DOCTORAL THESIS

L

Blind Equalization Using Third–Order Moments

Patrik Pääjärvi

Blind Equalization Using Third-Order Moments

Patrik Pääjärvi

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

Supervisor:

Professor James P. LeBlanc

ABSTRACT

The use of third-order moments in blind linear equalization has been studied with emphasis on their performance in on-line methods of low complexity. Blind equalization has widespread use in areas such as digital communications, acoustics, geophysical exploration, image processing and general measurement problems. The objective is to recover a desired, unknown source signal from distorted observations using a linear filter. In contrast to classical methods of deconvolution or trained equalization, no model of the distorting system or temporal observations of the source are assumed. Instead, knowledge of the higher-order source statistics are exploited to find an appropriate equalizer setting through some iterative numerical procedure.

The application field and research area have traditionally focused on methods based on fourth-order statistics. Examples are various kurtosis-maximization approaches and the widely used constant modulus algorithm (CMA). While qualifying for blind methods, third-order statistics have attracted less attention, mainly due to the reason that they can only be used when the source signal is asymmetric, i.e. when the probability density is skewed. As a consequence of the more restricted usability of third-order methods, mostly experimental research results can be found, with little explanation for their performance.

This work provides analytic and numerical results motivating why third-order methods should generally be chosen over their fourth-order counterparts when possible. It is shown that they possess improved convergence properties and robustness to noise, and that they lend themselves to efficient implementation on digital real-time hardware. These combined features make third-order methods an interesting option for on-line blind equalization.

CONTENTS

Acknowledgements

Patrik Pääjärvi
Luleå, May 2008

Part I

CHAPTER 1

Inverse Systems, Deconvolution and Equalization

In this chapter, the concept of an inverse system is presented at a medium level of detail. Examples are given illustrating applications of inverse systems using methods of deconvolution and equalization.
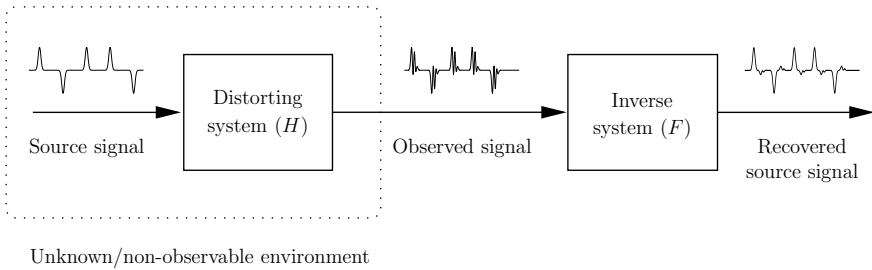
Figure 1.1: The principle of source signal recovery using an inverse system.

1.1 Recovering an Unknown Signal from Distorted Measurements

A situation arising in many areas of signal processing is that of measuring or observing a distorted version of an unknown signal of interest. Fig. 1.1 illustrates this with a block diagram model. A *source* signal becomes distorted as it propagates through some physical system H before being observed at the system's output. Depending on the context, the source signal and distorting system are either completely unknown or only

partially known/observable. The goal here is to recover the source using an *inverse system* F, designed to 'undo' the distorting effects of H. Unless stated otherwise, the system H is assumed to belong to the broad class of *linear time-invariant (LTI) systems* [1]. Non-linear systems and their inverses (if they exist) are beyond the scope of this work.

Ideally, F completely inverts the action due to H so that the source is perfectly recovered at its output. However, a perfectly inverting system is usually not feasible in practice, and so only an approximate inversion is pursued. The inverse system may be in the form of an electric circuit, a digital filter, or a computer process or algorithm.

Example: Loading Coils on Long-Distance Telephone Lines One of the first uses of inverse systems dates back to the late 19th century and the early days of telephony. Voice communication over long twisted-pair wires was plagued by distortion which degraded the quality of speech, and essentially limited the practical length of a cable to around 1 200 miles (1 900 km, or the distance between Boston and Chicago) [2]. In 1899, George A. Campbell demonstrated how inductance coils, connected at periodically spaced intervals along the cable, significantly reduced the distortion, thereby allowing much longer telephone lines. The design and exact placement of such *loading coils* was based on theoretical models of transmission lines. The coils compensated for the capacitance in the long cables, thus acting as simple inverting systems. □

1.2 Deconvolution and Equalization

As in the example above regarding loading coils, constructing the inverse system may become relatively straightforward if an accurate *model* of the distorting system is available, i.e. a mathematical or physical relationship between the input and output signal. Designing an inverse system using an existing model of the distorting system is commonly referred to as *deconvolution* [1]. Unfortunately, many practically encountered systems are too complex to model analytically, or simply have unknown characteristics.

If the distorting system input is somehow accessible, the characteristics of H may be figured out by observing the resulting output when a *known* signal is input to the system. This process is referred to as *system identification* [1], and produces a model of H from which an inverse system can be designed. Alternatively, the injected input signal is used as a template for the output of a perfect inverse system. F may then be directly synthesized by making its output resemble this known source signal as closely as possible. This is the principle behind *equalization* [3].

Example: Modem Communication Computer communications over switched telephone networks use modems that translate a stream of bits into an alternating waveform suited for transmission over wire lines. However, typical telephone lines were originally intended for adequate voice communication and not high-speed data transfer. As a

consequence, transmissions at bit rates above 2 400 bits/second suffer distortion caused mainly by bandwidth constraints enforced in the network [3]. If not properly compensated for, this distortion introduces severe bit errors in the transmissions.

A key component in modems is therefore an inverse system called an *equalizer*, in the form of a digital filter used to undo as much of the telephone channel distortion as possible. To succeed with this, the equalizer needs to closely emulate the inverse characteristics of the channel, and so an accurate model of it is needed. Matters are complicated by the fact that connections from one point to another in a switched telephone network may take different routes each time a call is made. As a consequence, the character of the distortion may vary greatly between dial-ups, but remains relatively fixed once a connection is established. Therefore, the equalizer needs to have adjustable parameters and hence the ability to adapt itself to various channel conditions [4].

Fig. 1.2 illustrates how such an *adaptive equalizer* is used to combat distortion from an initially unknown channel. At startup (before regular data transfer begins), the transmitting modem generates and sends a pre-determined *training signal* which subsequently arrives distorted at the receiving modem [5]. This distorted reception is first filtered by the equalizer and then subtracted from a locally generated replica of the training signal to form an *error signal*, reflecting the degree of mismatch between the equalizer output and the training signal. If the equalizer was optimally adjusted to invert the channel, the transmitted training signal would be perfectly recovered at its output, making the error signal become zero. In reality, the initial equalizer parameterization will probably be far from optimal, and the error signal will be nonzero. Through a parameter adjustment algorithm, the equalizer is iteratively adapted to minimize the power of the error signal during the course of the training period. After this, the adjustable parameters are frozen, and regular data transfer may commence with a low probability of bit error. □
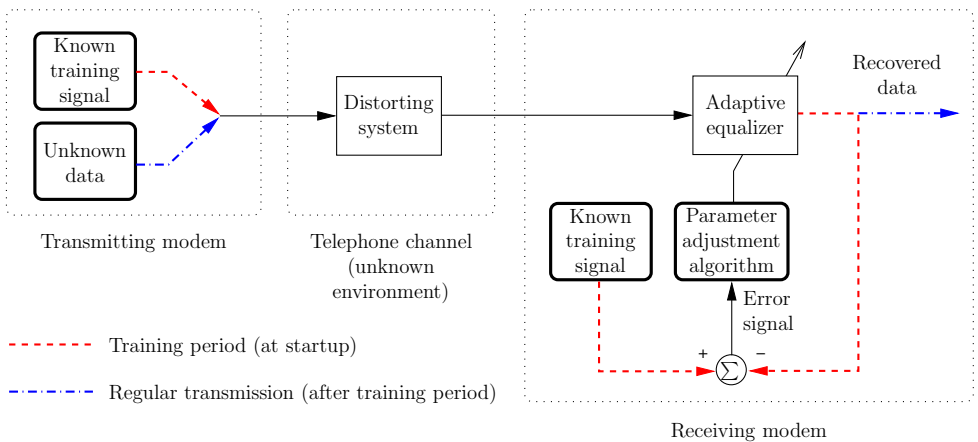
Figure 1.2: *The principle of adaptive equalization in modem communication.*

Example: Mobile Communication   A wireless communication channel between a cell phone and a base station is also impaired with distortion that causes bit errors if not properly compensated for. However; the distortion on a terrestrial radio channel is not due to bandwidth limitations, but to a phenomenon known as *multipath propagation* [3], illustrated in Fig. 1.3. As the radio signal transmitted from a base-station antenna propagates through the air, it is reflected off surrounding objects such as mountains and buildings. It reaches the cell phone (situated in the moving car) as a series of individually delayed and attenuated replicas.
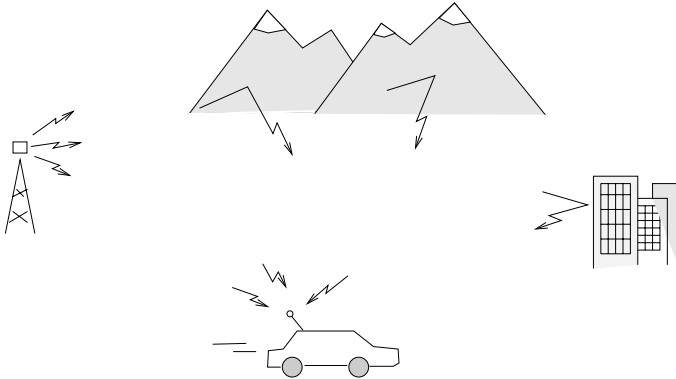
Figure 1.3: Multipath propagation in mobile communication.

Although of different origin than the distortion on a telephone line, multipath propagation effects can also be combatted with an adaptive equalizer. Unfortunately, as the mobile unit is moving, the surroundings change, causing the channel characteristics to become *time varying*. Hence; a single training period at communication startup (as in the initialization of a modem equalizer) is not sufficient here as the channel variations will make a fixed equalizer setting perform poorly after a short while [5].

A possible solution to the time-variation problem is to periodically interrupt data transfers for re-transmission of the training signal, thus letting the equalizer regularly re-adapt itself to current channel conditions. However, such repeated training comes at the cost of a reduced overall data throughput. □

1.3   Blind Deconvolution and Equalization

In cases where the distorting system is unknown and its input is not accessible for system identification or equalizer training purposes, finding the inverse system becomes significantly more challenging. Nevertheless, for some classes of systems and source signals,

the problem can still be solved using techniques for *blind deconvolution* [6]. The common approach is to rely on knowledge about some specific, measurable *features* or *characteristics* of the source signal. If these source features are altered by the distorting system in a way that can be measured, this may serve as guidance as to how the inverse system parameters should be chosen.

The terms 'blind deconvolution' and 'blind equalization' are often used interchangeably. As is customary, 'equalization' is here used in the context of on-line, or real-time problems, and 'deconvolution' will refer to off-line, or post-processing methods [7].

Example: Seismic Reflection Surveying   A method used by geophysicists to model the geology of the earth's subsurface without excavation is *seismic reflection surveying* [8], illustrated in Fig. 1.4. The subsurface is modeled as a layered medium, in which the layers consist of different materials, e.g. soil, water, oil or gas. The goal is to obtain a rough characterization of the subsurface structure by detecting these layers and determining their respective depths. A seismic source consisting of a pneumatic hammer, a hydraulic vibrator or dynamite, produces pulses, or *seismic wavelets*, that propagate down into the ground as acoustic waves. At the boundary between two layers, the change in acoustic impedance causes a reflection of the waveform back to the surface, where a geophone converts the ground motions into an electrical signal. This *seismic trace* is recorded for post-processing on a computer.
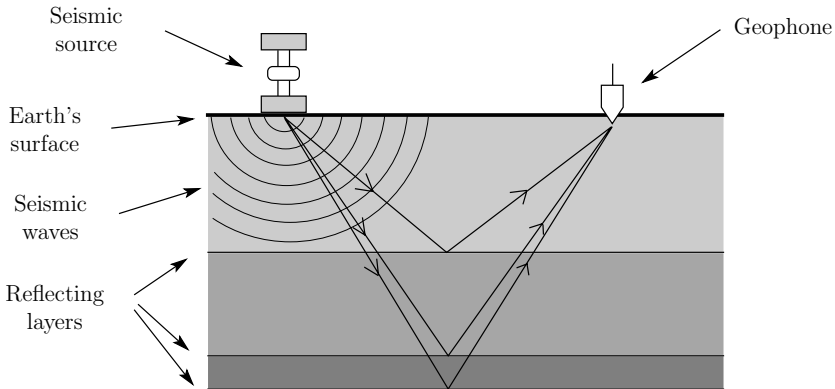
Figure 1.4: Illustration of seismic reflection surveying.

If the source was able to produce a perfect impulse of infinitely short duration, the resulting seismic trace would be an ideal 'signature' of the subsurface structure, from which the reflecting layer structure could be identified. However, the seismic wavelet has an imperfect, unknown shape that distorts, or 'smears' the signature, making the individual reflections difficult to resolve. Blind deconvolution methods may be used in post-processing to undo most of the dispersive effects of the seismic wavelet, and recover a

cleaner signature from which the earth layer reflections can be identified. A synthetic example is illustrated in Fig. 1.5 with three reflective layers. Research on seismic reflection surveying has made important contributions to the understanding of blind deconvolution principles which can be translated and applied to other application areas [9, 10]. □
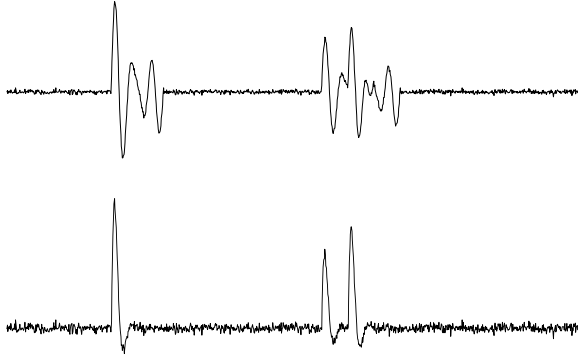
Figure 1.5: Synthetic example of a seismic trace (top) and deconvolved subsurface reflections (bottom).

Example: Blind Image Restoration   Digital images acquired from devices such as cameras, telescopes or x-ray machines are sometimes distorted, or *blurred*, where the blurring is caused by imperfections in the imaging system or transmission medium. This fits the model shown in Fig. 1.1 with the source signal representing the original image to be captured, H the 'blur' (i.e. the image distortion) and the observed signal representing the captured image. The objective of *blind image restoration* is to recover the original image through post-processing of the captured image using blind deconvolution methods [11].

The motivation behind blind image restoration is often that a high-quality image capture is impossible or not practical. For example; in astronomical imaging, pictures of objects in space are acquired from a telescope using a very long exposure time, during which turbulence in the atmosphere unavoidably blurs the image and calls for blind restoration. In medical x-ray imaging, sharp pictures require high beam intensity, which becomes dangerous for the patient. Therefore, improving image quality through post-processing is preferred for safety reasons [11].   □

Example: Blind Equalization   In the mobile communication example, the equalizer is initially adapted to an unknown distorting channel and then regularly re-trained

over time to track the time-varying channel characteristics. In some cases, repeated training is considered too costly as ordinary data transfer must stall while the training signal is sent. In other situations, such as in point-to-multipoint communication or broadcast networks, synchronizing training periods between a transmitter and a receiver is simply not possible [5].

Blind equalization provides a way to adapt the equalizer to the channel *without* the need of training. The concept is shown in Fig. 1.6. In the absence of a training signal, the equalizer parameter adjustments are based solely on the output signal through an *objective function*. This function is chosen to quantify the amount of distortion present in the equalizer output by measuring how altered some *known features or characteristics* of the source signal have become through the transmission and reception process. The equalizer parameters are tuned according to this function to ultimately minimize the distortion in the received data, similar to how a photographer adjusts the focus on a camera. As no training periods are needed, adaptation can be performed continuously, so that the equalizer may constantly track slow channel variations. □
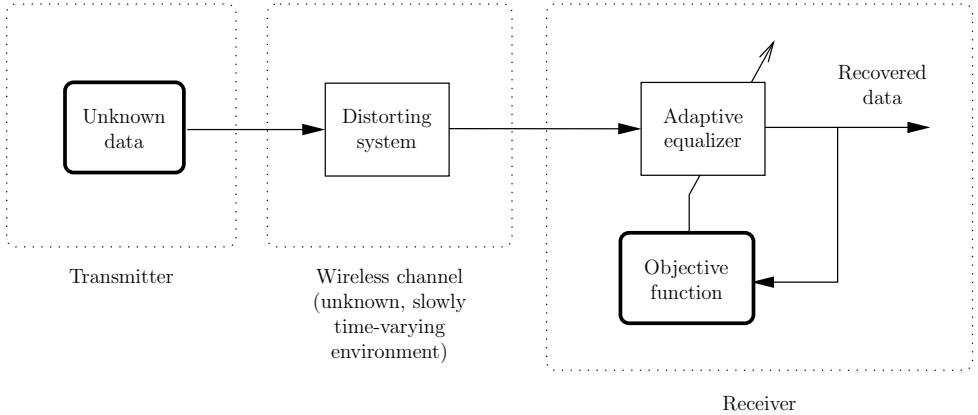
Figure 1.6: The principle of blind equalization.

The research in Part II of this thesis focuses on on-line methods for blind linear equalization. The fundamentals of linear equalization are given in the next chapter, and an overview of blind equalization is provided in Chapter 3. The research contributions of Part II are summarized in Chapter 4.

Linear Equalization Fundamentals

Here, the basic concepts of (non-blind) linear equalization are presented, in which access to the desired equalizer output is assumed. The problem setting is angled towards, but not necessarily limited to, applications in digital communication. First, zero forcing (ZF) and minimum mean-squared error (MMSE) equalization are considered without any practical restrictions imposed on the linear equalizer. Then, a practical MMSE equalizer of finite impulse response (FIR) type is considered, along with the classical method of implementing it, the *least-mean-square (LMS) algorithm*. Finally, adaptive equalization of a communication channel is addressed. Background knowledge in basic signals and systems theory [12] and stochastic processes [13] is hereafter assumed.

2.1 The Linear Equalization Model
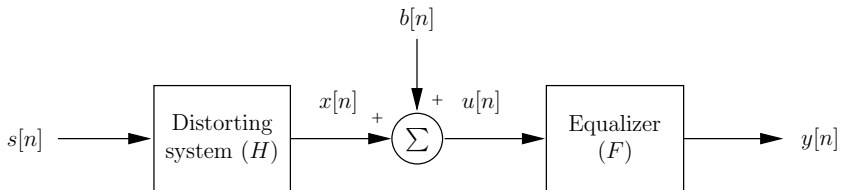
Figure 2.1: Model of the linear equalization problem.

A discrete-time model of the linear equalization problem is shown in Fig. 2.1. The following model assumptions are made throughout this chapter.

- The source signal $s[n]$ is a discrete-time, real, stationary stochastic process with zero mean and discrete-time power spectrum $P_s(\Omega)$.

9

- The distorting system is linear and time invariant (LTI) with discrete-time transfer function $H(\Omega)$.

- $b[n]$ is noise, statistically independent of $s[n]$, modeled as a discrete-time, real, stationary stochastic process with zero mean and power spectrum $P_b(\Omega)$.

- $u[n] = x[n] + b[n]$ is the observed signal.

- The equalizer is an LTI system with discrete-time transfer function $F(\Omega)$.

Before enforcing any practical restrictions on the linear equalizer, ideal expressions for $F(\Omega)$ according to the *zero-forcing* (ZF) and *minimum mean-squared error* (MMSE) criteria are given.

2.2   Ideal ZF Equalizer

The objective of the ZF equalizer is to invert the distorting system while disregarding the effects of the noise $b[n]$. Hence, the goal is to recover the source at the equalizer output and achieve:
$$y[n] = s[n].$$
In absence of $b[n]$, this is accomplished by choosing the equalizer transfer function as

$$F_{\text{ZF}}(\Omega) = \frac{1}{H(\Omega)}, \tag{2.1}$$

which is simply the inverse transfer function of the distorting system. Clearly, the ZF equalizer only exists if $H(\Omega)$ has no spectral nulls (i.e. the system function $H(z)$ has no zeros on the unit circle). The expression (2.1) reveals a fundamental connection between equalization and system identification. As seen from the reciprocal relationship between F_{ZF} and H, the transfer characteristics of the system are fully incorporated into the ZF equalizer. Hence, in a theoretical sense, inverting a linear system is essentially the same problem as identifying it.

2.3   Ideal MMSE Equalizer

While theoretically providing perfect system inversion, the main drawback of the ZF equalizer is that noise, which is always present in practice, is not taken into account. In fact, the detrimental effects of the additive noise $b[n]$ are generally *enhanced* by the ZF equalizer [3], making it less suitable in some applications. For instance; in digital communications, the figure of merit of ultimate interest is the bit-error rate, not how well the equalizer inverts H. With noise present, perfect source recovery becomes impossible. Instead, a natural performance measure is obtained by statistically quantifying the 'closeness' between $y[n]$ and $s[n]$ in terms of the mean-squared error

$$\text{E}\left\{(s[n] - y[n])^2\right\}, \tag{2.2}$$

where $\mathrm{E}\{\cdot\}$ denotes expectation. Choosing the equalizer transfer function to minimize (2.2) results in the *minimum mean-squared error* (MMSE) equalizer

$$F_{\mathrm{MMSE}}(\Omega) = \frac{\mathrm{P}_s(\Omega)H^*(\Omega)}{\mathrm{P}_s(\Omega)|H(\Omega)|^2 + \mathrm{P}_b(\Omega)}, \tag{2.3}$$

where * denotes complex conjugation. Unlike the ZF equalizer, the MMSE equalizer takes the spectral densities of both source signal and noise into account to offer the best possible trade-off between system inversion and noise suppression. Indeed, if $\mathrm{P}_b(\Omega)$ is set to zero in (2.3), $F_{\mathrm{MMSE}}(\Omega)$ reduces to $F_{\mathrm{ZF}}(\Omega)$. The benefits of the MMSE equalizer normally makes it the first-hand choice over the ZF equalizer.

2.4 Practical MMSE Equalizer

(2.3) gives the transfer function of the ideal, generally non-causal MMSE equalizer. A practical linear equalizer, on the other hand, must be stable and causal. It will be assumed throughout the following that the practical equalizer of interest is of finite impulse response (FIR) type. FIR filters feature guaranteed bounded-input bounded-output (BIBO) stability, ease of analysis, and simple realization on finite-precision hardware [12]. An N-tap FIR equalizer is represented by its parameter vector

$$\mathbf{f} \triangleq \begin{bmatrix} f_0 & f_1 & \cdots & f_{N-1} \end{bmatrix}^{\mathrm{T}}.$$

Given an observed sequence of N input samples collected in the equalizer regressor vector

$$\mathbf{u}_n \triangleq \begin{bmatrix} u[n] & u[n-1] & \cdots & u[n-N+1] \end{bmatrix}^{\mathrm{T}},$$

the output signal from the FIR equalizer becomes $y[n] = \mathbf{f}^{\mathrm{T}}\mathbf{u}_n$. The MSE criterion (2.2) is now impractical since the FIR equalizer is necessarily causal. A *delayed* version of the source signal must instead serve as the *desired output*

$$d[n] \triangleq s[n-\Delta],$$

where Δ is a finite, positive integer. The MSE to be minimized is now defined as

$$\xi \triangleq \mathrm{E}\left\{e^2[n]\right\}, \tag{2.4}$$

where $e[n]$ is the *error signal*

$$e[n] \triangleq d[n] - y[n].$$

Hence, ξ reflects the mean power of the error signal. Note that given observations of $\mathbf{u}_n$ and $d[n]$, (2.4) is a function of the equalizer parameters,

$$\xi = \xi(\mathbf{f}) = \mathrm{E}\left\{(d[n] - \mathbf{f}^{\mathrm{T}}\mathbf{u}_n)^2\right\}.$$

By further defining the input autocorrelation matrix $\mathbf{R}$ and the cross-correlation vector $\mathbf{p}$ as

$$\mathbf{R} \triangleq \mathrm{E}\{\mathbf{u}_n\mathbf{u}_n^{\mathrm{T}}\}, \quad \mathbf{p} \triangleq \mathrm{E}\{d[n]\mathbf{u}_n\},$$

the unique parameter vector $\mathbf{f}_W$ minimizing ξ becomes

$$\mathbf{f}_W \triangleq \mathbf{R}^{-1}\mathbf{p}, \tag{2.5}$$

assuming the inverse of $\mathbf{R}$ exists. $\mathbf{f}_W$ is often referred to as the *Wiener* filter [14].

   While providing a closed-form expression for the FIR MMSE equalizer, (2.5) has practical drawbacks. The invertibility of $\mathbf{R}$ cannot always be assumed in practice, for example if the power spectrum of $u[n]$ has frequency bands with zero power [15]. Furthermore, even if $\mathbf{R}$ is non-singular, the computational resources and numerical precision needed for the direct matrix inversion may be considered too high. Recursive procedures such as the *recursive least squares* (RLS) algorithm [15] or its variants may then be preferred.

   If low computational complexity is of greater concern than obtaining the optimum parameter vector, *iterative* minimization of (2.4) based on a gradient search method is a standard approach. The workhorse of iterative MSE minimization is the *least-mean-square*, or LMS algorithm [14].

2.5   The LMS Algorithm

Let $\mathbf{f}_n$ represent an *adaptive* FIR equalizer at time n,

$$\mathbf{f}_n \triangleq \left[\begin{array}{ccc} f_0^{[n]} & \cdots & f_{N-1}^{[n]} \end{array} \right]^{\mathrm{T}}.$$

The *gradient vector* of ξ with respect to $\mathbf{f}_n$, representing the direction of most rapid increase of ξ, is

$$\boldsymbol{\nabla}_\xi(\mathbf{f}_n) \triangleq \frac{\partial\xi}{\partial\mathbf{f}_n} = \left[\begin{array}{c} \dfrac{\partial\xi}{\partial f_0^{[n]}} \\ \vdots \\ \dfrac{\partial\xi}{\partial f_{N-1}^{[n]}} \end{array} \right] = -2\mathrm{E}\left\{ e[n]\mathbf{u}_n \right\}. \tag{2.6}$$

The MSE can be iteratively minimized by taking small steps in the *negative* direction of (2.6),

$$\mathbf{f}_{n+1} = \mathbf{f}_n - \mu\boldsymbol{\nabla}_\xi(\mathbf{f}_n) = \mathbf{f}_n + 2\mu\mathrm{E}\left\{ e[n]\mathbf{u}_n \right\}, \tag{2.7}$$

where μ is a small, positive stepsize[1]. In practical implementations, simple gradient estimates are obtained by dropping the expectation operator in the right-hand side of (2.7). This results in the *LMS algorithm*:

$$\mathbf{f}_{n+1} = \mathbf{f}_n + 2\mu e[n]\mathbf{u}_n. \tag{2.8}$$

The update procedure (2.8) is an example of a *stochastic gradient search* in which an instantaneous sample value $\alpha[n]$ is used as an estimate of $\mathrm{E}\{\alpha[n]\}$, resulting in an algorithm with a very low computational cost. Although such estimates may have relatively high variance, a small value on the stepsize μ makes the adaptation behave like a low-pass filter that reduces rapid variations over iterations.

