INSTITUTE OF PHYSICS PUBLISHING PHYSIOLOGICAL MEASUREMENT

Physiol. Meas. **26** (2005) R15–R39 doi:10.1088/0967-3334/26/1/R02

TOPICAL REVIEW

Independent component analysis for biomedical signals

Christopher J James and Christian W Hesse

Signal Processing and Control Group, ISVR, University of Southampton, University Road, Southampton SO17 1BJ, UK

E-mail: c.james@soton.ac.uk

Received 7 September 2004, accepted for publication 19 November 2004
Published 20 December 2004
Online at stacks.iop.org/PM/26/R15

Abstract

Independent component analysis (ICA) is increasing in popularity in the field of biomedical signal processing. It is generally used when it is required to separate measured multi-channel biomedical signals into their constituent underlying components. The use of ICA has been facilitated in part by the free availability of toolboxes that implement popular flavours of the techniques. Fundamentally ICA in biomedicine involves the extraction and separation of statistically independent sources underlying multiple measurements of biomedical signals. Technical advances in algorithmic developments implementing ICA are reviewed along with new directions in the field. These advances are specifically summarized with applications to biomedical signals in mind. The basic assumptions that are made when applying ICA are discussed, along with their implications when applied particularly to biomedical signals. ICA as a specific embodiment of blind source separation (BSS) is also discussed, and as a consequence the criterion used for establishing independence between sources is reviewed and this leads to the introduction of ICA/BSS techniques based on time, frequency and joint time–frequency decomposition of the data. Finally, advanced implementations of ICA are illustrated as applied to neurophysiologic signals in the form of electro-magnetic brain signals data.

Keywords: independent component analysis, ICA, blind source separation, BSS, biomedical signal and pattern processing

(Some figures in this article are in colour only in the electronic version)

1. Introduction

In the field of biomedical signal processing the ultimate aim is that of extracting information underlying a set of biosignal measurements made over time. Generally the signals are

electromagnetic (EM) measurements, although other physical and/or chemical quantities can also be measured. Such a set of multi-channel measurements is usually recorded using a known spatial distribution of the recording sensors with respect to the human body, hence giving rise to a set of temporally and spatially correlated measurements. The information inherent in the measurements depends on the specific application domain (which of course influences the number and position of recording sensors or electrodes). The signal(s) of interest is seldom recorded in isolation and is generally mixed with other ongoing 'background' activity and sensor noise, and is almost certainly contaminated by artifacts of either physiological or environmental origins; furthermore, the signal-to-noise (SNR) ratio of the desired signal is generally quite poor. When a clinician views measured biomedical signals, through training and experience he/she generally looks for distinct patterns of activity with particular spatial distributions—exactly what the clinician is looking for depends on the application domain. One viewpoint is that the recorded data contain measurements of a finite set of separate, overlapping (both in space and in time) activities which are either being generated by thebody or are artifactual in nature. So in essence the recorded data contain mixtures of distinct sources of activity which are contaminated by both artifacts and sensor noise, which are themselves sources in their own right. It could then be said that a clinician attempts to unmix these sources visually using human reasoning to be able to arrive at a conclusion or diagnosis. It would be of great benefit to clinicians if it were possible to automate the analysis of biomedical signals to do the following:

1. To be able to unmix and isolate a set of biomedical signal measurements into their constituent components or sources.
2. To provide information as to the number of distinct sources underlying the measurements.
3. To provide the spatial distribution of each source along with the time series of the source itself.
4. To be able to track changes in the number, spatial distribution and morphology of the sources over time.

This automation may be a simple artifact extraction algorithm, e.g. automatic removal of ocular artifacts from ongoing measurements of brain signals such as the electroencephalogram (EEG) or magnetoencephalogram (MEG) (Jung *et al* 1998, James and Gibson 2003), or the detection of event-related regions of activity in functional magnetic resonance imaging (fMRI) experiments (McKeown *et al* 1998).

Within the above context, the technique of independent component analysis (ICA) provides a tool which can go some way towards providing a solution to the requirements listed above. ICA is a technique which essentially extracts a set of underlying sources or components from a set of random variables, measurements or signals. The technique typically uses a large set of observed multivariate data to define a generative model for the observed data. The components are assumed to be mixed, either linearly or nonlinearly, and the components themselves—along with the mixing system—are assumed to be unknown. Fundamentally, it is assumed that the sources are *mutually independent*. ICA de-mixes or extracts these sources by exploiting this independence of the sources underlying the measured data and is a more powerful technique than classical methods such as principal component analysis (PCA).

