynomials in the equation system (10) will set an upper bound on the number of solutions, i.e. the number of stationary points on the objective surface. A large number of stationary points generally implies a large number of saddle points, which can 'stall' filter adaption.

Solving (10) for the third-moment objective function (6) leads to the system of equations

$$\sum_i f_i^2 \, \mathcal{R}_{m-i}^{\;\;0} + \sum_{i \neq j} f_i f_j \, \mathcal{R}_{m-i}^{\;\;j-i} = 0, \tag{11}$$

for $m, i, j = 0 \ldots N$, with the third moment of u_n defined as $\mathcal{R}_j^i = E\{u_n u_{n-i} u_{n-j}\}$. The highest polynomial degree of (11) is 2, which gives a Bezout [9] upper bound on the number of solutions, i.e. the number of stationary points on the function surface, equal to 2^{N+1}. The corresponding system of equations for the fourth-moment objective function (7) is

$$\sum_i f_i^3 \, \mathcal{R}_{\;\;m-i}^{\;0\;\;0} + 3 \sum_{i \neq j} f_i^2 f_j \mathcal{R}_{\;\;m-i}^{\;j-i\;\;0} + \sum_{i \neq j \neq k} f_i f_j f_k \mathcal{R}_{\;\;m-i}^{\;k-i\;\;j-i} = 0, \tag{12}$$

for $m, i, j, k = 0 \ldots N$, with the fourth moment of u_n defined as $\mathcal{R}_{\;k}^{i\;\;j} = E\{u_n u_{n-i} u_{n-j} u_{n-k}\}$. The highest polynomial degree of (12) is 3, giving a Bezout upper bound on the number of solutions equal to 3^{N+1}. Note that, in general, the moments of u_n in (11) and (12) may depend on n (e.g. for a time-varying system $\mathbf{c}$), despite the notation used. This is not essential here, since a dependence on n only implies that the *shape* of the function surface changes over time. The upper bounds on the number of stationary points are still constant.

Even for moderate filter orders, the maximum number of stationary points on the third-moment function surface is considerably smaller than on the corresponding fourth-moment surface. As previously noted for off-line (block-mode) algorithms in [8], lower polynomial order of score functions gives the benefit of a 'simpler' objective surface, which, in general, implies fewer saddle points. Since an excessive number of saddle points can 'stall' a gradient search, a simpler objective surface will therefore in general allow for faster adaption of such algorithms. This is of special importance in applications where the unknown system $\mathbf{c}$ is time-varying, and the deconvolution filter needs to 'track' changes in the system.

3.2 Gaussian Noise Effects on the Objective Function Surface

3.2.1 Objective Surface Analysis

In the presence of additive white Gaussian noise, as described in the model in Section 2, and with the filter output decomposed into the sum of d_n and v_n, as in (5), the third-moment objective function (6) at time n becomes

$$\Phi^{(3)}(y_n) = E\left\{d_n^3\right\}. \tag{13}$$

Since all odd moments of the Gaussian disturbance are zero, (13) depends solely on the filtered signal d_n, and not on the disturbance z_n. Thus, the function surface of the third-moment objective function is preserved in the presence of Gaussian noise. The corresponding expression for the fourth-moment objective function (7) in the presence of Gaussian noise is

$$\Phi^{(4)}(y_n) = E\left\{d_n^4\right\} + 3\left(\sigma_z^2\right)^2 \parallel \mathbf{f} \parallel^4 + 6\,\sigma_z^2 E\left\{d_n^2\right\} \parallel \mathbf{f} \parallel^2. \tag{14}$$

The Gaussian noise introduces two additional terms to the 'signal' (first) term. Under Assumption A5, the second term does not depend on $\mathbf{f}$, and will therefore not change the location of the stationary points. The third term, on the other hand, which depends on $\mathbf{f}$ through d_n, will alter the location of the stationary points. Since the local maximum points have moved under the influence of noise, the ability of the algorithm to invert $\mathbf{c}$ has been reduced.

3.2.2 Gradient Analysis

With the filter output defined as in (5), the gradient of the objective function can be expressed as

$$\nabla_\Phi = \nabla_{\Phi(d)} + \nabla_{\Phi(d,v)}. \tag{15}$$

$\nabla_{\Phi(d)}$ is the 'signal' component of the gradient due to the filtered source signal d_n. $\nabla_{\Phi(d,v)}$ is the perturbation of the gradient caused by the Gaussian noise. Taking the gradients of $\Phi^{(3)}$ and $\Phi^{(4)}$ with respect to $\mathbf{f}$ and separating them according to (15) yields

$$\nabla_{\Phi^{(3)}} \propto \nabla_{\Phi^{(3)}(d)}, \tag{16}$$

$$\nabla_{\Phi^{(4)}} \propto \nabla_{\Phi^{(4)}(d)} + 3\left[\left(\sigma_z^2\right)^2 \parallel \mathbf{f} \parallel^2 + \sigma_z^2 E\left\{d_n^2\right\}\right]\mathbf{f} + 3\,\sigma_z^2 \parallel \mathbf{f} \parallel^2 E\left\{d_n \mathbf{x}_n\right\}. \tag{17}$$