---

[1]It is assumed here for simplicity that iterations are performed at the sampling rate, which is not necessary in practice.

2.5.1 Benefits and Drawbacks of the LMS Algorithm

Unlike closed-form calculation of the Wiener filter through (2.5), a rank-deficient auto-correlation matrix $\mathbf{R}$ does not cause immediate method breakdown in LMS. MSE is still minimized under the given conditions although the solution vector is no longer unique, which can cause unlimited growth of one or several of the parameters in $\mathbf{f}_n$. Furthermore, with a fixed stepsize μ, only convergence to a *neighborhood* in parameter space around the optimum solution vector is achieved, in which $\mathbf{f}_n$ 'rattles around' during the final phase of adaptation. While a smaller stepsize makes this *misadjustment* smaller, it also reduces the convergence rate. Hence, the size of μ essentially trades off convergence rate for misadjustment. However, choosing μ too large may cause adaptation to become unstable.

Several modified versions of the LMS algorithm, such as *leaky LMS* and *normalized LMS* have been developed to address the issues mentioned above. Also, even though (2.8) is already very easy to implement, there are further simplified variants like *sign-error LMS, sign-data LMS* and *sign-sign LMS* that give up some performance for even lower computational cost [15].

2.5.2 Graphical Interpretation of Iterative MSE Minimization

A useful abstraction of iterative MSE minimization can be obtained by considering $\xi(\mathbf{f})$ to be an N-dimensional hypersurface parameterized by the equalizer coefficients $f_0, \ldots, f_{N-1}$. While hard to visualize in higher dimensions, the *two*-dimensional surface $\xi(f_0, f_1)$ can be drawn in a three-axis coordinate system. Assuming $\mathbf{R}$ has full rank, this surface assumes the shape of a (possibly elongated) bowl, as illustrated in Fig. 2.2. LMS adaptation of a two-tap FIR equalizer corresponds to traversing such a bowl in the estimated direction of steepest descent. An example of an LMS trajectory starting at the setting $f_0 = f_1 = -2$ is shown in Fig. 2.2 along with its projection onto the parameter plane. The trajectory heads towards the bottom of the bowl which is aligned at the Wiener setting, marked by ($\star$).

In higher dimensions, the shape of the MSE surface generalizes to a 'hyperbowl' with its bottom aligned at the corresponding Wiener setting $\mathbf{f}_W$. Hence, regardless of where iterative MSE minimization starts, convergence towards $\mathbf{f}_W$ is always achieved for a well-conditioned problem of arbitrary dimension. This useful property is known as *global convergence* and is discussed in Chapter 3.

2.6 Adaptive Equalization of a Communication Channel

Iterative MSE minimization is now put in the context of adaptive equalization of a time-varying digital communication channel. A discrete-time model of a baseband-equivalent communication system is shown in Fig. 2.3.
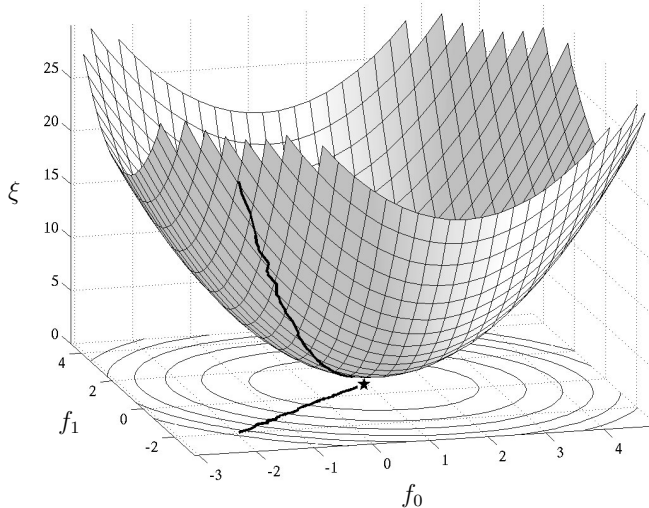
Figure 2.2: Example of a two-dimensional MSE surface $\xi(f_0, f_1)$ with its level curves shown on the parameter plane. An LMS trajectory starting at the point $(-2, -2)$ is shown along with its projection onto the plane. The Wiener parameter setting $\mathbf{f}_W = (0.93, 0.67)$ is marked by $(\star)$.

The transmitter sends a stream of independent, identically distributed (i.i.d.) digital symbols $s[n]$ over a channel modeled as a linear distorting system H with an unknown and slowly time-varying transfer function, and additive noise $b[n]$. In the ideal case of H being a pure delay, transmitted symbols could be estimated one-by-one at the receiver using a *decision device*, which produces a stream of symbol estimates $\hat{s}[n]$ from noisy observations. In reality, the non-ideal channel transfer function may cause samples in the received signal to interfere with each other, making symbol-by-symbol detection with a low bit-error rate impossible in such cases [3]. This harmful *intersymbol interference* (ISI) is here combatted using the adaptive equalizer F. The MSE criterion is chosen as ZF equalization generally amplifies the noise $b[n]$, thereby counteracting the removal of ISI in terms of reliable symbol estimation. Iterative minimization of MSE using an adaptive equalizer is normally preferred for the following reasons:

- Iterative, continuous adaptation of the equalizer makes it possible to track slow variations in the time-varying channel.

- High data transfer rates require fast, numerically stable algorithms with low complexity to meet real-time constraints.

- A battery-powered receiver, such as a cell phone, needs computationally efficient methods for prolonged operation time.

At communication startup, the LMS algorithm is employed during an initial training period in which pre-determined data sequences, known to both the transmitter and
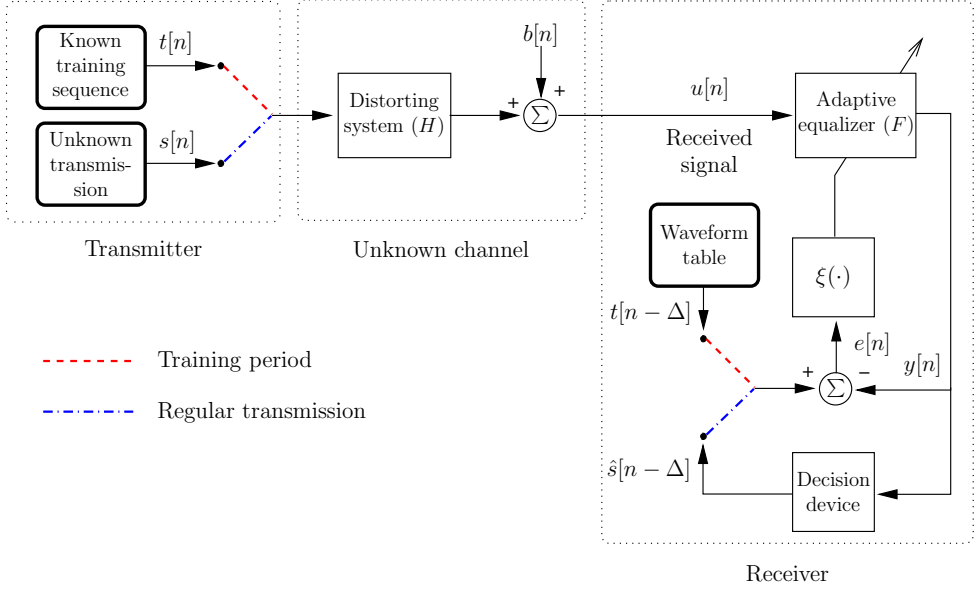
Figure 2.3: Model of adaptive equalization of a communication channel.

receiver, are used to drive equalizer parameters towards an optimum setting, thereby removing as much as feasible of the channel-induced ISI. After a successful training period, the output samples $\hat{s}[n]$ from the decision device are hopefully correct with such a high probability that they may assume the roles of the now absent training signal samples $t[n]$. Hence, an adequate error signal $e[n]$ can still be formed, and adaptation may proceed during regular data transmission in *decision-directed* (DD) mode [16, 14].

If initiated from a sufficiently good equalizer setting, DD adaptation is capable of tracking slow, moderate variations in the channel transfer function. However; drastic variations, such as those resulting from a deep *fade* [3] or a transmission outage, can cause symbol decisions $\hat{s}[n]$ to suddenly become too unreliable for successful DD operation. When this happens, the equalizer parameters must be re-positioned at a good setting before switching back to DD mode [3, 17]. This is because DD methods do not possess the important property of global convergence, meaning that an optimum setting is reached from any initialization. In cases where the receiver cannot rely on repeated training for this, or even on training altogether, a *blind* adaptation method is needed.

2.7 Equalization Without a Desired Signal

If a training signal or, more generally, a desired equalizer output is not available to guide equalizer adaptation, recovery of the source signal becomes more challenging. The objective of a blind method is to restore some known characteristic or statistical property

of the source in the equalizer output. For instance; in the digital communication example in Section 2.6, the transmitted signal is known to be i.i.d. and therefore an uncorrelated, or *white* sequence of samples. The channel H, not being a pure delay, imparts correlation (ISI) into the signal, making it colored (non-white). An appealing strategy might then be to tune the equalizer parameters to make the output signal as white as possible. The rationale behind this would be that restoring the known *second-order statistics* of the source, namely its autocorrelation, hopefully recovers the transmitted signal.

Unfortunately, a whitening filter is generally insufficient for blind equalization of a system H as it only equalizes the magnitude response $|H(\Omega)|$, and not the phase response $\angle H(\Omega)$ [18]. The reason for this is that the phase response of an LTI system is not preserved in the output second-order statistics from which 'whiteness' is defined. This can be seen from the expression of the power spectrum $P_x(\Omega)$ of $x[n]$ (i.e. the Fourier transform of the autocorrelation) in Fig. 2.1,

$$P_x(\Omega) = P_s(\Omega)|H(\Omega)|^2. \tag{2.9}$$

Substituting $H(\Omega)$ in (2.9) for $\widetilde{H}(\Omega) \triangleq H(\Omega)A(\Omega)$, where $A(\Omega)$ is an allpass filter with unit magnitude and arbitrary phase response, leaves $P_x(\Omega)$ unchanged. Hence, second-order statistics cannot distinguish between $H(\Omega)$ and $\widetilde{H}(\Omega)$, making the whitening filter non-unique.

The only exceptions to the insufficiency of whitening are when H is known to be either minimum- or maximum phase, so that a unique relationship exists between $|H(\Omega)|$ and $\angle H(\Omega)$ [12]. For general mixed-phase systems, blind equalization methods must be based on statistics of orders higher than two.

Overview of Blind Equalization Methods and Strategies

As concluded in Chapter 2, ordinary (non-blind) linear equalization commonly uses the MSE criterion, which is based on second-order statistics in the form of autocorrelations and cross-correlations. When the desired equalizer output is not available, second-order statistics are no longer sufficient. Instead, *higher-order statistics* need to be employed, where 'higher' means greater than two.

This chapter gives an overview of blind equalization methods and motivates their use of higher-order statistics. The important concept of global convergence is also discussed.

3.1 Higher-Order Moments

A foundation for many blind equalization methods is *higher-order moments*. The pth-order moment $\mathcal{M}^p$ of a random variable A is [13]

$$\mathcal{M}^p(A) \triangleq \mathrm{E}\{A^p\}. \tag{3.1}$$

For example, $\mathcal{M}^1(A) = \mathrm{E}\{A\}$ is the mean, and $\mathcal{M}^2(A) - (\mathcal{M}^1(A))^2 = \mathrm{E}\{A^2\} - (\mathrm{E}\{A\})^2$ the variance of A. A generalization is the pth-*order moment function* $\mathcal{R}_a$ of a real, stationary process $a[n]$, defined as [18]

$$\mathcal{R}_a[k_1, \dots k_{p-1}] \triangleq \mathrm{E}\big\{a[n]\, a[n+k_1] \cdots a[n+k_{p-1}]\big\}.$$

3.1.1 Blind System Identification with Higher-Order Moments

As seen earlier, the second-order moment function $\mathcal{R}_a[k] = \mathrm{E}\{a[n]\, a[n+k]\}$ is the autocorrelation of the process from which the power spectrum is obtained by taking the Fourier transform

$$\mathrm{P}_a(\Omega) \triangleq \sum_{k=-\infty}^{\infty} \mathcal{R}_a[k] e^{-j\Omega k}.$$

By extending this for the third-order moment function

$$\mathcal{R}_a[k_1,\, k_2] = \mathrm{E}\{a[n]\, a[n+k_1]\, a[n+k_2]\},$$

in a similar fashion the *bispectrum* of $a[n]$ is obtained from the two-dimensional Fourier transform [19]

$$\mathrm{B}_a(\Omega_1,\, \Omega_2) \triangleq \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} \mathcal{R}_a[k_1,\, k_2] e^{-j(\Omega_1 k_1 + \Omega_2 k_2)}.$$

Consider again the model in Fig. 2.1. As seen earlier, the power spectrum of the system output signal $x[n]$ given in (2.9) is strictly real, and so the phase response of H cannot be identified from $\mathrm{P}_x(\Omega)$. On the other hand, the bispectrum of $x[n]$ is [19]

$$\mathrm{B}_x(\Omega_1,\, \Omega_2) = \mathrm{B}_s(\Omega_1,\, \Omega_2) H(\Omega_1) H(\Omega_2) H^*(\Omega_1 + \Omega_2),$$

which is generally a complex function from which both the magnitude and phase of H can be identified if $\mathrm{B}_s(\Omega_1,\, \Omega_2)$ is known. Therefore, third-order statistics provide the possibility of blind identification of an LTI system from knowledge of the third-order statistics of the input and observations of the output. Similarly, while the shortcomings of $\mathrm{P}_x(\Omega)$ carry over to the time domain in $\mathcal{R}_x[k]$, the impulse response $h[n]$ of H is identifiable from the third-order moment function under certain conditions [20]. In fact; a closed-form expression for the impulse response $h[n]$ of an FIR system in terms of third-order statistics was given in [21] as

$$h[n] = \frac{\mathcal{R}_y[p, n]}{\mathcal{R}_y[-p, -p]}, \quad \text{for } n = 0,\, \ldots,\, p, \tag{3.2}$$

where $y[n]$ is the system output and p the FIR order. Unfortunately, (3.2) turns out to have limited practical use as the model order p must be known, and estimating the moments $\mathcal{R}_y[p, n]$ comes with large variances, even without noise present [20, 18, 22]. Nevertheless, it demonstrates the usability of third-order statistics for blind system identification, which can be generalized to higher orders. In [23], it was proven that an LTI system could be identified up to a scale factor and a delay from any higher-order statistic under broad conditions. In light of this and the relation between system identification and equalization mentioned in Section 2.2, it is not surprising that higher-order moments provide means of solving the blind equalization problem as well.

3.2 The Blind Equalization Model

The blind equalization model is shown in Fig. 3.1. The following assumptions are made:

- The source signal $s[n]$ is a real, zero-mean, stationary i.i.d. process, where each sample is a realization of a random variable S with known, non-Gaussian probability density function (pdf) ν_S.

- H is an LTI system with impulse response $h[n]$.

- The noise $b[n]$ is real, stationary and independent of $s[n]$.

- $u[n] = x[n] + b[n]$ is the observed signal.

- The linear equalizer F has impulse response $f[n]$.

- The equalizer output samples $y[n]$ are realizations of a random variable Y with pdf ν_Y.
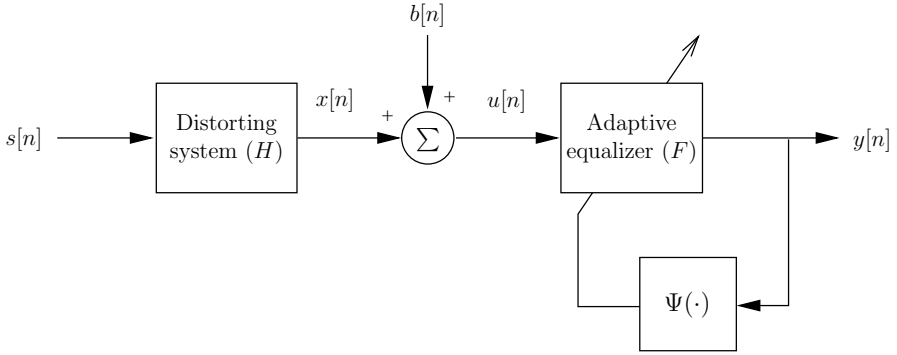
Figure 3.1: Model of the blind equalization problem.

In practice, the equalizer is usually adaptive of FIR type with coefficient vector $\mathbf{f}_n$. To guide the adaptation of $\mathbf{f}_n$, the *objective function* Ψ assumes the role of the MSE (2.4) which can no longer be formed in absence of training or a desired output. Note that for a given observation of the equalizer input, Ψ is a function of $\mathbf{f}_n$. The objective function is either minimized or maximized (depending on its design) in a gradient search

$$\mathbf{f}_{n+1} = \mathbf{f}_n \pm \mu \boldsymbol{\nabla}_\Psi(\mathbf{f}_n), \tag{3.3}$$

where μ is a small, positive stepsize and

$$\boldsymbol{\nabla}_\Psi(\mathbf{f}_n) \triangleq \frac{\partial \Psi}{\partial \mathbf{f}_n}$$

the gradient of Ψ with respect to $\mathbf{f}_n$. Using (3.3) with a positive sign is termed a *gradient ascent*, and with a negative sign a *gradient descent*. The *attraction points* of $\Psi(\mathbf{f}_n)$ are thus the maximum points of $\Psi(\mathbf{f}_n)$ in ascents, and the minimum points in descents.

Unlike the rather natural choice of the MSE (2.4) for trained equalizers, many different objective functions have been proposed and analyzed for various blind problems. The choice of Ψ usually has a significant impact on the performance of the method. The next section gives an overview of some classical approaches in digital communications.

3.3   Blind Equalization Methods in Digital Communications

In the adaptive equalization example in Section 2.6, decision-directed adaptation was used to track a time-varying channel once a sufficiently good initial setting had been supplied with the use of a training signal. DD equalization was introduced in 1966 by Lucky [16]. Although it is generally not regarded as a blind method, the objective function of decision-directed LMS (DD-LMS) is obtained by replacing the desired signal $d[n] = s[n - \Delta]$ in the MSE expression (2.4) with the detected symbols $\hat{s}[n - \Delta]$ from the decision device,

$$\Psi_{\mathrm{DD}} \triangleq \mathrm{E}\left\{(\hat{s}[n - \Delta] - y[n])^2\right\}. \tag{3.4}$$

Minimization of Ψ_{DD} minimizes average squared deviations of $y[n]$ from the detected symbols and equals MMSE equalization using (2.4) if all symbol decisions are correct. In [17], it was proven that optimum convergence of DD-LMS is guaranteed in absence of noise under *open-eye* conditions, i.e. if the algorithm is initiated from a state in which the residual ISI level is low enough to give correct symbol decisions. If initialized in a closed-eye state, optimum convergence cannot be guaranteed. Therefore, it is reasonable to request from a blind equalizer the ability to always reach an open-eye state regardless of initialization, so that a switch to DD mode can safely be made with convergence towards an MMSE setting to follow under mild channel variations [24].

Addressing the poor performance of (3.4) under closed-eye conditions, a modified version was proposed by Sato in 1975 [25]. Sato's method uses the sign of the decision device outputs, instead of the actual outputs, to guide adaptation. It is therefore more robust to decision errors and recovers from more adverse initial conditions. On the other hand, the convergence rate is quite slow [7]. Sato's algorithm was further analyzed by Benveniste, Goursat and Ruget [26], who showed that it belonged to a broad class of objective functions suitable for blind equalization. Benveniste and Goursat later proposed a method featuring a 'smooth automatic switching' from the Sato objective to DD-LMS mode [27], combining the robustness of the Sato method with the attractive convergence properties of decision-directed equalizers.

An important contribution was made in 1980 by Godard [28] who was looking for objective functions that measured ISI on quadrature amplitude modulated (QAM) signals independent of carrier phase, to achieve equalization independently from carrier phase recovery. Godard suggested minimization of 'dispersions of order p', with the pth-order dispersion defined as

$$\mathcal{D}^{(p)} \triangleq \mathrm{E}\left\{(|y[n]|^p - \gamma_p)^2\right\},$$

where γ_p is a positive constant essentially controlling the equalizer gain. Godard mainly analyzed the cases $p = 1$ and $p = 2$ as higher orders gave poor performance. Expanding $\mathcal{D}^{(2)}$ gives

$$\mathcal{D}^{(2)} = \mathrm{E}\left\{|y[n]|^4\right\} - 2\gamma_2 \mathrm{E}\left\{|y[n]|^2\right\} + \gamma_2^2. \tag{3.5}$$

Note that the first and second terms of (3.5) contain the fourth and second moments respectively of the output signal magnitude. Hence, this rather heuristic objective function

turns out to be based on higher-order moments.

The same objective was suggested by Treichler and Agee [29] for frequency modulation (FM) and quadrature phase-shift keying (QPSK). Their motivation to using (3.5) was to restore the 'constant modulus' property of such signals which is altered under ISI.

Minimization of $\mathcal{D}^{(2)}$ through a stochastic gradient descent is also known as the *constant modulus algorithm* (CMA), and is the most widely explored blind method in digital communications, especially when applied to fractionally spaced (oversampled) equalization (FSE) [30]. Despite its name, the use of CMA is not limited to constant modulus signals, as was indeed demonstrated by Godard in his original proposal [28]. Two exhaustive tutorials on CMA with emphasis on FSE are found in [31, 30]. A very attractive feature, early observed by both Godard [28] and Treichler/Agee [29], is that the resulting parameter settings at convergence lie close to Wiener settings. In [32, 33], an upper bound on the MSE of CMA was established, confirming that CMA equalizers do indeed give good performance in terms of MSE, and therefore in bit-error rate.

Noting that the popular constant modulus objective (3.5) is based on higher-order moments, their use in blind equalization is now further motivated.

3.4 Matching Probability Densities

Consider the model in Fig. 3.1. Temporarily assume the noise $b[n]$ to be zero and define the convolution

$$t[n] \triangleq h[n] * f[n] \tag{3.6}$$

as the *global system response*, i.e. the combined impulse response of H and F. For a blind equalizer, perfect equalization in the noiseless case is defined to hold when

$$y[n] = \pm\alpha s[n - \Delta], \tag{3.7}$$

where α is a positive scale factor and Δ an integer delay. The sign, scale and delay ambiguities are normally inherent in the problem but rarely of any practical concern [7]. In terms of the global response, (3.7) corresponds to

$$t[n] = \pm\alpha\delta[n - \Delta], \tag{3.8}$$

where $\delta[n]$ is the Kronecker delta function[1]. The following observation was made by Benveniste, Goursat and Ruget [26]. Under the assumptions listed in Section 3.2, $b[n]$ being zero and the source pdf ν_S being symmetric:

$$\text{if } \nu_Y = \nu_S, \text{ then } t[n] = \pm\delta[n - \Delta]. \tag{3.9}$$

In other words, if the pdf's of the source and equalizer output signal are equal, F inverts H up to a sign and a delay ambiguity.

The implication of (3.9) is that adjusting the equalizer parameters until the output samples have the same pdf as the source solves the blind equalization problem. The role

---

[1]$\delta[n] = 1$ if $n = 0$, and 0 if $n \neq 0$.

of the objective function Ψ should therefore be to distinguish the source pdf ν_S from that of a filtered version. Under certain conditions, the pdf of a random variable is uniquely determined by its moments (3.1) of all orders [13]. Hence, higher-order moments make suitable building blocks for objective functions. Furthermore, (3.9) suggests that only the marginal distribution of Y is of interest. Therefore, Ψ can be a memoryless function of $y[n]$ [27].

An important remark is that (3.9) does *not* hold if the source pdf ν_S is Gaussian, in which case ν_Y is also Gaussian for *any* $t[n]$ [13]. The pdf of a Gaussian random variable is completely characterized by its first and second moments [13], and so higher-order moments cannot be used to distinguish between different filtered versions of Gaussian signals of equal mean and variance. Therefore, the blind equalization problem generally has no solution when the source is Gaussian.

While not being allowed in the source signal, 'Gaussianity' provides another handle on the blind equalization problem, as discussed next.

3.5   Minimizing Gaussianity

A different approach started with a paper by Wiggins [9], who deconvolved seismic traces by maximizing an objective function he called the 'Varimax norm',

$$\Psi_{\text{Wig}} \triangleq \frac{\dfrac{1}{N}\displaystyle\sum_{n=1}^{N} y[n]^4}{\left(\dfrac{1}{N}\displaystyle\sum_{n=1}^{N} y[n]^2\right)^2}, \tag{3.10}$$

which can be recognized as a sample estimate of the *kurtosis* $\mathcal{K}$, or normalized fourth moment[2] of Y,

$$\mathcal{K}(Y) \triangleq \frac{\mathcal{M}^4(Y)}{(\mathcal{M}^2(Y))^2}.$$

As illustrated in Fig. 1.5 in Chapter 1, seismic deconvolution seeks to undo time dispersion in recorded traces to restore the 'spiky' appearance of an underlying reflection sequence. Wiggins therefore looked for an objective function quantifying the amount of time dispersion of impulses ('spikes') in a trace, and found that (3.10) had this feature. He termed his heuristic method *Minimum Entropy Deconvolution* (MED).

In a later contribution, Donoho [10] provided a statistical foundation to MED, explaining why Wiggins' method generally was successful. Donoho argued that, by the central limit theorem [13], a filtered version of a non-Gaussian i.i.d. process appears 'more Gaussian' than the source itself. The principle is demonstrated in Fig. 3.2, where a sequence of independent samples with equally probable values ± 1 is filtered through

---

[2]Alternative definitions of kurtosis exist, such as $\mathcal{M}^4(Y) - 3(\mathcal{M}^2(Y))^2$ [18] or $\mathcal{M}^4(Y)/(\mathcal{M}^2(Y))^2 - 3$ [34].

a single-pole filter. Comparing the histograms of input and output signals, a heuristic increase in Gaussianity is apparent.

(a) Source signal.

(b) Histogram of source.

(c) Filter output signal.