Algorithms that could successfully perform linear ICA appeared in the early to mid 1980s (Herault and Jutten 1986) and in the 1990s (Comon 1994, Bell and Sejnowski 1995) there were a large number of papers in the literature utilizing ICA for many applications. Application fields of ICA include digital imaging, economic and financial markets and psychometric

testing, along with analysis of biomedical signals such as neurophysiologic and cardiac signals as well as analysis of biomedical images such as in fMRI (McKeown *et al* 1998). ICA is also used in feature extraction, where underlying sources are used as a basis to represent the measured data (James and D'Alimonte 2004). Most applications of ICA are to multi-channel (or ensemble) time series measurements, in this case the term blind source separation (BSS) can be used as a more generic identifier of the separation process and more will be said about this relationship of ICA to BSS in later sections of this review. In practice, the observed time series is generally a band-limited mixture of the *actual* signals corrupted by noise. A more common application of BSS is the separation of multiple speakers from a set of parallel time series measurements made from across a number of microphones (the so-called *cocktail party* problem). Due to the assumption of the independence of the sources, ICA is greatly used in artifact rejection techniques. In general all ICA/BSS algorithms require multi-channel data to inform the process and as a rule cannot be applied to single-channel measurements—although techniques are available to extend ICA to the single channel case (Hyvärinen *et al* 2001, pp 355–70, James and Lowe 2003). However, whilst more channels generally imply more information for the algorithms to work with, this usually comes at a cost of greater processing time and brings with it issues related to the assumed number of underlying sources due to the square-mixing assumption that will be discussed in the next section.

In general, approaches to solving the ICA/BSS problem have arisen from unsupervised learning algorithms in the neural network field (Herault and Jutten 1986), as well as from the advanced statistics and signal processing fields (Donoho 1981, Shalvi and Weinstein 1990). In the context of biomedical signal processing we prefer to view the BSS process as a data-driven approach to extracting information from multiple measurements of a number of underlying sources. ICA generalizes the BSS problem to random variables (such that the ordering in time of the variables is not relevant). In general, in order to use particular ICA/BSS algorithms there are a few strong general assumptions that must be made about the sources themselves and the source mixing conditions before these can be applied to the measured data and any proper sense made of the results.

This review will concentrate on reviewing the ICA *concept* at a practical level and in particular within a biomedical signal processing framework—rather than reviewing *applications* of ICA to biomedicine. We will convey the assumptions made when applying ICA to any data, and what are the limitations of this as a consequence. We overview the different criteria used to drive the more popular algorithms for ICA in the literature and then explore ICA within the wider BSS context and review techniques which make use of additional characteristics of the data, such as the time and time-frequency structure of the measured data inherent in the multi-channel time-series data, and the many advanced capabilities this affords—especially in the biomedical field. We then look at advanced innovations that can be applied to this BSS problem making ICA a much more powerful tool in the biomedical signal-processing arena. In order to demonstrate the different capabilities of ICA we target the neurophysiological field where we highlight a number of our different applications of ICA/BSS to obtain a variety of information from the measured EM brain signals data.

In the literature there are a number of informative texts and edited paper collections, which provide a good basis of the underlying mathematics for ICA. These include, but are not limited to: Lee (1998), Nandi (1999), Girolami (2000), Haykin (2000), Hyvärinen *et al* (2001), Roberts and Everson (2001), Cichocki and Amari (2002). There have also been a number of international workshops and conferences on ICA: Cardoso *et al* (1999), Pajunen and Karhunen (2000), ICA (1999, 2000, 2001, 2003, 2004).

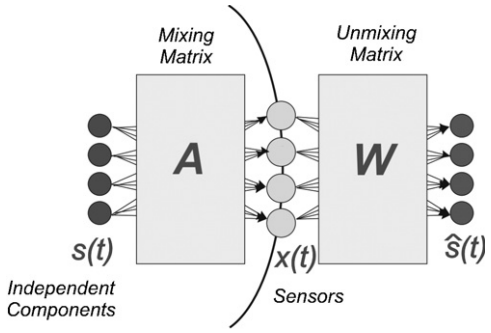

Figure 1. The general ICA process is described pictorially where the measurements at the sensors, $\mathbf{x}(t)$, are assumed to be composed of a linear mixture of the independent sources, $\mathbf{s}(t)$, i.e. that $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$. ICA produces an unmixing matrix $\mathbf{W}$, which unmixes the measurements to give estimates of the independent sources $\hat{\mathbf{s}}(t)$. Note that this oversimplification assumes: (a) linear mixing of the sources, (b) a stationary mixing matrix and (c) noiseless mixing—this is in order to make the BSS problem more tractable.