At 'true' local maximum points, the signal gradients $\nabla_{\Phi^{(3)}(d)}$ and $\nabla_{\Phi^{(4)}(d)}$ are zero. As indicated by (13), the function surface of $\Phi^{(3)}$ is not affected by noise. Therefore, an instantaneous estimate of $\nabla_{\Phi^{(3)}}$ (obtained by dropping the expectation operator in (8)) will be unbiased in the presence of Gaussian noise. For $\Phi^{(4)}$, the noise causes a perturbation of the gradient in the direction of $E\left\{d_n \mathbf{x}_n\right\}$, causing a corresponding instantaneous estimate of $\nabla_{\Phi^{(4)}}$ to become biased. This adversely affects the algorithms ability to invert

the unknown system, as indicated by (14). Although the perturbation in the direction of $\mathbf{f}$ does not introduce a bias under unit-norm constraints, a large noise variance may have a negative effect on the convergence rate of the algorithm on finite-precision machines.

4 Experimental Results From a Simulated Ultra-Wideband Radio Channel

Wireless communication over Ultra-Wideband (UWB) radio channels has attracted interest in recent years. One of the proposed signaling formats for UWB communication is Impulse Radio (IR) [5], which consists of pulse-position modulated pulses of extremely short duration, typically on the order of a nanosecond, transmitted without the use of a sinusoidal carrier. The short pulses used give IR signals a bandwidth from near DC to several gigahertz, giving them good material-penetrating abilities and resolvable multipath delays down to about 30 cm. To allow for multiple user access, an additional pseudo-random time-hopping modulation scheme is used. This reduces the risk of catastrophic collisions with other IR transmitters, and also avoids interference with coexisting narrow-band signals by 'spreading' the spectrum of the signal [5].

Although the large bandwidth of IR signals makes them robust to fading, the large multipath spread of a typical indoor UWB channel is likely to cause intersymbol interference (ISI) at higher data rates [7], [10]. The asymmetry of typical IR signal pulses motivates the use of a blind adaptive linear equalizer based on third-order moment maximization to mitigate ISI.

A numerical experiment was conducted in which the two objective functions (6) and (7) were used to implement two fractionally spaced, adaptive linear equalizers for an UWB channel. IR signals with independent, identically distributed symbols were simulated based on the model described in [5], using a pulse shape

$$\omega(t) = \left[1 - 4\pi(t/\tau_m)^2\right] \exp\left[-2\pi(t/\tau_m)^2\right], \tag{18}$$

with $\tau_m = 0.2333$, giving a pulse duration of about one nanosecond. The pulse shape is shown in Fig. 2. The sampling interval was chosen to give each pulse a support of 15 samples, based on results from [11]. The IR signals used binary orthogonal modulation at a bit rate of 10 Mbits/second. An UWB channel impulse response with a rich multipath spread up to approximately 200 nanoseconds was synthesized with the aid of a recipe from [12]. Although only a single transmitter was simulated, the interference from a large number of adjacent transmitters can in many situations be modeled as a Gaussian random process [5].

The receiver structure consisted of a filter matched to (18) followed by the linear equalizers. The two FIR equalizers of order $N = 400$ were implemented with adaption using third-order moment and fourth-order moment maximization respectively. The equalizers were recursively updated at the symbol instants, using instantaneous estimates

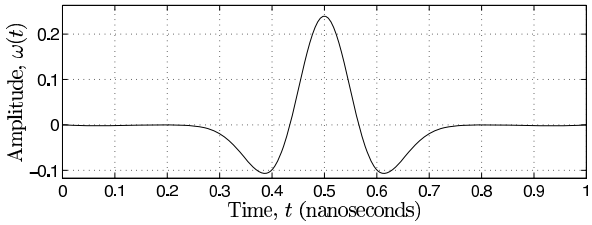

Figure 2: The signal pulse shape $\omega(t)$ used in the experiment. $\omega(t) = \left[1 - 4\pi(t/\tau_m)^2\right] \exp\left[-2\pi(t/\tau_m)^2\right]$, with $\tau_m = 0.2333$.