(d) Histogram of filter output with superimposed Gaussian pdf curve.
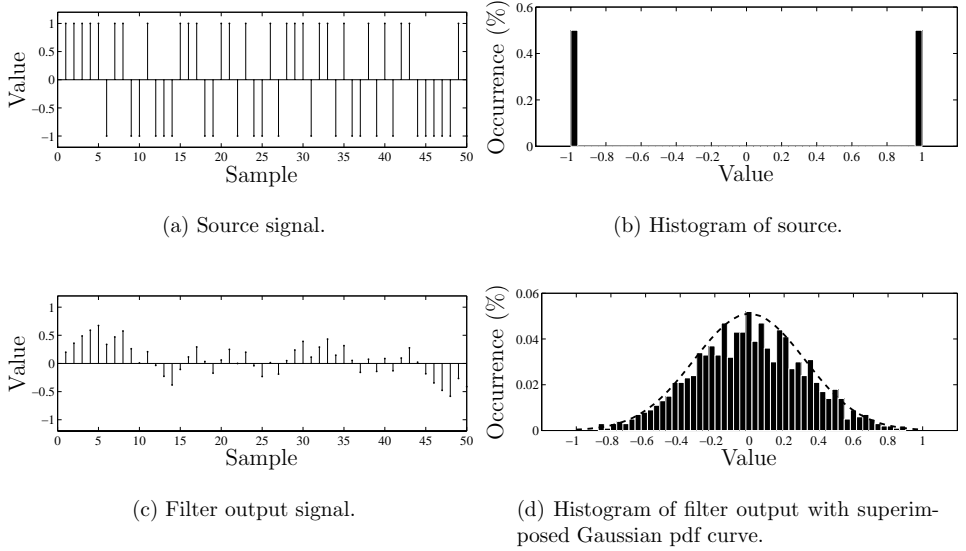
Figure 3.2: Example of the increase in Gaussianity due to filtering. The i.i.d. source signal (a) with histogram (b) is filtered through a single-pole filter, yielding the output (c) with histogram (d). A Gaussian pdf curve is superimposed on the output histogram.

Using an objective function to measure the Gaussianity of a random variable, Donoho suggested adjusting the equalizer parameters to minimize the Gaussianity of the output. He further showed that Wiggins' objective (3.10) does indeed qualify for this task, and hence justified its use in practice. He also concluded that general higher-order statistics can be used to reflect the amount of Gaussianity of a random variable.

3.6   The Popularity of Fourth-Order Moments

Both strategies of matching probability densities (Section 3.4) and minimizing Gaussianity (Section 3.5) can be adopted with the use of higher-order moments. The first strategy, making $\nu_Y = \nu_S$, generally involves the use of all moments. On the other hand, the works of Wiggins and Donoho show that it is actually sufficient to only use the fourth and second moments. A similar claim was later made by Shalvi and Weinstein [35], who proposed maximization of the fourth moment under constrained equalizer norm.

While generalizations to other orders are possible [36, 34], fourth-order methods have

dominated the application field and research area. The main reasons for their popularity
are the following:

- **The use of odd-order moments is more restrictive:** All odd-order moments
  of a *symmetric signal*, i.e. a signal having a symmetric marginal pdf, are zero [37].
  This is illustrated in Fig. 3.3 for third moments and continuous pdf's. If an i.i.d.
  signal is symmetric, then so is any filtered variant of it. Hence, filtering does not
  affect the odd moments of a symmetric source, which therefore precludes their use in
  this case. On the other hand, as shown in Fig. 3.3, an *asymmetric* source (having an
  asymmetric marginal pdf), have non-zero odd moments which are generally altered
  under filtering. Therefore, asymmetric sources allow the use of *either* odd or even
  moments. However, symmetric sources are more common, for example in digital
  communication where almost all modulation types result in symmetric densities [3].
  One reason for the traditional dominance of fourth-order methods is hence their
  versatility [20, 18].

- **Orders > 4 provide no added benefits:** Moments of orders higher than four do
  not give any extra information and are more difficult to estimate accurately from
  noisy observations [38, 34].

- **The popularity of CMA:** Much of the evolution of blind methods has been
  driven by a need to improve the bit-rate performance of digital communication
  receivers. With CMA being the blind equalization method of choice for modern
  QAM systems, the understanding of CMA behavior (and hence of fourth-order
  methods) has been, and continues to be of interest [5].
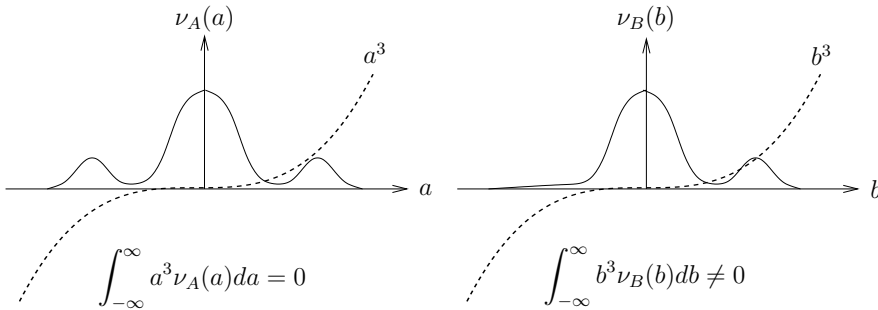
*Figure 3.3: Illustration of the third-order moments of a random variable A with symmetric pdf
$\nu_A(a)$, and a random variable B with asymmetric pdf $\nu_B(b)$. By definition, $\mathcal{M}^3(A) = \mathrm{E}\{A^3\} = \int_{-\infty}^{\infty} a^3 \nu_A(a) da$.*

A consequence of the historical bias towards fourth-order methods is that relatively
little research has investigated the use of *third* moments. Nevertheless, their selection
over fourth moments turns out to be well justified in the case of an asymmetric source.

This is the point of interest in this thesis, and the research contributions are summarized in Chapter 4 and supplied in detail in Part II.

3.7 Global Convergence

A highly desired feature of any objective function is *global convergence*, which guarantees that a gradient search converges towards a point of good equalizer performance from any initialization under ideal conditions. Objectives that do not possess a global convergence property should either be modified or combined with a suitable initialization strategy to avoid convergence to poor parameter settings.

Recall the well-behaved MSE surface in Fig. 2.2. The only attraction point is the single global minimum at the bottom of the bowl, corresponding to the Wiener setting with optimum MSE performance. As no attraction points of poor performance exist on this surface, the MSE objective (2.4) is globally convergent. In contrast, general objective surfaces of blind methods are not bowl-shaped but have several *stationary points* where the gradient vector is zero, i.e. maxima, minima and saddle points. Therefore, there may be several attraction points, and the initialization determines which one is reached (compare with dropping a marble into an egg box instead of a bowl). Global convergence requires all attraction points to correspond to settings of satisfactory performance.

The shape of a specific objective surface $\Psi(\mathbf{f})$ depends on the statistics of the source signal and the distorting system transfer function $H(\Omega)$. While the source statistics are usually assumed to be known, the system is not, and so assessment of 'good' and 'bad' settings in equalizer parameter space cannot be done without specifying $H(\Omega)$. On the other hand, such points are well defined in terms of the global system response $t[n]$ (3.6). Recall the condition (3.8) on $t[n]$ for perfect equalization. Assume the global response to be of finite duration and represented by the N-tap FIR coefficient vector

$$\mathbf{t} \triangleq \begin{bmatrix} t_0 & \cdots & t_{N-1} \end{bmatrix}^{\mathrm{T}}.$$

Formulating the condition (3.8) in terms of $\mathbf{t}$ gives the points of perfect equalization in $\mathbf{t}$-space as all single-element settings

$$\mathbf{t} = \pm \begin{bmatrix} 0 & \cdots & 0 & \alpha & 0 & \cdots & 0 \end{bmatrix}^{\mathrm{T}}, \tag{3.11}$$

where the position of the single nonzero element is determined by the delay Δ. Points in the form of (3.11) are also referred to as zero forcing (ZF) settings. An objective function having all its attraction points at ZF settings is hence globally convergent.

In practice, perfect equalization is not always pursued. For example; when an initialization to a decision-directed algorithm is to be supplied by a blind method (see Section 2.6), a small amount of residual distortion is acceptable. This means that attraction points at settings *approximately* in the form of (3.11) (e.g. having several small taps dominated by one large tap), could also be tolerated. On the other hand, attraction points *far* from ZF settings (e.g. with all taps equal) correspond to heavy residual distortion, and are not acceptable unless the probability of convergence to such points can be proven to be small.

Fig. 3.4 illustrates the location of some points at different coordinates (t_0, t_1, t_2) in global three-tap space. ZF settings lie on the coordinate axes, while points far from all axes correspond to heavy distortion.
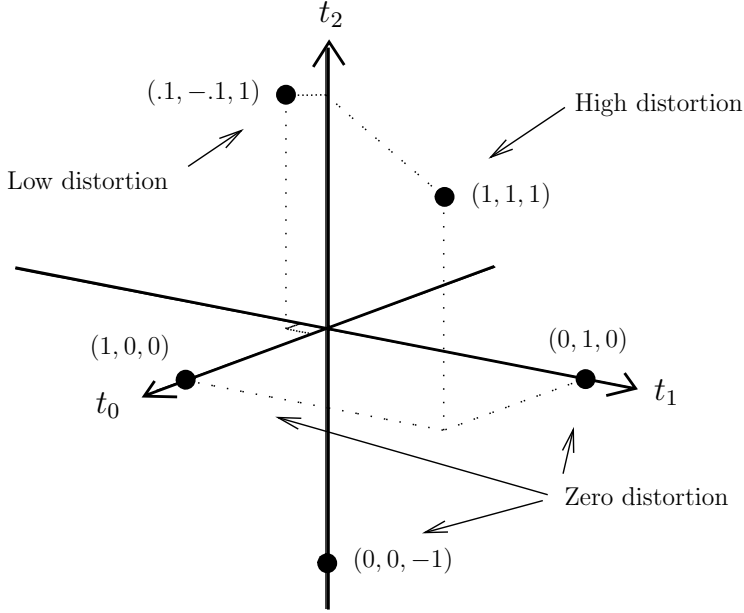
Figure 3.4: *The location of different settings (t_0, t_1, t_2) in three-tap global parameter space. Two positive ZF points, $(1, 0, 0)$, $(0, 1, 0)$, and one negative, $(0, 0, -1)$, are marked. The approximate ZF setting $(.1, -.1, 1)$ might be an acceptable attraction point in many cases, unlike the point $(1, 1, 1)$ which corresponds to maximum distortion.*

Research Contributions

"Man får inte vara rädd för lite kalkyler."

Reinhold Näslund

This chapter summarizes the research contributions of the thesis and discusses their impact. Detailed results are found in Papers A-E in Part II.

4.1 Focus Area: Third-Order Moments

The use of higher-order moments in blind equalization was introduced in Chapter 3. As discussed in Section 3.6, the research field has been dominated by investigation of fourth-order methods. Third-order methods were of early interest for blind system identification, but have since been somewhat left in the shadow. Out of the relatively few application examples, some worth mentioning are: [39, 40] (digital communications), [41] (ultrasonics), [42] (mechanical signals) and [43] (underwater acoustics). While these contributions employ third-order methods for different applications, little analytic results are presented to explain their performance. Derivations of the global convergence of objectives based on general higher-order moment magnitudes are given in [36]. Still, third-order methods are not as well understood as fourth-order methods, particularly CMA whose properties are well documented [31, 30].

The historical lack of attention to third-order methods is probably due to the somewhat restrictive requirement of an asymmetric source pdf, which precludes their use in most of the the digital communication area, where much of the blind equalization research has been conducted. The authors of [39, 40, 44] suggest using a transformation to enforce asymmetry into the symbol constellation, so that third-order methods can be used with symmetrical modulation formats. However, the effect of such a transformation on the probability of bit error is potentially harmful [40]. Interestingly, a modulation type being asymmetric by construction, *pulse-position modulation* (PPM) [45], has recently come into consideration for *ultra wideband (UWB) radio* [46, 47] where blind linear

equalization may prove useful to ensure high bit-rate operation [48]. Papers D and E consider third-moment based blind equalization of PPM signaling, while experimental results are also found in Papers B and C.

Motivated by a perceived knowledge gap in the understanding of third-moments in blind equalization, the research herein presents both analytic and numerical results to partially fill this gap. The methods of interest are on-line algorithms of low computational cost, e.g. stochastic-gradient searches as used in CMA, maximizing third-order moments ($\mathcal{M}^3$) of the equalizer output. To assess the general behavior of such methods, a characterization of the *objective surface topology* is valuable, i.e. determining the types and locations of all stationary points. This helps predict the performance of gradient-search methods as follows:

- If all attraction points correspond to good settings (e.g. points of perfect equalization), the objective function is globally convergent.

- The existence and location of any *false optimum*, i.e. an attraction point corresponding to a sub-optimum parameter setting, can aid in specifying an objective modification, an equalizer initialization strategy, or methods to detect ill-convergence.

- A large number of saddle points generally has a negative impact on convergence speed, as the slope of the objective function approaches zero in their vicinity, causing gradient searches to 'stall'. Convergence rate issues with CMA have been related to saddle points [49, 30].

4.2   Topology of the Third-Moment Surface

The surface topology of third-moment objectives is studied in equalizer parameter space in Papers A and B, and in global parameter space in Papers D and E. It is interesting to compare the topology of the third-moment surface with that of CMA, which is based on fourth moments and for which the performance and surface characteristics are well known [31, 30]. In [50], the stationary points of CMA are shown to lie on a manifold topologically equivalent to a sphere. For third-moment maximization, a norm constraint is enforced on the equalizer parameter vector in practice to avoid numerical instability (see Section 4.5). This effectively restricts all settings (and hence stationary points) to have constant norm, and hence reside on the sphere as well.

A visual comparison of the results for CMA from [50] with those for third moments in Paper E[1] is given in Figures 4.1 and 4.2 respectively for three-tap global system parameter vectors on the two-dimensional S_2 sphere. The curved arrows indicate the direction along the gradient vector field that adaptation trajectories would ideally take. Each stationary point is marked and labeled according to the number of attractive dimensions in the vector field as follows [51]:

---

[1]The results in Paper E are derived for PPM signaling, but generalize to i.i.d. sources as considered in [50].

- *Sink*, marked by (●). Two attractive, zero repulsive dimensions (attraction point).

- *Saddle point*, marked by (**s**). One attractive, one repulsive dimension.

- *Source*, marked by (○). Zero attractive, two repulsive dimensions (repulsive point).

The number of sinks, saddles and sources for both objectives are listed in Table 4.1. The following conclusions can be drawn from the table and a visual comparison of the spheres:

- Compared to the CMA surface, the third-moment surface has a 'simpler' structure with fewer stationary points and, specifically, fewer saddle points. The convergence rate should therefore generally be in favor of third-order methods.

- The stationary points of CMA are distributed all across the surface. In contrast, all stationary points on the third-moment sphere are confined to the two octants enclosed by dashed lines in Fig. 4.1. These distributions of points persist into higher dimensions. In general, for a length-N global response, the volume ratio of the two hyperoctants to the total parameter space is 2^{-N+1}. Hence, the simplicity of the third-moment surface compared to CMA becomes even more accentuated in practice, and enables faster convergence properties.

- While three of the attraction points lie at perfect equalization (ZF) settings, a single false optimum exists on the third-order surface (the sink at the back of the sphere in Fig. 4.1(b)). However, as established in Paper E, this is the only false optimum, regardless of dimension. Furthermore, the probability of ill-convergence proves to be very small in practice, as it only occurs if the starting point lies within the octant containing it. As discussed above, the size of this region practically vanishes in higher dimensions. Also, Paper E shows that the *character* of this region makes it an even more unlikely starting point in reality.

Table 4.1: *The number of stationary points on the S_2 sphere for the third moment (TM) and CMA objectives.*

Objective	Sinks	Saddles	Sources
TM	4	6	4
CMA	6	12	8

The benefits of a simple objective surface are demonstrated in numerical experiments in Papers A and B, showing that third-moment based gradient searches converge faster than corresponding fourth-moment versions.

(a) Front of third-moment sphere.   (b) Back of third-moment sphere.
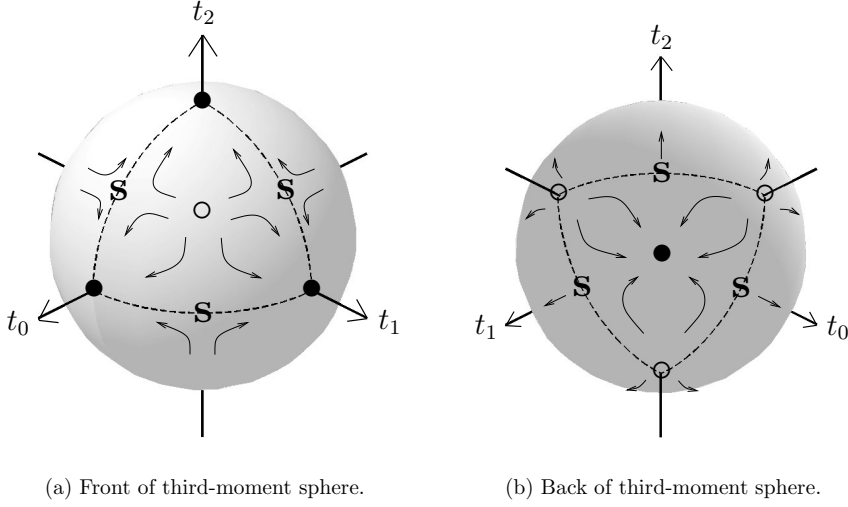
Figure 4.1: Vector field and stationary points of the third-moment objective surface on the sphere in three-tap global parameter space, ($\bullet$) = sink, ($\mathbf{s}$) = saddle point, ($\circ$) = source.

(a) Front of CMA sphere.   (b) Back of CMA sphere.
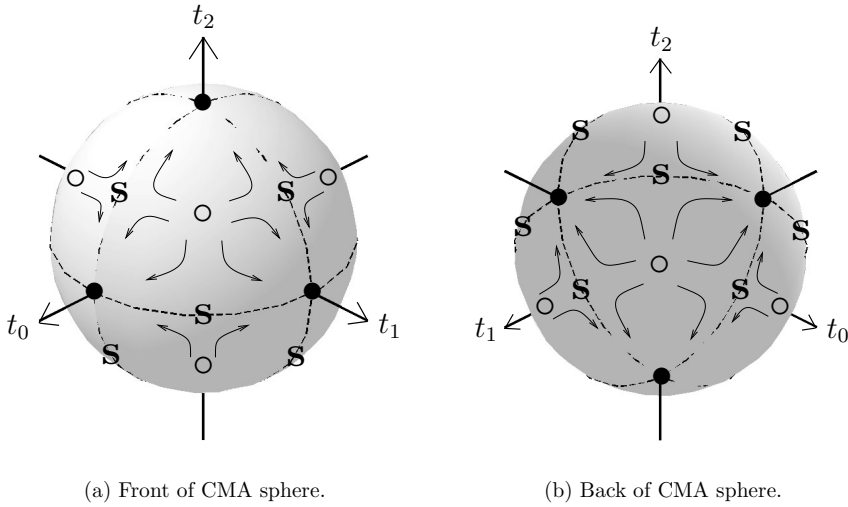
Figure 4.2: Vector field and stationary points of the CMA objective surface on the sphere in three-tap global parameter space, ($\bullet$) = sink, ($\mathbf{s}$) = saddle point, ($\circ$) = source.

4.3 Equivalent Polytope

Another interesting abstraction made in [50] is to relate the sphere-topology of CMA to a polytope by introducing the following associations:

- Sinks ↔ vertices (faces of dimension zero).

- Saddle points ↔ edges (faces of dimension one).

- Sources ↔ faces (faces of dimension two).

The polytope representation is possible since the topology of the objective surface places constraints on the gradient vector field and, therefore on the relative number of maxima, minima and saddle points [51]. The equivalent polytope of the three-tap third-moment surface, with four vertices, six edges and four faces (confirm with Table 4.1 and Fig. 4.1) is a tetrahedron. In higher dimensions, it generalizes to a hypertetrahedron, or simplex. The equivalent polytope of a three-tap CMA system, having six vertices, twelve edges and eight faces (see Table 4.1 and Fig. 4.2), is an octahedron, generalizing to a cross polytope in higher dimensions. The polytopes of the third moment and CMA objectives, situated in the same coordinate system as used for the spheres earlier, are shown in Fig. 4.3 and Fig. 4.4 respectively. Note how the sinks from Fig. 4.1 and Fig. 4.2 line up with the polytope vertices. The relative simplicity of the third-order objective surface topology is again visually apparent from comparing the tetrahedron to the octahedron.
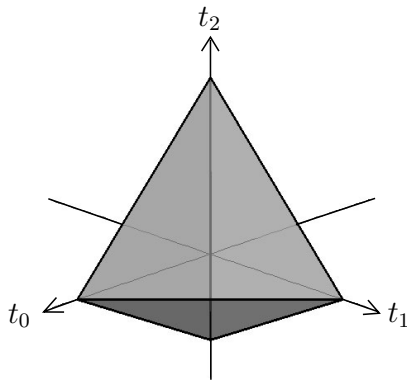
Figure 4.3: The topologically equivalent polytope (tetrahedron) of the three-tap third-moment surface.
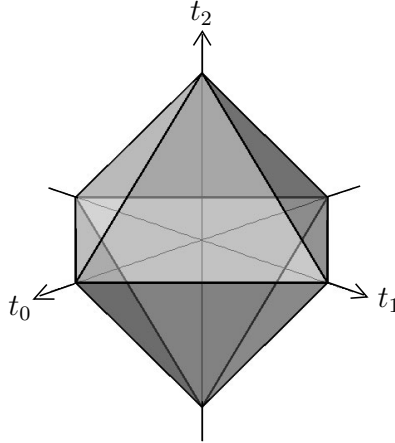
Figure 4.4: The topologically equivalent polytope (octahedron) of the three-tap CMA surface.

4.4   Robustness to Gaussian Noise

In Paper B, it is shown that, unlike fourth-moment stationary points, the third-moment stationary points are not displaced by additive white Gaussian noise. Hence, estimates of the third-moment gradient are not biased by white Gaussian noise, while the corresponding fourth-moment gradient estimates are. This inherent robustness to noise further emphasizes the advantages of third-order methods over their fourth-order counterparts, especially in lower-SNR settings.

4.5   A Low-Cost, Self-Stabilizing Gradient Search

Maximization of a third-moment objective can be done in a stochastic gradient search with a very low computational cost, comparable to that of an LMS or CMA algorithm. However, the equalizer parameter vector requires frequent normalization to avoid numerical instability. Such vector normalizations do not affect the equalization performance but require several additional multiply-adds, as well as divisions and square-root operations which are costly on a fixed-point architecture such as a digital signal processor (DSP). A division- and square-root free algorithm is proposed and analyzed in Paper C, providing third-moment maximization with automatically constrained equalizer norm. Hence, third-moment maximization lends itself to efficient implementation on real-time digital hardware.

4.6   Conclusions

The results outlined here show that third-order methods possess features which make them an interesting and viable alternative to the traditionally employed fourth-order methods in applications where the source signal is asymmetric. In fact, the benefits of faster convergence and better robustness to noise actually suggest that an asymmetric source should be *exploited* by choosing third-order methods.

Summary of Appended Papers

Paper A - Skewness Maximization for Impulsive Sources in Blind Deconvolution

A comparison between the objective surfaces of kurtosis and skewness (normalized third-order moment) is done in equalizer parameter space. It is shown that the skewness surface has fewer stationary points in general and therefore fewer saddle points, which would on average give faster convergence of gradient searches. A numerical experiment is supplied where both objectives are used to deconvolve recorded sound from a running diesel engine, for which the source signal has an asymmetric, impulsive character. The experiment confirms the faster convergence of the third-order method.

Paper B - Online Adaptive Blind Deconvolution Based on Third-Order Moments

Two simple objective functions are compared, the third- and fourth-order moments respectively of the equalizer output. The third-order objective is demonstrated to provide faster convergence on average. It is also shown that estimates of the third-moment gradient are unbiased under white Gaussian noise, while the fourth-order estimates are not. Equalization of a pulse-position modulated (PPM) signal on a simulated ultra wideband (UWB) radio channel is performed, highlighting the faster convergence of the third moment and demonstrating its use in on-line algorithms.

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

Blind equalization by maximizing third-order moments was demonstrated in Paper B to be viable method for on-line applications. However, a significant part of the computational cost in implementations comes from the frequent normalizations of the equalizer coefficient vector, which are necessary to maintain numerical stability. The extra computational cost of these normalizations may be quite large in practice, requiring division and square-root operations as well as several extra multiply-adds. In Paper C, a self-stabilizing algorithm is proposed and analyzed which gives third-moment maximization under constant filter norm, using only multiplications and additions. The computational

simplicity makes this algorithm well suited for realization on real-time, fixed-point hardware such as a digital signal processor (DSP). Experimental results are supplied, featuring blind equalization of a simulated UWB radio channel with PPM signaling.

Paper D - Blind Equalization of PPM Signals Using Third-Order Moments

PPM has attracted attention as being a suitable signaling format for UWB radio. A challenge in the deployment of UWB systems is how to ensure reliable high-rate transmissions in presence of the rich multipath spread of typical indoor radio channels. Paper D proposes the use of a blind, linear equalizer to combat the resulting ISI. Being asymmetric by construction, PPM signals allow the use of third-order moment maximization. Analysis of the global convergence properties under PPM signaling is supplied for a low-dimensional example, showing the existence of a false optimum point which signals the need for further study regarding the probability of ill-convergence and/or an initialization strategy.

Paper E - Blind Linear Equalization of PPM Signals Using Third-Order Moments

Third-moment maximization under PPM signaling is further investigated in Paper E. Global convergence analysis shows that a single false optimum point exists for arbitrary length of the global system response. However, the probability of ill-convergence is proven to be vanishingly small in practice, providing confidence in the usability of third-order methods. It is also demonstrated that maximization of the magnitude of third moments is a globally convergent method for PPM. A simulation experiment using a widely accepted UWB channel model demonstrates that third-order methods can be used to provide an initial setting of sufficiently low residual ISI for safe switch to a decision-directed algorithm.

REFERENCES

[1] J. G. Proakis and D. G. Manolakis, *Introduction to Digital Signal Processing.* New York, NY: Macmillan, 1988.

[2] J. E. Brittain, "The introduction of the loading coil: George A. Campbell and Michael I. Pupin," *Technology and Culture*, vol. 11, no. 1, pp. 36–57, Jan. 1970.

[3] J. G. Proakis, *Digital Communications*, 4th ed. New York, NY: McGraw-Hill, 2001.

[4] S. U. H. Qureshi, "Adaptive equalization," *Proc. IEEE*, vol. 73, no. 9, pp. 1349–1387, Sep. 1985.

[5] J. R. Treichler, M. G. Larimore, and J. C. Harp, "Practical blind demodulators for high-order QAM signals," *Proc. IEEE*, vol. 86, no. 10, pp. 1907–1926, Oct. 1998.

[6] S. Haykin, Ed., *Blind Deconvolution.* Englewood Cliffs, New Jersey: Prentice-Hall, 1994.