2. Independent component analysis fundamentals

Fundamentally, the basic BSS problem that ICA attempts to solve assumes a set of m measured data points at time instant t, $\mathbf{x}(t) = [x_1(t), x_2(t), \ldots, x_m(t)]^{\mathrm{T}}$ to be a combination of n unknown underlying sources $\mathbf{s}(t) = [s_1(t), s_2(t), \ldots, s_n(t)]^{\mathrm{T}}$. The mixing of the sources is generally assumed to be linear, and the mixing matrix describing the linear combination of the $\mathbf{s}(t)$ is given by the full rank $n \times m$ matrix $\mathbf{A}$ such that

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t); \tag{1}$$

it is also generally assumed that the number of underlying sources is less than or equal to the number of measurement channels ($n \leqslant m$).

The task of the ICA algorithms is to recover the original sources $\mathbf{s}$(t) from just the observations $\mathbf{x}$(t) and this generally translates to that of finding a separating or de-mixing matrix $\mathbf{W}$ such that

$$\hat{\mathbf{s}}(t) = \mathbf{W}\mathbf{x}(t), \tag{2}$$

given the set of observed values in $\mathbf{x}(t)$ and where $\hat{\mathbf{s}}(t)$ are the resulting *estimates* of the underlying sources. This idealistic representation of the ICA problem is described pictorially in figure 1.

2.1. Basic assumptions

In reality the basic mixing model assumed in equation (1) is simplistic and assumed for the ease of implementation. In fact, the more general mixing model which makes no assumptions on the linearity of the mixing and allows additive noise, may be a more realistic model for a system in general, i.e.

$$\mathbf{x}(t) = \mathbf{f}\{\mathbf{s}(t)\} + \mathbf{n}(t), \tag{3}$$

where $\mathbf{f}$ can be any unknown function and $\mathbf{n}(t)$ is additive sensor noise corrupting the measurements $\mathbf{x}(t)$ (generally assumed to be i.i.d. spatially and temporally white noise, or possibly temporally colored noise).

As now in (3) the BSS problem is that of obtaining a demixing matrix (mapping) by inverting $\mathbf{f}$ whilst not having information on the properties of either $\mathbf{s}$ or $\mathbf{n}$ (or $\mathbf{f}$); it can be appreciated that without making any assumptions about the nature of the data, noise or mixing

process (or at the very least without some *a priori* knowledge about each) the BSS problem will remain quite intractable. This is why basic assumptions are made when formulating ICA algorithms in order to make the problem more tractable. It turns out that in a biomedical signals context, most of these basic assumptions still make the technique attractive and viable. Some of the more apparent assumptions made are listed next.

Linear mixing. The first traditional assumption for ICA algorithms is that of *linear* mixing which reduces equation (3) to

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t), \tag{4}$$

where $\mathbf{A}$ is the linear mixing matrix described earlier. In a biomedical signals context, linear mixing assumes (generally instantaneous) mixing of the sources using simple linear superposition of the attenuated sources at the measurement channel—for the most part a reasonable assumption to make. For the most part assuming instantaneous mixing is perfectly legitimate as this assumes that transmission through the mixing medium is instantaneous—this holds for such applications as fMRI and EM brain signals. Quantities such as sound signals measured through microphones then do become an issue, however, as this assumes convolutive mixing. Whilst the linear mixing assumption makes the BSS problem less intractable we are still faced with the problem of having $n \times m$ unknown quantities to be estimated from just m known measured signals.

Noiseless mixing. The next assumption that is generally made in the majority of the traditional ICA models is that the observations $\mathbf{x}(t)$ are noiseless (or at least that the noise term $\mathbf{n}(t)$ is negligible), i.e. equation (4) reduces to equation (1). Whilst this is probably less realistic in practical terms (i.e. biosignals are measured with no sensor noise) it allows ICA algorithms to separate sources of interest even if the separate sources themselves remain contaminated by the measurement noise.

Square mixing. So far it has been assumed that the mixing matrix $\mathbf{A}$ may be non-square ($n \times m$); in fact, in the case of physiologic signal analysis it is likely that the number of underlying sources n is *less* than the number of measurement channels m in use. However, for most of the popular ICA algorithms if it is assumed that the number of sources underlying the measurement signals is less than the number of measurement channels this brings with it a model order selection problem in trying to establish the optimum value for n. In fact most classic ICA algorithms assume a *square-mixing* matrix, i.e. $m = n$, this makes the BSS problem more tractable. From a biomedical signal analysis perspective the square-mixing assumption is sometimes less than desirable, particularly in situations where high-density measurements are made over relatively short periods of time such as in most MEG recordings or fMRI. The probability of there being as many sources as measurement channels in these situations is less likely. For this reason most researchers apply data-reduction techniques to the data prior to ICA (Hyvärinen *et al* 2001, pp 125–44) although this may be ill advised in certain situations.

