1

Introduction to Blind Signal Processing:
Problems and Applications

The fundamental problem of communication is that of reproducing at
one point either exactly or approximately a message selected at
another point. (Claude Shannon, 1948)

In this book, we describe various approaches, methods and techniques to
blind and semi-blind signal processing, especially principal and independent
component analysis, blind source separation, blind source extraction, multi-
channel blind deconvolution and equalization of source signals when the
measured sensor signals are contaminated by additive noise. Emphasis is
placed on an information-theoretical unifying approach, adaptive filtering
models and the development of simple and efficient associated on-line adaptive
nonlinear learning algorithms.

We derive, review and extend the existing adaptive algorithms for blind and
semi-blind signal processing with a particular focus on robust algorithms with
equivariant properties in order to considerably reduce the bias caused by
measurement noise, interferences and other parasitic effects. Moreover, novel
adaptive systems and associated learning algorithms are presented for
estimation of source signals and reduction of influence of noise. We discuss the
optimal choice of nonlinear score functions for various signals and noise
distributions, e.g., Gaussian, Laplacian and uniformly-distributed noise assuming
a generalized Gaussian distribution and other models. Extensive computer
simulations have confirmed the usefulness and superior performance of the
developed algorithms. Some of the research results presented in this book are
new and are presented here for the first time.
1.1 PROBLEM FORMULATIONS – AN OVERVIEW

1.1.1 Generalized Blind Signal Processing Problem

A fairly general blind signal processing (BSP) problem can be formulated as follows. We observe records of sensor signals \( x(t) = [x_1(t), x_2(t), \ldots, x_m(t)]^T \) from a MIMO (multiple-input/multiple-output) nonlinear dynamical system. The objective is to find an inverse system, termed a reconstruction system, neural network or an adaptive inverse system, if it exists and is stable, in order to estimate the primary source signals \( s(t) = [s_1(t), s_2(t), \ldots, s_n(t)]^T \). This estimation is performed on the basis of the output signals \( y(t) = [y_1(t), y_2(t), \ldots, y_n(t)]^T \) and sensor signals as well as some a priori knowledge of the mixing system. Preferably, the inverse system should be adaptive in such a way that it has some tracking capability in nonstationary environments (see Fig. 1.1). Instead of estimating the source signals directly, it is sometimes more convenient to identify an unknown mixing and filtering dynamical system first (e.g., when the inverse system does not exist or the number of observations is less than the number of source signals) and then estimate source signals implicitly by exploiting some a priori information about the system and applying a suitable optimization procedure.

In many cases, source signals are simultaneously linearly filtered and mixed. The aim is to process these observations in such a way that the original source signals are extracted by the adaptive system. The problems of separating and estimating the original source waveforms from the sensor array, without knowing the transmission channel characteristics and the sources can be expressed briefly as a number of related problems: Independent Components Analysis (ICA), Blind Source Separation (BSS), Blind Signal Extraction (BSE) or Multichannel Blind Deconvolution (MBD). Roughly speaking, they can be formulated as the problems of separating or estimating the waveforms of the original sources from an array of sensors or transducers without knowing the characteristics of the transmission channels.

There appears to be something magical about blind signal processing; we are estimating the original source signals without knowing the parameters of mixing and/or filtering processes. It is difficult to imagine that one can estimate this at all. In fact, without some a priori knowledge, it is not possible to uniquely estimate the original source signals. However, one can usually estimate them up to certain indeterminacies. In mathematical terms these indeterminacies and ambiguities can be expressed as arbitrary scaling, permutation and delay of estimated source signals. These indeterminacies preserve, however, the waveforms of original sources. Although these indeterminacies seem to be rather severe limitations, in a great number of applications these limitations are not essential, since the most relevant information about the source signals is contained in the waveforms of the source signals and not in their amplitudes or order in which they are arranged in the output of the system. For some dynamical models, however, there is no guarantee that the estimated or extracted signals have exactly the same waveforms as the source signals, and then the requirements

1Single-input single-output (SISO) or single-input/multiple-output (SIMO) are special cases.
must be sometimes further relaxed to the extent that the extracted waveforms are distorted (filtered or convolved) versions of the primary source signals [176, 1286] (see Fig. 1.1).

We would like to emphasize the essential difference between the standard inverse identification problem and the blind or semi-blind signal processing task. In a basic linear identification or inverse system problem we have access to the input (source) signals (see Fig. 1.2(a)). Our objective is to estimate a delayed (or more generally smoothed or filtered) version of the inverse system of a linear dynamical system (plant) by minimizing the mean square error between the delayed (or model-reference) source signals and the output signals.
INTRODUCTION TO BLIND SIGNAL PROCESSING: PROBLEMS AND APPLICATIONS

(a) Conceptual model of inverse system problem. (b) Model-reference for adaptive inverse control. For the switch in position 1, the system performs a standard adaptive inverse by minimizing the norm of error vector $e$, for switch in position 2, the system estimates errors blindly.

In BSP problems we do not have access to source signals (which are usually assumed to be statistically independent), so we attempt, for example, to design an appropriate nonlinear filter that estimates desired signals as illustrated in the case of an inverse system in Fig. 1.2 (a). Similarly, in the basic adaptive inverse control problem [1295], we attempt to estimate a form of adaptive controller whose transfer function is the inverse (in some sense) of that of the plant itself. The objective of such an adaptive system is to make the
plant to directly follow the input signals (commands). A vector of error signals defined as the difference between the plant outputs and the reference inputs are used by an adaptive learning algorithm to adjust parameters of the linear controller. Usually, it is desirable that the plant outputs do not track the input source (command) signals themselves but rather track a delayed or smoothed (filtered) version of the input signals represented in Fig. 1.2 by transfer function \( M(z) \). It should be noted that in the general case the global system consisting of the cascade of the controller and the plant after convergence should model a dynamical response of the reference model \( M(z) \) (see Fig. 1.2).

1.1.2 Instantaneous Blind Source Separation and Independent Component Analysis

In blind signal processing problems, the mixing and filtering processes of the unknown input sources \( s_j(k) \) \((j = 1, 2, \ldots, n)\) may have different mathematical or physical models, depending on specific applications.

In the simplest case, \( m \) mixed signals \( x_i(k) \) \((i = 1, 2, \ldots, m)\) are linear combinations of \( n \) (typically \( m \geq n \)) unknown mutually statistically independent, zero-mean source signals \( s_j(k) \), and are noise-contaminated (see Fig. 1.3). This can be written as

\[
    x_i(k) = \sum_{j=1}^{n} h_{ij} s_j(k) + \nu_i(k), \quad (i = 1, 2, \ldots, m) \tag{1.1}
\]

or in the matrix notation

\[
    \mathbf{x}(k) = \mathbf{H} \mathbf{s}(k) + \mathbf{\nu}(k), \tag{1.2}
\]

where \( \mathbf{x}(k) = [x_1(k), x_2(k), \ldots, x_m(k)]^T \) is a vector of sensor signals, \( \mathbf{s}(k) = [s_1(k), s_2(k), \ldots, s_n(k)]^T \) is a vector of sources, \( \mathbf{\nu}(k) = [\nu_1(k), \nu_2(k), \ldots, \nu_m(k)]^T \) is a vector of additive noise, and \( \mathbf{H} \) is an unknown full rank \( m \times n \) mixing matrix. In other words, it is assumed that the signals received by an array of sensors (e.g., microphones, antennas, transducers) are weighted sums (linear mixtures) of primary sources. These sources are typically time-varying, zero-mean, mutually statistically independent and totally unknown as is the case of arrays of sensors for communications or speech signals.

In general, it is assumed that the number of source signals \( n \) is unknown unless stated otherwise. It is assumed that only the sensor vector \( \mathbf{x}(k) \) is available and it is necessary to design a feed-forward or recurrent neural network and an associated adaptive learning algorithm that enables estimation of sources, identification of the mixing matrix \( \mathbf{H} \) and/or separating matrix \( \mathbf{W} \) with good tracking abilities (see Fig. 1.3).

The above problems are often referred to as BSS (blind source separation) and/or ICA (independent component analysis): the BSS of a random vector \( \mathbf{x} = [x_1, x_2, \ldots, x_m]^T \) is obtained by finding an \( n \times m \), full rank, linear transformation (separating) matrix \( \mathbf{W} \) such that the output signal vector \( \mathbf{y} = [y_1, y_2, \ldots, y_n]^T \), defined by \( \mathbf{y} = \mathbf{W} \mathbf{x} \), contains components that are as independent as possible, as measured by an information-theoretic cost function such as the Kullback-Leibler divergence or other criteria like sparseness, smoothness or linear predictability. In other words, it is required to adapt the weights \( w_{ij} \) of the \( n \times m \) matrix \( \mathbf{W} \) of the linear system \( \mathbf{y}(k) = \mathbf{W} \mathbf{x}(k) \) (often referred to as a single-layer feed-forward neural network) to combine the observations \( x_i(k) \) to generate estimates of the
Fig. 1.3 Block diagram illustrating the basic linear instantaneous blind source separation (BSS) problem: (a) General block diagram represented by vectors and matrices, (b) detailed architecture. In general, the number of sensors can be larger, equal to or less than the number of sources. The number of sources is unknown and can change in time [265, 276].

source signals

\[ \hat{s}_j(k) = y_j(k) = \sum_{i=1}^{m} w_{ji} x_i(k), \quad (j = 1, 2, \ldots, n). \quad (1.3) \]

The optimal weights correspond to the statistical independence of the output signals \( y_j(k) \) (see Fig 1.3).