[7] S. Bellini, "Bussgang techniques for blind deconvolution and equalization," in *Blind Deconvolution*, S. Haykin, Ed. Englewood Cliffs, New Jersey: PTR Prentice Hall, 1994.

[8] P. Kearey and M. Brooks, *An Introduction to Geophysical Exploration*, 2nd ed. Oxford 0X2 0EL, England: Blackwell Scientific Publications, 1991.

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

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

[11] D. Kundur and D. Hatzinakos, "Blind image deconvolution," *IEEE Signal Process. Mag.*, vol. 13, no. 3, pp. 43–64, May 1996.

[12] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing.* Englewood Cliffs, New Jersey: Prentice-Hall, 1989.

[13] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, 2nd ed. Singapore: McGraw-Hill, 1984.

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

[15] 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.

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

[17] O. Macchi and E. Eweda, "Convergence analysis of self-adaptive equalizers," *IEEE Trans. Inf. Theory*, vol. IT-30, no. 2, pp. 161–176, Mar. 1984.

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

[19] C. L. Nikias and M. R. Raghuveer, "Bispectrum estimation: A digital signal processing framework," *Proc. IEEE*, vol. 75, no. 7, pp. 869–891, Jul. 1987.

[20] J. M. Mendel, "Tutorial on higher-order statistics (spectra) in signal processing and system theory: Theoretical results and some applications," *Proc. IEEE*, vol. 79, no. 3, pp. 278–305, Mar. 1991.

[21] G. B. Giannakis, "Cumulants: A powerful tool in signal processing," *Proc. IEEE*, vol. 75, no. 9, pp. 1333–1334, Sep. 1987.

[22] A. K. Nandi, "Blind identification of FIR systems using third order cumulants," *Signal Processing*, vol. 39, pp. 131–147, Sep. 1994.

[23] K. S. Lii and M. Rosenblatt, "Deconvolution and estimation of transfer function phase and coefficients for nongaussian linear processes," *The Annals of Statistics*, vol. 10, no. 4, pp. 1195–1208, Dec. 1982.

[24] Z. Ding, C. R. Johnson, Jr., and R. A. Kennedy, "On the admissibility of blind adaptive equalizers," in *Acoustics, Speech, and Signal Processing, 1990. Proceedings. (ICASSP-90). IEEE International Conference on*, vol. 3, Albuquerque, NM, Apr. 1990, pp. 1707–1710.

[25] Y. Sato, "A method of self-recovering equalization for multilevel amplitude-modulation systems," *IEEE Trans. Commun.*, vol. COM-23, no. 6, pp. 679–682, Jun. 1975.

[26] 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. Autom. Control*, vol. AC-25, no. 3, pp. 385–399, Jun. 1980.

[27] A. Benveniste and M. Goursat, "Blind equalizers," *IEEE Trans. Commun.*, vol. COM-32, no. 8, pp. 871–883, Aug. 1984.

[28] 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.

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

[30] C. R. Johnson, Jr., P. Schniter, I. Fijalkow, L. Tong, J. D. Behm, M. G. Larimore, D. R. Brown, R. A. Casas, T. J. Endres, S. Lambotharan, A. Touzni, H. H. Zeng, M. Green, and J. R. Treichler, "The core of FSE-CMA behavior theory," in *Unsupervised Adaptive Filtering, Volume II: Blind Deconvolution*, S. Haykin, Ed. New York, NY: Wiley Interscience, 2000.

[31] C. R. Johnson, Jr., P. Schniter, T. J. Endres, J. D. Behm, D. R. Brown, and R. A. Casas, "Blind equalization using the constant modulus criterion: A review," *Proc. IEEE*, vol. 86, no. 10, pp. 1927–1950, Oct. 1998.

[32] H. H. Zeng, L. Tong, and C. R. Johnson, Jr., "Relationships between the constant modulus and Wiener receivers," *IEEE Trans. Inf. Theory*, vol. 44, no. 4, pp. 1523–1538, Jul. 1998.

[33] H. H. Zeng, L. Tong, and C. R. Johnson, Jr., "An analysis of constant modulus receivers," *IEEE Trans. Signal Process.*, vol. 47, no. 11, pp. 2990–2999, Nov. 1999.

[34] J. A. Cadzow, "Blind deconvolution via cumulant extrema," *IEEE Signal Process. Mag.*, vol. 13, no. 3, pp. 24–42, May 1996.

[35] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)," *IEEE Trans. Inf. Theory*, vol. 36, no. 2, pp. 312–321, Mar. 1990.

[36] O. Shalvi and E. Weinstein, "Universal methods for blind deconvolution," in *Blind Deconvolution*, S. Haykin, Ed. Englewood Cliffs, New Jersey: PTR Prentice Hall, 1994.

[37] M. R. Raghuveer, "Third-order statistics: Issue of PDF asymmetry," *IEEE Trans. Signal Process.*, vol. 43, no. 7, pp. 1736–1738, Jul. 1995.

[38] D. Hatzinakos and C. L. Nikias, "Blind equalization based on higher-order statistics (H.O.S.)," in *Blind Deconvolution*, S. Haykin, Ed. Englewood Cliffs, New Jersey: PTR Prentice Hall, 1994.

[39] F.-C. Zheng, S. McLaughlin, and B. Mulgrew, "Blind equalisation of multilevel PAM series via higher-order cumulant fitting," in *IEEE International Conference on Communications 1991 (ICC '91) Conference Record*, vol. 3, Denver, CO, Jun. 1991, pp. 1393–1397.

[40] F.-C. Zheng, S. McLaughlin, and B. Mulgrew, "Blind equalization of nonminimum phase channels: Higher order cumulant based algorithm," *IEEE Trans. Signal Process.*, vol. 41, no. 2, pp. 681–691, Feb. 1993.

[41] A. K. Nandi, D. Mämpel, and B. Roscher, "Blind deconvolution of ultrasonic signals in nondestructive testing applications," *IEEE Trans. Signal Process.*, vol. 45, no. 5, pp. 1382–1390, May 1997.

[42] J. Y. Lee and A. K. Nandi, "Blind deconvolution of impacting signals using higher-order statistics," *Mechanical Systems and Signal Processing*, vol. 12, no. 2, pp. 357–371, Mar. 1998.

[43] M. K. Broadhead, L. A. Pflug, and R. L. Field, "Use of higher order statistics in source signature estimation," *J. Acoust. Soc. Am.*, vol. 107, no. 5, pp. 2576–2585, May 2000.

[44] F. B. Ueng and Y. T. Su, "Adaptive blind equalization using second- and higher order statistics," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 1, pp. 132–140, Jan. 1995.

[45] J. M. Kahn and J. R. Barry, "Wireless infrared communications," *Proc. IEEE*, vol. 85, no. 2, pp. 265–298, Feb. 1997.

[46] 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.

[47] L. Yang and G. B. Giannakis, "Ultra-wideband communications: An idea whose time has come," *IEEE Signal Process. Mag.*, vol. 21, no. 6, pp. 26–54, Nov. 2004.

[48] A. G. Klein, C. R. Johnson, Jr., and P. Duhamel, "On blind equalization of biorthogonal signaling," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1421–1435, Apr. 2007.

[49] S. Lambotharan, J. Chambers, and C. R. Johnson, Jr., "Attraction of saddles and slow convergence in CMA adaptation," *Signal Processing*, vol. 59, no. 3, pp. 335–340, Jun. 1997.

[50] J. P. LeBlanc and C. R. Johnson, Jr., "Global CMA error surface characteristics, source statistic effects: Polytopes and manifolds," in *Digital Signal Processing Proceedings, 1997. DSP 97., 1997 13th International Conference on*, vol. 1, Santorini, Greece, Jul. 1997, pp. 131–134.

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

Part II

Skewness Maximization for Impulsive Sources in Blind Deconvolution

Authors:
Patrik Pääjärvi and James P. LeBlanc

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
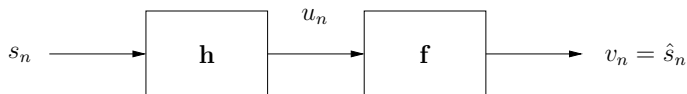
Figure 1: Block diagram of a deconvolution problem.

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

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

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,

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

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 adaptations.

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 mth 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 multimodal (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 mth filter coefficient, f_m, and equating to zero, we obtain the following; for $m, i = 0 \ldots N$,

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

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$^{\text{nd}}$ and 4$^{\text{th}}$ moments of u_n are defined as

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

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 mth 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$^{\text{rd}}$ moment of u_n is defined as

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

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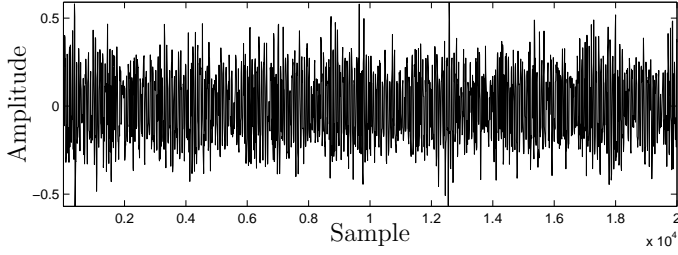
Figure 2: Measurement data from a sound recording of a diesel engine.

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.
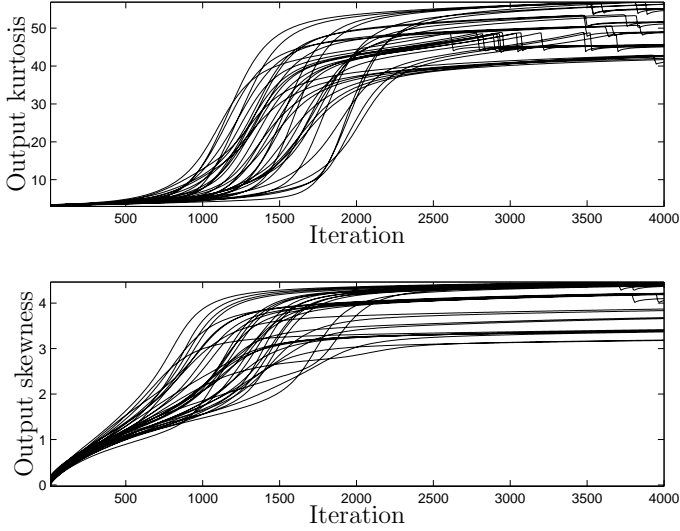
Figure 3: Kurtosis (top) and skewness (bottom) versus filter iteration.
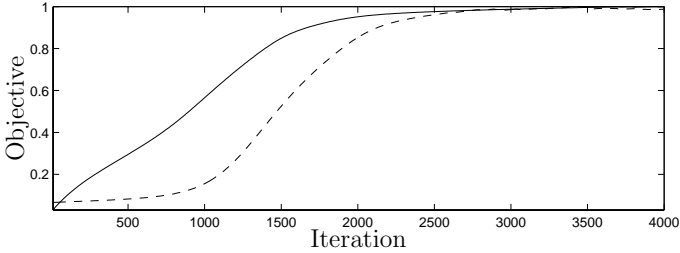
Figure 4: Kurtosis (dashed) and skewness (solid) versus filter iteration.

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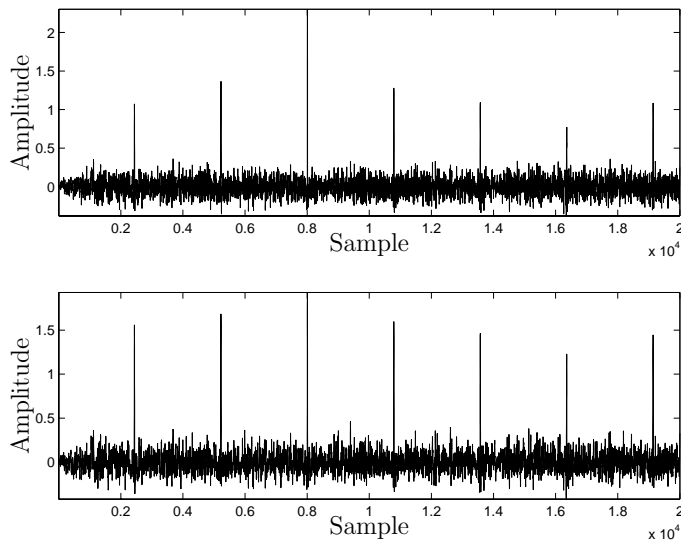
Figure 5: Kurtosis (top) and skewness (bottom) deconvolution filter outputs.

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 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.

Kurtosis: 8 minima (×), 12 saddles (s), 6 maxima (•)
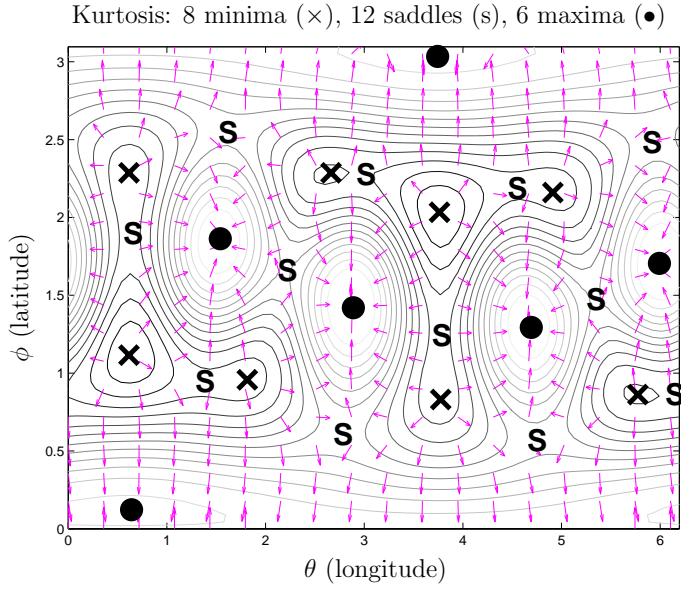
Figure 6: Kurtosis error surface.

Skewness: 4 minima (×), 6 saddles (s), 4 maxima (•)
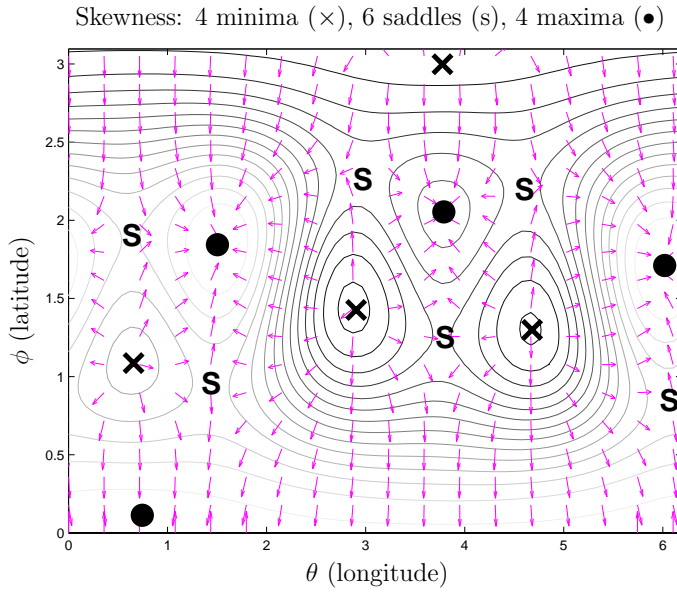
Figure 7: Skewness error surface.

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, Jul. 2003.

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

Online Adaptive Blind Deconvolution Based on Third-Order Moments

Authors:
Patrik Pääjärvi and James P. LeBlanc

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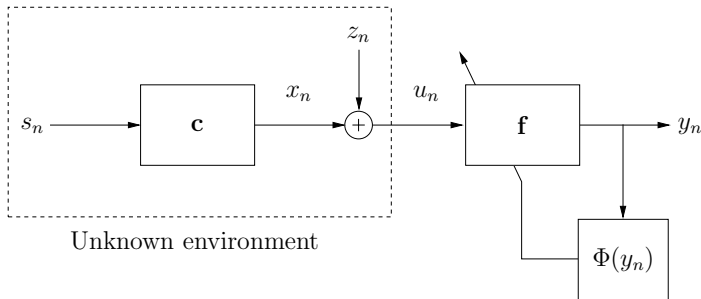
Figure 1: The signal model of a general blind deconvolution problem.

Adaptive blind deconvolution is used for equalization or identification of unknown systems when only the output of the system can be observed. Figure 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 $\mathbf{c}$ or the source signal s_n. The system output x_n plus an additive disturbance z_n gives the observed signal u_n. The un-

known, possibly time-varying system **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 **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 **f**. In general, **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

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

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 Figure 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

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

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 constant.

3   Comparative Performance Analysis of 3RD- and 4TH-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

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

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

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

for $m, i, j, k = 0 \ldots N$, with the fourth moment of u_n defined as $\mathcal{R}_{\substack{i \\ k}}^{\,i} = E\{u_n u_{n-i} u_{n-j} u_{n-k}\}$. 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 is 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 UWB 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

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

with $\tau_m = 0.2333$, giving a pulse duration of about one nanosecond. The pulse shape is shown in Figure 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 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. Figure 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.
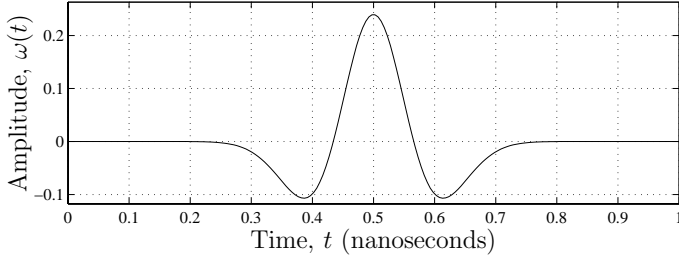
Figure 2: The pulse shape $\omega(t)$ used in the experiment. $\omega(t) = \left[1 - 4\pi(t/\tau_m)^2\right] \exp\left[-2\pi(t/\tau_m)^2\right]$, with $\tau_m = 0.2333$.
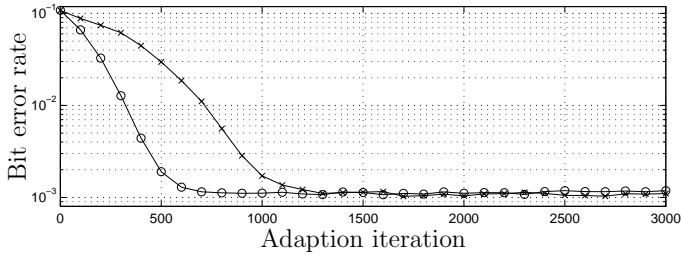
Figure 3: Bit error rate versus adaption iteration for third-moment ($\circ$) and fourth-moment ($\times$) objective functions, averaged over 13 runs. SNR per bit = 11dB.

As seen in Figure 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 Process.*, 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., Sep. 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, Jul. 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, Jun. 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, Jun. 2004, pp. 2648–2652.

[11] V. Lottici, A. D'Andrea, and U. Mengali, "Channel estimation for ultra-wideband communications," *IEEE J. Sel. 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. Sel. 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

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 the filter norm.

1   Introduction
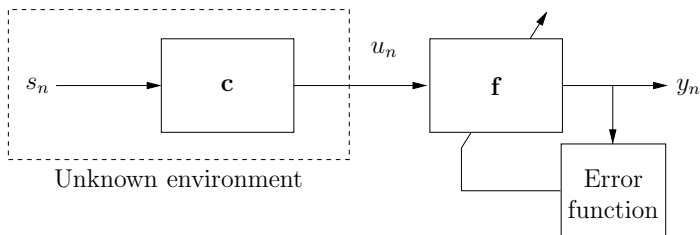
Figure 1: Model of general adaptive blind deconvolution.

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 iteratively 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 **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 based on 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 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 necessarily being a white sequence.

If s_n is known to have a non-zero third-order 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 moment

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

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 the stochastic variable x. While easier to estimate than skewness, (2) is not scale invariant in x. That is, $\mathcal{O}(x) \neq \mathcal{O}(kx)$ for $k \neq 1$. As a consequence, standard gradient-ascent algorithms based on (2) give a rapid increase in filter norm over iterations. In fact, such problems 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 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* $\|\mathbf{f}_n\|$ 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 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 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 $\alpha\mathbf{f}$, $\alpha > 1$. Therefore (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 Nth-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 αy_n of the filter output signal; the 'quality' of deconvolution is not affected. 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 $\mathbf{f}_n$,

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

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

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

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 Nth-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

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

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, maintaining 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 Nth-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 next section considers 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

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

The necessary conditions 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.

While the assumptions of time-invariance in (17) and (18) are not realistic over a larger span of iterations (in fact, note that S_y is proportional to the objective function to be maximized), (17) and (18) 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

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

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, $S_y = \mathrm{E}\left\{y_n^3\right\}$ 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 $\mathbf{f}_n$ maximizing 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 and a sampling rate of 15 samples per nanosecond. Equalizers of order $N = 400$ were generated using (11) for three different stepsizes, each 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.

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 $\|\mathbf{f}_n\|$.
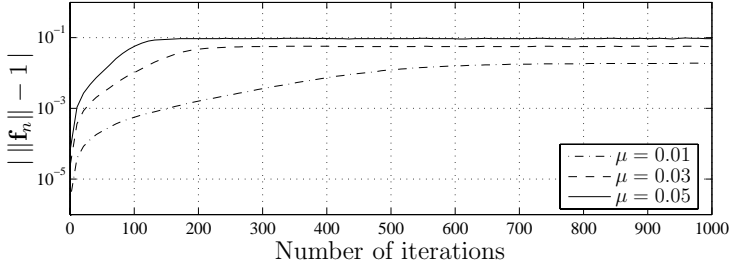
Figure 2: Absolute deviation of $\|\mathbf{f}_n\|$ from unity versus iteration number.
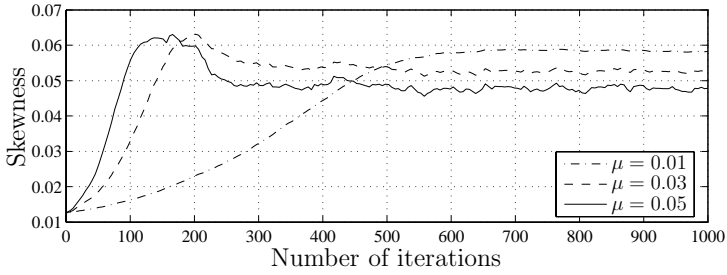
Figure 3: Estimated skewness of filter output versus iteration.

6   Conclusion

A computationally efficient algorithm for norm-constrained gradient acsent has been studied for blind deconvolution of asymmetric source signals using a third-order moment based objective function. The results indicate that the algorithm maintains approximately unit filter norm for reasonable choices of adaptation stepsize. The condition on adaptation stepsize insuring a stable filter norm is trivial to calculate and verify. 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, Jun. 2004, pp. 304–307.

[4] P. Pääjärvi and J. P. LeBlanc, "Online adaptive blind deconvolution based on third-order moments," *IEEE Signal Process. Lett.*, vol. 12, pp. 863–866, Dec. 2005.

[5] H. Mathis and S. C. Douglas, "Bussgang blind deconvolution for impulsive signals," *Proc. IEEE*, vol. 51, no. 7, pp. 1905–1915, Jul. 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 Process.*, vol. 48, no. 6, pp. 1843–1847, Jun. 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, Jul. 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, Jun. 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. Sel. Areas Commun.*, vol. 20, no. 6, pp. 1247–1257, Aug. 2002.

Paper D

Blind Equalization of PPM Signals Using Third-Order Moments

Authors:
Patrik Pääjärvi and James P. LeBlanc

Blind Equalization of PPM Signals Using Third-Order Moments

Patrik Pääjärvi and James P. LeBlanc

Abstract

In pulse-position modulation (PPM) signaling, the time location of short-duration pulses are used to convey information over a communication channel. For successful noncoherent reception, the channel duration must be short compared to the symbol interval. This paper analyzes the use of third moments in a blind adaptive equalizer setting to limit the effective delay spread of the channel. Results detail the global convergence properties of the proposed method, showing that the parameters approach ISI-free settings under general conditions.

1   Introduction
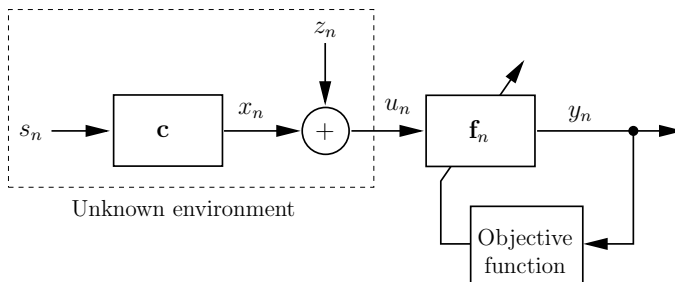
Figure 1: Discrete-time blind adaptive equalization model.

Pulse-position modulation (PPM) signaling is based on transmitting short-duration pulses of constant magnitude, with the information being conveyed in the time location of the individual pulses. While historically mainly used in optical communication, the recent evolution of ultra-wideband (UWB) radio has spawned renewed interest in PPM as a proposed signal for wideband, carrier-free wireless communication over short range [1, 2].

Recent work on low-complexity noncoherent receivers for PPM show promising results in cases where the transfer rate is low compared to the maximum delay spread of the multipath channel [3]. With increasing demands for higher data rates, it becomes

83

necessary to limit the effects of excessive channel length at the receiver. This paper proposes the use of a blind adaptive equalizer to allow noncoherent reception of high-rate PPM signals without estimating channel coefficients.

Blind adaptive equalization is used to reduce intersymbol interference (ISI) over a dispersive channel in situations where sending a training sequence is not possible. The target of adaptation is to remove enough ISI to provide a good initial setting for decision-directed (DD) equalization. With reliable estimates of the transmitted symbols, DD equalization can quickly move parameters close to the minimum mean squared error (MMSE) setting [4].

The principle of blind equalization is shown in Figure 1. The channel $\mathbf{c}$ filters the transmitted source s with added noise z, forming the observed (received) signal u. Adjustment of the equalizer $\mathbf{f}_n$ is governed by an *objective function* in the form of a non-linear function of the output y. If the source signal is a sequence of i.i.d. samples, the objective function may be chosen from several broad classes [5]. *Higher-order moments* (order > 2) of y are often used since they are easy to estimate from instantaneous sample values. The pth moment of a random variable ξ is defined as

$$\mathcal{M}_p(\xi) \triangleq \mathrm{E}\left\{\xi^p\right\}.$$

Fourth moments are a popular choice and form the basis of the widely known constant modulus algorithm (CMA) [6, 7, 8].

When s is not an i.i.d. sequence, the objective function should reflect some characteristic of s to be emphasized in y. *Asymmetric* signals, i.e. signals characterized by positive *skewness* (normalized third moment), allow the use of third moments in blind equalization. While having applications in various other fields [9, 10, 11], third moments are rarely considered in digital communication since almost all modulation types yield *symmetric* signals with zero skewness, prohibiting the use of odd moments. PPM signals, however, are a notable exception, being asymmetric by construction. Previous studies have demonstrated that exploiting asymmetry by choosing third moments over fourth moments, when possible, generally leads to faster adaptation and improved robustness to noise [12].