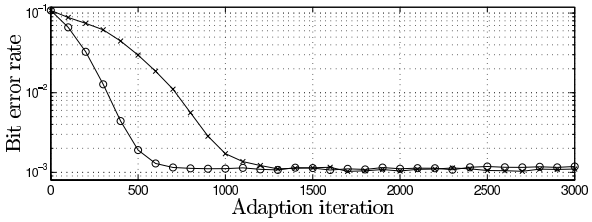

Figure 3: Bit error rate versus adaption iteration for third-moment (∘) and fourth-moment (×) objective functions, averaged over 13 runs. SNR per bit = 11dB.

of (8) and (9) respectively, starting from the customary 'center-tap' initialization. The individual stepsizes of adaption, $\mu^{(3)} = 9.5 \cdot 10^{-3}$ for the third-moment algorithm and $\mu^{(4)} = 6 \cdot 10^{-3}$ for the fourth-moment algorithm, were chosen so that both algorithms gave equal bit error rate performance at convergence. Fig. 3 shows the bit error rate versus adaption iteration for the third- and fourth-order moment based objective functions. The curves show the average results from 13 runs for a signal-to-noise ratio per bit of 11dB.

As seen in Fig. 3, the algorithm that uses the third-order moment objective function converges approximately twice as fast as the corresponding fourth-order moment version. This confirms the results in Section 3.1, namely that the lower order polynomial structure of the third-order moment results in a 'simpler' function surface. In general, this should imply faster convergence of filter adaption, which is important for the algorithm's ability to track a time-varying channel. Since typical indoor UWB channels are indeed time-varying, third-order moment based blind deconvolution, with its ability to exploit the source asymmetry, seems to be a suitable option for UWB channel equalization.

5 Conclusion

We have compared the performance of two objective functions for adaptive blind deconvolution based on third-order moments and fourth-order moments respectively. Asymmetric source signals offer opportunities to use objective functions based on third-order

moments, as an alternative to the commonly used fourth-order moments. Both the analytical and the experimental results indicate that a lower order objective function results in fewer stationary points on the objective function surface, which in general allows for faster convergence of on-line blind adaptive algorithms. The analysis of gradient estimation in the presence of Gaussian noise further highlights the advantages of using third-order moments. The faster convergence and increased robustness to additive Gaussian noise makes third-order-moment based methods interesting candidates for blind adaptive equalization in Ultra-Wideband communication.

References

[1] R. A. Wiggins, "Minimum entropy deconvolution," *Geoexploration*, no. 16, pp. 21–35, 1978.

[2] D. L. Donoho, "On minimum entropy deconvolution," in *Applied Time Series Analysis*, D. F. Findley, Ed. New York: Academic Press, 1981.

[3] D. N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commun.*, vol. COM-28, no. 11, pp. 1867–1875, Nov. 1980.

[4] J. R. Treichler and B. G. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, no. 2, pp. 459–472, Apr. 1983.

[5] M. Z. Win and R. A. Scholtz, "Ultra-wide bandwidth time-hopping spread-spectrum impulse radio for wireless multiple access communications," *IEEE Trans. Commun.*, vol. 48, no. 4, pp. 679–691, Apr. 2000.

[6] R. A. Scholtz, R. Weaver, E. Homier, J. Lee, P. Hilmes, A. Taha, and R. Wilson, "UWB radio deployment challenges," in *Proc. IEEE Int. Symp. Personal, Indoor, Mobile Radio Communications*, vol. 1, London, U.K., Sept. 2000, pp. 620–625.

[7] D. Porcino and W. Hirt, "Ultra-wideband radio technology: Potential and challenges ahead," *IEEE Commun. Mag.*, vol. 41, no. 7, pp. 66–74, July 2003.

[8] P. Pääjärvi and J. P. LeBlanc, "Skewness maximization for impulsive sources in blind deconvolution," in *Proc. IEEE Nordic Signal Proc. Symp.*, Espoo, Finland, June 2004, pp. 304–307.

[9] E. W. Weisstein, *CRC Concise Encyclopedia of Mathematics*. Chapman & Hall/CRC, 1999.

[10] A. G. Klein and C. R. Johnson, Jr., "MMSE decision feedback equalization of pulse position modulated signals," in *Communications, 2004 IEEE International Conference on*, vol. 5, Paris, France, June 2004, pp. 2648–2652.

[11] V. Lottici, A. D'Andrea, and U. Mengali, "Channel estimation for ultra-wideband communications," *IEEE J. Select. Areas Commun.*, vol. 20, no. 9, pp. 1638–1645, Dec. 2002.

[12] D. Cassioli, M. Z. Win, and A. F. Molisch, "The ultra-wide bandwidth indoor channel: From statistical model to simulations," *IEEE J. Select. Areas Commun.*, vol. 20, no. 6, pp. 1247–1257, Aug. 2002.

Paper C

Computationally Efficient Norm-Constrained Adaptive Blind Deconvolution Using Third-Order Moments

Authors:
Patrik Pääjärvi and James P. LeBlanc

Reformatted version of paper submitted to:
2006 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2006), Toulouse, France

Computationally Efficient Norm-Constrained Adaptive Blind Deconvolution Using Third-Order Moments

Patrik Pääjärvi and James P. LeBlanc

Abstract

Third-order central moments have been shown to be well suited as objective functions for blind deconvolution of impulsive signals. On-line implementations of such algorithms may suffer from increasing filter norm, forcing adaptation under constrained filter norm. This paper extends a previously known efficient algorithm with self-stabilizing properties to the case of using a third-order moment objective function. New results herein use averaging analysis to determine adaptation stepsize conditions for asymptotic stability of filter norm.