Stationary mixing. Another common assumption is that the statistics of the mixing matrix $\mathbf{A}$ do not change with time—i.e. the assumption of stationarity of the mixing matrix. In terms of biomedical signals this means that the physics of the mixing of the sources as measured by the sensors is not changing—this may not be the case in situations where, for example, electrocardiogram (ECG) is recorded on chest electrodes and the electrodes move over time due to breathing. However, in EM brain signal recordings the assumption of a stationary mixing matrix can be interpreted as the fixed biophysical structure of the brain itself whilst the sources distributed within this structure change their intensity over time, which is perfectly plausible.

Statistical independence of the sources. By far the most important assumption made in applying ICA is that the sources are mutually *independent*. Statistical independence is a stronger assumption than uncorrelatedness, and while statistically independent sources are necessarily uncorrelated, the converse does not follow. Two random variables are statistically independent if there is a joint distribution of functions of these variables. This means, for example, that independent variables are uncorrelated and have no higher order correlations. In the case of time-series data it is assumed that each source is generated by a random process which is independent of the random processes generating the other sources.

Thus, the BSS problem can be made further tractable by allowing the introduction of a set of algorithms that can take advantage of this independence of the sources. The assumption of independence of the sources can be quite obvious in some situations , for example, when used in artifact rejection separating brain signals from, say, 50 Hz line noise or ocular artifact, and similarly when separating fetal electrocardiogram (FECG) from maternal ECG (MECG) through trans-abdominal recordings (De Lathauwer *et al* 2000).

Figure 2 depicts an example of four sources consisting of 'dummy' waveforms (figure 2(a)). Figure 2(b) depicts the measurement signals after the sources are linearly mixed using a square random mixing matrix. As the mixing matrix is square there are as many measurement channels as there are sources. The channels of figure 2(b) alone were then used to extract the recovered sources of figure 2(c) through an appropriate ICA algorithm. The assumptions in this case were that there were as many sources as measurement channels, that the mixing was linear and noiseless, and that the sources were statistically independent of each other.

When placed in a physiologic analysis setting the utility of ICA in light of the above assumptions should be assessed on an individual application basis. However, for the most part, ICA can still be a very useful technique simplifying assumptions notwithstanding.

2.2. Higher-order statistics based methods

Some of the original and most commonly cited ICA algorithms performing a BSS of statistically independent sources are based on techniques involving higher-order statistics (HOS), several such implementations can be found in the literature (Makeig *et al* 1997, Comon 1994, Bell and Sejnowski 1995, Hyvärinen and Oja 1997, James and Lowe 2001). In this implementation the measurements $\mathbf{x}(t)$ of equation (1) are assumed to be observations of random variables, where temporal ordering is irrelevant and which are generated as a linear mixture of statistically independent sources.

When seeking statistical independence in sources, it is not enough to obtain uncorrelatedness between the sources, which is what PCA does; statistical independence is based on HOS—although decorrelating the measured data is generally a useful first step. It turns out that it is possible to obtain an estimate $\hat{\mathbf{s}}(t)$ of the sources $\mathbf{s}(t)$ iff the sources $\mathbf{s}(t)$ are *non-Guassian*. In practice it is enough to try and make the estimates $\hat{\mathbf{s}}(t)$ as non-Gaussian as possible as, according to the central limit theorem, sums of non-Gaussian random variables are closer to Gaussian than the originals. In this way looking for independent sources is equivalent to looking for non-Gaussian sources. This also highlights a potential limitation of the method when used in a biomedical signal processing setting as ICA using this technique can only resolve independent sources which have non-Gaussian distributions (or at most only one source with a Gaussian distribution).

We list next three of the most popular and widely referenced techniques for implementing ICA using a HOS approach.