While fourth-order methods, especially CMA, have been subject of much study in the past [8], the details of third-order methods have attracted less attention. This paper examines global convergence properties of a third-moment based blind equalization scheme under PPM signaling. It is shown that the proposed method is insensitive to white Gaussian noise, and moves the parameters of a sufficiently long equalizer close to ISI-free settings.

2 Notation and Model Description

2.1 M-PPM Signal

An M-PPM signal ς is made up of consecutive *symbol frames* of M samples ($M \geq 3$). To each frame, one of M equally probable symbols $\{a_k\}_{k=1\ldots M}$ is mapped by assigning unity

to the kth sample and zero to the remaining $\text{M} - 1$ samples. An example with $\text{M} = 4$ and the three consecutive symbols a_2, a_4 and a_1 starting at frame N would give

$$\varsigma = \{ \cdots \; \underbrace{0\,1\,0\,0}_{N} \; \underbrace{0\,0\,0\,1}_{N+1} \; \underbrace{1\,0\,0\,0}_{N+2} \; \cdots \}.$$
$$\textit{frame:}$$

The transmitted sequence s is then formed by subtracting off the mean to achieve a DC-free signal, $s \triangleq \varsigma - \text{E}\{\varsigma\} = \varsigma - \frac{1}{\text{M}}$. The autocorrelation of s is

$$\text{E}\{s_n s_{n+m}\} = \begin{cases} \dfrac{\text{M}-1}{\text{M}^2}, & m = 0 \\[2mm] \dfrac{|m|}{\text{M}^3} - \dfrac{1}{\text{M}^2}, & |m| = 1 \dots \text{M}-1 \\[2mm] 0, & |m| \geq \text{M}. \end{cases}$$

From the construction of ς, samples within the same frame are dependent, while samples from distinct frames are independent. Hence, s is correlated up to lags $\text{M} - 1$. The skewness of s is

$$\frac{\text{E}\{s^3\}}{(\text{E}\{s^2\})^{3/2}} = \frac{\text{M}-2}{\sqrt{\text{M}-1}},$$

which is non-zero if $\text{M} \geq 3$, and increases strictly with increasing M. Restricting M to be greater than or equal to three therefore guarantees s to be an asymmetric signal.

In practice, an enhanced PPM signaling format incorporating pseudo-random time-hopping would be used in multiple-access systems [1]. The simple form of s will be used in this paper, since the addition of time-hopping can be approximately modeled by a larger M. The influence of M on the performance of the proposed blind equalization method is investigated in Sections 4 and 5.

2.2   Channel and Equalizer

Referring to Figure 1, the adaptive equalizer $\mathbf{f}_n$ is FIR of length N_f, denoted at time n by the $N_f \times 1$ column vector $\mathbf{f}_n \triangleq [f_{0n} \; f_{1n} \; \cdots \; f_{(N_f-1)n}]^{\text{T}}$. The impulse response of the channel $\mathbf{c}$ is assumed to have the substantial part of its energy concentrated to the N_c first samples. It is therefore modeled as an FIR system, represented by the $N_c \times 1$ vector $\mathbf{c} \triangleq [c_0 \; c_1 \; \cdots \; c_{N_c-1}]^{\text{T}}$, with c_j being the jth sample of the impulse response. $\mathbf{c}$ represents the aggregate response of the transmitter's pulse-shaping filter, the wireless channel and possible receiver filters. Also, define the $N_f \times N_t$ channel convolution matrix

$$\mathbf{C} \triangleq \begin{bmatrix} c_0 & \cdots & c_{N_c-1} & & \\ & \ddots & & \ddots & \\ & & c_0 & \cdots & c_{N_c-1} \end{bmatrix}$$

where $N_t = N_c + N_f - 1$, and the $N_t \times 1$ combined channel-equalizer response vector

$$\mathbf{t}_n \triangleq \mathbf{C}^{\text{T}} \mathbf{f}_n. \tag{1}$$

3   Blind Equalization Strategy

Perfect equalization is said to hold if no ISI is present in y after equalization. In practice, a sufficient requirement is

$$y_n \approx \alpha s_{n-\delta} \tag{2}$$

for some scale factor α and delay δ. Adaptation may switch to decision-directed mode once enough ISI has been removed during the initial adaptation. This is the reason why strict equality is not necessary in (2).

The equalizer under investigation is iteratively adjusted to maximize the third moment of y through the gradient ascent algorithm

$$\mathbf{f}_{n+1} = \mathbf{f}_n + \mu \widehat{\boldsymbol{\nabla}}_n, \tag{3a}$$

$$\mathbf{f}_{n+1} \leftarrow \frac{\mathbf{f}_{n+1}}{\|\mathbf{f}_{n+1}\|} \tag{3b}$$

where μ is a small stepsize, $\| \cdot \|$ is the Euclidean norm and $\widehat{\boldsymbol{\nabla}}_n$ is some instantaneous estimate of the gradient $\boldsymbol{\nabla}$ of $\mathcal{M}_3$ with respect to $\mathbf{f}_n$,

$$\boldsymbol{\nabla} \triangleq \frac{\partial \mathcal{M}_3(y)}{\partial \mathbf{f}_n}$$

The normalization (3b) is necessary to keep the algorithm numerically stable; since increasing the amplitude of y by 'stretching' $\mathbf{f}_n$ increases $\mathcal{M}_3(y)$ without improving equalization. Without (3b), the norm of $\mathbf{f}_n$ grows rapidly over iterations.

3.1   Why use $\mathcal{M}_3$?

A necessary condition in order for the algorithm (3) to be successful in equalizing channels with M-PPM signals, is that a high value of the objective $\mathcal{M}_3(y)$ corresponds to low levels of ISI in y. While the next section will show that this is indeed the case, some rationales for using third moments are described.

- With s asymmetric, the channel output x will appear more symmetric if $\mathbf{c}$ is non-trivial, due to 180° phase shifts from reflections. Therefore, adapting $\mathbf{f}_n$ to **promote asymmetry** in y should recover s.

- s has an impulsive appearance (i.e. its pdf has a positive heavy tail). Such impulsiveness is characterized by large positive skewness. The observed signal u generally appears less impulsive (with small or zero skewness). The objective function should hence **promote impulsiveness** in y.

3.2   Topological View of Equalizer Adaptation

Due to the normalization step (3b), we can view the adaptation algorithm (3) as traversing the $(N_f - 1)$-dimensional unit sphere in search of maximum points, i.e. points where

∇ vanishes and the Hessian is negative definite. With $\mathcal{M}_3$ being a cubic function of $\mathbf{f}_n$, the objective becomes a multimodal surface on the sphere. Optimization over such surfaces has been widely explored for CMA (see for example [8] and references therein). Some general observations are worth mentioning.

- There is usually no single global maximum point, but several local maxima with equal ISI reduction performance. With μ small, the length and initialization of $\mathbf{f}_n$ determines which one is approached.

- 'False' or 'bad' local maxima may also exist where adaptation can get stuck at high ISI parameter settings. These may be avoided by choosing a 'good' initial setting of $\mathbf{f}_n$.

- Adaptation may get stalled near saddle points (points with zero gradient and indefinite Hessian) during adaptation.

- Additive noise can perturb the objective surface, moving maxima away from their noise-free locations.

- Compared to i.i.d. sources, temporal correlation in s may move desired maxima and even give rise to new, unwanted ones. This effect has previously been observed for CMA [13].

3.3 Convergence in Combined Channel-Equalizer Space

Blind equalization by maximizing $\mathcal{M}_3(y)$ will be considered justified if (2) is obtained at maximum points. However, with the channel being unknown, generally such equalizer settings can not be formulated. Following the style of [14] and many others, we instead examine the objective surface in combined channel-equalizer space.

With $\mathbf{t}_n \in \mathbb{R}^{N_t}$, the points corresponding to zero ISI are

$$\mathbf{t}_n = \mathbf{e}_\delta \triangleq [0 \cdots 0 \, \alpha \, 0 \cdots 0]^{\mathrm{T}},$$

with the non-zero α occurring at element δ. This would give $y_n = \alpha s_{n-\delta}$. Perfect equalization is thus achieved if (3) converges to $\mathbf{e}_\delta$ for some $1 \leq \delta \leq N_t$. As stated earlier, (2) is usually sufficient so that convergence to a neighborhood of $\mathbf{e}_\delta$ will do.

It should be pointed out that; unless $\mathbf{f}_n$ is infinitely long or fractionally spaced (oversampled), not all of $\mathbf{t}_n$-space ($\mathbb{R}^{N_t}$) can be reached for an arbitrary channel. So it may not be possible in practice to reach $\mathbf{e}_\delta$ for all values of δ. However, it is commonly argued that if the equalizer is 'long enough', there are at least some $\mathbf{e}_\delta$ to which adaptation may converge [14].

4 Objective Surface Analysis

In this Section, the global convergence properties of the proposed blind equalization algorithm are investigated. First, it is shown that the objective function is invariant to
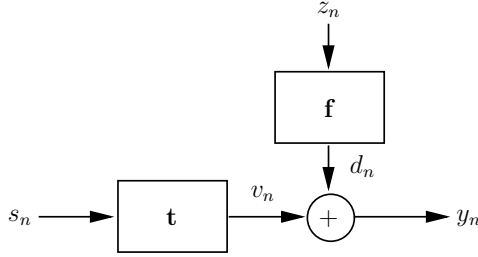
Figure 2: Combined channel-equalizer model.

white Gaussian noise. Second, equations for the stationary points on the objective surface are derived for arbitrary lengths of combined channel-equalizer response $\mathbf{t}_n$. Finally, the objective surface is visualized for a three-tap $\mathbf{t}_n$-vector to gain some insight into the behavior of the method.

4.1 Assumptions

The following assumptions are made throughout the remainder.

A1) All sequences and vectors are real.

A2) The channel noise z is zero-mean, white and Gaussian, independent on s.

A3) The adaptation stepsize μ is small, so that $\mathbf{f}_n$ may be regarded as approximately constant compared to all signals.

A4) The channel is assumed to be slowly varying compared to the time index n. Thus, $\mathbf{C}$ is also regarded as constant.

A5) $\mathbf{f}_n$ has infinite length.

Under Assumptions A3 and A4, $\mathbf{f}_n$ and $\mathbf{t}_n$ are hereafter regarded as constant vectors, denoted by $\mathbf{f}$ and $\mathbf{t}$ respectively. A5 simplifies analysis and is standard in analysis of blind systems (see for example [14]).

Keeping in mind the practical limitations on $\mathbf{f}$ mentioned in Section 3.3, the combined channel-equalizer model in Figure 2 is introduced for purpose of analysis. The noise d is z filtered through $\mathbf{f}$. With z white and Gaussian, d will be colored Gaussian noise, independent on the noise-free channel output v.

4.2 Noise Invariance of $\mathcal{M}_3$

First, we demonstrate that the third moment is unaffected by white Gaussian noise. $\mathcal{M}_3(y)$ expressed in terms of v and d is

$$\mathcal{M}_3(y) = \mathrm{E}\left\{(v+d)^3\right\} = \mathrm{E}\left\{v^3 + d^3 + 3v^2d + 3vd^2\right\}.$$

With v being zero-mean, and with d zero-mean, Gaussian and independent of v, we have

$$\mathcal{M}_3(y) = \mathrm{E}\left\{v^3\right\}.$$

The objective surface is not perturbed by z, and therefore white Gaussian noise does not introduce any bias in gradient estimates used in the adaptation algorithm (3a).

4.3   Stationary Points for General N_t

Equations for the stationary points of the objective surface are now derived for arbitrary lengths of $\mathbf{t}$. As seen above, the noise d may be disregarded in the following.

With the combined channel-equalizer response $\mathbf{t}$ defined in (1), let

$$\mathbf{s}_n \triangleq \left[s_n \, s_{n-1} \, \cdots \, s_{n-N_t+1}\right]^{\mathrm{T}}$$

so that $y_n = v_n = \mathbf{t}^{\mathrm{T}}\mathbf{s}_n = \mathbf{f}^{\mathrm{T}}\mathbf{C}\mathbf{s}_n$. Expressing the third moment of y_n in terms of s_n yields $\mathcal{M}_3(y_n) = \mathrm{E}\left\{y_n^3\right\} = \mathrm{E}\left\{\left(\mathbf{f}^{\mathrm{T}}\mathbf{C}\,\mathbf{s}_n\right)^3\right\}$. The stationary points of the objective surface with respect to the equalizer parameters are found by solving

$$\boldsymbol{\nabla} = \frac{\partial \mathcal{M}_3(y_n)}{\partial \mathbf{f}} = \mathrm{E}\left\{y_n^2\mathbf{C}\mathbf{s}_n\right\} = \mathbf{C}\mathrm{E}\left\{\left(\mathbf{t}^{\mathrm{T}}\mathbf{s}_n\right)^2\mathbf{s}_n\right\} = \mathbf{0} \tag{4}$$

where Assumption A4 ($\mathbf{C}$ constant) has been used[1]. Invoking Assumption A5 in the case where the channel has no spectral nulls (i.e. an inverse exists), (4) has only the trivial solution

$$\mathrm{E}\left\{\left(\mathbf{t}^{\mathrm{T}}\mathbf{s}_n\right)^2\mathbf{s}_n\right\} = \mathbf{0}. \tag{5}$$

This implies that there is a one-to-one correspondence between stationary points in $\mathbf{f}$-space and $\mathbf{t}$-space.

Solving (5) for the stationary points is possible since the statistics of s are known. However, as pointed out in Section 3, $\mathcal{M}_3(y)$ only attains maxima if the norm of $\mathbf{f}$ is restricted. Hence, $\|\mathbf{t}\|$ will be similarly constrained. Such a restriction causes no practical or theoretical limitations, since if $\mathbf{t}^*$ is a stationary point of $\mathcal{M}_3$, then so is $k\mathbf{t}^*$ for any $k \in \mathbb{R}$. We therefore solve (5) for $\mathbf{t}$ constrained to the manifold

$$S_{N_t-1} \triangleq \left\{\mathbf{t} \in \mathbb{R}^{N_t} \mid \|\mathbf{t}\|^2 = 1\right\},$$

i.e. the unit sphere in $\mathbb{R}^{N_t}$. Using the method of Lagrange multipliers, solving (5) under the constraint $\mathbf{t} \in S_{N_t-1}$ leads to

$$\mathrm{E}\left\{\left(\mathbf{t}^{\mathrm{T}}\mathbf{s}_n\right)^2\mathbf{s}_n\right\} + 2\lambda\mathbf{t} = \mathbf{0}, \tag{6}$$

where $\lambda \in \mathbb{R}$ is a constant Lagrange multiplier. Expanding (6), and using the assumption of $\mathbf{t}$ being constant, leads to the following system of N_t non-linear equations in the N_t

---

[1]A constant scale factor not affecting the results has been left out for notational simplicity.

Table 4.1: Stationary points $(t_0\,t_1\,t_2)$ on S_2 for large M.

Class	Solution subset	Type	Legend
C_1^+	$\{(1\,0\,0)\,(0\,1\,0)\,(0\,0\,1)\}$	Max	$*$
C_2^+	$\frac{1}{\sqrt{2}}\{(1\,1\,0)\,(1\,0\,1)\,(0\,1\,1)\}$	Saddle	$\triangle$
C_3^+	$\frac{1}{\sqrt{3}}\{(1\,1\,1)\}$	Min	$\circ$
C_1^-	$\{(\text{-}1\,0\,0)\,(0\,\text{-}1\,0)\,(0\,0\,\text{-}1)\}$	Min	$\bullet$
C_2^-	$\frac{1}{\sqrt{2}}\{(\text{-}1\,\text{-}1\,0)\,(\text{-}1\,0\,\text{-}1)\,(0\,\text{-}1\,\text{-}1)\}$	Saddle	$\triangledown$
C_3^-	$\frac{1}{\sqrt{3}}\{(\text{-}1\,\text{-}1\,\text{-}1)\}$	Max	$\star$

parameters $\{t_m\}_{m=0\ldots N_t-1}$:

$$\sum_{i=0}^{N_t-1} \mathcal{R}(0, i-m)\, t_i^2 + \sum_{\substack{i,j=0 \\ (j\neq i)}}^{N_t-1} \mathcal{R}(i-j, i-m)\, t_i t_j + 2\lambda t_m = 0, \tag{7}$$

where $\mathcal{R}(p,q) \triangleq \mathrm{E}\{s_n s_{n+p} s_{n+q}\}$ are generalized third moments of s. With s being an M-PPM signal, these moments are functions of M.

While solving (7) analytically is difficult (even for moderate values of N_t); in the limit of large M, $\mathcal{R}(0,0)$ becomes the dominating monomial coefficient and (7) approaches

$$\mathcal{R}(0,0)\, t_m^2 + 2\lambda t_m = 0, \tag{8}$$

for $m = 0 \ldots N_t - 1$. Since all N_t equations of the system (8) share the same monomial coefficients, finding all $2^{N_t+1} - 2$ solutions for $\mathbf{t} \in S_{N_t-1}$ is straightforward. Specifically, the N_t points $\mathbf{e}_\delta$ of perfect equalization are indeed solutions.

Although no $\mathbf{e}_\delta$ is a solution to the 'true' system (7), the approximation for large M suggests that stationary points exist in neighborhoods of the ISI-free points. This is confirmed in an experiment in Section 5 along with the following low-dimensional example.

4.4   Stationary Points for $N_t = 3$

An improved understanding of the algorithm can be gained by visualizing the objective surface for $N_t = 3$ with $\mathbf{t}$ on the two-dimensional unit sphere S_2. The $2^4 - 2 = 14$ solutions to the approximated system (8) are divided into six disjoint classes $\{C_N^+, C_N^-\}_{N=1,2,3}$, where C_N^+ (C_N^-) is the subset of solutions with N positive (negative) non-zero taps. The stationary points are listed in Table 4.1 with indicated type (max / min / saddle). Also in the table is a legend connecting the points to Figure 3, showing a contour plot of the objective surface on S_2 for M=32 along with gradient arrows to illustrate its curvature. From the contour plot, the following observations are made:

- The approximated (in the limit of large M) stationary points marked in the plot all lie very close to the true ones.
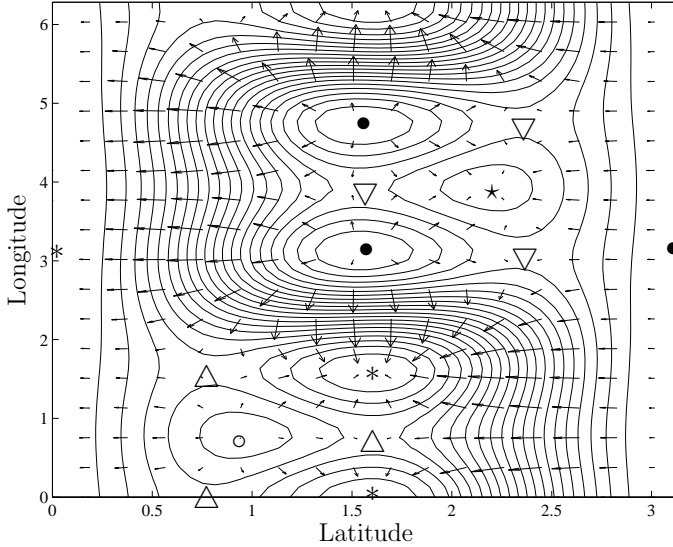
Figure 3: Objective surface on S_2 for M=32. Stationary points are marked by class membership according to Table 4.1.

- The points of perfect equalization ($*$) are global maxima.

- A single 'bad' local maximum point ($\star$) corresponding to severe ISI exists.

This low-dimensional example indicates that equalization of M-PPM signals by maximizing $\mathcal{M}_3$ does indeed lead to ISI reduction. However, the existence of the 'bad' local maxima ($\star$) shows that adaptation could converge to a point of very poor performance, which places emphasis on initialization of the equalizer or misconvergence detection capabilities.

5 Numerical Experiment

A numerical experiment was done to support the results from Section 4 suggesting that local maxima of the proposed method coincide with the points of perfect equalization $\mathbf{t} = \mathbf{e}_\delta$ in the limit of large M.

M-PPM signals for different values of M between 3 and 64 were generated. For each signal, a gradient ascent over a 100-tap $\mathbf{t}$-vector was performed using a small stepsize of $\mu = 5 \times 10^{-4}$, starting at $\mathbf{t} = \mathbf{e}_{50}$. After 10^6 iterations (enough for all trajectories to reach convergence), the Euclidean distance from the starting point to the estimated convergence point (obtained by averaging over the final 10^4 $\mathbf{t}$-vectors) was calculated. This provides a measure of how far $\mathbf{t} = \mathbf{e}_{50}$ is from the nearest local maxima for different M.
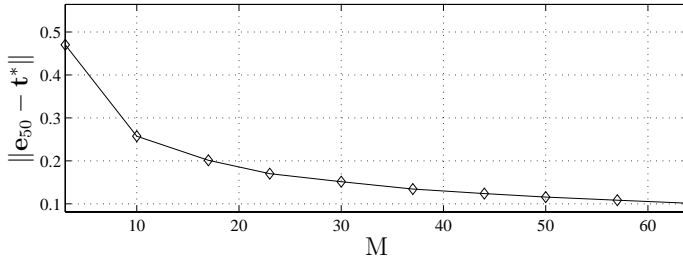
Figure 4: Euclidean distance from $\mathbf{e}_{50}$ to nearest local maxima $\mathbf{t}^*$ versus M.

The result is shown in Figure 4, where the distance from a point of perfect equalization to the nearest local maxima decreases as M increases.

6   Conclusions

This paper shows the usefulness of third-order moments in blind adaptive equalization with M-PPM signaling, by demonstrating that attraction points exist close to zero-ISI settings. The invariance to white Gaussian noise was shown, and the proximity of convergence points to perfect equalization points was demonstrated in a numerical experiment. The existence of undesired (or 'false') local maxima has also been noted, which indicates a need for an initialization or problem detection strategy.

References

[1] 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.

[2] L. Yang and G. B. Giannakis, "Ultra-wideband communications: An idea whose time has come," *IEEE Signal Process. Mag.*, vol. 21, no. 6, pp. 26–54, Nov. 2004.

[3] C. Carbonelli and U. Mengali, "M-PPM noncoherent receivers for UWB applications," *IEEE Trans. Wireless Commun.*, vol. 5, no. 8, pp. 2285–2294, Aug. 2006.

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

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

[6] 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.

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

[8] C. R. Johnson, Jr., P. Schniter, T. J. Endres, J. D. Behm, D. R. Brown, and R. A. Casas, "Blind equalization using the constant modulus criterion: A review," *Proc. IEEE*, vol. 86, no. 10, pp. 1927–1950, Oct. 1998.

[9] M. K. Broadhead, L. A. Pflug, and R. L. Field, "Use of higher order statistics in source signature estimation," *J. Acoust. Soc. Am.*, vol. 107, no. 5, pp. 2576–2585, May 2000.

[10] J. Y. Lee and A. K. Nandi, "Blind deconvolution of impacting signals using higher-order statistics," *Mechanical Systems and Signal Processing*, vol. 12, no. 2, pp. 357–371, Mar. 1998.

[11] A. K. Nandi, D. Mämpel, and B. Roscher, "Blind deconvolution of ultrasonic signals in nondestructive testing applications," *IEEE Trans. Signal Process.*, vol. 45, no. 5, pp. 1382–1390, May 1997.

[12] P. Pääjärvi and J. P. LeBlanc, "Online adaptive blind deconvolution based on third-order moments," *IEEE Signal Process. Lett.*, vol. 12, pp. 863–866, Dec. 2005.

[13] 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.

[14] 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. Autom. Control*, vol. AC-25, no. 3, pp. 385–399, Jun. 1980.

Blind Linear Equalization of PPM Signals Using Third-Order Moments

Authors:
Patrik Pääjärvi and James P. LeBlanc

Reformatted version of paper submitted to:
IEEE Transactions on Signal Processing, March 4, 2008.

Blind Linear Equalization of PPM Signals Using Third-Order Moments

Patrik Pääjärvi and James P. LeBlanc

Abstract

Pulse position modulation (PPM) has evolved from a signaling format mainly used in optical communications, to becoming a candidate transmission type for indoor wireless communication, following the development of ultra-wideband radio. Enhanced with time-hopping schemes, the robustness of PPM to multipath fading and multiple-access interference has been noted. One major challenge is how to ensure reliable symbol detection in presence of the rich multipath propagation that plagues typical indoor channels and causes intersymbol interference (ISI). In this work, a blind linear equalizer is used to combat ISI without the need for training data. Exploiting the asymmetry of the PPM signal, parameter adaptation is based on third-order moment maximization, which provides improved convergence rate and noise insensitivity compared to traditional fourth-order methods. It is shown that a stochastic gradient search with low complexity is globally convergent to zero-ISI settings. Analysis takes into account the practical conditions of time-hopped PPM and channel noise.

1 Introduction

Pulse position modulation (PPM) is a digital signaling format traditionally used in optical communication [1], in which information is conveyed in the time position of very narrow pulses of equal shape and polarity. Following the advent of ultra-wideband (UWB) radio, PPM has attracted attention as being a candidate format for carrier-free, high-speed wireless communication over short range. Enhanced with time-hopping (TH) techniques, PPM signals are robust to both multipath fading and multiuser interference [2, 3, 4].

Assuming that transmitter-receiver synchronization has been achieved, symbol detection is accomplished by estimating the time-of-arrival of received individual pulses. For the AWGN channel with insignificant dispersion, the maximum-likelihood (ML) detector compares the received signal energy in each of the possible time slots constituting a symbol, and selects the largest [5, 6]. Such non-coherent receivers are attractive considering their low complexity. Unfortunately, a realistic wireless channel will be far from dispersion-free, making non-coherent PPM receivers useful only at low bit rates [4, 7, 8, 6]. For example, the multipath spread of a typical UWB indoor channel may range up to around 200 nanoseconds [7, 8], causing both intrasymbol and intersymbol interference (whose combined effect is hereafter referred to as ISI) at higher bit rates. While offering optimum performance in the presence of ISI, the ML sequence detector requires channel

estimates and generally becomes far too complex, directing focus to suboptimal receivers [9].

A RAKE receiver could be used to exploit the multipath, regarding it as channel diversity. However, for a typical channel with rich multipath, a RAKE becomes very complex since a large number of fingers is potentially needed for sufficient energy capture [10]. Another option is to use an estimate of the received pulse shape (transmitted waveform convolved with the channel response) as a template in a correlation receiver. Estimating the pulse shape can be accomplished using transmit reference (TR) schemes [11] or training sequences [12], though a time-varying channel with a long multipath spread makes it difficult to accurately estimate and store such a template. Furthermore, information throughput is reduced by both TR schemes and training sequences as they occupy a part of the available transmission time. Adopting the approach of attempting to reduce ISI to ease symbol detection, decision feedback equalizers (DFEs) have been proposed, [9, 13], in which the cyclostationary nature of the PPM signal is exploited in a MIMO structure with blocks of parallel DFEs.