**Remark 1.1** In this book, unless otherwise mentioned, we assume that the source signals (and consequently output signals) are zero-mean. A non-zero-mean source can be modeled by zero-mean source with an additional constant source. This constant source can be usually detected but its amplitude cannot be recovered without some a priori knowledge.
There are several definitions of ICA. In this book, depending on the problem, we use different definitions given below.

**Definition 1.1 (Temporal ICA)** The ICA of a noisy random vector \( x(k) \in \mathbb{R}^m \) is obtained by finding an \( n \times m \) (with \( m \geq n \)), full rank separating matrix \( W \) such that the output signal vector
\[
y(k) = [y_1(k), y_2(k), \ldots, y_n(k)]^T
\]
defined by
\[
y(k) = W x(k),
\]
contains the estimated source components \( s(k) \in \mathbb{R}^n \) that are as independent as possible, evaluated by an information-theoretic cost function such as minima of Kullback-Leibler divergence.

**Definition 1.2** For a random noisy vector \( x(k) \) defined by
\[
x(k) = H s(k) + \nu(k),
\]
where \( H \) is an \( m \times n \) mixing matrix, \( s(k) = [s_1(k), s_2(k), \ldots, s_n(k)]^T \) is a source vector of statistically independent signals, and \( \nu(k) = [\nu_1(k), \nu_2(k), \ldots, \nu_m(k)]^T \) is a vector of uncorrelated noise terms, ICA is obtained by estimating both the mixing matrix \( H \) and the independent components \( s(k) = [s_1(k), s_2(k), \ldots, s_n(k)]^T \).

**Definition 1.3** The ICA task is formulated as estimation of all the source signals and their numbers and/or identification of a mixing matrix \( \hat{H} \) or its pseudo-inverse separating matrix \( W = \hat{H}^+ \) assuming only the statistical independence of the primary sources and linear independence of columns of \( H \).

The mixing (ICA) model can be represented in a batch form as
\[
X = HS,
\]
where \( X = [x(1), x(2), \ldots, x(N)]^T \in \mathbb{R}^{m \times N} \) and \( S = [s(1), s(2), \ldots, s(N)]^T \in \mathbb{R}^{n \times N} \). In many applications, especially where the number of ICs is large and they have sparse (or other specific) distributions, it is more convenient to use the following equivalent form:
\[
X^T = S^T H^T.
\]

By taking the transpose, we simply intercache the roles of the mixing matrix \( H = [h_1, h_2, \ldots, h_n] \) and the ICs \( S = [s(1), s(2), \ldots, s(N)]^T \), thus the vectors of the matrix \( H^T \) can be considered as independent components and the matrix \( S^T \) as the mixing matrix and vice-versa. In the standard temporal ICA model, it is usually assumed that ICs \( s(k) \) are time signals and the mixing matrix \( H \) is a fixed matrix without imposing any constraints on its elements. In the spatio-temporal ICA, the distinction between ICs and the mixing matrix is completely abolished \[1114, 598\]. In other words, the same or similar assumptions are made on the ICs and the mixing matrix. In contrast to the conventional ICA the spatio-temporal ICA maximizes the degree of independence over time and space.

**Definition 1.4 (Spatio-temporal ICA)** The spatio-temporal ICA of random matrix \( X^T = S^T H^T \) is obtained by estimating both the unknown matrices \( S \) and \( H \) in such a way that
rows of $S$ and columns of $H$ be as independent as possible and both $S$ and $H$ consist of
the same or very similar statistical properties (e.g., the Laplacian distribution or sparse
representation).

The real-world sensor data often build up complex nonlinear structures, so applying ICA
to global data may lead to poor results. Instead, applying ICA to all available data, we
can preprocess this data by grouping it into clusters or sub-bands with specific features
and then apply ICA individually to each cluster or sub-band separately. The preprocessing
stage of suitable grouping or clustering of data is responsible for an overall coarse nonlinear
representation of the data, while the linear ICA models of individual clusters are used for
describing local features of the data.

**Definition 1.5 (Local ICA)** In local ICA the available sensor data are suitably prepro-
cessed, by grouping them into clusters in space, or in the time, frequency or in the time-
frequency domain, and then applying linear ICA to each cluster locally. More generally, an
optimal local ICA can be implemented as the result of mutual interaction of two processes:
A suitable clustering process and the application of the ICA process to each cluster.

A globally linear model, as implied by conventional ICA, may be insufficient to represent
multivariate data in many situations. A combination of several local ICA’s can provide a
suitable approach in such cases. An important question is then how to find an appropriate
partitioning of the data space together with a proper choice of the local numbers of inde-
dependent components (IC’s).

Despite the success of using standard ICA in many applications, the basic assumptions of
ICA may not hold hence some caution should be taken when using standard ICA to analyze
real world problems, especially in biomedical signal processing. In fact, by definition, the
standard ICA algorithms are not able to estimate statistically dependent original sources,
that is, when the independence assumption is violated. A natural extension and general-
ization of ICA is multiresolution subband decomposition ICA (MSD-ICA) which relaxes
considerably the assumption regarding mutual independence of primarily sources. The key
idea in this approach is the assumption that the wide-band source signals are dependent,
however some narrow band subcomponents are independent In other words, we assume that
each unknown source can be modeled or represented as a sum of narrow-band sub-signals
(sub-components):

$$s_i(k) = s_{i1}(k) + s_{i2}(k) + \cdots + s_{iK}(k).$$

(1.8)

The basic concept of MSD-ICA is to divide the sensor signal spectra into their subspectra
or subbands, and then to treat those subspectra individually for the purpose at hand. The
subband signals can be ranked and processed independently. Let us assume that only a
 certain set of sub-components is independent. Provided that for some of the frequency
subbands (at least one) all sub-components, say $\{s_{ip}(k)\}_{i=1}^n$, are mutually independent or
temporally decorrelated, then we can easily estimate the mixing or separating system (under
condition that these subbands can be identified by some a priori knowledge or detected by
some self-adaptive process) by simply applying any standard ICA algorithm, however not
for all available raw sensor data but only for suitably preprocessed (band pass filtered) sensor signals. Such explanation can be summarized as follows.

**Definition 1.6 (Multiresolution Subband Decomposition ICA)** The MSD-ICA can be formulated as a task of estimation of the mixing matrix \( H \) on the basis of suitable multiresolution subband decomposition of sensors signals and by applying a classical ICA (instead for raw sensor data) for one or several preselected subbands for which source sub-components are independent.

In one of the most simplest cases, source signals can be modeled or decomposed into their low- and high-frequency sub-components:

\[
s_i(k) = s_{iL}(k) + s_{iH}(k) \quad (i = 1, 2, \ldots, n).
\]

In practice, the high-frequency sub-components \( s_{iH}(k) \) are often found to be mutually independent. In such a case in order to separate the original sources \( s_i(k) \), we can use a High Pass Filter (HPF) to extract high frequency sub-components and then apply any standard ICA algorithm to such preprocessed sensor (observed) signals. In the preprocessing stage, more sophisticated methods, such as block transforms, multirate subband filter bank or wavelet transforms, can be applied.

In many blind signal separation problems, one may want to estimate only one or several desired components with particular statistical features or properties, but discard the rest of the uninteresting sources and noises. For such problems, we can define Blind Signal Extraction (BSE) (see Chapter 5 for more detail and algorithms).

**Definition 1.7 (Blind Signal Extraction)** BSE is formulated as a problem of estimation of one source or a selected number of the sources (smaller than \( n \)) with particular desired properties or characteristics, sequentially one by one or “one shot” estimation of a specific group of sources. Equivalently the problem is formulated as an identification of the corresponding vector(s) \( \hat{h}_j \) of the mixing matrix \( \hat{H} \) and/or their pseudo-inverses \( w_j \) which are rows of the separating matrix \( W = \hat{H}^+ \), assuming only the statistical independence of its primary sources and linear independence of columns of \( H \).

**Remark 1.2** It is worth emphasizing that in the literature, BSS/BSE and ICA terms are often confused or interchanged, although they refer to the same or similar models and are solved with the similar algorithms under the assumption that the primary sources are mutually independent. However, in the general case, especially for real-world problems, the objective for ICA and BSS are somewhat different. In fact, the objective of BSS is to estimate the original source signals even if they are not completely mutually statistically independent, while the objective of ICA is to determine a transformation which assures that the output signals are as independent as possible. It should be noted that ICA methods use higher-order statistics (HOS) in many cases, while BSS methods are apt to use only second order statistics (SOS). The second order methods assume that sources have some temporal structure, while the higher order methods assume their mutual independence. Thus, the second statistics methods, generally do not perform independent component analysis. Another difference is that the higher-order statistics methods can not be applied to Gaussian signals
while second order methods do not have such constraints. In fact, BSS methods do not really replace ICA and vice versa, since each approach is based on different criteria, assumptions and often different objectives.

Although many different source separation algorithms are available, their principles can be summarized by the following four approaches (see Fig. 1.4):

- The most popular approach exploits as the cost function some measure of signals independence, non-Gaussianity or sparseness. When original sources are assumed to be statistically independent without a temporal structure, the higher-order statistics (HOS) are essential (implicitly or explicitly) to solve the BSS problem. In such a case, the method does not allow more than one Gaussian source (see Chapters 5 and 6 for more detail).