1 Introduction

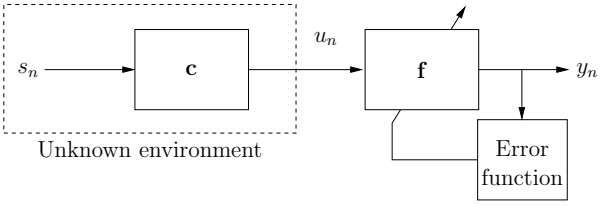

Figure 1: Model of a general adaptive blind deconvolution.

Blind deconvolution is used for identification or equalization of unknown systems in situations where only the system output can be observed. A general discrete-time model is shown in Figure 1, where n denotes a time index, s_n the unknown source, and $\mathbf{c}$ the unknown system. The object is to find the deconvolution filter $\mathbf{f}$ that approximately inverts the unknown system, so that y_n becomes an estimate of s_n.

If the deconvolution filter $\mathbf{f}$ is continuously adjusted according to some *error function*, we get an *adaptive* blind deconvolution setting. The error function (corresponding to the error signal of the standard LMS algorithm) is related to the gradient of an *objective*

function of the filter output y_n. Adaptation of $\mathbf{f}$ is aimed at maximizing the objective function.

Minimum Entropy Deconvolution (MED) methods [1], [2] are based on the idea that; given an uncorrelated sequence s_n, the probability distribution of u_n is closer to a Gaussian distribution compared to that of s_n. This consequence of the central limit theorem allows for blind deconvolution by discriminating the distribution of y_n from a Gaussian distribution. An objective function for adaptive MED should therefore be a measure of 'how Gaussian' y_n is. *Higher-order central moments* (order greater than two) are popular measures of Gaussianity, especially the *kurtosis* (normalized fourth-order moment).

Apart from their ability to measure Gaussianity, higher-order moments can also be used to describe how heavy-tailed the probability density function (PDF) of a signal is. A signal with a heavy-tailed PDF has a 'spiky' appearance. This type of distribution characterization allows for blind deconvolution without the assumption of the source signal being a white sequence.

If s_n is known to have a non-zero third-order central moment, this *asymmetry* allows for exploitation of *skewness* as an objective function, as an alternative to kurtosis. The skewness of a stochastic variable x is the normalized third-order central moment

$$\mathcal{S}(x) = \frac{\mathrm{E}\{x^3\}}{\left(\mathrm{E}\{x^2\}\right)^{3/2}}, \tag{1}$$

where $\mathrm{E}\{\cdot\}$ denotes expectation. Since all odd-order moments of a signal with symmetric PDF are zero, the use of odd-order moments such as (1) is restricted to asymmetric signals.

In previous work, skewness has been used for blind deconvolution of impulsive signals (i.e. asymmetric signals dominated by positive 'spikes'). When compared to kurtosis, skewness generally gives faster convergence of algorithms, and is less sensitive to additive white Gaussian noise [3], [4]. This motivates why exploitation of signal asymmetry using skewness may be preferable to kurtosis-based methods.

Due to the relative complexity of its gradient equation, (1) may not be suitable as an objective function for real-time applications requiring minimal computational cost. A more computationally efficient function is

$$\mathcal{O}(x) = \frac{1}{3}\mathrm{E}\left\{x^3\right\}, \tag{2}$$

a scaled version of the third-order moment of x. While easier to estimate than skewness, the third-order moment is not scale invariant in x. That is, $\mathcal{O}(x) \neq \mathcal{O}(kx)$ for $k \neq 1$. As a consequence, any gradient ascent algorithm based on (2) will lead to a rapid increase in deconvolution filter norm over iterations. In fact, problems of increasing filter norm arise for general choices of objective functions when impulsive signals are deconvolved [5]. Increasing filter norm causes numerical problems in implementations, especially on fixed-point architectures. Therefore, a blind deconvolution algorithm maximizing the third-order moment must work under *constrained filter norm*.

An overview of several norm-constrained gradient adaptation algorithms can be found in [6] and [7], mainly considering objective functions of the form

$$J = \pm\frac{1}{p}\mathrm{E}\left\{|y_n|^p\right\},$$

where p is a positive integer. Since these functions are sign-invariant in their argument, they are unable to exploit asymmetry in y_n.

In this paper, one of the algorithms from [6] and [7] is studied when (2) is the specific function to be maximized. The work in [6] and [7] is extended, using averaging analysis, to determine conditions for asymptotic numerical stability. The computational cost of this algorithm is compared to other commonly employed methods.