Neglecting cyclostationarity and instead focusing on ease of implementation, this work considers the use of a single, linear equalizer to combat channel dispersion. In this context, a *blind* adaptive equalizer is highly attractive since it does not need a training sequence for adjusting its parameters, allowing the receiver to track slow channel variations or recover from abrupt changes without sacrificing information throughput.

The performance, reliability and complexity of a blind equalizer largely depends on the choice of adaptation method and objective function. A common approach that can provide low complexity comparable to that of the LMS algorithm[14], is to optimize a function based on higher-order moments of the equalizer output through a stochastic gradient search. The pth-order moment of a random variable ξ is defined as $\mathcal{M}_p(\xi) \triangleq \mathrm{E}\{\xi^p\}$. Arbitrary $p \geq 3$ may be used as long as the corresponding moment of a transmitted i.i.d. signal is nonzero and finite [15, 16]. Moments of order greater than four are rarely considered since they provide no real benefits over the choices $p = 3$ or $p = 4$. The popular Godard, or constant modulus algorithm (CMA) [17, 18] is based on fourth-order moments, and has been the subject of extensive research in the past (see [19] and references therein). While a nonzero fourth moment can be assumed; for a stochastic process to possess a nonzero third moment, the probability density function (pdf) of its samples must be asymmetric. Since almost every digital modulation format yields a signal with symmetric pdf, third moments have found little use in blind equalization. A notable exception is found in [20], in which the transmitted signal is pre-transformed to induce asymmetry in the pdf, thereby allowing the use of a third-order method. In the context of digital communication, PPM is a rare example of a modulation type resulting in an asymmetric pdf.

While asymmetry allows the use of *either* third or fourth-order methods, earlier work has demonstrated that choosing $p = 3$ over $p = 4$ provides faster convergence of adaptation algorithms and better robustness to noise [21]. These advantages motivate that asymmetry should be exploited using third moments. In the field of blind system identification, third-order methods have been studied [22, 23, 24, 25] and applied to areas such

as mechanics[26], ultrasonics [27] and underwater acoustics [28].

In this article, third moments are applied to blind equalization under PPM signaling, where the standard assumption of an i.i.d. source signal is not valid. Research addressing non-i.i.d. sources has shown that periodicity, temporal correlation or unequal distributions over samples may deform the objective surface enough to produce false optima (i.e. local maxima) with poor ISI reduction for methods that are otherwise globally convergent with an i.i.d. source, such as CMA [29, 19].

By characterizing the objective surfaces of two proposed functions, it is investigated whether a stochastic gradient search will risk converging to a 'false' optimum point (local maxima) of poor equalizer performance. Analytic and experimental results are presented, taking into account practical conditions like time-hopped PPM formats and additive noise.

The rest of the paper is organized as follows. Section 2 defines the notation and models used for PPM signals, channel and equalizer. Section 3 discusses the blind equalization method and defines two objective functions which are analyzed in Section 4. Simulation results and conclusions are provided in Sections 5 and 6 respectively.

2   Model Description and Notation
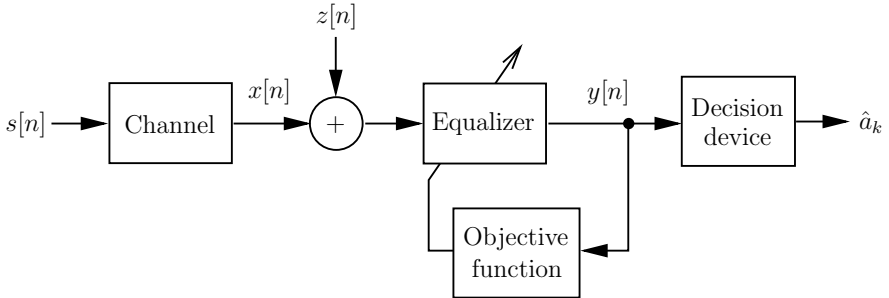
Figure 1: Discrete-time model of a communication system with blind equalization.

2.1   General Model Description

A discrete-time signal model of a communication system with a blind adaptive equalizer is shown in Fig. 1, with n denoting the sample index and k the symbol index. The received signal is the convolution x of the transmitted PPM signal s with the channel plus a disturbance z. The channel model incorporates the aggregated response of transmitter filters, the transmission medium, and receiver front-end and noise-whitening filters. The additive disturbance z is the sum of thermal noise and multiple-access interference which, assuming a large number of users are active, is approximately Gaussian distributed [3].

Hence, z is modeled as white Gaussian noise, independent of s. The equalizer output y is fed to a decision device producing a sequence $\{\hat{a}_k\}$ of data symbol estimates. All signals and system responses are assumed to be real valued.

2.2 Channel and Equalizer

The channel is assumed to be a linear system with an impulse response having most of its energy concentrated to the first L_c samples. It is therefore modeled as FIR, represented by the $L_c \times 1$ vector $\mathbf{c} \triangleq [c_0\ c_1\ \cdots\ c_{L_c-1}]^\mathrm{T}$, with c_i being the ith sample of the impulse response. Although $\mathbf{c}$ will be time-varying in practice, the variations are assumed to be slow compared to the sample index n which is therefore omitted in the notation.

The equalizer is FIR of length L_f, denoted by the $L_f \times 1$ vector $\mathbf{f} \triangleq [f_0\ f_1\ \cdots\ f_{L_f-1}]^\mathrm{T}$. It is adaptive and therefore time-varying, although slowly so that the sample index is again left out for notational simplicity. From $\mathbf{c}$ and $\mathbf{f}$, define the $L \times 1$ combined channel-equalizer response vector,

$$\mathbf{t} = [t_0\ t_1\ \cdots\ t_{L-1}]^\mathrm{T} \triangleq \mathbf{C}^\mathrm{T}\mathbf{f}, \tag{1}$$

where $L = L_c + L_f - 1$ and $\mathbf{C}$ is the $L_f \times L$ channel convolution matrix

$$\mathbf{C} \triangleq \begin{bmatrix} c_0 & \cdots & c_{L_c-1} & & \\ & \ddots & & \ddots & \\ & & c_0 & \cdots & c_{L_c-1} \end{bmatrix}.$$

Hence, $\mathbf{t}$ represents the aggregated response of channel and equalizer.

2.3 M-PPM Signal

In an ordinary time-continuous M-PPM signal $\zeta(t)$, the time axis is divided into consecutive frames of length T_f seconds. Each frame is, in turn, divided into M time slots, or 'chips', of equal lengths T_f/M seconds. By transmitting a narrow pulse in one chip and nothing in the others, each frame conveys one of M possible data symbols. A simple discrete-time model $\zeta[n]$ is obtained by sampling $\zeta(t)$ once per chip. Assuming the width of a pulse does not exceed the chip length, the discrete-time M-PPM signal model is then

$$\zeta[n] = \sum_{k=-\infty}^{\infty} \delta\left[n - kM - a_k\right], \tag{2}$$

where $\{a_k\}$ is the data symbol sequence with $a_k \in \{0,\ 1,\ \ldots,\ M-1\}$ representing the kth transmitted symbol and $\delta[n]$ is the Kronecker delta function with value 1 when $n = 0$ and value 0 when $n \neq 0$. Throughout the following, $\{a_k\}$ is considered to be a sequence of independent, uniformly distributed symbols.

2.4   Time-Hopped PPM Signal

The M-PPM signal (2) is not suitable in a wireless multi-user scenario since the probability of catastrophic collisions is high with many users transmitting at the same time [3]. An enhanced format with *time-hopping* (TH) uses a longer frame length with much more than M chips and adds an additional time shift to each transmitted pulse according to a pseudo-random, user-specific TH pattern.

Assuming a sampling rate of one sample per chip as in (2), a discrete-time TH-PPM signal can be modeled as [2, 3],

$$\zeta^{\mathrm{TH}}[n] = \sum_{k=-\infty}^{\infty} \delta\left[n - kN_f - c_k - a_{\lfloor k/N_s \rfloor}\right], \tag{3}$$

where N_f is the frame length in samples, assumed to be much larger than M. $\{c_k\}$ is a periodic TH sequence of period N_p with $c_k \in \{0, 1, \ldots, N_h - 1\}$. $\{a_k\}$ is the sequence of information-bearing time shifts with $a_k \in \{0, 1, \ldots, M - 1\}$. $\lfloor x \rfloor$ denotes the integer part of x, so that $a_{\lfloor k/N_s \rfloor}$ means that the same symbol is repeated over N_s frames to achieve a higher effective SNR under transmission power constraints, at the cost of a lower symbol rate. To prohibit pulses from ending up in trailing frames after added time-hops, it is necessary to bound the TH span to $N_h \le N_f - M + 1$.

2.5   Modeling TH-PPM with M-PPM

As TH-PPM is a format that currently attracts attention from researchers and system designers, it is also the one of interest in this article, despite its relative complexity compared to M-PPM. While the M-PPM signal (2) is parameterized solely by the symbol alphabet size M, specifying the TH-PPM signal (3) requires four additional parameter choices. Fortunately, for purposes of the equalizer convergence analysis in Section 4, TH-PPM with frame length N_f is modeled well by an M-PPM signal with $M = N_f$ under the reasonable assumption of a uniformly distributed TH pattern with a large span (N_h close to $N_f - M + 1$) and a long period (N_p large). With a large N_h, the frame length is maximally exploited to avoid interuser interference [3]. Combined with a long TH period N_p, the transmitted signal energy is spread evenly over a wide frequency band [30], which is of importance in UWB radio.

To conclude, M-PPM (2) with $M = N_f$ is hereafter used in the analysis to model a practical TH-PPM signal with a frame length of N_f samples. A typical value of N_f, and hence for M in the following, may range from a hundred up to a thousand [3].

2.6   Model and Statistics of the Transmitted Signal

With the PPM signal defined as in (2), the transmitted signal s is formed by subtracting the mean from ζ to achieve a zero-mean (DC-free) signal,

$$s[n] \triangleq \zeta[n] - \mathrm{E}\{\zeta[n]\} = \zeta[n] - \frac{1}{M}.$$

The variance of s is $\sigma_s^2 = (M-1)/M^2$. No power normalization factor is incorporated into the signal model since, although the mean power of s depends on M, this will be irrelevant for the coming analysis. The normalized third moment, or *skewness*, of s, which reflects the asymmetry in the pdf, is

$$\frac{\mathrm{E}\{s^3\}}{(\mathrm{E}\{s^2\})^{3/2}} = \frac{M-2}{\sqrt{M-1}},$$

which is positive and strictly increasing with M for $M \geq 3$. Since M is here assumed to be on the order of a hundred or more, the transmitted signal definitely has positive skewness, enabling the use of third-order blind methods in the following.

The M-PPM signal (2) is a cyclostationary process. Hence, the autocorrelation $r_s(n,m) \triangleq \mathrm{E}\{s[n]s[n+m]\}$ is not independent on n, but periodic with period M so that $r_s(n,m) = r_s(n+kM,m)$ for integer k. The cyclostationarity can be exploited by sub-sampling the received signal into M parallel streams and applying a bank of M equalizers, as suggested in [9, 13]. In this work however, we target a low-complex solution and investigate the performance of a single linear equalizer, for which the cyclostationarity only complicates the analysis. To eliminate the time dependence in r_s, the *time average autocorrelation* $\bar{r}_s$ can be formed by averaging over a period [31],

$$\bar{r}_s(m) \triangleq \frac{1}{M} \sum_{n=0}^{M-1} r_s(n,m) = \begin{cases} \dfrac{M-1}{M^2}, & m = 0 \\ \dfrac{|m|-M}{M^3}, & |m| = 1, \ldots, M-1 \\ 0, & |m| \geq M. \end{cases}$$

From $\bar{r}_s$ it is seen that, in an average sense, the transmitted signal is correlated for lags $|m| < M$ and therefore not an i.i.d. process.

3 Blind Equalization Strategy and Method

3.1 Zero-Forcing Equalization

The FIR equalizer is working in blind mode (without training data) where zero-forcing (ZF) equalization is the natural criterion [32, 33]. The ZF objective is to invert the channel and recover an ISI-free, delayed version of s at the equalizer output. That is, to obtain $y[n] = \alpha s[n-\Delta] + p[n]$, where α and Δ is an arbitrary scale and delay respectively and $p[n]$ is colored Gaussian noise ($z[n]$ filtered through the equalizer). In practice, it is sufficient to only approximately invert the channel, allowing some residual ISI to be present at the equalizer output. The practical equalization goal then becomes

$$y[n] \approx \alpha s[n-\Delta] + p[n]. \tag{4}$$

The reason why strict equality in (4) is not pursued is that adaptation may switch to decision-directed mode (such as DD-LMS [34]) as soon as the equalizer parameters are

sufficiently close to a ZF setting. DD mode converges faster than blind mode and may operate under the advantageous minimum mean square error (MMSE) criterion, but is not suitable during a start-up phase or the recovery from abrupt changes in the channel, when symbol estimates $\hat{a}_k$ are typically unreliable [34]. Furthermore; as only a fractionally spaced (oversampled) FIR equalizer can perfectly equalize a channel of finite length [19], (4) is the reasonable target for baud-spaced[1] blind equalization.

3.2   Steepest-Ascent Adaptation

Adaptation of the equalizer is performed by iteratively adjusting the parameter vector $\mathbf{f}$ in the estimated direction of the gradient of a suitable objective function $\mathcal{O}$ with respect to $\mathbf{f}$, i.e. using a steepest-ascent method. With $\mathbf{f}^{(i)}$ denoting the equalizer parameters after i iterations, a steepest-ascent algorithm with parameter vector normalization is

$$\mathbf{f}^{(i+1)} = \mathbf{f}^{(i)} + \mu\widehat{\boldsymbol{\nabla}}(\mathbf{f}^{(i)}), \tag{5a}$$

$$\mathbf{f}^{(i+1)} \leftarrow \frac{\mathbf{f}^{(i+1)}}{\|\mathbf{f}^{(i+1)}\|_2}, \tag{5b}$$

where μ is a small, positive stepsize, $\|\cdot\|_2$ is the Euclidean norm, and $\widehat{\boldsymbol{\nabla}}(\mathbf{f}^{(i)})$ is an estimate of the objective function gradient with respect to $\mathbf{f}^{(i)}$,

$$\boldsymbol{\nabla}(\mathbf{f}^{(i)}) \triangleq \frac{\partial\,\mathcal{O}}{\partial\,\mathbf{f}^{(i)}}.$$

The normalization step (5b) does not affect the equalizer's performance in terms of ISI reduction, but prohibits a drift or a rapid increase of its gain. As the division operation in (5b) can be computationally costly on fixed-point architectures, a division-free algorithm has been proposed and developed in [35], giving third-moment based steepest ascent without the need for the extra normalization step.

3.3   Equalization Objective in Combined Channel-Equalizer Space

Following the customary approach in the study of blind equalizers; convergence analysis of the proposed method is done in combined channel-equalizer parameter space, rather than in equalizer parameter space, since it is then possible to judge the 'goodness' or 'badness' of a certain setting for a general channel. The steepest-ascent algorithm (5) is then regarded as a search for a maximum point on the objective surface $\mathcal{O}(\mathbf{t})$ in $\mathbf{t}$-space. Note that there is generally no one-to-one correspondence between points in $\mathbf{t}$-space and $\mathbf{f}$-space if the equalizer has finite length. Nevertheless, for purposes of analysis, we may regard $\mathbf{f}$ as being infinitely long, since in practice the extrema of $\mathcal{O}(\mathbf{t})$ and $\mathcal{O}(\mathbf{f})$ will coincide if the actual FIR equalizer is long enough and the channel has no spectral nulls [33].

---

[1]For PPM, 'baud-spaced' here refers to a system sampled once per chip.

The blind equalization target (4) formulated in terms of $\mathbf{t}$ becomes

$$\mathbf{t} \approx \mathbf{e}_\Delta \triangleq [0 \cdots 0 \; \alpha \; 0 \cdots 0]^{\mathrm{T}}, \tag{6}$$

where the nonzero element with arbitrary value α occurs at index Δ. The set of vectors

$$T_{\mathrm{ZF}} = \{\mathbf{e}_\Delta\}_{\Delta = 0, \, \dots, \, L-1} \tag{7}$$

are the global system responses corresponding to zero-forcing settings. A globally convergent method is one for which T_{ZF} are global maximum points on the objective surface, with no global or local maxima elsewhere.

3.4   Objective Function Selection

The objective function of a blind method is generally chosen to reflect some characteristic in the source signal s that is suppressed by a non-trivial channel, so that $\mathcal{O}(x) < \mathcal{O}(s)$ with x being the noise-free part of the received signal as in Fig. 1. If the source is a non-Gaussian i.i.d. process; the pdf of x will be closer to a Gaussian compared to the pdf of s, as a result of the central limit theorem since x is a linear combination of samples from s. Donoho [16] therefore suggested higher-order moments as objective functions, based on their ability to quantify the level of 'non-Gaussianity'. Various methods based on higher-order moments have since been investigated under the assumption of the source being i.i.d. (see, for example [36, 33] and references therein).

Here, we propose to use an objective function based on third moments, despite the observation in Section 2.6 that a PPM sequence is not i.i.d. Motivations for this choice are given below.

- The pdf of the transmitted signal is positively asymmetric (as established in Section 2.6), implying that the third moment of s is nonzero and positive. On the other hand, the pdf of x will appear more symmetric, due to 180° phase shifts from reflections in the channel, causing the third moment of x to be closer to zero. Fig. 2 shows a typical impulse response of an indoor UWB channel synthesized using a script provided by the IEEE P802.15.3a working group [37]. As the randomly polarized channel reflections *reduce* asymmetry, the equalizer should *promote* asymmetry and hence maximize the third moment of its output.

- A PPM signal is *sparse*, i.e. impulsive with a low duty cycle. Sparseness is characterized by large skewness. The dispersion (i.e. ISI) enforced by the channel (see Fig. 2) reduces sparseness and therefore skewness. Intuitively, the third moment should therefore be inversely proportional to the amount of ISI in a PPM signal.

- By using third moments instead of fourth moments, the asymmetry in the PPM signal is exploited to improve equalizer convergence rate. Earlier work has shown that third-moment based methods provide faster convergence of adaptation compared to fourth-order methods [21] in cases where both types can be employed, such as PPM signaling.

- With z being white Gaussian noise independent on s, it is easy to show that $\mathcal{M}_3(y) = \mathcal{M}_3(x * \mathbf{f})$. In words, the third moment of the equalizer output is 'blind' to the channel noise and introduces no bias in estimates of the objective function gradient [21]. This is a feature of odd-order moments only, and not of even-order moments such as $\mathcal{M}_4$ used in CMA.

- Algorithms with low computational costs are possible, since $\mathcal{M}_3(y)$ is a memoryless function of y. Its gradient can therefore be estimated from instantaneous sample values, as done in the LMS and CMA algorithms.
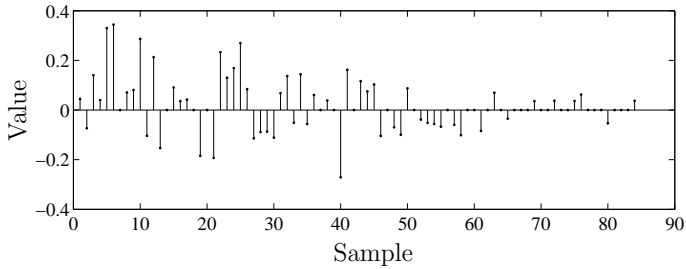
Figure 2: A synthesized impulse response of an indoor UWB channel.

3.5 The Objective Functions Under Study and a Motivational Experiment

To introduce and motivate the choices of objective functions, consider the following experiment with the first function chosen to be the third moment of y,

$$\mathcal{O}_{\mathrm{TM}} \triangleq \mathrm{E}\left\{y^3\right\}.$$

A source signal was formed from an M-PPM sequence using $M = 64$, with white Gaussian noise added at a signal-to-noise ratio of 10 dB. Two stochastic gradient ascents in $\mathbf{t}$-parameter space were conducted over a length-three global response $\mathbf{t} = [t_0 \ t_1 \ t_2]^{\mathrm{T}}$, starting from two arbitrarily selected initial settings of opposite signs; $\mathbf{t}_a = [0.725 \ \ 0.653 \ \ 0.218]^{\mathrm{T}}$ and $-\mathbf{t}_a$. Iteration was performed at each sampling instant using a stepsize $\mu = 0.02$ with the gradient estimated from instantaneous sample values. Fig. 3 shows the resulting $\mathbf{t}$ settings well into convergence after 50000 iterations. Starting at $\mathbf{t}_a$ results in an essentially ISI-free global response, while the search starting at $-\mathbf{t}_a$ converges to a maximum-ISI setting!

Noting the ill-convergence of $\mathcal{O}_{\mathrm{TM}}$ above, the experiment was repeated using an amended objective function, namely the magnitude of the third moment of y,

$$\mathcal{O}_{\mathrm{MAG}} \triangleq \left|\mathrm{E}\left\{y^3\right\}\right|.$$

(a) End point when starting from t_a.   (b) End point when starting from $-t_a$.
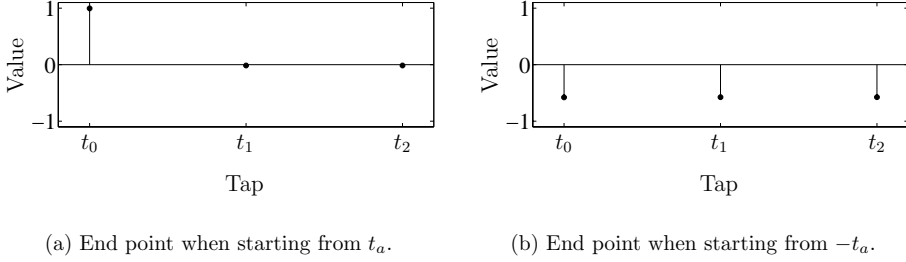
Figure 3: Stochastic gradient ascent end points for $\mathcal{O}_{\mathrm{TM}}$.

The resulting $\mathbf{t}$ settings corresponding to starting points $\mathbf{t}_a$ and $-\mathbf{t}_a$ are shown in Fig. 4. Here, the gradient searches end near two different ZF points of opposite signs.

(a) End point when starting from t_a.   (b) End point when starting from $-t_a$.
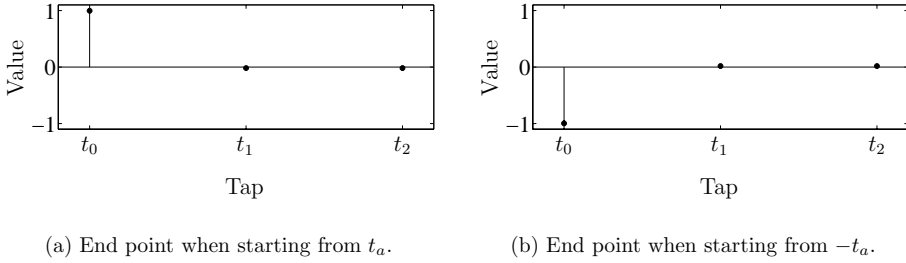
Figure 4: Stochastic gradient ascent end points for $\mathcal{O}_{\mathrm{MAG}}$.

This simple experiment reveals the possibility of ill-convergence of $\mathcal{O}_{\mathrm{TM}}$ under PPM signaling. While no such behavior is observed with the modified objective $\mathcal{O}_{\mathrm{MAG}}$, its blindness to polarity in the equalizer output, and hence in $\mathbf{t}$, is noted. This sign ambiguity is of no concern however, as information is not conveyed in the amplitude of the PPM signal.

In the next section, the global convergence properties of $\mathcal{O}_{\mathrm{TM}}$ and $\mathcal{O}_{\mathrm{MAG}}$ are examined for arbitrary lengths of $\mathbf{t}$.

4   Objective Surface Analysis

To analyze the global convergence properties of an objective function, all stationary points on the objective surface $\mathcal{O}(\mathbf{t})$ must be found and classified as maxima, minima or saddles. For the proposed functions $\mathcal{O}_{\mathrm{TM}}$ and $\mathcal{O}_{\mathrm{MAG}}$, first note their invariance to additive white Gaussian noise.

Referring to Fig. 1, let $y = v + p$, where v is the noise-free channel output x filtered by the equalizer and p is the correspondingly filtered version of z. As z is zero-mean, white Gaussian noise, independent on x, it follows that

$$\mathcal{O}_{\text{TM}}(y) = \mathrm{E}\left\{(v + p)^3\right\} = \mathrm{E}\left\{v^3\right\}.$$

The noise invariance of $\mathcal{O}_{\text{MAG}}$ follows trivially from its definition. Hence, neither of the two objective surfaces are perturbed by the noise. Since only the static properties of the surfaces (i.e. the location of stationary points and curvature) are investigated here, noise is disregarded throughout the remainder of the analysis.

Secondly, only *norm-constrained* steepest ascent adaptation is considered, i.e. maximizing the objective functions under constant norm of $\mathbf{f}$. Without this restriction, the third moment given any parameter vector $\mathbf{f}$ is improved by simply increasing its norm. Hence, maximum points only exist on norm-constrained objective surfaces. The normalization step (5b) in the steepest ascent algorithm accomplishes this in practice. As constraining $\|\mathbf{f}\|_2$ also implies a constraint on $\|\mathbf{t}\|_2$ for a fixed channel; in the following, all feasible parameter vectors are assumed to lie on the $L - 1$-dimensional unit sphere

$$S_{L-1} \triangleq \left\{\mathbf{t} \in \mathbb{R}^L \;\middle|\; \|\mathbf{t}\|_2^2 = 1\right\}.$$

There is no loss of generality from this constraint since, if $\mathbf{t}^*$ is a stationary point of $\mathcal{O}_{\text{TM}}$ or $\mathcal{O}_{\text{MAG}}$ on the unit sphere, then $k\,\mathbf{t}^*$ is a stationary point on the sphere of radius $|k|$ for any $k \in \mathbb{R}$.