- If sources have temporal structures, then each source has non-vanishing temporal correlation, and less restrictive conditions than statistical independence can be used, namely, second-order statistics (SOS) are sufficient to estimate the mixing matrix and sources. Along this line, several methods have been developed [1169, 1165, 859, 92]. Note that these SOS methods do not allow the separation of sources with identical power spectra shapes or i.i.d. (independent and identically distributed) sources (see Chapter 4).

- The third approach exploits nonstationarity (NS) properties and second order statistics (SOS). Mainly, we are interested in the second-order nonstationarity in the sense that source variances vary in time. The nonstationarity was first taken into account by Matsuoka et al. [857] and it was shown that a simple decorrelation technique is able to perform the BSS task. In contrast to other approaches, the nonstationarity information based methods allow the separation of colored Gaussian sources with identical
power spectra shapes. However, they do not allow the separation of sources with identical nonstationarity properties. There are some recent works on nonstationary source separation [224, 225, 975] (see Chapters 4, 6, and 8).

• The fourth approach exploits the various diversities of signals, typically, time, frequency, (spectral or “time coherence”) and/or time-frequency diversities, or more generally, joint space-time-frequency (STF) diversity.

Remark 1.3 In fact, the concept of space-time-frequency diversities are widely used in wireless communications systems. Signals can be separated easily if they do not overlap in either the time-, the frequency- or the time-frequency domain (see Fig. 1.3 and Fig. 1.6). When signals do not overlap in the time-domain then one signal stops (is silent) before another one begins. Such signals are easily separated when a receiver is accessible only while the signal of interest is sent. This multiple access method is called TDMA (Time Division Multiple Access). If two or more signals do not overlap in the frequency domain, then they can be separated with bandpass filters as is illustrated in Fig. 1.5. The method based on this principle is called FDMA (Frequency Division Multiple Access). Both TDMA and FDMA are used in many modern digital communication systems [481]. Of course, if the source power spectra overlap, the spectral diversity is not sufficient to extract sources, therefore, we need to exploit other kinds of diversity. If the source signals have different time-frequency diversity and time-frequency signatures of the sources do not (completely) overlap then still they can be extracted from one (or more) sensor signal by masking individual source signals or interference in the time-frequency domain and then synthesized from time-frequency domain as illustrated in Fig. 1.5. However, in such cases some a priori information about source signals is necessary. Therefore, separation is not completely blind but only semi-blind.

More sophisticated or advanced approaches use combinations or integration of all the above mentioned approaches: HOS, SOS, NS and STF (Space-Time-Frequency) diversity, in order to separate or extract sources with various statistical properties and to reduce the influence of noise and undesirable interferences. Methods that exploit either the temporal structure of sources (mainly second-order correlations) and/or the nonstationarity of sources, lead to the second-order BSS methods. In contrast to BSS methods based on HOS, all the second-order statistics based methods do not have to infer the probability distributions of sources or nonlinear activation functions.

1.1.3 Independent Component Analysis for Noisy Data

As the estimation of a separating (unmixing) matrix $W$ and a mixing matrix $\hat{H}$ in the presence of noise is rather difficult; the majority of past research efforts have been devoted to only the noiseless case, where $\nu(k) = 0$. One of the objectives of this book is to present promising novel approaches and associated algorithms that are more robust with respect to noise and/or that can reduce the noise in the estimated output vector $y(k)$. Usually, it is

---

2By diversities we mean usually different characteristics or features of the signals.
assumed that the source signals and additive noise components are statistically independent.

In some models described in this book, it is assumed that sources of additive noise are incorporated as though they were unknown source signals. In other words, the effect of incident noise fields impinging on several sensors may be considered to be equivalent to additional sources, and thus are subject to the same separation process as the desired signals. Of course, there may be more than one noise source. However, for the separation of noise sources, at most one noise source may have a Gaussian distribution, and all other sources must have non-Gaussian distributions. It may well be that one is not interested in separation of the noise sources.

In general, the problem of noise cancellation is difficult and even impossible to treat because we have \((m + n)\) unknown source signals \((n\) sources and \(m\) noise signals, see Fig. 1.5). Various signal processing methods have been developed for noise cancelling.
Fig. 1.6 Illustration of exploiting time-frequency diversity in BSS. (a) Original unknown source signals and available mixed signal (horizontal axis represents time scaled in milliseconds). (b) Time-frequency representation of the mixed signal. Due to non-overlapping time-frequency signatures of the sources by masking and synthesis (inverse transform), we are able to extract the desired sources.

and with some modifications they can be applied to noise cancellation in BSS. In many practical situations, we can measure or model the environmental noise. Such noise is termed reference noise (denoted by $\nu_R$ in Fig1.7). For example, in the acoustic “cocktail party” problem, we can measure or record the environmental noise by using an isolated microphone. In a similar way, noise in biomedical applications can be measured by appro-
Fig. 1.7 Standard model for noise cancellation in a single channel using a nonlinear adaptive filter or neural network.

appropriately placed auxiliary sensors (or electrodes). The noise $\nu_R(k)$ may influence each sensor in some unknown manner due to environmental effects; hence, such effects as delays, reverberations, echo, nonlinear distortions etc. may occur. It may be assumed that the reference noise is processed by some unknown dynamical system before reaching the sensors. In a simple case, a convolutive model of noise is assumed where the reference noise is processed by some FIR filters (see Fig. 1.8). In this case, two learning processes are performed simultaneously: An un-supervised learning procedure performing blind separation and a supervised learning algorithm performing noise reduction [268]. This approach has been successfully applied to the elimination of noise under the assumption that the reference noise is available [268, 674].

In a traditional linear Finite Impulse Response (FIR) adaptive noise cancellation filter, the noise is estimated as a weighted sum of delayed samples of the reference interference. However, the linear adaptive noise cancellation systems mentioned above may not achieve an acceptable level of cancellation of noise in many real world situations when interference signals are related to the measured reference signals in a complex dynamic and nonlinear way.

In many applications, especially in biomedical signal processing, the sensor signals are corrupted by various interference and noise sources. Efficient interference and noise cancellation usually require nonlinear adaptive processing of the observed signals. In this book, we describe various neural network models and associated on-line adaptive learning algorithms for noise and interference cancellation. In particular, we propose to use the Hyper Radial Basis Function Network (HRBFN) with all of its parameters being fully adaptive. Moreover, we examine Amari-Hopfield recurrent neural networks [261]. We study the problem from the perspective of optimal signal estimation and nonlinear adaptive systems. Our mathematical analysis and computer simulations demonstrate that such neural networks
1.1.4 Multichannel Blind Deconvolution and Separation

A single channel convolution and deconvolution process is illustrated in Fig. 1.9

\[
s(k) \quad \text{Convolution} \quad h_p \quad x(k) = \sum_p h_p s(k-p) \quad \text{Deconvolution} \quad w_p \quad y(k) = \sum_p w_p x(k-p)
\]

\[
s(k) \quad \text{Cascade system} \quad g_p = w_p \ast h_p \quad y(k) = \sum_p g_p s(k-p)
\]

Fig. 1.9 Diagram illustrating the single channel convolution and inverse deconvolution process.
A multichannel blind deconvolution problem can be considered as a natural extension or generalization of the instantaneous blind separation problem (see Fig.1.10). In the multidimensional blind deconvolution problem, an \( m \)-dimensional vector of received discrete-time signals \( \mathbf{x}(k) = [x_1(k), x_2(k), \ldots, x_m(k)]^T \) at time \( k \) is assumed to be produced from an \( n \)-dimensional vector of source signals \( \mathbf{s}(k) = [s_1(k), s_2(k), \ldots, s_n(k)]^T \), \( m \geq n \), by using a stable mixture model \[39, 21, 254, 615\]

\[
\mathbf{x}(k) = \sum_{p=-\infty}^{\infty} \mathbf{H}_p \mathbf{s}(k-p) = \mathbf{H}_k * \mathbf{s}(k), \quad \text{with} \quad \sum_{p=-\infty}^{\infty} \|\mathbf{H}_p\| < \infty, \quad (1.10)
\]

where \( * \) denotes the convolution operator and \( \mathbf{H}_p \) is an \((m \times n)\) matrix of mixing coefficients at time-lag \( p \).
Define
\[ H(z) = \sum_{p=-\infty}^{\infty} H_p z^{-p} \]  
(1.11)

where \( z^{-1} \) denotes the unit time-delay (backward shift) operator (i.e. \( z^{-p}[s_i(k)] = s_i(k-p) \)). It should be noted that if \( z \) is replaced with the complex variable \( \tilde{z} = \exp(-\sigma + j\omega T) \), then \( H(\tilde{z}) \) is the \( Z \)-transform of \( \{H_p\} \), i.e., it is the system matrix transfer function \([380][481]\).

Using (1.11), (1.10) may be rewritten as
\[ x(k) = [H(z)] \mathbf{s}(k). \]  
(1.12)

The goal of multichannel deconvolution is to calculate the possibly scaled and time-delayed (or filtered) versions of the source signals from the received signals by using approximate knowledge of the source signal distributions and statistics. Typically, every source signal \( s_i(k) \) is an i.i.d. (independent and identically-distributed) sequence that is independent of all the other source sequences.