2 Notation

Referring to Figure 1, the deconvolution filter $\mathbf{f}$ is an adaptive, real FIR filter of order N, represented at time n by its coefficient vector $\mathbf{f}_n \triangleq [f_{0n}\ f_{1n}\ \cdots\ f_{Nn}]^{\mathrm{T}}$. The *norm* of $\mathbf{f}_n$ is defined as the Euclidean, or ℓ_2-norm. Denoting the filter regressor by the vector of real samples $\mathbf{u}_n \triangleq [u_n\ u_{n-1}\ \cdots\ u_{n-N}]^{\mathrm{T}}$, the filter output becomes the vector inner product $y_n = \mathbf{f}_n^{\mathrm{T}}\mathbf{u}_n$.

The objective function to be maximized is the third-order central moment of the filter output y_n,

$$\mathcal{O}(y_n) \triangleq \frac{1}{3}\mathrm{E}\left\{y_n^3\right\} = \frac{1}{3}\mathrm{E}\left\{\left(\mathbf{f}_n^{\mathrm{T}}\mathbf{u}_n\right)^3\right\}. \tag{3}$$

Throughout the remainder of this paper, the operation count associated with implementations of each of the presented algorithms are taken under the assumption that all expectations of the form $\mathrm{E}\{x_n\}$ are estimated by instantaneous values x_n, as is customary for on-line applications.

3 Adaptation Under Constrained Filter Norm

3.1 Adaptation Using Steepest Ascent

Adaptation by *steepest-ascent* is used to adjust the filter to maximize the objective (3),

$$\mathbf{f}_{n+1} = \mathbf{f}_n + \mu\,\boldsymbol{\nabla}_n, \tag{4}$$

where μ is a small positive stepsize and $\boldsymbol{\nabla}_n$ is the gradient of $\mathcal{O}$ with respect to $\mathbf{f}_n$,

$$\boldsymbol{\nabla}_n \triangleq \frac{\partial\,\mathcal{O}}{\partial\,\mathbf{f}_n} = \frac{\partial\,\mathcal{O}}{\partial\,y_n}\frac{\partial\,y_n}{\partial\,\mathbf{f}_n} = \mathrm{E}\left\{y_n^2\,\mathbf{u}_n\right\}. \tag{5}$$

Using (4), $\mathbf{f}_n$ is iteratively adjusted until $\mathcal{O}$ attains a local maximum. Note that for any number α and any vector $\mathbf{f}$,

$$\mathcal{O}(\alpha\,\mathbf{f}) = \alpha^3\,\mathcal{O}(\mathbf{f}).$$

Hence, for any filter vector $\mathbf{f}$, we can improve $\mathcal{O}$ with the vector $\alpha\,\mathbf{f}$ if $\alpha > 1$. This indicates that (4) will never converge since $\boldsymbol{\nabla}_n$ will never approach zero. Instead, the norm of $\mathbf{f}_n$ will rapidly increase over iterations. A simple way to deal with this is to combine (4) with a frequent normalization procedure,

$$\mathbf{f}_{n+1} \leftarrow \frac{\mathbf{f}_{n+1}}{\|\mathbf{f}_{n+1}\|}. \tag{6}$$

While this would keep $\|\mathbf{f}_n\| = 1$ over iterations, the computational cost associated with combining (4) and (6) is relatively large; on the order of $4N$ operations per iteration for an N^{th}-order filter. Therefore, alternative ways to do steepest ascent under constrained filter norm are desired.

3.2 Orthogonal Gradient Decomposition

Recognize that a scaling $\alpha\,\mathbf{f}_n$ only results in a scaling $\alpha\,y_n$ of the filter output signal; the 'quality' of deconvolution is not changed. A reasonable approach would therefore be to avoid updating $\mathbf{f}_n$ in the radial direction.

Consider a decomposition of $\boldsymbol{\nabla}_n$ into $\boldsymbol{\nabla}_n = \mathbf{R}_n + \mathbf{P}_n$, where $\mathbf{R}_n$ is the orthogonal projection of $\boldsymbol{\nabla}_n$ onto $\mathbf{f}_n$,

$$\mathbf{R}_n \triangleq \frac{\boldsymbol{\nabla}_n^{\mathrm{T}}\mathbf{f}_n}{\|\mathbf{f}_n\|^2}\mathbf{f}_n. \tag{7}$$

Then modify the steepest-ascent algorithm to only update $\mathbf{f}_n$ in non-radial directions,

$$\mathbf{f}_{n+1} = \mathbf{f}_n + \mu\,\mathbf{P}_n = \mathbf{f}_n + \mu\,[\boldsymbol{\nabla}_n - \mathbf{R}_n]\,. \tag{8}$$

This algorithm can be viewed as a search for local maximum points of the objective function in the *tangent space of the hypersphere* $\|\mathbf{f}\| = \|\mathbf{f}_n\|$ at $\mathbf{f} = \mathbf{f}_n$. Unlike the standard algorithm (4), the modified version is expected to converge to points at which $\mathbf{P}_n$ approaches zero.