4.1   Localization of $\mathcal{O}_{\text{TM}}$ Stationary Points

Using the method of Lagrange multipliers [38], the stationary points of $\mathcal{O}_{\text{TM}}$ under unit-norm $\mathbf{t}$ are the stationary points of the Lagrangian function

$$\mathscr{L}_{\text{TM}}(\mathbf{t}, \lambda) \triangleq \mathcal{O}_{\text{TM}}(\mathbf{t}) - \lambda d(\mathbf{t}), \tag{8}$$

where $\lambda \in \mathbb{R}$ is a Lagrange multiplier and $d(\mathbf{t})$ is the constraint function

$$d(\mathbf{t}) \triangleq \|\mathbf{t}\|_2^2 - 1 = \mathbf{t}^{\text{T}}\mathbf{t} - 1. \tag{9}$$

Solving for the gradient of $\mathscr{L}_{\text{TM}}$ equal to zero leads to the set of equations

$$\frac{\partial}{\partial \mathbf{t}}\mathscr{L}_{\text{TM}}(\mathbf{t}, \lambda) = \frac{\partial \mathcal{O}_{\text{TM}}}{\partial \mathbf{t}} - 2\lambda\mathbf{t} = \mathbf{0}, \tag{10}$$

$$\frac{\partial}{\partial \lambda}\mathscr{L}_{\text{TM}}(\mathbf{t}, \lambda) = -d(\mathbf{t}) = 0, \tag{11}$$

where $\mathbf{0}$ is the $L \times 1$ zero vector. The stationary points of $\mathcal{O}_{\text{TM}}$ on S_{L-1} are the solutions to (10) under the unit-norm constraint (11). With the combined channel-equalizer response $\mathbf{t}$ defined in (1) and disregarding noise, let $\mathbf{s}_n \triangleq \left[\, s[n]\, s[n-1]\, \cdots\, s[n-L+1]\,\right]^{\text{T}}$ so that $y[n] = \mathbf{t}^{\text{T}}\mathbf{s}_n$. (10) becomes

$$3\mathrm{E}\left\{\left(\mathbf{t}^{\text{T}}\mathbf{s}_n\right)^2 \mathbf{s}_n\right\} - 2\lambda\mathbf{t} = \mathbf{0}. \tag{12}$$

Expanding (12) yields a system of L equations in the L parameters $\{t_m\}_{m=0,\,\ldots,\,L-1}$:

$$3\sum_{i=0}^{L-1}\sum_{j=0}^{L-1}t_i t_j \mathcal{R}(i-j,i-m) - 2\lambda t_m = 0, \tag{13}$$

for $m = 0,\,\ldots,\,L-1$, where $\mathcal{R}(p,q)$ is the time-averaged generalized third moment of s,

$$\mathcal{R}(p,q) \triangleq \frac{1}{M}\sum_{n=0}^{M-1} \mathrm{E}\left\{s[n]s[n+p]s[n+q]\right\}. \tag{14}$$

The equation system (13) is non-linear in the parameters and difficult to solve for general values of L. However, as M is assumed to be large (as discussed in Section 2.5), it is shown in Appendix A that $\mathcal{R}(0,0)$ becomes the dominating monomial coefficient so that (13) approaches

$$3\mathcal{R}(0,0)\,t_m^2 - 2\lambda t_m = 0, \tag{15}$$

for $m = 0,\,\ldots,\,L-1$ and M large. Under the unit-norm constraint (11), the $2^{L+1}-2$ solution vectors to (15) are the stationary points of $\mathcal{O}_{\mathrm{TM}}(\mathbf{t})$ under approximation for large M. The set of stationary points can be partitioned into disjoint subsets as follows. Denote by $\mathbf{t}^*$ an arbitrary solution to (15) under (11) and define the subsets

$$\mathcal{C}_K^+ \triangleq \{\mathbf{t}^* \in S_{L-1} \mid \mathbf{t}^*\text{has exactly } K \text{ non-zero,equal positive elements}\},$$

$$\mathcal{C}_K^- \triangleq \{\mathbf{t}^* \in S_{L-1} \mid \mathbf{t}^*\text{has exactly } K \text{ non-zero,equal negative elements}\},$$

for $K = 1,\,\ldots,\,L$. The non-zero elements of points in $\mathcal{C}_K^+$ and $\mathcal{C}_K^-$ has values $1/\sqrt{K}$ and $-1/\sqrt{K}$ respectively. For example, the union $\mathcal{C}_1^+ \cup \mathcal{C}_1^-$ correspond to the ZF points T_{ZF} defined in (7), while the single points in $\mathcal{C}_L^+$ and $\mathcal{C}_L^-$ correspond to maximum ISI.

A more coarse localization of stationary points will prove useful in the following. Recognize that $\mathcal{C}_K^+ \subset \mathcal{H}^+$ and $\mathcal{C}_K^- \subset \mathcal{H}^-$, where

$$\mathcal{H}^+ \triangleq \{\mathbf{t} \in S_{L-1} \mid t_i \geq 0, \quad \forall i\},$$

$$\mathcal{H}^- \triangleq \{\mathbf{t} \in S_{L-1} \mid t_i \leq 0, \quad \forall i\}.$$

$\mathcal{H}^+$ and $\mathcal{H}^-$ are the disjoint subsets of S_{L-1} containing vectors with all elements non-negative and non-positive respectively.

4.2   Classification of $\mathcal{O}_{\mathrm{TM}}$ Stationary Points

Each stationary point can be classified as a maxima, minima or saddle point depending on the constraint curvature of the Lagrangian at that point. For this, define the Hessian matrix $\mathbf{H}_{\mathrm{TM}}$ of the Lagrangian (8) with respect to $\mathbf{t}$ as

$$\mathbf{H}_{\mathrm{TM}}(\mathbf{t},\lambda) \triangleq \frac{\partial^2 \mathscr{L}_{\mathrm{TM}}}{\partial\mathbf{t}\partial\mathbf{t}^{\mathrm{T}}} = \begin{bmatrix} \dfrac{\partial^2 \mathscr{L}_{\mathrm{TM}}}{\partial t_0^2} & \cdots & \dfrac{\partial^2 \mathscr{L}_{\mathrm{TM}}}{\partial t_0 \partial t_{L-1}} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 \mathscr{L}_{\mathrm{TM}}}{\partial t_{L-1}\partial t_0} & \cdots & \dfrac{\partial^2 \mathscr{L}_{\mathrm{TM}}}{\partial t_{L-1}^2} \end{bmatrix}.$$

Differentiating (12) with respect to $\mathbf{t}$ and expanding yields the Hessian as

$$\mathbf{H}_{\mathrm{TM}}(\mathbf{t}, \lambda) = 6 \begin{bmatrix} \displaystyle\sum_i t_i \mathcal{R}(i, i) & \cdots & \displaystyle\sum_i t_i \mathcal{R}(i, i - L + 1) \\ \vdots & \ddots & \vdots \\ \displaystyle\sum_i t_i \mathcal{R}(i - L + 1, i) & \cdots & \displaystyle\sum_i t_i \mathcal{R}(i - L + 1, i - L + 1) \end{bmatrix} - 2\lambda \mathbf{I}_L, \quad (16)$$

where summations over i are taken from 0 to $L - 1$ and $\mathbf{I}_L$ is the $L \times L$ identity matrix.

In Appendix B, the Hessian is evaluated under the approximation of large M. The results are shown in Table 5.1. As seen, the positive ZF solutions in $\mathcal{C}_1^+$ are indeed maxima, while the negative ones in $\mathcal{C}_1^-$ are minima. A critical observation to be made from Table 5.1 is the existence of a single false optimum, the maximum in $\mathcal{C}_L^-$. This is the maximum-ISI point $\mathbf{t}^\times \triangleq \frac{1}{\sqrt{L}} \begin{bmatrix} -1 & -1 & \cdots & -1 \end{bmatrix}^{\mathrm{T}}$, recognized from the example in Section 3.5 for $L = 3$. Depending on initialization, a steepest ascent on $\mathcal{O}_{\mathrm{TM}}$ hence either converges to a ZF point or the maximum-ISI point $\mathbf{t}^\times$.

Table 5.1: Stationary points of $\mathcal{O}_{\mathrm{TM}}$ on S_{L-1} for large M.

Subset	# of points	Character	Type
$\mathcal{C}_1^+$	L	Zero ISI	Maxima
$\mathcal{C}_1^-$	L	Zero ISI	Minima
$\displaystyle\bigcup_{i=2}^{L-1} \mathcal{C}_i^+$	$2^L - L - 2$	Low to severe ISI	Saddles
$\displaystyle\bigcup_{i=2}^{L-1} \mathcal{C}_i^-$	$2^L - L - 2$	Low to severe ISI	Saddles
$\mathcal{C}_L^+$	1	Max ISI	Minimum
$\mathcal{C}_L^-$	1	Max ISI	Maximum

4.3 Conditions for Ill-Convergence of $\mathcal{O}_{\mathrm{TM}}$

Noting the existence of the false optimum $\mathbf{t}^\times$, finding the conditions under which a steepest ascent converges to this point becomes of prime interest. Ill-convergence will occur if the equalizer taps are initialized in such a way that the corresponding global parameter vector lies within the region of attraction to $\mathbf{t}^\times$. As the $\mathbf{t}$-vector corresponding to any given initial choice of $\mathbf{f}$ depends on the channel (which is unknown), the probability of ill-convergence is mainly affected by the size and character of the region of attraction, both of which will now be derived.

As noted in Section 4.1, $\mathbf{t}^\times$ is interior to the region $\mathcal{H}^-$, constituting the subset of vectors $\mathbf{t}$ on S_{L-1} with all elements non-positive. In Appendix C, it is established that

$\mathcal{H}^-$ actually *is* the region of attraction to $\mathbf{t}^\times$, making the rest of the unit sphere the region of attraction to ZF-points. This implies that starting from an equalizer setting such that at least one tap in the corresponding global response is positive (i.e. starting outside of $\mathcal{H}^-$) is sufficient to ensure convergence to a zero-forcing setting.

For example, in the low-dimensional case $L = 3$, $\mathcal{H}^-$ is one octant of S_2, or one 8th of the total parameter space. For general L, the volume ratio of the hyperoctant $\mathcal{H}^-$ to the total space is 2^{-L}, which is vanishingly small for any practically encountered global system response length. This, combined with the character of $\mathcal{H}^-$ (all $\mathbf{t}$-elements non-positive) makes it a very unlikely starting point. Indeed, since a blind equalizer is usually initialized with a single non-zero and positive coefficient in $\mathbf{f}$, a starting point in $\mathcal{H}^-$ only occurs if all nonzero taps of the channel impulse response are negative, which can be ruled out in practice.

4.4   Visualization of $\mathcal{O}_{\text{TM}}$ for $L = 3$

To get an improved insight into the topology of the third-moment objective surface, Fig. 5 shows the locations of all stationary points on the S_2 unit sphere. The regions $\mathcal{H}^+$ and $\mathcal{H}^-$ are the octants enclosed by dashed lines in Fig. 5.5(a) ($\mathcal{H}^+$) and 5.5(b) ($\mathcal{H}^-$) respectively. Table 5.2 lists all the stationary points for this example with a plot legend to the figure. Some gradient field lines are sketched to highlight the curvature on the sphere. The steepest-ascent trajectories corresponding to the experiments in Section 3.5 are also shown. Note how the trajectory starting at $-t_a$ inside of $\mathcal{H}^-$ converges to the false optimum $\mathbf{t}^\times = \frac{1}{\sqrt{3}}[\text{-}1\ \text{-}1\ \text{-}1]^{\text{T}}$ of Fig. 5.3(b).

Table 5.2: *Stationary points $[t_0\ t_1\ t_2]$ of $\mathcal{O}_{\text{TM}}$ on S_2 for large M.*

Subset	Subset elements	Type	Legend
C_1^+	$\{[1\,0\,0]\,[0\,1\,0]\,[0\,0\,1]\}$	Maxima	•
C_2^+	$\frac{1}{\sqrt{2}}\{[1\,1\,0]\,[1\,0\,1]\,[0\,1\,1]\}$	Saddles	■
C_3^+	$\frac{1}{\sqrt{3}}\{[1\,1\,1]\}$	Minimum	▲
C_1^-	$\{[\text{-}1\,0\,0]\,[0\,\text{-}1\,0]\,[0\,0\,\text{-}1]\}$	Minima	△
C_2^-	$\frac{1}{\sqrt{2}}\{[\text{-}1\,\text{-}1\,0]\,[\text{-}1\,0\,\text{-}1]\,[0\,\text{-}1\,\text{-}1]\}$	Saddles	□
C_3^-	$\frac{1}{\sqrt{3}}\{[\text{-}1\,\text{-}1\,\text{-}1]\}$	Maximum	○

4.5   Localization of $\mathcal{O}_{\text{MAG}}$ Stationary Points

As for $\mathcal{O}_{\text{TM}}$, the stationary points of the magnitude-third moment objective function $\mathcal{O}_{\text{MAG}}$ on the unit sphere are now found and classified using Lagrange multipliers. The
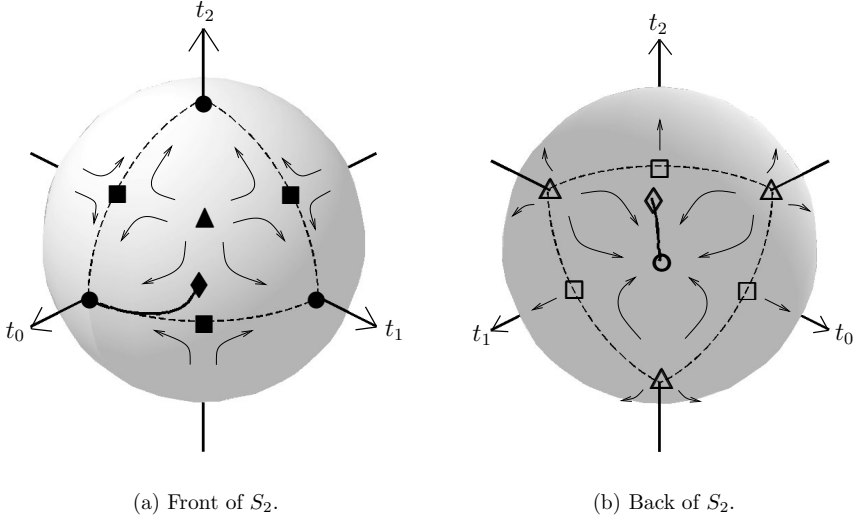
(a) Front of S_2.   (b) Back of S_2.

Figure 5: Locations of $\mathcal{O}_{\mathrm{TM}}$ stationary points on the S_2 unit sphere. The curved arrows indicate the direction of the gradient vector field. Steepest-ascent trajectories starting from points t_a (marked by ◆) and $-t_a$ (marked by ◇) are also shown. For a plot legend, see Table 5.2.

Lagrangian function becomes

$$\mathscr{L}_{\mathrm{MAG}}(\mathbf{t}, \lambda) \triangleq \mathcal{O}_{\mathrm{MAG}} - \lambda d(\mathbf{t}), \tag{17}$$

where $d(\mathbf{t})$ is the unit-norm constraint function defined in (9). Taking the gradient of $\mathscr{L}_{\mathrm{MAG}}$ with respect to $\mathbf{t}$, expanding, and equating to zero yields, for $m = 0, \ldots, L-1$,

$$3\,\mathrm{sgn}\,[\mathcal{O}_{\mathrm{TM}}(\mathbf{t})] \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} t_i t_j \mathcal{R}(i-j, i-m) - 2\lambda t_m = 0, \tag{18}$$

where $\mathcal{O}_{\mathrm{TM}}(\mathbf{t}) = \mathrm{E}\left\{(\mathbf{t}^{\mathrm{T}}\mathbf{s}_n)^3\right\}$ and $\mathrm{sgn}(x)$ is the signum function which is 1 if $x > 0$, -1 if $x < 0$ and undefined if $x = 0$. The stationary points are thus the solution vectors to the system (18) under the constraint $d(\mathbf{t}) = 0$. As for $\mathcal{O}_{\mathrm{TM}}$, an approximation is taken for large M (all generalized third moments but $\mathcal{R}(0,0)$ are zero), under which (18) tends to

$$3\,\mathrm{sgn}\left(\sum_{i=0}^{L-1} t_i^3\right)\mathcal{R}(0,0)\, t_m^2 - 2\lambda t_m = 0, \tag{19}$$

for $m = 0, \ldots, L-1$, where it was used that for large M,

$$\mathcal{O}_{\mathrm{TM}}(\mathbf{t}) \approx \widetilde{\mathcal{O}}_{\mathrm{TM}}(\mathbf{t}) \triangleq \mathcal{R}(0,0) \sum_{i=0}^{L-1} t_i^3 \tag{20}$$

and $\mathcal{R}(0,0) > 0$ (see Appendix A). From (19), the set of stationary points of $\mathcal{O}_{\mathrm{MAG}}$ can be found and partitioned into the subsets

$$\widehat{\mathcal{C}}_K^+ \triangleq \{\mathbf{t}^* \in S_{L-1} \mid \mathbf{t}^*\text{has exactly } K \text{ non-zero,equal positive elements}\},$$

$$\widehat{\mathcal{C}}_K^- \triangleq \{\mathbf{t}^* \in S_{L-1} \mid \mathbf{t}^*\text{has exactly } K \text{ non-zero,equal negative elements}\},$$

for $K = 1, \ldots, L$. As for $\mathcal{O}_{\mathrm{TM}}$, note that $\widehat{\mathcal{C}}_K^+ \subset \mathcal{H}^+$ and $\widehat{\mathcal{C}}_K^- \subset \mathcal{H}^-$. Furthermore; the set of vectors $\mathcal{Z} \triangleq \left\{\mathbf{t} \in S_{L-1} \middle| \sum_{i=0}^{L-1} t_i^3 = 0\right\}$, the intersection of the sphere with the surface $\sum_{i=0}^{L-1} t_i^3 = 0$, forms a region on the sphere for which (19) has no solutions since the signum function is undefined there. However, note that by (20), $\mathcal{O}_{\mathrm{TM}}$, and therefore $\mathcal{O}_{\mathrm{MAG}}$, is zero in this region for large M. Since $\mathcal{O}_{\mathrm{MAG}} \geq 0$ for all $\mathbf{t}$, $\mathcal{Z}$ is a continuous[2] (uncountable) set of global minima.

4.6   Classification of $\mathcal{O}_{\mathrm{MAG}}$ Stationary Points

The Hessian matrix $\mathbf{H}_{\mathrm{MAG}}$ of the Lagrangian (8) with respect to $\mathbf{t}$ is obtained from (18) as

$$\mathbf{H}_{\mathrm{MAG}}(\mathbf{t}, \lambda) \triangleq \frac{\partial^2 \mathscr{L}_{\mathrm{MAG}}}{\partial \mathbf{t} \partial \mathbf{t}^{\mathrm{T}}} =$$

$$6\,\mathrm{sgn}\left[\mathcal{O}_{\mathrm{TM}}(\mathbf{t})\right] \begin{bmatrix} \sum_i t_i \mathcal{R}(i,i) & \cdots & \sum_i t_i \mathcal{R}(i, i-L+1) \\ \vdots & \ddots & \vdots \\ \sum_i t_i \mathcal{R}(i-L+1, i) & \cdots & \sum_i t_i \mathcal{R}(i-L+1, i-L+1) \end{bmatrix} - 2\lambda \mathbf{I}_L.$$

In Appendix D, the Hessian is evaluated under approximation of large M. The results are listed in Table 5.3. The only maxima are the ZF points, as the single point in $\widehat{\mathcal{C}}_L^-$ is not a maximum but a minimum here. This is expected, since $\mathcal{O}_{\mathrm{MAG}}(\mathbf{t} \in \mathcal{H}^-) = \mathcal{O}_{\mathrm{MAG}}(\mathbf{t} \in \mathcal{H}^+) = \mathcal{O}_{\mathrm{TM}}(\mathbf{t} \in \mathcal{H}^+)$. In words, the stationary points of $\mathcal{O}_{\mathrm{MAG}}$ in $\mathcal{H}^-$ have the same locations and characters as those in $\mathcal{H}^+$, which in turn are identical to the stationary points of $\mathcal{O}_{\mathrm{TM}}$ in $\mathcal{H}^+$, where no false optimum exists. Hence, the method is globally convergent to zero-ISI settings. Note from the example in Section 3.5 how the trajectory starting in $-t_a$ converges to the closest ZF point of negative polarity.

5   Simulation Results

The functionality of both the $\mathcal{O}_{\mathrm{TM}}$ and the $\mathcal{O}_{\mathrm{MAG}}$ objective is demonstrated in the following numerical experiment. 15 different channel impulse responses were generated from the specifications of the widely accepted channel model CM3, provided by the IEEE P802.15.3a working group for wireless personal area networks [37]. CM3 models indoor

---

[2]Except in the trivial case $L = 2$ where $\mathcal{Z}$ only contains two discrete points.

Table 5.3: Stationary points of $\mathcal{O}_{\mathrm{MAG}}$ on S_{L-1} for large M.

Subset	# of points	Character	Type
$\widehat{\mathcal{C}}_1^+$	L	Zero ISI	Maxima
$\widehat{\mathcal{C}}_1^-$	L	Zero ISI	Maxima
$\bigcup_{i=2}^{L-1} \widehat{\mathcal{C}}_i^+$	$2^L - L - 2$	Low to severe ISI	Saddles
$\bigcup_{i=2}^{L-1} \widehat{\mathcal{C}}_i^-$	$2^L - L - 2$	Low to severe ISI	Saddles
$\widehat{\mathcal{C}}_L^+$	1	Max ISI	Minimum
$\widehat{\mathcal{C}}_L^-$	1	Max ISI	Minimum

UWB channels for transmitter-receiver distances of 4-10 meters with obstructed line-of-sight, featuring significant dispersion. In the script attached to [37], the sampling interval was chosen to be 1 nanosecond. In each generated channel realization, taps of magnitude less than 20 dB below the tap of largest magnitude were zeroed. Also, any leading or trailing sequences of zeros were removed. One of the resulting responses is shown in Fig. 2 in Section 3.

TH-PPM signals were generated with parameters $M = 2$, $N_f = 100$, $N_s = 10$, $N_h = 99$ and $N_p = 50$. At the selected sampling interval, this corresponds to a transmission rate of 1 Mbit/s. After convolution with the channel responses, white Gaussian noise was added with a SNR per bit $E_b/N_0 = 10$ dB. This models a received signal in agreement with Fig. 1.

The objectives $\mathcal{O}_{\mathrm{TM}}$ and $\mathcal{O}_{\mathrm{MAG}}$ were then used to adapt equalizers of length $L_f = 100$ taps using the steepest-ascent algorithm (5), starting from the customary center-spike initialization with the tap at index 51 equal to one and all other taps zero. Gradient estimates were formed as $\widehat{\boldsymbol{\nabla}}(\mathbf{f}^{(i)}) = \mathrm{sgn}(y_i)y_i^2\mathbf{u}_i$ for $\mathcal{O}_{\mathrm{MAG}}$ and $\widehat{\boldsymbol{\nabla}}(\mathbf{f}^{(i)}) = y_i^2\mathbf{u}_i$ for $\mathcal{O}_{\mathrm{TM}}$, with $y_i = \mathbf{f}^{(i)\mathrm{T}}\mathbf{u}_i$ and the $L_f \times 1$ regressor vectors $\mathbf{u}_i$ obtained by averaging over N_s received frames, thereby exploiting the symbol repetition pattern to improve the SNR during adaptation. Equal stepsizes $\mu = 0.001$ were used for both objectives, which on average gave equal ISI reduction performance upon convergence. Assuming perfect transmitter-receiver synchronization and the TH sequence $\{c_k\}$ known to the receiver, equalizer updates for each received symbol were performed only at the M chips where a pulse was possibly transmitted. The experiment was repeated for all 15 channel realizations, and the average resulting ISI versus iteration number is shown in Fig. 6. ISI is defined from the combined channel-equalizer response as

$$\mathrm{ISI}(\mathbf{t}) \triangleq \frac{1}{t_{lc}^2}\left(\sum_{i=0}^{L-1} t_i^2 - t_{lc}^2\right),$$

where t_{lc} is the leading cursor, i.e. the tap with largest absolute value. As seen in Fig.

6, $\mathcal{O}_{\mathrm{MAG}}$ converges slightly faster than $\mathcal{O}_{\mathrm{TM}}$. This is expected since $\mathcal{O}_{\mathrm{MAG}}$ has more maxima, resulting in shorter trajectories from starting points to convergence points on average.
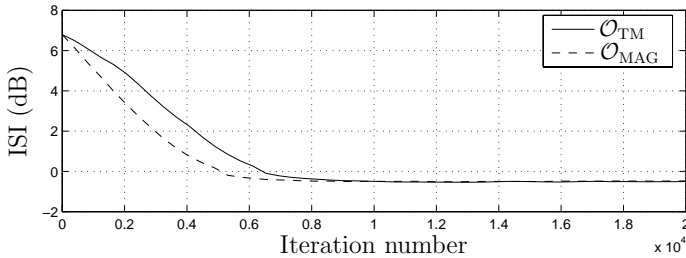

Figure 6: ISI versus iteration number for $\mathcal{O}_{\mathrm{TM}}$ and $\mathcal{O}_{\mathrm{MAG}}$. Average result from 15 different channel realizations.

The lengths of the channel responses ranged from 48 to 89 taps, averaging at 72. Because of the center-spike initialization of the equalizer, Fig. 6 also reveals that the average channel ISI prior to equalization was roughly 7 dB. As seen, the average ISI of the equalized systems steadily decreases, converging to around -0.5 dB. To evaluate the success of a hypothetical switch to a decision-directed algorithm, the symbol error rate was evaluated for each channel after convergence of the blind methods. Symbol decisions were made by comparing signal energies over the M pulse-conveying chips (after averaging over N_s symbol-repeated frames) and selecting the largest. As a rule of thumb, successful decision-directed convergence requires a decision error rate below 0.1 [39, 17]. This was achieved for all channels at $E_b/N_0 = 10$ dB, with the worst channel giving a decision error rate of 0.0043 for both objectives. The conclusion from this simulation is that both objectives are capable of providing good start-up conditions for decision-directed methods.

6   Conclusions

The third-order moment has been proven to be a sufficient statistic of a time-hopped PPM signal for blind channel equalization. An objective function based on the third-moment magnitude provides a globally convergent blind method of low computational complexity. The possibility of ill-convergence of the regular (non-magnitude) third moment was also investigated and deemed highly unlikely in practice, making it a reliable alternative of even lower complexity. As third-moment based methods feature fast convergence and good robustness to noise compared to other common blind methods, they are an interesting choice for applications featuring TH-PPM signals, such as UWB radio. The functionality of two proposed objectives was demonstrated in a simulation using a widely accepted UWB channel model and a simple stochastic gradient search.

A   Approximation of Gradient Equations for Large M

As defined in (14), the time-averaged generalized third moments $\mathcal{R}(p,q)$ of a zero-mean M-PPM signal s are; for $|p| < M$, $|q| < M$ and $|p - q| < M$,

$$
\mathcal{R}(p,q) = \begin{cases}
\dfrac{1}{M^3}(M-1)(M-2), & p = q = 0, \\[2mm]
\dfrac{1}{M^3}\left(\dfrac{|p|}{M} - 1\right)(M-2), & q = 0 \text{ and } p \neq 0, \\[2mm]
\dfrac{2}{M^3} - \dfrac{2}{M^4}\max\left(|p|, |q|\right), & p \neq q \neq 0 \text{ and } \operatorname{sgn}(p) = \operatorname{sgn}(q), \\[2mm]
\dfrac{2}{M^3} - \dfrac{2}{M^4}|p - q|, & p \neq q \neq 0 \text{ and } \operatorname{sgn}(p) \neq \operatorname{sgn}(q).
\end{cases}
$$

For lags $|p| \geq M$ or $|q| \geq M$ or $|p - q| \geq M$, $\mathcal{R}(p,q)$ is zero. Also, $\mathcal{R}(p,0) = \mathcal{R}(0,p) = \mathcal{R}(p,p)$. Recall from Section 2.6 that $M \geq 3$, which implies that $\mathcal{R}(0,0) > 0$.