In order to recover the source signals, we can use the neural network models depicted in Fig.1.3 (b) and Fig.1.10 but the synaptic weights should be generalized to filters (e.g., FIR or IIR) as is illustrated in Fig.1.11. In this book, many such extensions and generalizations are described.

Let us consider briefly one example of such a generalization: A standard multichannel blind deconvolution where each weight \([38][217][615]\)
\[ W_{ji}(z, k) = \sum_{p=0}^{M} w_{jip}(k) z^{-p} \]  
(1.13)

is described by a multichannel finite-duration impulse response (FIR) adaptive filter at discrete-time \( k \) \([615][660]\).

We will consider a stable feed-forward model that estimates the source signals directly by using a truncated version of a doubly-infinite multichannel equalizer of the form \([615]\) (see Fig.1.11 (a))
\[ y_j(k) = \sum_{i=1}^{m} \sum_{p=-\infty}^{\infty} w_{jip} x_i(k-p), \quad (j = 1, 2, \ldots, n) \]  
(1.14)

or in the compact matrix form as
\[ y(k) = \sum_{p=-\infty}^{\infty} W_p(k) x(k-p) = W_p(k) \ast x(k) = [W(z, k)] x(k), \]  
(1.15)

where \( y(k) = [y_1(k), y_2(k), \ldots, y_n(k)]^T \) is an \( n \)-dimensional vector of outputs and \( W(k) = \{W_p(k), -\infty \leq p \leq \infty\} \) is a sequence of \( n \times m \) coefficient matrices used at time \( k \), and the matrix transfer function is given by
\[ W(z, k) = \sum_{p=-\infty}^{\infty} W_p(k) z^{-p}. \]  
(1.16)
The goal of adaptive blind deconvolution or equalization is then to adjust $W(z, k)$ such that the global system be described as

$$\lim_{k \to \infty} G(z, k) = W(z, k) H(z) = P D(z),$$

where $P$ is an $n \times n$ permutation matrix, $D(z)$ is an $n \times n$ diagonal matrix whose $(i, i)$-th entry is $c_i z^{-\Delta_i}$, $c_i$ is a non-zero scalar factor, and $\Delta_i$ is an integer delay. We assume that both $H(z)$ and $W(z, k)$ are stable with non-zero eigenvalues on the unit circle $|z| = 1$. In addition, the derivatives of quantities with respect to $W(z, k)$ can be understood as a series of matrices indexed by the lag $p$ of $W_p(k)$.

Fig. 1.11(b) and (c) show alternative neural network models with the weights in the form of stable constrained infinite impulse response (IIR) filters. In these models, the weights

---

**Fig. 1.11** Basic models of synaptic weights for the feed-forward adaptive system (neural network) shown in Fig. 1.3: (a) Basic FIR filter model, (b) Gamma filter model, (c) Laguerre filter model.
Fig. 1.12  Block diagram illustrating the sequential blind extraction of sources or independent components. Synaptic weights $w_{ij}$ can be time-variable coefficients or adaptive filters (see Fig. 1.11).

$W_{ji}$ are generalized to real- or complex-valued Gamma [999, 998] or Laguerre filters (see Fig. 1.11(b) and (c)) or other structures such as state-space models (see Fig. 1.13) which may have some useful properties [31, 1368, 1384]. In all these models, it is assumed that only the sensor vector $x(k)$ is available and it is necessary to design a feed-forward or recurrent neural network and an associated adaptive learning algorithm that enables estimation of the source signals.

1.1.5 Blind Extraction of Signals

There are two main approaches to solve the problem of blind separation and deconvolution. The first approach, which was mentioned briefly in previous sections, is to simultaneously separate all sources. In the second one, we extract sources sequentially in a blind fashion, one by one, rather than separating them all simultaneously. In many applications, a large number of sensors (electrodes, microphones or transducers) are available but only a very few source signals are subjects of interest. For example, in the EEG or MEG devices, we observe typically more than 64 sensor signals, but only a few source signals are interesting; the rest can be considered as interfering noise. In another example, the cocktail party problem, it is usually essential to extract the voices of specific persons rather than separate all the source signals available from a large array of microphones. For such applications it is essential to develop reliable, robust and effective learning algorithms which enable us to extract only a small number of source signals that are potentially interesting and contain useful information (see Fig. 1.12). This problem is the subject of Chapter 5. The blind signal extraction approach may have several advantages over simultaneous blind separation/deconvolution, such as.

- Signals can be extracted in a specified order according to the statistical features of the source signals, e.g., in the order determined by absolute values of generalized normalized kurtosis. Blind extraction of sources can be considered as a generalization of PCA (principal components analysis), where decorrelated output signals are extracted according to the decreasing order of their variances.
• Only “interesting” signals need to be extracted. For example, if the source signals are mixed with a large number of Gaussian noise terms, we may extract only specific signals which possess some desired statistical properties.

• The available learning algorithms for BSE are purely local and biologically plausible. In fact, the learning algorithms derived below can be considered as extensions or modifications of the Hebbian/anti-Hebbian learning rule. Typically, they are simpler than those of instantaneous blind source separation.

In summary, blind signal extraction is a useful approach when our objective is to extract several source signals with specific statistical properties from a large number of mixtures. Extraction of a single source is closely related to the problem of blind deconvolution. In blind signal extraction (BSE), our objective is to extract the source signals sequentially, i.e. one by one, rather than to separate all of them simultaneously. This procedure is called the sequential blind signal extraction in contrast to the simultaneous blind signal separation (BSS). Sequential blind signal extraction can be performed by using a cascade neural network similar to the one used for the extraction of principal components. However, in contrast to PCA, the optimization criteria for BSE are different. A single processing unit (artificial neuron) is used in the first step to extract one source signal with specified statistical properties. In the next step, a deflation technique can be used to eliminate the already extracted signals from the mixtures.

1.1.6 Generalized Multichannel Blind Deconvolution – State Space Models

In the general case, linear dynamical mixing and demixing systems can be described by state-space models. In fact, any stable mixing dynamical system can be described as (see Fig. 1.13)

\[
\begin{align*}
\xi(k+1) &= \bar{A} \xi(k) + \bar{B} s(k) + \bar{N} \nu_P(k), \\
x(k) &= \bar{C} \xi(k) + \bar{D} s(k) + \nu(k),
\end{align*}
\]

(1.18) \hspace{1cm} (1.19)

where \( \xi \in \mathbb{R}^r \) is the state vector of the system, \( s(k) \in \mathbb{R}^n \) is a vector of unknown input signals (assumed to be zero-mean, non-Gaussian independent and identically distributed (i.i.d.) and mutually (spatially) independent), \( x(k) \) is an available vector of sensor signals, \( \nu_P(k) \) is the vector of process noise, \( \nu(k) \) is the vector of output noise, and the state matrices have dimensions: \( \bar{A} \in \mathbb{R}^{r \times r} \) is a state matrix, \( \bar{B} \in \mathbb{R}^{r \times n} \) an input mixing matrix, \( \bar{C} \in \mathbb{R}^{m \times r} \) an output mixing matrix, \( \bar{D} \in \mathbb{R}^{m \times n} \) an input-output mixing matrix and \( \bar{N} \in \mathbb{R}^{r \times p} \) is a noise matrix. The transfer function is an \( m \times n \) matrix of the form

\[
H(z) = \bar{C} (z I - \bar{A})^{-1} \bar{B} + \bar{D},
\]

(1.20)

where \( z^{-1} \) is a delay operator (i.e., \( z^{-1} x(k) = x(k - 1) \)).

Analogously, we can assume that the demixing model is another linear state-space system described as (see Fig. 1.13)

\[
\begin{align*}
\xi(k+1) &= A \xi(k) + B x(k) + L \nu_R(k), \\
y(k) &= C \xi(k) + D x(k),
\end{align*}
\]

(1.21) \hspace{1cm} (1.22)
where the unknown state-space matrices, respectively have the dimension: $A \in \mathbb{R}^{M \times M}$, $B \in \mathbb{R}^{M \times m}$, $C \in \mathbb{R}^{n \times M}$, $D \in \mathbb{R}^{n \times m}$, $L \in \mathbb{R}^{M \times m}$, with $M \geq r$ (i.e., the order of the demixing system should be at least the same or larger than the order of the mixing system). It is easy to see that the linear state-space model is an extension of the instantaneous blind source separation model. In the special case when the matrices $\overline{A}, \overline{B}, \overline{C}$ in the mixing model and $A, B, C$ in the demixing model are null matrices, the problem is simplified to the standard ICA problem. In general, the matrices $\Theta = [A, B, C, D, L]$ are parameters to be determined in a learning process on the basis of knowledge of the sequence $x(k)$ and some a priori knowledge about the system. The transfer function of the demixing model is $W(z) = C(zI - A)^{-1}B + D$. We formulate the dynamical blind separation problem as a task to recover original source signals from the observations $x(k)$ without a priori knowledge of the source signals or the state-space matrices $[\overline{A}, \overline{B}, \overline{C}, \overline{D}]$, by assuming, for example, that the sources are mutually independent, zero-mean signals. Other assumptions
such as smoothness or linear predictability of sources can also be used. We also usually assume that the output signals $y(k) = [y_1(k), y_2(k), \ldots, y_n(k)]^T$ will recover the source signals for the noiseless case in the following sense

$$y(k) = [W(z)H(z)]s(k) = [D(z)]P s(k), \quad (1.23)$$

where $P$ is an $n \times n$ generalized permutation matrix which consists of $n$ nonzero elements and only one nonzero element in each column and $D(z) = \text{diag}\{D_{11}(z), D_{22}, \ldots, D_{nn}(z)\}$ is a diagonal matrix with transfer functions $D_{ii}(z)$ of shaping filters. In some applications, such as equalization problems, it is required that $D_{ii}(z) = \lambda_i z^{-\tau_i}$, where $\lambda_i$ is a non-zero constant scaling factor and $\tau_i$ is an positive integer delay (i.e., constant scaling factors and/or pure delays are only acceptable).