Ideally, the search for local maximum points should be restricted to some hypersphere, $\|\mathbf{f}_n\| = C$, to ensure that the filter norm stays fixed. For (8), it is straightforward to show that $\|\mathbf{f}_{n+1}\| \geq \|\mathbf{f}_n\|$. Hence, although the growth in $\|\mathbf{f}_n\|$ will not be as rapid as for the standard algorithm, this modified gradient ascent must be combined with an infrequent normalization of $\mathbf{f}_n$. Even without normalization, the operation count per iteration for an implementation of (8) is on the order of $4N$ for an N^{th}-order filter. Hence, this algorithm offers no computational savings.

3.3 Pseudo-Orthogonal Gradient Decomposition

A slight modification of (8) is achieved if the factor $1/\|\mathbf{f}_n\|^2$ is neglected in (7). Define

$$\widetilde{\mathbf{R}}_n \triangleq \left(\boldsymbol{\nabla}_n^{\mathrm{T}}\mathbf{f}_n\right)\mathbf{f}_n \quad \text{and} \quad \widetilde{\mathbf{P}}_n \triangleq \boldsymbol{\nabla}_n - \widetilde{\mathbf{R}}_n,$$

and do the filter adaptation as

$$\mathbf{f}_{n+1} = \mathbf{f}_n + \mu \widetilde{\mathbf{P}}_n = \left(1 - \mu \boldsymbol{\nabla}_n^{\mathrm{T}} \mathbf{f}_n\right) \mathbf{f}_n + \mu \boldsymbol{\nabla}_n. \tag{9}$$

As noted in [6] and [7], if $\boldsymbol{\nabla}_n^{\mathrm{T}} \mathbf{f}_n > 0$, this algorithm operates in a stable manner and maintains approximately unit filter norm. Absolute convergence of the algorithm will, however, ultimately depend on μ. From the definitions of the gradient (5) and the objective function (3), we find that

$$\boldsymbol{\nabla}_n^{\mathrm{T}} \mathbf{f}_n = \mathrm{E}\left\{y_n^3\right\} = 3\,\mathcal{O}(y_n), \tag{10}$$

i.e. the quantity $\boldsymbol{\nabla}_n^{\mathrm{T}} \mathbf{f}_n$ is proportional to the objective function to be maximized by the algorithm. Although $\mathcal{O}(y_n) > 0$ cannot be guaranteed for all n, the algorithm will most likely tend towards a positive objective over iterations for a well-conditioned problem.

Using (10), (9) can be rewritten as

$$\mathbf{f}_{n+1} = \left(1 - \mu\, \mathrm{E}\left\{y_n^3\right\}\right) \mathbf{f}_n + \mu \boldsymbol{\nabla}_n, \tag{11}$$

which exposes the algorithms simplicity. The computational cost of implementing this algorithm with an N^{th}-order filter is on the order of $3N$ operations per iteration. Furthermore, (11) contains only multiplications and additions (i.e. no divisions), making it highly suitable for implementation on fixed-point digital signal processors, which are specialized at performing such arithmetic operations.

The following section analyzes the asymptotical behavior of this algorithm and derives a sufficient condition on μ for numerical stability.

4 Asymptotic Stability of The Pseudo-Orthogonal Gradient Decomposition Algorithm

To analyze the behavior of $\|\mathbf{f}_n\|$ over iterations in the algorithm (11), define

$$\varepsilon_n \triangleq \|\mathbf{f}_n\|^2 - 1 \tag{12}$$

as the deviation of $\|\mathbf{f}_n\|^2$ from unity at time n. Multiplying both sides of (11) with their transposes and subtracting off one, gives after rearranging terms

$$\varepsilon_{n+1} = \left(1 - \mu\, 2\mathrm{E}\left\{y_n^3\right\}\right) \varepsilon_n + \mu^2 \|\widetilde{\mathbf{P}}_n\|^2. \tag{13}$$

This expression describes how the norm of $\mathbf{f}_n$ deviates from unity over iterations. The goal is to derive sufficient conditions on μ such that $\varepsilon_n \to 0$ as $n \to \infty$.

Note that (13) is a difference equation of the form

$$\varepsilon_{n+1} = \varepsilon_n + \mu\, g(n, \varepsilon_n, \mu), \tag{14}$$

where g is a nonlinear, stochastic and time-varying function. Assuming that the stepsize μ is small, (14) may be approximated by the *averaged system*

$$\overline{\varepsilon}_{n+1} = \overline{\varepsilon}_n + \mu\, g_{av}(\overline{\varepsilon}_n), \tag{15}$$

where

$$g_{av}(\varepsilon) = \mathrm{E}\left\{g(n, \varepsilon, 0)\right\}|_{\varepsilon\,=\,\text{constant}}\,. \tag{16}$$

The conditions necessary for the approximation of (14) with the averaged system (15) to be valid are essentially that, over a fixed time interval; g_{av} is time invariant, $|\overline{\varepsilon}|$, $|g_{av}|$ are bounded, and g and the difference $g - g_{av}$ fulfill global Lipschitz conditions in ε and μ. Refer to [8, Ch. 9] for details.