To achieve an approximation of the gradient equations (13) in the limit of large M; it is noted that, for non-zero p and q,

$$
\lim_{M \to \infty} \frac{\mathcal{R}(0,0)}{\mathcal{R}(p,0)} = -\infty,
$$

$$
\lim_{M \to \infty} \frac{\mathcal{R}(0,0)}{\mathcal{R}(p,q)} = \infty.
$$

Specifically; the ratio between the moment of largest magnitude, $\mathcal{R}(0,0)$, and the second largest, $\mathcal{R}(1,0)$, is $|\mathcal{R}(0,0)|/|\mathcal{R}(1,0)| = M$. Hence; for large M, the term involving the center moment $\mathcal{R}(0,0)$ will, for example, dominate all other terms in (13), resulting in the approximated system (15).

To confirm the proximity of actual (non-approximated) stationary points to approximated points under large M, consider the following numerical experiment where the distance from a ZF point to the nearest 'true' maximum was measured. Ten M-PPM signals were generated for values of M between 3 and 128. For each signal, an $\mathcal{O}_{\text{TM}}$-based stochastic gradient ascent in $\mathbf{t}$-space with stepsize $\mu = 0.0005$ was conducted for a 100-tap $\mathbf{t}$-vector, starting at the unit-norm ZF point $\mathbf{e}_{50}$ (as defined by (6) with $\alpha = 1$ at index 50). After 1.5×10^6 iterations, convergence to a small neighborhood around the nearest maximum point $\tilde{\mathbf{t}}$ was reached. This point was then estimated by averaging the $\mathbf{t}$-vectors from the final 10^4 iterations, and the Euclidean distance from $\tilde{\mathbf{t}}$ to the starting point $\mathbf{e}_{50}$ was calculated. The results are shown in Fig. 7. As expected, the true maximum approaches the closest ZF point asymptotically as M increases.
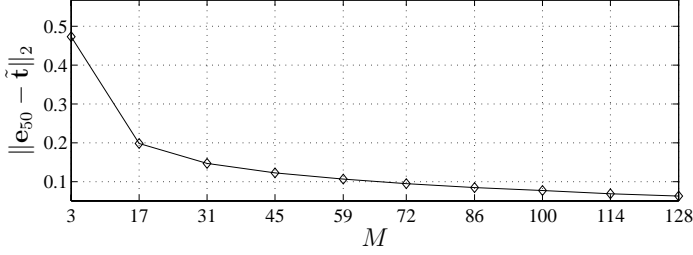
Figure 7: *Euclidean distance from* $\mathbf{e}_{50}$ *to nearest local maxima* $\tilde{\mathbf{t}}$ *versus* M.

B   Characterization of $\mathcal{O}_{\mathrm{TM}}$ Stationary Points for Large M

Here, the Hessian $\mathbf{H}_{\mathrm{TM}}$ is evaluated under approximation for large M to classify the stationary points of the third-moment objective $\mathcal{O}_{\mathrm{TM}}$ on the unit sphere. As shown in Appendix A, $\mathcal{R}(0,0)$ will dominate all other moments $\mathcal{R}(p,q)$ as $M \to \infty$. The Hessian $\mathbf{H}_{\mathrm{TM}}$ (16) then approaches the diagonal form

$$\widetilde{\mathbf{H}}_{\mathrm{TM}}(\mathbf{t}, \lambda) = 6\,\mathcal{R}(0,0)\,\mathrm{diag}\{t_0,\,\ldots,t_{L-1}\} - 2\lambda\mathbf{I}_L.$$

Define $\mathbf{a}$ to be the gradient of the constraint function (9),

$$\mathbf{a}(\mathbf{t}) \triangleq \frac{\partial d}{\partial \mathbf{t}} = 2\mathbf{t}, \tag{B.21}$$

and let $\mathbf{v}(\mathbf{t})$ be a non-zero vector orthogonal to $\mathbf{a}(\mathbf{t})$,

$$\mathbf{v}(\mathbf{t})^{\mathrm{T}}\mathbf{a}(\mathbf{t}) = 0. \tag{B.22}$$

The set of all non-zero $\mathbf{v}$ satisfying (B.22) constitute the *feasible directions* of optimization at $\mathbf{t}$ [38], i.e. all vectors in the tangent plane of the unit sphere at $\mathbf{t}$. Finally, define the quadratic form

$$Q(\mathbf{t}, \lambda) \triangleq \mathbf{v}(\mathbf{t})^{\mathrm{T}}\widetilde{\mathbf{H}}_{\mathrm{TM}}(\mathbf{t}, \lambda)\mathbf{v}(\mathbf{t}). \tag{B.23}$$

A stationary point $\mathbf{t}^*$ with corresponding Lagrange multiplier λ^* can now be classified as a maximum, minimum or saddle point of the constrained objective function by evaluating $Q(\mathbf{t}^*, \lambda^*)$ for all $\mathbf{v}$ satisfying (B.22) according to

$$Q(\mathbf{t}^*, \lambda^*) < 0 \quad \rightarrow \quad \mathbf{t}^* \text{ maximum}$$

$$Q(\mathbf{t}^*, \lambda^*) > 0 \quad \rightarrow \quad \mathbf{t}^* \text{ minimum}$$

$$Q(\mathbf{t}^*, \lambda^*) \text{ indefinite} \quad \rightarrow \quad \mathbf{t}^* \text{ saddle}.$$

B.1   Characterization of Points in Subsets $\mathcal{C}_K^+$

Let $\mathbf{t}_K^* = [t_0^* \, t_1^* \, \cdots \, t_{L-1}^*]^{\mathrm{T}}$ be any point in one of the subsets $\mathcal{C}_K^+$, $1 \leq K \leq L$. Then, $\mathbf{t}_K^*$ has exactly K nonzero, positive elements with equal values $1/\sqrt{K}$, and corresponding positive Lagrange multiplier $\lambda_K^* = \frac{3\mathcal{R}(0,0)}{2\sqrt{K}}$. Next, define $\mathcal{A}$ to be the set of indices corresponding to the nonzero elements in $\mathbf{t}_K^*$, so that t_i^* has value $1/\sqrt{K}$ if $i \in \mathcal{A}$ and value 0 if $i \notin \mathcal{A}$. Any feasible direction $\mathbf{v}(\mathbf{t}_K^*) = [v_0^* \, v_1^* \, \cdots \, v_{L-1}^*]^{\mathrm{T}}$ at $\mathbf{t} = \mathbf{t}_K^*$ must then satisfy

$$\mathbf{v}(\mathbf{t}_K^*)^{\mathrm{T}} \mathbf{a}(\mathbf{t}_K^*) = \sum_{i=0}^{L-1} v_i^* t_i^* = \frac{1}{\sqrt{K}} \sum_{i \in \mathcal{A}} v_i^* = 0,$$

which implies that

$$\sum_{i \in \mathcal{A}} v_i^* = 0. \tag{B.24}$$

Evaluating (B.23) now gives

$$Q(\mathbf{t}_K^*, \lambda_K^*) = 3\mathcal{R}(0,0) \left(2 \sum_{i=0}^{L-1} (v_i^*)^2 t_i^* - \frac{1}{\sqrt{K}} \sum_{i=0}^{L-1} (v_i^*)^2 \right),$$

which can be simplified to

$$Q(\mathbf{t}_K^*, \lambda_K^*) = \frac{3\mathcal{R}(0,0)}{\sqrt{K}} \left(\sum_{i \in \mathcal{A}} (v_i^*)^2 - \sum_{i \notin \mathcal{A}} (v_i^*)^2 \right). \tag{B.25}$$

1) Points in $\mathcal{C}_1^+$:   Any point $\mathbf{t}_1^*$ in $\mathcal{C}_1^+$ has exactly one nonzero element. Let q be the index of this element, so that $\mathcal{A} = q$. Evaluating (B.25) gives

$$Q(\mathbf{t}_1^*, \lambda_1^*) = 3\mathcal{R}(0,0) \left((v_q^*)^2 - \sum_{i \neq q} (v_i^*)^2 \right) = -3\mathcal{R}(0,0) \sum_{i \neq q} (v_i^*)^2,$$

where the last equality follows from v_q^* being zero by (B.24). Since $\mathcal{R}(0,0) > 0$, the quadratic form is strictly negative for all non-zero $\mathbf{v}(\mathbf{t}_1^*)$. Hence, all points in $\mathcal{C}_1^+$ are maxima.

2) Points in $\mathcal{C}_2^+$, ..., $\mathcal{C}_{L-1}^+$:   Let $\mathbf{t}_K^*$ be a stationary point with exactly K non-zero elements where $2 \leq K \leq L-1$. Then $\mathcal{A}$ contains at least 2 and at most $L-1$ indices. As a consequence, (B.25) is indefinite for any $2 \leq K \leq L-1$, meaning that all points in the union of subsets $\bigcup_{i=2}^{L-1} \mathcal{C}_i^+$ are saddles.

3) Points in $\mathcal{C}_L^+$:   For the single stationary point $\mathbf{t}_L^*$ with all elements nonzero, $\mathcal{A} = \{1, 2, \ldots, L\}$, and (B.25) simplifies to

$$Q(\mathbf{t}_L^*, \lambda_L^*) = \frac{3\mathcal{R}(0,0)}{\sqrt{L}} \sum_{i=0}^{L-1} (v_i^*)^2,$$

which is strictly positive for all nonzero $\mathbf{v}(\mathbf{t}_L^*)$. The point $\mathbf{t}_L^*$ is therefore a minimum.

B.2 Characterization of Points in Subsets $\mathcal{C}_K^-$

A stationary point $\mathbf{t}_K^*$ belonging to any of $\mathcal{C}_K^-$, $1 \leq K \leq L$ has exactly K nonzero, negative elements with equal values $-1/\sqrt{K}$, and negative Lagrange multiplier $\lambda_K^* = -\frac{3\mathcal{R}(0,0)}{2\sqrt{K}}$. As above, let $\mathcal{A}$ be the set of indices corresponding to the non-zero elements in $\mathbf{t}_K^*$. Feasible directions $\mathbf{v}(\mathbf{t}_K^*)$ must still satisfy (B.24). The quadratic form (B.23) here becomes

$$Q(\mathbf{t}_K^*, \lambda_K^*) = -\frac{3\mathcal{R}(0,0)}{\sqrt{K}} \left(\sum_{i \in \mathcal{A}} (v_i^*)^2 - \sum_{i \notin \mathcal{A}} (v_i^*)^2 \right),$$

which is simply the negative of the quadratic form (B.25), causing the characteristics of points in subsets $\mathcal{C}_K^-$ to become 'reversed' compared to points in $\mathcal{C}_K^+$. Hence, points in $\mathcal{C}_1^-$ are minima, points in $\mathcal{C}_K^-$, $2 \leq K \leq L-1$ are saddles, and the single point in $\mathcal{C}_L^-$ is indeed a maximum.

C Region of Attraction to the False Optimum Point of $\mathcal{O}_{\mathrm{TM}}$

The region of attraction to the false optimum point of $\mathcal{O}_{\mathrm{TM}}$ can be found by considering the behavior of an ideal, noise-free steepest-ascent algorithm in $\mathbf{t}$-space. Denote by $\mathbf{t}^{(i)} = [t_0(i) \quad t_1(i) \quad \cdots \quad t_{L-1}(i)]^{\mathrm{T}}$ the combined channel-equalizer response after i iterations. Let $\boldsymbol{\nabla}(\mathbf{t})$ be the gradient of $\mathcal{O}_{\mathrm{TM}}$ at $\mathbf{t}$. Under approximation of large M, the gradient approaches, using (20),

$$\widetilde{\boldsymbol{\nabla}}(\mathbf{t}) = 3\,\mathcal{R}(0,0)\,[t_0^2 \quad t_1^2 \quad \cdots \quad t_{L-1}^2]^{\mathrm{T}},$$

and the steepest-ascent algorithm in $\mathbf{t}$-space becomes

$$\mathbf{t}^{(i+1)} = \mathbf{t}^{(i)} + \mu \widetilde{\boldsymbol{\nabla}}(\mathbf{t}^{(i)}), \qquad (\text{C.26a})$$

$$\mathbf{t}^{(i+1)} \leftarrow \frac{\mathbf{t}^{(i+1)}}{\|\mathbf{t}^{(i+1)}\|_2}, \qquad (\text{C.26b})$$

where $0 < \mu \ll 1$. The element-wise update of $\mathbf{t}$ in (C.26a) is, for element $j = 0, \ldots, L-1$,

$$t_j(i+1) = t_j(i) + \mu\, t_j^2(i) = t_j(i)\,[1 + \mu\, t_j(i)]. \qquad (\text{C.27})$$

From (C.27), note the following relation between a tap before and after an iteration,

$$t_j(i) > 0 \quad \text{implies} \quad t_j(i+1) > t_j(i) > 0,$$

$$t_j(i) < 0 \quad \text{implies} \quad 0 > t_j(i+1) > t_j(i),$$

$$t_j(i) = 0 \quad \text{implies} \quad t_j(i+1) = 0.$$

In words; positive taps remain positive and increase in magnitude, negative taps remain negative and decrease in magnitude, and zero-valued taps remain zero. In practice, taps only have value zero momentarily due to additive noise and imperfect gradient estimates.

1) Case: $\mathbf{t}^{(i)} \in \mathcal{H}^+$: Assume $\mathbf{t}^{(i)} \in \mathcal{H}^+$ (all elements non-negative) and let $t_m(i)$ and $t_r(i)$ be two arbitrary taps such that $t_m(i) > t_r(i) > 0$. Comparing sizes over one iteration gives

$$\frac{t_m(i+1)}{t_r(i+1)} = \frac{t_m(i)}{t_r(i)} \frac{1 + \mu\, t_m(i)}{1 + \mu\, t_r(i)} > \frac{t_m(i)}{t_r(i)}.$$

Hence, larger taps outgrow smaller taps over iterations. With the normalization step (C.26b), the unique tap of largest magnitude goes to 1 and all the others go to zero as $i \to \infty$ (confirm with the example in Section 3.5), so that $\mathbf{t}^{(i)}$ tends to a ZF point. If the largest tap is not unique, $\mathbf{t}^{(i)}$ tends to a saddle point which is an unstable equilibria from which it escapes if given a small perturbation.

2) Case: $\mathbf{t}^{(i)} \in \mathcal{H}^-$: Assume $\mathbf{t}^{(i)} \in \mathcal{H}^-$ (all elements non-positive) and let $t_m(i)$ and $t_r(i)$ be two arbitrary taps with magnitude relation $|t_m(i)| > |t_r(i)| > 0$. Then,

$$\frac{|t_m(i+1)|}{|t_r(i+1)|} = \frac{|t_m(i)|}{|t_r(i)|} \frac{|1 + \mu\, t_m(i)|}{|1 + \mu\, t_r(i)|} = \frac{|t_m(i)|}{|t_r(i)|} \frac{|1 - \mu\, |t_m(i)||}{|1 - \mu\, |t_r(i)||} < \frac{|t_m(i)|}{|t_r(i)|},$$

so that tap magnitudes are leveled over iterations. A zero-valued tap is an unstable state, and with the normalization step (C.26b), all taps tend to equal value $-1/\sqrt{L}$. Therefore, a steepest ascent starting anywhere in $\mathcal{H}^-$ converges to the false optimum $\mathbf{t}^\times$.

3) Case: $\mathbf{t}^{(i)} \in S_{L-1} - \{\mathcal{H}^+ \bigcup \mathcal{H}^-\}$: Assume $\mathbf{t}^{(i)}$ is outside both $\mathcal{H}^+$ and $\mathcal{H}^-$, so that there is at least one positive and one negative element in $\mathbf{t}^{(i)}$. Let $t_m(i)$ and $t_r(i)$ be any two elements such that $t_m(i) > 0$, $t_r(i) \neq 0$ and $t_m(i) > t_r(i)$. Then,

$$\frac{|t_m(i+1)|}{|t_r(i+1)|} = \frac{|t_m(i)|}{|t_r(i)|} \frac{|1 + \mu\, t_m(i)|}{|1 + \mu\, t_r(i)|} > \frac{|t_m(i)|}{|t_r(i)|}.$$

Hence, larger taps outgrow smaller ones, as in $\mathcal{H}^+$. With normalization (C.26b), the unique tap of largest magnitude goes to 1 and all the others go to zero as $i \to \infty$, and $\mathbf{t}^{(i)}$ tends to a ZF point. A zero-valued tap is an unstable state, and so is the event of non-unique largest taps.

Concluding, the only region of the sphere from which ill-convergence occurs is $\mathcal{H}^-$ which is therefore the region of attraction to $\mathbf{t}^\times$.

D   Characterization of $\mathcal{O}_{\mathrm{MAG}}$ Stationary Points for Large M

The derivations that follow can be identified as a special case of those found in [33, App. A]. They are included for comparison with the results for $\mathcal{O}_{\mathrm{TM}}$.

As for $\mathbf{H}_{\mathrm{TM}}$, let $M \to \infty$, under which the Hessian matrix $\mathbf{H}_{\mathrm{MAG}}$ approaches the diagonal form

$$\widetilde{\mathbf{H}}_{\mathrm{MAG}}(\mathbf{t}, \lambda) = 6 \operatorname{sgn}\left(\sum_{i=0}^{L-1} t_i^3\right) \mathcal{R}(0,0) \operatorname{diag}\{t_0, \ldots, t_{L-1}\} - 2\lambda\mathbf{I}_L.$$

With the constraint function gradient $\mathbf{a}$ and feasible direction vectors $\mathbf{v}$ defined in (B.21) and (B.22) respectively, define the quadratic form

$$\widehat{Q}(\mathbf{t}, \lambda) \triangleq \mathbf{v}(\mathbf{t})^{\mathrm{T}}\widetilde{\mathbf{H}}_{\mathrm{MAG}}(\mathbf{t}, \lambda)\mathbf{v}(\mathbf{t}), \tag{D.28}$$

from which the curvature of $\mathcal{O}_{\mathrm{MAG}}$ on the unit sphere can be determined in the same manner as in Appendix B.

D.1 Characterization of Points in Subsets $\widehat{\mathcal{C}}_K^+$

Let $\mathbf{t}_K^* = [t_0^*\ t_1^*\ \cdots\ t_{L-1}^*]^{\mathrm{T}}$ be any point belonging to one of the subsets $\widehat{\mathcal{C}}_K^+$, $1 \leq K \leq L$. Then, $\mathbf{t}_K^*$ has exactly K non-zero, positive elements with equal values $1/\sqrt{K}$, and corresponding positive Lagrange multiplier $\lambda_K^* = \frac{3\mathcal{R}(0,0)}{2\sqrt{K}}$. As for $\mathcal{O}_{\mathrm{TM}}$, feasible directions $\mathbf{v}$ must still satisfy (B.24) and the quadratic form (D.28) evolves into (B.25). Therefore, stationary points in subsets $\widehat{\mathcal{C}}_K^+$ have corresponding character as those in $\mathcal{C}_K^+$ for $\mathcal{O}_{\mathrm{TM}}$; points in $\widehat{\mathcal{C}}_1^+$ are maxima, points in $\widehat{\mathcal{C}}_K^+$, $2 \leq K \leq L-1$ are saddles, and the single point in $\widehat{\mathcal{C}}_L^+$ is a minimum.

D.2 Characterization of Points in Subsets $\widehat{\mathcal{C}}_K^-$

With $\mathbf{t}_K^* = [t_0^*\ t_1^*\ \cdots\ t_{L-1}^*]^{\mathrm{T}}$ being any point belonging to one of $\widehat{\mathcal{C}}_K^-$, $1 \leq K \leq L$, $\mathbf{t}_K^*$ has exactly K non-zero, negative elements with equal values $-1/\sqrt{K}$ and *positive* Lagrange multiplier $\lambda_K^* = \frac{3\mathcal{R}(0,0)}{2\sqrt{K}}$. (B.24) still applies, and the quadratic form evolves into (B.25). Hence, points in $\widehat{\mathcal{C}}_1^-$ are maxima, points in $\widehat{\mathcal{C}}_K^-$, $2 \leq K \leq L-1$ are saddles, and the single point in $\widehat{\mathcal{C}}_L^-$ is a maximum.

References

[1] J. M. Kahn and J. R. Barry, "Wireless infrared communications," *Proc. IEEE*, vol. 85, no. 2, pp. 265–298, Feb. 1997.

[2] M. Z. Win and R. A. Scholtz, "Impulse radio: How it works," *IEEE Commun. Lett.*, vol. 2, no. 2, pp. 36–38, Feb. 1998.

[3] 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.

[4] M. Z. Win and R. A. Scholtz, "On the robustness of ultra-wide bandwidth signals in dense multiptah environments," *IEEE Commun. Lett.*, vol. 2, no. 2, pp. 51–53, Feb. 1998.

[5] M. Weisenhorn and W. Hirt, "Robust noncoherent receiver exploiting UWB channel properties," in *Proc. IEEE UWBST*, May 2004, pp. 156–160.

[6] C. Carbonelli and U. Mengali, "M-PPM noncoherent receivers for UWB applications," *IEEE Trans. Wireless Commun.*, vol. 5, no. 8, pp. 2285–2294, Aug. 2006.

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

[8] J.-M. Cramer, R. A. Scholtz, and M. Z. Win, "Evaluation of an ultra-wide-band propagation channel," *IEEE Trans. Antennas Propag.*, vol. 50, no. 5, pp. 561–570, May 2002.

[9] J. R. Barry, "Sequence detection and equalization for pulse-position modulation," in *IEEE International Conference on Communications (ICC 94) Conference Record*, vol. 3, May 1994, pp. 1561–1565.

[10] M. Z. Win and R. A. Scholtz, "On the energy capture of ultrawide bandwidth signals in dense multipath environments," *IEEE Commun. Lett.*, vol. 2, no. 9, pp. 245–247, Sep. 1998.

[11] R. Hoctor and H. Tomlinson, "Delay-hopped transmitted-reference RF communications," in *Proc. IEEE UWBST*, 2002, pp. 265–269.

[12] Z. Tian and G. B. Giannakis, "A GLRT approach to data-aided timing acquisition in UWB radios - part I: Algorithms," *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 2956–2967, Nov. 2005.

[13] 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, Jun. 2004, pp. 2648–2652.

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

[15] K. S. Lii and M. Rosenblatt, "Deconvolution and estimation of transfer function phase and coefficients for nongaussian linear processes," *The Annals of Statistics*, vol. 10, no. 4, pp. 1195–1208, Dec. 1982.

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

[17] 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.

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

[19] C. R. Johnson, Jr., P. Schniter, T. J. Endres, J. D. Behm, D. R. Brown, and R. A. Casas, "Blind equalization using the constant modulus criterion: A review," *Proc. IEEE*, vol. 86, no. 10, pp. 1927–1950, Oct. 1998.

[20] F.-C. Zheng, S. McLaughlin, and B. Mulgrew, "Blind equalization of nonminimum phase channels: Higher order cumulant based algorithm," *IEEE Trans. Signal Process.*, vol. 41, no. 2, pp. 681–691, Feb. 1993.

[21] P. Pääjärvi and J. P. LeBlanc, "Online adaptive blind deconvolution based on third-order moments," *IEEE Signal Process. Lett.*, vol. 12, pp. 863–866, Dec. 2005.

[22] G. B. Giannakis, "Cumulants: A powerful tool in signal processing," *Proc. IEEE*, vol. 75, no. 9, pp. 1333–1334, Sep. 1987.

[23] C. L. Nikias and M. R. Raghuveer, "Bispectrum estimation: A digital signal processing framework," *Proc. IEEE*, vol. 75, no. 7, pp. 869–891, Jul. 1987.

[24] J. M. Mendel, "Tutorial on higher-order statistics (spectra) in signal processing and system theory: Theoretical results and some applications," *Proc. IEEE*, vol. 79, no. 3, pp. 278–305, Mar. 1991.

[25] A. K. Nandi, "Blind identification of FIR systems using third order cumulants," *Signal Processing*, vol. 39, pp. 131–147, Sep. 1994.

[26] J. Y. Lee and A. K. Nandi, "Blind deconvolution of impacting signals using higher-order statistics," *Mechanical Systems and Signal Processing*, vol. 12, no. 2, pp. 357–371, Mar. 1998.

[27] A. K. Nandi, D. Mämpel, and B. Roscher, "Blind deconvolution of ultrasonic signals in nondestructive testing applications," *IEEE Trans. Signal Process.*, vol. 45, no. 5, pp. 1382–1390, May 1997.

[28] M. K. Broadhead, L. A. Pflug, and R. L. Field, "Use of higher order statistics in source signature estimation," *J. Acoust. Soc. Am.*, vol. 107, no. 5, pp. 2576–2585, May 2000.

[29] 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.

[30] M. Z. Win, "A unified spectral analysis of generalized time-hopping spread-spectrum signals in the presence of timing jitter," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 9, pp. 1664–1676, Dec. 2002.

[31] J. G. Proakis, *Digital Communications*, 4th ed. New York, NY: McGraw-Hill, 2001.

[32] 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. Autom. Control*, vol. AC-25, no. 3, pp. 385–399, Jun. 1980.

[33] O. Shalvi and E. Weinstein, "Universal methods for blind deconvolution," in *Blind Deconvolution*, S. Haykin, Ed. Englewood Cliffs, New Jersey: PTR Prentice Hall, 1994.

[34] O. Macchi and E. Eweda, "Convergence analysis of self-adaptive equalizers," *IEEE Trans. Inf. Theory*, vol. IT-30, no. 2, pp. 161–176, Mar. 1984.

[35] P. Pääjärvi and J. P. LeBlanc, "Computationally efficient norm-constrained adaptive blind deconvolution using third-order moments," in *Acoustics, Speech, and Signal Processing, 2006. Proceedings. (ICASSP 2006). IEEE International Conference on*, Toulouse, France, May 2006, pp. III–752 – III–752.

[36] J. A. Cadzow, "Blind deconvolution via cumulant extrema," *IEEE Signal Process. Mag.*, vol. 13, no. 3, pp. 24–42, May 1996.

[37] J. Foerster, "Channel modeling sub-committee report final IEEE P802.15-02/490r1-SG3a," Feb. 2003.

[38] R. Fletcher, *Practical Methods of Optimization*, 2nd ed.   West Sussex PO19 8SQ, England: Wiley, 1987.

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