A question arising here is whether matrices $[A, B, C, D]$ exist for the demixing model shown in Fig.1.13 such that the transfer function $W(z)$ satisfies (1.23). The answer is affirmative [1365] [1384] [1385]. It will be shown later that if there is a filter $W_*(z)$, which is the inverse of $H(z)$ in the sense of (1.23), then for the given specific matrices $[A, B]$, there are matrices $[C, D]$, such that the transfer matrix $W(z)$ satisfies equation (1.23).

**Remark 1.4** It should be noted that in general case, we can assume that $D$ is an $m \times m$ square matrix, i.e., the number of outputs of the system is equal to the number of sensors, although in practice the number of sources can be less than the number of sensors ($m \geq n$). Such a model is justified by two facts. First of all, the number of sources is generally unknown and may change over time. Secondly, in practice we have additive noise signals that can be considered as auxiliary unknown sources; therefore, it is also reasonable to extract these noise signals. In the ideal noiseless case, the redundant $(m-n)$ output signals $y_j$ should decay to zero during adaptive learning process and then only $n$ outputs will correspond to the recovered sources.

### 1.1.7 Nonlinear State Space Models – Semi-Blind Signal Processing

The above linear state-space demixing and filtering model is relatively easy to generalize into a flexible nonlinear model as (see Fig.1.14)

$$\xi(k) = f[x(k), \xi(k)], \quad (1.24)$$

$$y(k) = C(k) \xi(k) + D(k) x(k), \quad (1.25)$$

where $\xi(k) = [\xi_1(k), \xi_2(k), \ldots, \xi_M(k)]^T$ is the state vector, $x(k) = [x_1(k), x_2(k), \ldots, x_m(k)]^T$ is an available vector of sensor signals, $f[\xi(k), \xi(k)]$ is an $M$-dimensional vector of nonlinear functions (with $\underline{x}(k) = [x^T(k), x^T(k), \ldots, x^T(k-L_x)]^T$, $\underline{\xi}(k) = [\xi^T(k), \xi^T(k-1), \ldots, \xi^T(k-L_x)]^T$, $y(k) = [y_1(k), y_2(k), \ldots, y_n(k)]^T$ is the vector of output signals, and $C \in \mathbb{R}^{M \times n}$ and $D \in \mathbb{R}^{n \times m}$ are output matrices. It should be noted that equation (1.24) describes the nonlinear autoregressive moving average (NARMA) model while the output model (1.25) is linear. Our objective will be to estimate the output matrices $C$ and $D$, as well as to identify the NARMA model by using a neural network on the basis of sensor signals $x(k)$ and source (desired) signals $s(k)$ (which are available for short-time windows).
In order to solve this challenging and difficult problem, we attempt to apply a semi-blind approach, i.e., we combine both supervised and un-supervised learning algorithms. Such an approach is justified in many practical applications. For example, for MEG or EEG, we can use a phantom of the human head with known artificial source excitations located in specific places inside of the phantom. For the cocktail party problem, in some case is possible to record for short-time windows original test speech sources. These short-time window training sources enable us to determine, on the basis of a supervised algorithm, a suitable nonlinear demixing model and associated nonlinear basis functions of the neural network and their parameters.

However, we assume that the mixing system is a slowly time-varying system for which some parameters fluctuate slightly over time, mainly due to the change in localization of source signals in space. Furthermore, we assume that training sources are available only for short-time slots. During the time windows in which the training signals are not available, we can apply an unsupervised learning algorithm which performs a fine adjustment of the output matrices $C$ and $D$ (by keeping the nonlinear model fixed). In this way, we will be able to estimate continuously in time the source signals. An exemplary implementation of the nonlinear state-space model using the radial basis function (RBF) neural network is shown in Figure 1.15 (see Chapter 12 for detail).

### 1.1.8 Why State Space Demixing Models?

There are several essential reasons why the state-space models provide a useful and powerful approach in blind signal processing:

- The mixing and filtering processes of unknown input sources $s_j(k), \ (j = 1, 2, ..., n)$ may have different mathematical or physical models, depending on specific applications. The state-space demixing model is a flexible and universal linear model which
describes a wide class of stable dynamical systems including standard multichannel deconvolution models with finite impulse response (FIR) filters, Gamma filters or more general models: AR (autoregressive), MA (moving average) and ARMA (autoregressive moving average) models as special cases.

• Moreover, such a dynamical demixing model enables us to generate many canonical realizations of the same dynamical system by using equivalent transformations.

• It is easy to note that the linear state-space model is an extension of the instantaneous mixture blind source separation model.

• State-space models have two subsystems: A linear, memoryless output layer and a dynamical linear or nonlinear recurrent network, which can be identified or updated using different approaches [290, 291, 292, 1374].

1.2 POTENTIAL APPLICATIONS OF BLIND AND SEMI-BLIND SIGNAL PROCESSING

The problems of independent component analysis (ICA), blind separation and multichannel deconvolution of source signals have received wide attention in various fields such as
biomedical signal analysis and processing (EEG, MEG, ECG), geophysical data processing, data mining, speech enhancement, image recognition and wireless communications. In such applications a number of observations of sensor signals or data that are filtered superpositions of separate signals from different independent sources are available, and the objective is to process the observations in such a way that the outputs correspond to the separate primary source signals.

Acoustic applications are considered in situations where signals, from several microphones in a sound field produced by several speakers (the so-called cocktail-party problem) or from several acoustic transducers in an underwater sound field produced by engine noises of several ships (sonar problem) need to be processed. Radio and wireless communication examples include the observations corresponding to outputs of antenna array elements in response to several transmitters, and the observations may also include the effects of the mutual couplings of the elements. Other radio communication examples include the use of polarization multiplexing in microwave links. The maintenance of the orthogonality of the polarization cannot be perfect and there is still interference between the separate transmissions. Radar examples include the superposition of signals from different target modulating mechanisms as observed by multiple receivers whose elements are sensitive to different polarizations.

Let us consider some exemplary promising biomedical applications in more detail.

1.2.1 Biomedical Signal Processing

A great challenge in biomedical engineering is to non-invasively assess the physiological changes occurring in different internal organs of the human body (Figure 1.16(a)). These variations can be modeled and measured often as biomedical source signals that indicate the function or malfunction of various physiological systems. To extract the relevant information for diagnosis and therapy, expert knowledge in medicine and engineering is also required.

Biomedical source signals are usually weak, nonstationary signals and distorted by noise and interference. Moreover, they are usually mutually superimposed. Besides classical signal analysis tools (such as adaptive supervised filtering, parametric or non-parametric spectral estimation, time-frequency analysis, and higher-order statistics), intelligent blind signal processing techniques (IBSP) can be used for preprocessing, noise and artifact reduction, enhancement, detection and estimation of biomedical signals by taking into account their spatio-temporal correlation and mutual statistical dependence.

One successful and promising application domain of blind signal processing includes those biomedical signals acquired with multi-electrode devices: Electrocardiography (ECG), electromyography (EMG), electroencephalography (EEG) and magnetoencephalography (MEG).

Exemplary applications in biomedical problems include the following:

- Fetal electrocardiogram (ECG) extraction, i.e., removing/filtering maternal electrocardiogram signals and noise from fetal electrocardiogram signals.
- Enhancement of low-level ECG components.
- Separation of transplanted heart signals from residual original heart signals.
• Separation of heart sounds from gastrointestinal acoustic phenomena (bowel-sounds). Bowel sounds can be measured in a non-invasive way by using microphones or accelerometers positioned on the skin.

• Reduction or blind separation of heart sounds from lung sounds using multichannel blind deconvolution.

• Cancellation of artifacts and noise from electroencephalographic and magnetoencephalographic recordings.

• Enhancement of evoked potentials (EP) and categorization of detected brain signals. (The brain potentials evoked by sensory stimulations such as visual, acoustic or somatosensory are generally called evoked potentials).

• Detection and estimation of sleep-spindles. (Sleep-spindles are specific phenomena of electroencephalograms (EEG) appearing during sleep; they are characterized by a group of oscillations in the range 11.5-15 Hz).

• Decomposition of brain sources as independent components and then localizing them in time and space.

Let us consider in more detail, some exemplary promising biomedical applications.

1.2.2 Blind Separation of Electrocardiographic Signals of Fetus and Mother

The mechanical action of the heart muscles is stimulated by electrical depolarization and repolarization signals. These quasi-periodical signals project potential differences to the skin level which can be measured and visualized as functions of time using electrocardiogram (ECG). As for adults, it would also be possible to measure the electrical activity of a fetal heart [729,731]. The characteristics of a fetal electrocardiogram (FECG) can be very useful for determining if a fetus is developing or being delivered properly. These characteristics include an elevated heart rate that indicates fetal stress, cardiac arrythmia and ST segment depression which may indicate acidosis.