Although the expectation in (16) is taken with $\mu = 0$, we choose to regard g_{av} as a function of both ε and μ to investigate how the stepsize affects the asymptotical behavior of the algorithm. Comparing (13) with (14) gives

$$g(n, \varepsilon_n, \mu) = -2\mathrm{E}\left\{y_n^3\right\}\varepsilon_n + \mu\,\|\widetilde{\mathbf{P}}_n\|^2,$$

and the averaged system from (16) as

$$g_{av}(\varepsilon, \mu) = -2\,S_y\,\varepsilon + \mu\,\widetilde{P}^2,$$

where

$$S_y \;\triangleq\; \mathrm{E}\left\{y_n^3\right\}, \tag{17}$$

$$\widetilde{P}^2 \;\triangleq\; \mathrm{E}\left\{\|\widetilde{\mathbf{P}}_n\|^2\right\} \tag{18}$$

are assumed to be time invariant.

Although the time-invariance assumption on (17) and (18) is not realistic over a large span of iterations (in fact, note that S_y is proportional to the objective function to be maximized), they are approximately time invariant over limited number of iterations if μ is small.

For small values of μ, (13) may thus be approximated by

$$\overline{\varepsilon}_{n+1} = (1 - \mu 2 S_y)\,\overline{\varepsilon}_n + \mu^2 \widetilde{P}^2. \tag{19}$$

If $1 - \mu 2 S_y \neq 1$, (19) can be rewritten as

$$\overline{\varepsilon}_n = (1 - \mu 2 S_y)^n\,\overline{\varepsilon}_0 + \frac{\mu \widetilde{P}^2}{2 S_y}\left[1 - (1 - \mu 2 S_y)^n\right]. \tag{20}$$

Under the condition

$$|1 - 2\mu S_y| < 1, \tag{21}$$

the sequence (20) converges, and we get

$$\lim_{n\to\infty} \overline{\varepsilon}_n = \frac{\mu \widetilde{P}^2}{2S_y}. \tag{22}$$

Thus, the asymptotic deviation of $\|\mathbf{f}_n\|^2$ from unity is proportional to the stepsize μ. In general, $\mu \ll 1$, and so the algorithm (11), if stable, operates very close to unit filter norm.

Although the limit (22) is taken under the approximation of S_y and $\widetilde{P}^2$ being time invariant, it is suggested that (21) gives a sufficient condition for local convergence of ε_n over a limited span of iterations, over which time invariance assumptions hold.

Condition (21) can be rewritten as

$$0 < \mu\, S_y < 1. \tag{23}$$

Since μ is positive by definition, $S_y = \mathrm{E}\{y_n^3\}$ is required to be positive, which is expected at convergence for a well-conditioned problem. Furthermore, S_y is expected to slowly increase over iterations as the algorithm converges in $\mathbf{f}_n$. The quantity $\mu\, S_y$ could therefore be monitored during adaptation, and the stepsize decreased if necessary, in order to insure that (23) holds. This guidance on stepsize requires only a simple scalar multiplication and check.

Note that (23) only concerns stability in $\|\mathbf{f}_n\|$. A stepsize satisfying (23) is not guaranteed to give convergence to an $\mathbf{f}_n$ that maximizes the objective.

5 Experimental Results

In a numerical experiment, the algorithm (11) was used to implement an adaptive blind equalizer for a synthetic indoor Ultra-Wideband (UWB) communication channel with Impulse Radio signaling [9]. Such signals consist of pulse-position modulated impulses of extremely short duration, typically on the order of a nanosecond. Because of the large multipath spread of typical indoor UWB channels, intersymbol interference (ISI) is likely to occur at high data rates [10], [11]. Due to the impulsive nature of these signals, an adaptive blind equalizer based on third-order moments might be used to mitigate the effects of ISI.

The impulse response of an indoor UWB channel with a rich multipath spread of approximately 200 nanoseconds was synthesized with the aid of a recipe from [12], with additive white Gaussian noise at a signal-to-noise ratio per bit of 15dB. The Impulse Radio signals used binary orthogonal modulation at a bit rate of 10 Mbits/second, using a sampling rate of 15 samples per nanosecond. Equalizers of order $N = 400$ were generated using (11) for three different stepsizes over 1000 adaptation iterations. All expectation operations in (11) were estimated using instantaneous values. Figures 2 and 3 show, respectively, the resulting absolute deviation of $\|\mathbf{f}_n\|$ from unity and skewness versus iteration number. The plots show averaged results over 20 independent runs.