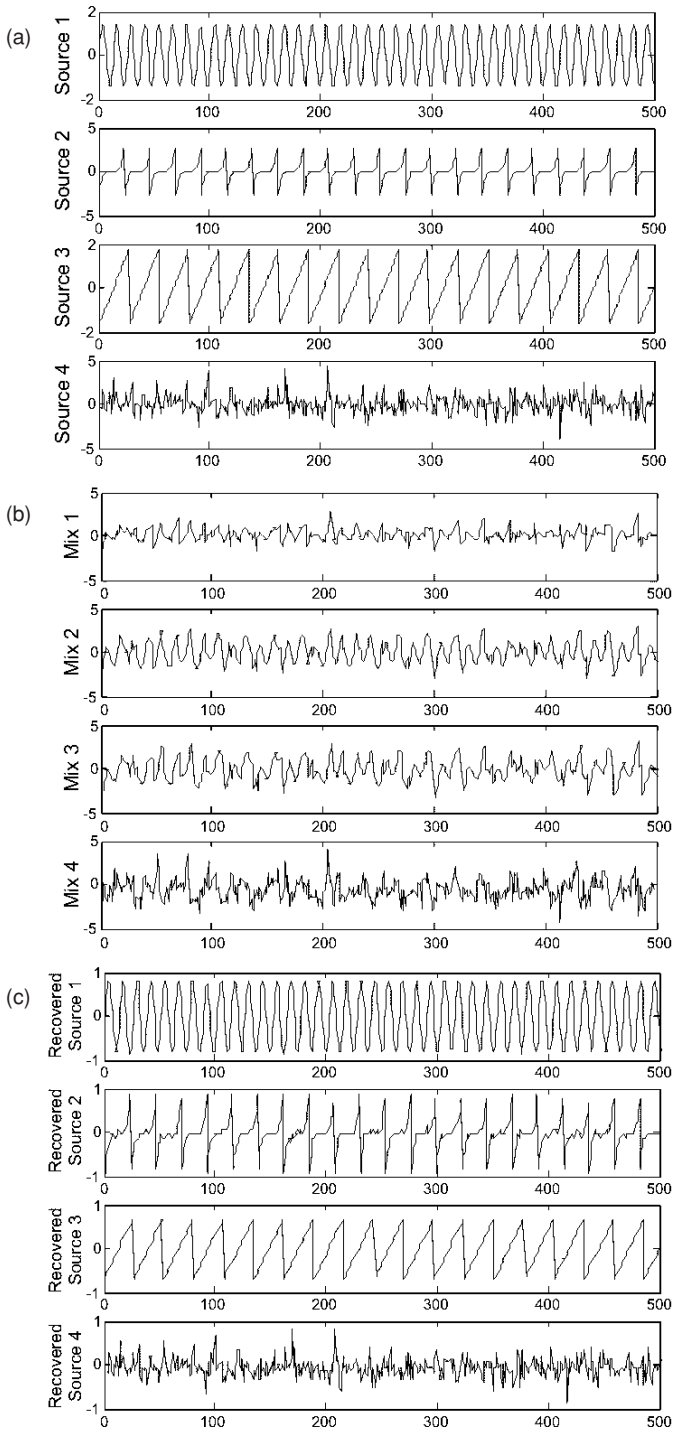

Figure 2. (a) Four underlying sources of a synthetic dataset. (b) The 'measurement signals' generated by linearly mixing the sources of (a). (c) Each of the recovered sources using the ICA process (the ordering of the sources was manually performed and is purely cosmetic).

Non-Gaussianity through kurtosis: FastICA. FastICA is one of the more referenced ICA techniques in the literature (Hyvärinen and Oja 1997) and it is distributed as a freely downloadable set of Matlab® functions from the internet (FastICA). FastICA attempts to separate underlying sources from the given measurement set based on their 'non-Gaussianity'. The simple premise behind FastICA is that the fast fixed-point iterative algorithm undertakes to find projections that maximize the non-Gaussianity of components by their kurtosis (the fourth-order cumulant given to a random variable). In other words, as kurtosis is identically zero for Gaussian distributed signals, the aim is to maximize the magnitude of the kurtosis to make the estimated sources as *non-Gaussian* (and hence as *independent*) as possible. Here, the ICA problem is posed as an optimization problem with the sources as its solution. The kurtosis that is used to describe the peakedness of a distribution is defined as

$$\text{kurt}(\mathbf{x}) = E\{\mathbf{x}^4\} - 3(E\{\mathbf{x}^2\})^2, \tag{5}$$

for a zero-mean random variable $\mathbf{x}$. Further details about the FastICA algorithm can be found in Hyvärinen and Oja (1997).

Non-Gaussianity through negentropy: infomax. Another common algorithm that implements ICA through attempting to discover non-Gaussianity of the sources is the Bell–Sejnowski algorithm (Bell and Sejnowski 1995, EEGLAB), where non-Gaussianity is measured using negentropy, which is based on the information-theoretic quantity of differential entropy. For random variables with equal variance but different distributions, Gaussian random variables have the largest entropy, i.e. contain the least information. Thus, negentropy (or differential entropy) is defined as the difference between the entropy of a Gaussian random variable with the same variance as the observed random variable, and the entropy of the observed variable. Negentropy is zero when the observed random variable is also Gaussian, and positive when the observed variable is non-Gaussian.

The Bell–Sejnowski algorithm is a neural network gradient-based algorithm whose learning rule is based on the principle of information maximization (infomax), and it maximizes the output entropy of a neural network with nonlinear outputs. The learning criterion is the maximum likelihood estimation of an ICA model (derived through the parameters of a neural network). In effect, it can be seen that ICA estimation by this criterion reduces to the maximization of the non-Gaussianity of the sources.