It is a non-trivial task to obtain an accurate and reliable FECG in a non-invasive fashion by using several electrodes. Problems develop due to the facts that the electrocardiogram (ECG) also contains a maternal electrocardiogram (MECG) which can be from one-half to one-thousandth the magnitude of the MECG. Moreover, the FECG will occasionally overlap the MECG and make it normally impossible to detect. Along with the MECG, extensive electromyographic (EMG) noise also interferes with the FECG and it can completely mask the FECG. The separation of fetal and maternal electrocardiograms from skin electrodes located on a pregnant woman’s body may be modeled as a Blind Signal Processing problem (see Figure 1.10). The recordings pick up a mixture of FECG, MECG contributions, and other interferences, such as maternal electromyogram (MEMG), power supply interference, thermal noise from the electrodes and other electronic equipment. In fact, BSP techniques can be successfully applied to efficiently solve this problem and the first results are very promising [231,233,888]. Ordinary filtering and signal processing techniques have great difficulties with this problem [1295].
Fig. 1.16 Exemplary biomedical applications of blind signal processing: (a) A multi-recording monitoring system for blind enhancement of sources, cancellation of noise, elimination of artifacts and detection of evoked potentials, (b) blind separation of the fetal electrocardiogram (FECG) and maternal electrocardiogram (MECG) from skin electrode signals recorded from a pregnant woman, (c) blind enhancement and independent components analysis of multichannel electromyographic (EMG) signals.
1.2.3 Enhancement and Decomposition of EMG Signals

The movement and positioning of limbs are controlled by electrical signals travelling back and forth between the central nervous system and the muscles. Electromyography is a technique of recording the electrical signals in the muscle (muscle action potentials). Electromyographic (EMG) signals recorded by a multi-electrode system provide important information about the brain motor system and the diagnosis of neuromuscular disorders that affect the brain, spinal cord, nerves or muscles. EMG signals, which are recorded simultaneously by several electrodes at low and moderate force levels can be composed of motor unit action potentials (MUAPs) generated by different motor units. The motor unit is the smallest functional unit of the muscle that can be voluntarily activated: It consists of a group of muscle fibers all innervated by the same motor neuron. In other words, MUAP consists of the spatial and temporal summation of all single fiber potentials innervated by the same motor neuron. The MUAP waveforms give information about the structural organization of the motor units [1098].

Blind signal processing techniques can be used for the enhancement of EMG signals. A more challenging problem is to apply BSS for decomposition of EMG signals into independent components and MUAPs. Such blind or semi-blind processing may be able to cluster MUAPs into groups of similar waveforms and provide important information about the brain motor system thus facilitating the assessment of neuromuscular pathology.

1.2.4 EEG and MEG Data Processing

Applications of BSP show special promise in the areas of non-invasive human brain imaging techniques to delineate the neural processes that underlie human cognition and sensorimotor functions.

To understand human neurophysiology, we rely on several types of non-invasive neuroimaging techniques. These techniques include electroencephalography (EEG), magnetoencephalography (MEG), anatomical magnetic resonance imaging (MRI) and functional MRI (fMRI). While each of these techniques is useful, there is no single technique that provides both the spatial and temporal resolution necessary to make inferences about the intracranial brain sources of activity.

Very recently, several research groups have demonstrated that the techniques and methods of blind source separation (BSS) are related to those currently used in electromagnetic source localization (ESL) [845]. This framework provides a methodology by which several different types of information can be combined to aid in making inferences about a problem. Neural activity in the cerebral cortex generates small electric currents which create potential differences on the surface of the scalp (detected by EEG) as well as very small magnetic fields which can be detected using SQUIDs (SuperConducting QUantum Interference Devices). The greatest benefit of MEG is that it provides information that is complementary to EEG. In addition, the magnetic fields (unlike the electric currents) are not distorted by the intervening biological mass. Under certain circumstances, this allows precise localization of the neural currents responsible for the measured magnetic field.

Here, we give a very brief introduction to EEG and MEG [1259, 1260]. When a region of neural tissue (consisting of about 100,000 neurons) is synchronously active, detectable
extracellular electric currents and magnetic fields are generated. These regions of activity can be modeled as “current dipoles” because they generate a dipolar electric current field in the surrounding volume of the head. These extracellular currents flow throughout the volume of the head and create potential differences on the surface of the head that can be detected with surface electrodes in a procedure called electroencephalography (EEG). One can also place super-conducting coils above the head and detect the magnetic fields generated by the activity in a procedure called magnetoencephalography (MEG).

If one knows the positions and orientations of the sources in the brain, one can calculate the patterns of electric potentials or magnetic fields on the surface of the head. This is called the forward problem. If otherwise one has only the patterns of electric potential or magnetic fields, then one needs to calculate the locations and orientations of the sources. This is called the inverse problem. Inverse problems are notoriously more difficult to solve than forward problems. In this case, given only the electric potentials and magnetic fields on the surface, there is no unique solution to the problem. The only hope is that there is some additional information available that can be used to constrain the infinite set of possible solutions to a single unique solution. This is where intelligent blind signal processing will be used.

The idea is that one must use all the available information to solve the problem. We will demonstrate this by focusing on an inverse problem, where we have information delivered from one or several devices, say EEG and/or MEG.

In Figure 1.17 we depicted three neural sources, represented in this case by equivalent current dipoles, in the cortical gray matter of the brain. The electrodes on the surface of the head detect the potential differences due to the extracellular currents generated by these active sources. The arrows merely demonstrate that each electrode detects some of the current flow from each neural source. The currents do not flow directly from the sources to the electrodes, but instead they flow throughout the volume of the entire head.

Determining active regions of the brain, given EEG/MEG measurements on the scalp is an important problem. An accurate and reliable solution to such a problem can give
information about higher brain functions and patient-specific cortical activity. However, estimating the location and distribution of electric current sources within the brain from EEG/MEG recording is an ill-posed problem, because there is no unique solution and the solution does not depend continuously on data. The ill-posedness of the problem and distortion of sensor signals by large noise sources makes finding a correct solution a challenging analytic and computational problem.

The ICA approach and blind signal extraction methods are promising techniques for the extraction of useful signals from the EEG/MEG recorded raw data. The EEG/MEG data can be first decomposed into useful signal and noise subspaces using standard techniques like local and robust PCA, SVD and nonlinear adaptive filtering. Next, we apply ICA algorithms to decompose the observed signals (signal subspace) into independent components. The ICA approach enables us to project each independent component (independent “brain source”) onto an activation map at the skull level. For each activation map, we can apply an EEG/MEG source localization procedure, looking only for a single dipole (or 2 dipole) per map. By localizing multiple dipoles independently, we can dramatically reduce the complexity of the computation and increase the likelihood of efficiently converging to the correct and reliable solution.

Figure 1.18 illustrates an example of a promising application of blind source separation and independent component analysis (ICA) algorithms for localization of the brain source signals activated after the auditory and somatosensory stimuli were applied simultaneously. In the MEG experiments performed in collaboration with the Helsinki University of Technology, Finland, the stimulus presented to the subject was produced with a sub-woofer, and the acoustic energy was transmitted to the shielded-room via a plastic tube with a balloon on the end [265]. The subject had his hands in contact with the balloon and sensed the vibration. In addition, the subject listened to the sound produced by the sub-woofer that provided auditory stimulation. Using ICA, we successfully extracted auditory and somatosensory evoked fields (AEF and SEF, respectively) and localized the corresponding brain sources [265] (see Figure 1.18).

1.2.5 Application of ICA/BSS for Noise and Interference Cancellation in Multi-sensory Biomedical Signals

The nervous systems of humans and animals must encode and process sensory information within the context of noise and interference, and the signals which are encoded (the images, sounds, etc.) have very specific statistical properties. One of the challenging tasks is how to reliably detect, enhance and localize very weak, nonstationary brain source signals corrupted by noise (e.g., evoked and event related potentials EP/ERP) by using EEG/MEG data.

Independent Component Analysis (ICA) and related methods like Adaptive Factor Analysis (AFA) are promising approaches for elimination of artifacts and noise from EEG/MEG data [260, 634]. In fact, for these applications, ICA/BSS techniques have been successfully applied to remove artifacts and noise including background brain activity, electrical activity of the heart, eye-blink and other muscle activity, and environmental noise efficiently.
However, most of the methods require manual detection, classification of interference components and the estimation of the cross-correlation between independent components and the reference signals corresponding to specific artifacts. One of the important problems is how to automatically detect, extract and eliminate noise and artifacts. Another related problem is how to classify independent “brain sources” and artifacts. The automatic on-line elimination of artifacts and other interference sources is especially important for extended recordings, e.g., EEG/MEG recording during sleep.