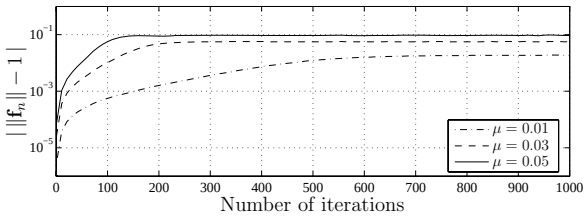

Figure 2: Absolute deviation of $\|\mathbf{f}_n\|$ from unity versus iteration number.

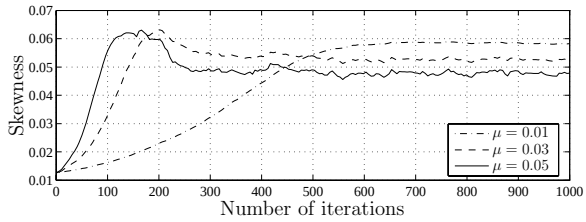

Figure 3: Estimated skewness of filter output versus iteration.

As seen in Figure 2, the deviation from unit norm at convergence increases with the stepsize, confirming the result (22) from Section 4. Figure 3 shows the convergence of the algorithm in terms of skewness. Note that a larger stepsize leads to faster convergence, but results in a smaller asymptotic skewness.

Experimental results also indicate that the stability condition (23) indeed can be monitored to indicate instability in $\|\mathbf{f}_n\|$. However, for stepsizes that give convergence in $\mathbf{f}_n$ (as seen in Figure 3), (23) is typically satisfied by a large margin. Thus, for reasonable choices of μ, the algorithm should be stable in $\|\mathbf{f}_n\|$.

6 Conclusion

A computationally efficient algorithm for constrained-norm gradient acsent has been studied for blind deconvolution. The results indicate that the algorithm maintains approximately unit filter norm for reasonable choices of adaptation stepsize. The condition on adaptation stepsize that insures a stable filter norm is trivial to calculate and verify. Therefore, this algorithm provides a way to perform blind deconvolution without the problem of increasing filter norm. The small computational cost, involving only multiplications and additions, makes it well suited for on-line implementation on fixed-point digital signal processors.

References

[1] R. A. Wiggins, "Minimum entropy deconvolution," *Geoexploration*, no. 16, pp. 21–35, 1978.

[2] D. L. Donoho, "On minimum entropy deconvolution," in *Applied Time Series Analysis*, D. F. Findley, Ed. New York: Academic Press, 1981.

[3] P. Pääjärvi and J. P. LeBlanc, "Skewness maximization for impulsive sources in blind deconvolution," in *Proc. IEEE Nordic Signal Proc. Symp.*, Espoo, Finland, June 2004, pp. 304–307.

[4] P. Pääjärvi and J. P. LeBlanc, "On-line adaptive blind deconvolution based on third-order moments," IEEE Signal Processing Lett., Dec. 2005, to appear.

[5] H. Mathis and S. C. Douglas, "Bussgang blind deconvolution for impulsive signals," *Proc. IEEE*, vol. 51, no. 7, pp. 1905–1915, July 2003.

[6] S. C. Douglas, Shun-ichi Amari, and S.-Y. Kung, "Gradient adaptation with unit-norm constraints," Southern Methodist Univ., Dallas, Texas, Tech. Rep. EE-99-003, 1999.

[7] S. C. Douglas, S. Amari, and S.-Y. Kung, "On gradient adaptation with unit-norm constraints," *IEEE Trans. Signal Processing*, vol. 48, no. 6, pp. 1843–1847, June 2000.

[8] V. Solo and X. Kong, *Adaptive Signal Processing Algorithms, Stability and Performance*. Englewood Cliffs, New Jersey: Prentice-Hall, 1995.

[9] M. Z. Win and R. A. Scholtz, "Ultra-wide bandwidth time-hopping spread-spectrum impulse radio for wireless multiple access communications," *IEEE Trans. Commun.*, vol. 48, no. 4, pp. 679–691, Apr. 2000.

[10] D. Porcino and W. Hirt, "Ultra-wideband radio technology: Potential and challenges ahead," *IEEE Commun. Mag.*, vol. 41, no. 7, pp. 66–74, July 2003.

[11] A. G. Klein and C. R. Johnson, Jr., "MMSE decision feedback equalization of pulse position modulated signals," in *Communications, 2004 IEEE International Conference on*, vol. 5, Paris, France, June 2004, pp. 2648–2652.

[12] D. Cassioli, M. Z. Win, and A. F. Molisch, "The ultra-wide bandwidth indoor channel: From statistical model to simulations," *IEEE J. Select. Areas Commun.*, vol. 20, no. 6, pp. 1247–1257, Aug. 2002.