As this is a neural network approach, and hence involves a gradient training algorithm, it suffers from the problems inherent with such learning in general. The choice of the nonlinearities used also affects the performance of the algorithm, although in practice for biomedical signals this choice is generally not critical to a successful implementation.

Joint approximate diagonalization of eigenmatrices: JADE. This approach is known as ICA by tensorial methods using higher-order cumulant tensors (Hyvärinen *et al* 2001, pp 229–37). The covariance matrix is the second-order cumulant tensor, and the fourth-order tensor is defined by the fourth-order cumulants. By performing an eigenvalue decomposition of the covariance matrix of the data, $\mathbf{C_x}$, the data are transformed such that the second-order correlations are zero. Similarly, as a generalization of this principle, fourth-order cumulant tensors can be used to make the fourth-order cumulants zero or as close to zero as possible. As with HOS methods described previously, reducing the fourth-order cumulants to zero in this way implies statistical independence of the sources and JADE is the algorithm that implements this. As the name implies, JADE involves the joint diagonalization of a number of matrices (i.e. attempts to make all off-diagonals zero or close to zero as possible) (Cardoso and Souloumiac 1993, 1996). Whilst useful in low dimensional problems, JADE has the limitation that it cannot be used with high dimensional data due to numerical reasons.

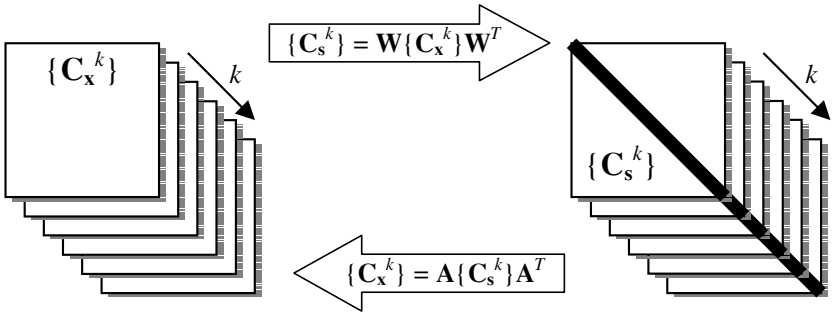

Figure 3. The relationship between the two covariance matrix stacks of $\mathbf{C}_\mathbf{x}^k$ and $\mathbf{C}_\mathbf{s}^k$. The mixing matrix $\mathbf{A}$ can be seen to link the covariance stack of the sources to the covariance stack of the measurements and vice versa with the unmixing matrix $\mathbf{W}$.

Overall, ICA based on HOS methods provides a good basis for solving the ICA problem. However, by far the biggest drawback in using these techniques in a biomedical signal processing setting is the requirement that the sources have non-Gaussian distributions as this is generally not known *a priori* although there are instances where some sources can be argued to be such, such as for example, bursts of rhythmic narrowband activity. It is also less intuitive to discard the time-series nature of the recorded data in calculating the underlying sources and using time-structure-based techniques (as described next) shows more advantages.

2.3. Time structure based methods

A completely different approach to ICA is given by considering not higher order correlations, or moments of the source waveform distribution, but rather by considering the time structure of the sources. The assumption of the independence of the sources has a very important and useful consequence: the source waveforms have no spatial temporal or spatial time–frequency correlations. The basic approach here is to capture the dependency structure of the observed signals using a set of square matrices (in the form of a *stack* of matrices), and then find the de-mixing matrix which is the joint diagonalizer of the stack, i.e.

$$\mathbf{C}_\mathbf{x}^k = \mathbf{A}\mathbf{C}_\mathbf{s}^k\mathbf{A}^\mathrm{T}, \tag{6}$$

where $\mathbf{C}_\mathbf{x}^k$ is the kth covariance matrix of the data $\mathbf{x}(t)$, $\mathbf{C}_\mathbf{s}^k$ the corresponding covariance matrix of the sources $\mathbf{s}(t)$ and $\mathbf{A}$ is the mixing matrix. Conversely, the source covariances are obtained from the data covariance through the inversion

$$\mathbf{C}_\mathbf{s}^k = \mathbf{W}\mathbf{C}_\mathbf{x}^k\mathbf{W}^\mathrm{T}, \tag{7}$$

where $\mathbf{W}$ is the unmixing matrix. These two equations hold in general, regardless of the nature of the matrices in the stack. The relationship between the two covariance matrix stacks of $\mathbf{C}_\mathbf{x}^k$ and $\mathbf{C}_\mathbf{s}^k$ are shown pictorially in figure 3, where the mixing matrix $\mathbf{A}$ links the covariance stack of the sources to the covariance stack of the measurements and vice versa with the unmixing matrix $\mathbf{W}$.