Evoked potentials (EPs) of the brain are meaningful for clinical diagnosis and they are important factors in understanding higher order mechanisms in the brain. The EPs are usually embedded within the ongoing EEG/MEG with a signal to noise ratio (SNR) less than 0 dB, making them very difficult to extract by using only a single trial. The traditional method of EPs extraction uses ensemble averaging to improve the SNR. This often requires hundreds or even thousands of trials to obtain a usable waveform. Therefore, it is important to develop novel techniques that can rapidly improve the SNR and reduce the number of trials required to a minimum. Traditional signal processing techniques, such as Wiener filtering, adaptive noise cancellation, latency-corrected averaging and invertible wavelet transform filtering, have recently been proposed for SNR improvements and ensemble reduction. However, these methods require a priori knowledge pertaining to the
nature of the signal \cite{539,1147}. Since EP signals are known to be nonstationary, sparse and changing their characteristics from trial to trial, it is essential to develop novel algorithms for enhancement of single trial EEG/MEG noisy data.

The formulation of the problem can be given in the following form: Denote by $x(k) = [x_1(k), x_2(k), ..., x_m(k)]^T$ the observed $m$-dimensional vector of noisy signals that must be “cleaned” from the noise and interference. Here we have two types of noise. The first is so called “inner” noise generated by some primary sources that cannot be observed directly but contained in the observations. They are mixtures of useful signals and random noise signals or other undesirable sources. The second type of noise is the sensor additive noise (observation errors) at the output of the measurement system. This noise is not directly measurable, either. Formally, we can write that an observed $m$-dimensional vector of sensor signals $x(k)$ is a mixture of source signals plus observation errors

$$x(k) = Hs(k) + \nu(k),$$

where $k = 0, 1, 2, ...$ is a discrete-time index; $H$ is a full rank $(m \times n)$ mixing matrix; $s(k) = [s_1(k), s_2(k), ..., s_n(k)]^T$ is an $n$-dimensional vector of sources containing useful signals and $\nu(k)$ is an $m$-dimensional vector of additive white noise. We also assume that some useful sources are not necessarily statistically independent. Therefore, we cannot achieve perfect separation of primary sources by using any ICA procedure. However, our purpose here is not the separation of the sources but the removal of independent or uncorrelated noisy sources.

Let us emphasize that the problem consists of cancellation of the noise sources and reduction of observation errors based only on information about observed vector $x(k)$.

A conceptual model for elimination of noise and other undesirable components from multi-sensory data is depicted in Figure 1.19. Firstly, ICA is performed using any robust (with respect to Gaussian noise) algorithm \cite{25,31,261,262,859} by a linear transformation of sensory data as $y(k) = Wx(k)$, where the vector $y(k)$ represents independent components. However, robust ICA methods allow us only to obtain an unbiased estimate of the unmixing matrix $W$. Furthermore, due to memoryless structure such methods by definition, cannot remove the additive noise. Noise removal can be performed using optional nonlinear adaptive filtering and nonlinear noise shaping (see Figure 1.20). In the next stage, we classify independent signals $\hat{y}_j(k)$ and then remove noise and undesirable components by switching corresponding switches “off”.

The projection of interesting or useful independent components (e.g., independent activation maps) $\hat{y}_j(k)$ back onto the sensors (electrodes) can be done by the transformation $\hat{x}(k) = W^+\hat{y}(k)$, where $W^+$ is the pseudo-inverse of the unmixing matrix $W$. In the typical case, where the number of independent components is equal to the number of sensors, we have $W^+ = W^{-1}$.

The standard adaptive noise and interference cancellation systems may be subdivided into the following classes \cite{555,557}:

1. **Noise cancellation** (see Figure 1.20). This term is normally referred to the case, when we have both the primary signal $y_j(k) = \hat{y}_j(k) + n_j(k)$ contaminated with noise and reference noise $n_j(k)$, which is correlated with the noise $n_j(k)$ but is independent of the primary signal $\hat{y}_j(k)$. By feeding the reference signal to the linear adaptive filter
Fig. 1.19  Conceptual models for removing undesirable components like noise and artifacts and enhancing multi-sensory (e.g., EEG/MEG) data: (a) Using expert decision and hard switches, (b) using soft switches (adaptive nonlinearities in time, frequency or time-frequency domain), (c) using nonlinear adaptive filters and hard switches \[287\] 163.
we are able to estimate or reconstruct the noise, then subtract it from the primary signal and thereby enhance the signal to noise ratio.

2. Deconvolution-reverberation and echo cancelling. This kind of interference cancelling is often referred to as echo cancelling, because it enables the removal of reverberations and echo from a single observed signal. A delayed version of the primary input signal is fed to the linear adaptive filter thus enabling the filter to reconstruct and remove reverberation from the primary signal. The deconvolver may also be used to cancel periodic interference components in the primary input such as power line interference, etc. The adaptive filter is able to extrapolate the periodic interference and subtract this component from the undelayed primary input (see Figure 1.20). This approach normally provides superior performance compared to standard notch or comb filtering techniques.

3. Line enhancement. In this case the objective is to estimate or extract a periodic or quasi periodic signal buried in noise. The adaptive filter receives the same input as the deconvolver, however, instead of subtracting the extrapolated periodic signal from the input, it outputs directly the enhanced signal (see Fig. 1.20).

4. Adaptive bandpass filtering. Often we may take advantage of some a priori knowledge regarding the bandwidth of the signal we wish to denoise. By bandpass filtering of the signal, we eliminate a part of the frequency range where the useful signal is weak and the noise is comparatively strong, thus enhancing the overall signal to noise ratio.

In a traditional linear Finite Impulse Response (FIR) adaptive noise cancellation filter, the noise is estimated as a weighted sum of the delayed samples of reference interference. However, for many real world problems (when interference signals are related to the measured reference signals in a complex dynamic and nonlinear way) the linear adaptive noise cancellation systems mentioned above may not achieve acceptable levels of noise cancellation. Optimum interference and noise cancellation usually requires nonlinear adaptive processing of the recorded and measured on-line signals [265, 268].

A common technique for noise reduction is to split the signal in two or more bands. The high-pass bands are subjected to a threshold nonlinearity that suppresses low amplitude values while retaining high amplitude values (see Fig. 1.20) [558, 554]. In addition to denoising and artifacts removal, ICA/BSS techniques can be used to decompose EEG/MEG data into separate components, each representing a physiologically distinct process or brain source. The main idea here is to apply localization and imaging methods to each of these components in turn. The decomposition is usually based on the underlying assumption of statistical independence between the activation of different cell assemblies involved. An alternative criterion for decomposition is temporal predictability or smoothness of components. These approaches lead to interesting and exciting new ways of investigating and analyzing brain data and developing new hypotheses how the neural assemblies communicate and process information. This is actually a very extensive and potentially promising research area, however these approaches still remain to be validated at least experimentally.
1.2.6 Cocktail Party Problem

The “cocktail party” problem can be described as the ability to focus one’s listening attention on a single talker among a cacophony of conversations and background noise. This problem has long been recognized as an interesting and challenging problem. Also known as the “cocktail party effect” or more technically, “multichannel blind deconvolution”, the problem of separating a set of mixtures of convolved (filtered) signals, detected by an array of microphones, into their original source signals is performed extremely well by the human brain. Over the years attempts have been made to capture this function by using assemblies of abstracted neurons or adaptive processing units.

Humans are able to concentrate on listening to one voice in the midst of other conversations and noise, but not all the mechanisms for this process are completely understood. This specialized listening ability may be because of characteristics of the human speech production system, auditory system, or high-level perceptual and language processing.

In the EEG/MEG brain source separation algorithms, we make the fundamental assumption that the recorded signals form an instantaneous mixture, meaning that all of the signals are time-aligned so that they enter the sensors simultaneously without any delay.

Consider now an application to speech separation in which the sounds are recorded in a typical room using an array of microphones (see Fig. 1.21). Each microphone will receive a direct copy of the sound source (at some propagation delay based on the location of both the sources and the microphone) as well as several reflected and modified (attenuated and delayed) copies of the sound sources (as the sound waves bounce off the walls and objects in the room).

The distortions of the recorded signals are dependent upon the reverberation and absorption characteristics of the room, as well as the objects within the room, and can be modeled as an impulse response in a linear system. The impulse response provides a model of all the possible paths that the sound sources take to arrive at the microphones.

To find a specific original sound source that was recorded with the microphones in a conference room, we must cancel out, or deconvolve, the rooms impulse response to the original sound source. Since we have no prior knowledge of what this impulse response
INTRODUCTION TO BLIND SIGNAL PROCESSING: PROBLEMS AND APPLICATIONS

of the room is, we call this process the multichannel blind deconvolution or cocktail party problem.

In the “cocktail party problem” our objective is to design intelligent adaptive systems and associated learning algorithms that have similar abilities to humans to focus attention on one conversation among the many that would be occurring concurrently in a hypothetical cocktail party.

1.2.7 Digital Communication Systems

Blind and semi-blind signal processing models and algorithms also arise in a wide variety of digital communications applications, for example, digital radio with diversity, dually polarized radio channels, high speed digital subscriber lines, multi-track digital magnetic recording, multiuser/multi-access communications systems, multi-sensor sonar/radar systems, to mention just a few. BSP algorithms are promising tools for a unified and optimal design of MIMO equalizers(filters/combiners for suppression of intersymbol interference (ISI), cochannel and adjacent channel interference (CCI and ACI) and multi-access interference (MAI). The state-of-the-art in this area incorporates complete knowledge of the MIMO transfer functions which is unrealistic for practical communication systems. The operating environment may consist of dispersive media involving multipath propagation.
and frequency-selective fading, the characteristics of which are unknown at the receiver. The blind signal processing methods may result in more effective and computationally efficient algorithms for a broad class of digital communication systems such as high-speed digital subscriber lines, multi-track digital magnetic recording and multiuser wireless communications [1000, 1094, 1095, 1201, 1208].