The index k is an index into the matrix stack and will have different interpretations depending on what quantities are being measured. For example, when the temporal dependency is captured through temporal correlation measured at different lags, k is an index into the cross-covariances at each lag, starting from $k = 0, 1, 2, \ldots$ etc, until a maximum number of lags are reached. So, for a maximum of L lags, there will be $L + 1$ matrices in the stack ($k = 0, 1, 2, \ldots, L$).

Estimating the mixing matrix. There are two ways of estimating the mixing matrix $\mathbf{A}$ with reference to a stack of correlation matrices. The most common approach is to estimate the demixing matrix $\mathbf{W}$ first—this is known as an *inverse* estimation method. Since $\mathbf{C_s}$ is supposed to be diagonal, we can optimize the coefficients of $\mathbf{W}$ in such a way as to make the matrix given by $\mathbf{W}\mathbf{C}_\mathbf{x}^k\mathbf{W}^\mathrm{T}$ as diagonal as possible. The diagonality of a matrix can be measured, for example, by the sum of the squared off-diagonal elements. Previous methods employed orthogonal constraints (Cardoso and Souloumiac 1993, 1996, Pham 2001); however, more advanced methods allow non-orthogonal diagonalization (Ziehe *et al* 2003) but require a square mixing matrix—which in turn requires pre-whitening. Conversely, *forward* estimation methods (e.g. Yeredor (2002)) have the advantage that they allow non-orthogonal and non-square mixing. However, these methods are not quite as efficient as some of the inverse methods, and still require some estimate of the number of sources.

ICA by temporal decorrelation. A very straightforward approach to ICA using time structure is based on temporal decorrelation. If we take expectations of the ICA model with respect to time and a set of time delays $\tau = 1, 2, 3, \ldots$, we obtain the cross-covariance function of the signal,

$$\mathrm{E}\{\mathbf{x}(t,\tau)\mathbf{x}^\mathrm{T}(t,\tau)\} = \mathbf{C}_\mathbf{x}^\tau = \mathbf{A}\mathbf{C}_\mathbf{s}^\tau\mathbf{A}^\mathrm{T}, \tag{8}$$

where the source cross-covariance function $\mathbf{C}_\mathbf{s}^\tau$ is a set of diagonal matrices due to the assumed statistical independence of the sources,

$$\mathbf{C}_\mathbf{s}^\tau = \mathbf{W}\mathbf{C}_\mathbf{x}^\tau\mathbf{W}^\mathrm{T}. \tag{9}$$

This approach was initially proposed for two time lags (Tong *et al* 1991, Molgedey and Schuster 1994) where diagonalization could be performed in the context of a joint eigenvalue decomposition, and subsequently extended to several time lags by Belouchrani *et al* (1997) and Ziehe and Müller (1998) where diagonalization has to be performed using iterative methods. Since this approach is based on the cross-covariance function, temporal decorrelation assumes that the source waveforms are stationary and have unique power spectra, which is unlikely to hold for long-term recordings of biomedical signals. However, the approach can be adapted to handle non-stationary signals if we assume that the auto-correlation function of the source activity is slowly varying in time, so that the sources are approximately stationary over short time windows. In this case, temporal correlations can be computed over short time windows and these can be used to estimate the mixing matrix in the usual way (Choi *et al* 2002, James and Hesse 2004c).

One issue with this method is the appropriate choice of the number of time lags to use to describe the spatio-temporal covariance of the data. There are no hard and fast rules for selecting an appropriate number of time lags. One possibility, however, would be to determine the average number of time lags of the data by fitting an autoregressive model to each channel using statistical model selection criteria to determine the necessary and sufficient number of lags. While this is a well-motivated and principled approach, it may be too costly in practice, and empirical observation of appropriate values of time lag may suffice.

It is worth noting that ICA by temporal decorrelation is robust in the presence of Gaussian noise that is spatially and temporally white (because for all lags other than lag zero, the noise cross-correlation function is zero). Also, this method is very flexible and uses intuitive and familiar concepts from time series analysis (Chatfield 1996), which makes it attractive for use in conjunction with conventional signal processing methods.

ICA by sub-band decorrelation. Another way in which BSS can be achieved using time structure is with reference to signal cross-correlations in different frequency bands, using

for example a bank of band-pass filters (Cichocki and Belouchrani 2001, Cichocki and Amari 2002). Again, due to assumed statistical independence, the sub-band cross-correlations of the source waveforms should be a set of diagonal matrices. This may be a very attractive method in biomedical signal analysis, as the activity of interest is in many instances restricted to well-defined frequency bands (e.g. rhythmic components of spontaneous EEG).