In Fig. 1.22 we have an illustration of multiple signal propagation in a wireless communication scenario; a number of users broadcast digitally modulated signals $s_1, s_2, \ldots, s_n$ towards a base station in a multi-path propagation environment. In other words, via multiple paths digital signals are received at an antenna array from many users. The transmitted signals interact with various objects in the physical region before reaching the antenna array or the base station. Each path follows a different direction, with some unknown propagation delay and attenuation. This phenomenon of receiving a superposition of many time-varying delayed signals is called multi-path fading.

Moreover, in some cellular networks, there is another additional source of distortion, so called co-channel interference. This interference may be caused by multiple users that share the same frequency and time slot. The level of interference depends on the propagation environment, mobile location and mobile transmission power. Each transmitted signal is susceptible to multiple interference, multi-user interference and additive noise. In addition, the channel may be time-varying due to user mobility. Advanced blind signal processing algorithms are required to extract desired signals from the interference noise. An even more challenging signal processing problem is the blind joint space-time separation and equalization of transmitted signals, i.e. to estimate source signals and their channels in the presence of other co-channel signals and noise without the use of a training set.

1.2.7.1 Why Blind? Blind signal processing techniques are promising because they require neither prior knowledge of the array response geometry nor any training signals in order to equalize the channels. Moreover, they are usually robust under severe multi-path fading en-
INTRODUCTION TO BLIND SIGNAL PROCESSING: PROBLEMS AND APPLICATIONS

Fig. 1.23  Blind extraction of binary image from superposition of several images [705].

environments. In situations where prior spatial knowledge or a set of short training sequences is available, the prior information can be incorporated in the semi-blind techniques applied.

There are several reasons to apply blind signal processing techniques [1093], [1141], [1142], [1143], such as

• Training examples for interference are often not available.
• In rapid time-varying channels, training may not be efficient.
• Capacity of the system can be increased by eliminating or reducing training sets.
• Multi-path fading during the training period may lead to poor source or channel estimations.
• Training in distributed systems requires synchronization and/or sending a training set each time a new link is to be set up. This may not be feasible in a multi-user scenario.

1.2.8  Image Restoration and Understanding

Image restoration involves the removal or minimization of degradation (blur, clutter, noise, interference etc.) in an image using a priori knowledge about the degradation phenomena. Blind restoration is the process of estimating both the true image and the blur from the degraded image characteristics, using only partial information about degradation sources and the imaging system.

Scientists and engineers are actively seeking to overcome the degradation of image quality caused by optical recording devices, atmospheric turbulence and other image degradation processes.

In many applications, it is necessary to extract or enhance the target image from an image corrupted or superimposed by other images. This is illustrated in Figure 1.23. In some applications, it is necessary to extract or separate all superimposed images as illustrated in Figure 1.24. In many instances, the degraded observation $g(x, y)$ can be modeled as a two-dimensional convolution of the true image $f(x, y)$ and the point-spread function (also called
The blurring function $h(x, y)$ of a linear shift-invariant system plus some additive noise $n(x, y)$. That is, $g(x, y) = f(x, y) * h(x, y) + n(x, y)$. In many situations, the point-spread function $h(x, y)$ is known explicitly. The goal of the general blind deconvolution problem is to recover convolved signals, when only a noisy version of their convolution is available along with some or no partial information about either signal. In practice, all blind deconvolution algorithms require some partial information to be known and some conditions to be satisfied. Our main interest concerns image enhancement, where the degradation involves a convolution process. Blind deconvolution is a technique that permits recovery of the target object from a set of “blurred” images in the presence of a poorly determined or unknown point spread function (PSF). Regular linear and non-linear deconvolution techniques require a known PSF. In many situations, the point-spread function is known explicitly prior to the image restoration process. In these cases, the recovery of the image is known as the classical linear image restoration problem. This problem has been thoroughly studied and a long list of restoration methods for this situation includes numerous well-known techniques, a few examples of which are inverse filtering, Wiener filtering, subspace filtering and least-squares filtering. However, there are numerous situations in which the point-spread function is not explicitly known, and the true image must be identified directly from the observed image $g(x, y)$ by using partial or no information about the true image and the point-spread function. In these cases, we have a more difficult problem of blind deconvolution of images. For the “blind” case a set of multiple images (data cube) of the same target object is preferable, each having dissimilar PSF’s. The blind deconvolution algorithm would be then able to
restore not only the target object but also the PSFs. A good estimate of the PSF is helpful for quicker convergence but is not necessary.

The algorithmic way of processing and analyzing digital images has developed powerful means to interpret specific kinds of images, but failed to provide general image understanding methods that work on all kinds of images. This is mostly due to the fact that every image can be interpreted in many ways, as long as we do not know anything about what we expect to be in it. Thus, we need to build models about the expected contents of images in order to be able to “understand” them. There are many successful applications of image processing; but they are almost always fragile in the sense that it is difficult to adapt them to slightly different forms of imagery or to slightly different circumstances. The aim of Image Understanding is to address this fundamental problem by providing a set of image processing competences within an architecture that can observe the performance of each process, reflect on them, and choose to use/reject certain processes.

Fig. 1.25 Illustration of image restoration problem: (a) Original image (unknown), (b) distorted (blurred) available image, (c) restored image using blind deconvolution approach, (d) final restored image obtained after smoothing (post-processing)\[330\][331].
Obviously, there is a wide gap between the nature of images and descriptions. It is the bridging of this gap that has kept researchers very busy over the last two decades in the fields of Artificial Intelligence, Scene Analysis, Image Analysis, Image Processing, and Computer Vision. Nowadays we summarize these fields as “Image Understanding” research.

In order to make the link between image data and domain descriptions, an intermediate level of description is introduced. It generally contains geometric information. Processing usually starts with some image processing, where noise and distortion are reduced and certain important aspects of the imagery are emphasized. Then, events are extracted from the images that characterize the information needed for description. Typically, these events are such as blobs, edges, lines, corners and regions. They are stored at the intermediate level of abstraction. These are referred to in the literature as “features”. Such descriptions are free of domain information - they are not specifically objects or entities of the domain of understanding, but they contain spatial and other information. It is the spatial/geometric (and other) information that can be analyzed in terms of the domain in order to interpret the images.

Image understanding is one of the most important and difficult tasks on the way towards what is known as artificial intelligence (AI). There is no working system yet which comes close to the capabilities of the human visual system. Some reasons are:

- Biological systems cannot be easily imitated.
- Specialized problem solving methods can hardly be generalized.
- The computational power needed for real-time digital image analysis exceeds the capacity of even the best workstation.

BSP algorithms, especially ICA/PCA, are promising approaches to Image Understanding. One of the ideas of a transform based image/signal description is to expand a signal by using a set of transform basis functions. A well-suited signal description allows us to extract characteristic signal properties which can be used for a variety of signal processing tasks, such as signal estimation, signal compression, or signal analysis. The suitability of an image transform in this context is connected to the efficiency of the transform in representing a given image, i.e. how many coefficients does a transform need to represent the image. The measure for efficiency is the sparseness of the transform coefficients, represented by the decay of the ordered coefficients from a given transform. The local singularities are characterized by location, orientation, and spatial extension. Finding a suitable signal transform for the description of linear singularities is the key for an analysis of the underlying information contained within natural images.

The question arises: How to efficiently describe images which contain a linear or nonlinear mixture of very different signal components. The application of classical signal transforms (such as the Fourier or wavelet transform) to such images is limited since there is no single dominant signal component that can be efficiently estimated with one transform. The idea of ICA or related decomposition approaches is to decompose the image to basic independent components and to start with a large set of independent components. For the image description, only those components that contribute to a sparse description are used. We want to have a small (sparse) number of large coefficients that condense the image information.
The reason for desiring a sparse representation is that under certain assumptions it will also reduce statistical dependencies among units: This provides a more efficient representation of the image structure.

One of the specific goals of research is to understand the coding strategies used by the human visual system for accomplishing tasks such as object recognition and scene analysis. In a task such as face recognition, much of the important information may be contained in the high-order relationships among the image pixels. ICA and related decomposition/separation techniques are able to recover signal components out of signal mixtures. Moreover, ICA/BSS image decomposition allows us to efficiently represent signal components in images. It also allows us to determine the “interesting” signal components in images (see Fig. 1.19). Therefore, ICA image decomposition is a promising tool for image analysis, reconstruction, and classification, as well as for feature detection and image indexing. Statistically independent basis images (e.g., for the faces) can be viewed as a set of independent (facial) features. Unlike PCA basis vectors, the ICA basis images are spatially localized. The representation consists of the coefficients for the linear combinations of basis images that comprised each image. Theories of sensory coding based on the idea of maximizing information transmission while eliminating statistical redundancy from the raw sensory signals have been successful in explaining several properties of neural responses in the visual system such as receptive fields in the visual cortex.

The long-term goal of the Image Understanding research is to develop computational theories and techniques for use in artificial vision systems for which the performance matches or exceeds that of humans, by analyzing sequence of images in space, time and frequency domains.