The success of this approach is obviously dependent on an appropriate choice of frequency sub-bands. For non-stationary signals, it may be desirable to consider sub-band correlations over shorter time windows. Because of the need to compute and store different band-pass filtered versions of the signal for determination of sub-band covariance, this method can be computationally quite expensive when performed off-line. One way of improving computational efficiency is to compute the sub-band cross-covariances in the *wavelet* domain using the discrete wavelet transform or wavelet packets (Koehler and Orglmeister 1998, Hesse and James 2004a, 2004b, 2004c). The wavelet-based methods also seem to work reasonably well for non-stationary signals.

ICA by time–frequency decorrelation. A very principled way of achieving BSS of non-stationary signals is based on representing the signal dependency structure in terms of a spatial time–frequency distribution (STDF) (Belouchrani and Amin 1998). Essentially, an STFD is an extension of a conventional time–frequency distribution (TFD) (e.g. Cohen (1995)) which considers the cross time–frequency distribution of two signals, computed for each pair of signals within the recording array. Joint diagonalization is then carried out on the STFD in the usual way (as described previously).

The sophistication of this approach notwithstanding, computation and diagonalization of signal STFDs can be prohibitive in terms of memory storage and processing for even a small number of channels and modest signal length. There are a few suggestions to be found in the literature that aim at increasing efficiency by doing spatial averaging over selected time–frequency regions (e.g. Amin *et al* (2003)), yet this brings with it the problem of choosing such regions appropriately, and moreover, make the method very similar to more efficient and simpler approaches based on windowed temporal or sub-band decorrelation.

In general it is apparent that ICA methods using time structure are very well suited to biomedical signal analysis, possibly more so than methods based on HOS. The temporal and time–frequency information utilized is clearly relevant in biomedical signals; whereas the assumptions regarding the (non-) Gaussianity of underlying sources that are necessary in HOS-based techniques cannot always be guaranteed or anticipated to hold. Whilst dealing with the spatio-temporal dynamics of biomedical signals brings with it issues such as those of stationarity, these techniques can be adapted to handle such eventualities. Moreover, we will show in section 3 how time structure based methods are readily adaptable to provide more advanced applications of ICA to biomedical signals.

2.4. Perceived limitations when using ICA in practice

These are limitations that are *perceived* because although they suggest ambiguities in the results of applying ICA, in reality these ambiguities or indeterminacies are not insurmountable and workarounds are quite easily implemented and so they do not actually impose a limitation on the accuracy of the BSS model estimation itself.

Ambiguities. Due to the nature of the BSS problem and the techniques used in solving ICA, there are generally a small number of ambiguities which apply to the sources that are extracted from the measured data, these are that: (a) neither energies nor signs of the sources can be

calculated and, (b) there is no ordering between the sources. Neither of these is particularly restrictive and for each there are workarounds.

In the case of the first; whilst the sources extracted, i.e. $\hat{\mathbf{s}}(t)$ in equation (2), are each normalized to unit variance, the columns of the mixing matrix $\mathbf{A}$ reflect the power of each component across the measurement space. This means that it is possible to calculate a quantity, such as RMS power, for each source from the columns of the mixing matrix which will then allow ranking of the sources according to RMS power. Such as shown here

$$p_j = \sqrt{\frac{1}{m}\sum_{i=1}^{n}\left(\mathbf{a}_i^j\right)^2}, \tag{10}$$

where $\mathbf{a}_i^j$ is the ith element of column j of the mixing matrix $\mathbf{A}$ and p_j is the RMS power for independent source j ($1 \leqslant j \leqslant n$).

Interpretation of results. The above limitations are symptomatic of a wider problem that is apparent in the application of ICA to real world data, especially in biomedical signals, and that there is little in the literature about how to *interpret* sources extracted after implementing ICA algorithms. This is generally a subjective process by the authors, choosing sources of interest based on subjective criterion usually related to the expected outcomes of the analysis. For example, sources of interest can be chosen by observing either the source time series, a frequency analysis of the sources or the distribution of the source over the measurement channels, as well as other relevant techniques (James and Lowe 2000a, 2000b). However, the choice of sources of interest remains highly subjective.

Model order selection. Correct determination of the number of sources is a problem in ICA, the importance of which is not immediately apparent from the literature. Especially when the number of channels exceeds the number of sources, as is the case in many biomedical applications such as high density EEG/MEG, and in fMRI especially, knowing how many sources to estimate can have major impact on the quality and accuracy of the ICA solution. When we try to estimate more sources than there actually are, ICA algorithms (those based on HOS, especially) will tend to 'overfit' the ICA model which leads to distortions as described in Hyvärinen *et al* (2001).

The most common approaches to estimating the number of sources underlying multi-channel measurements are based on the PCA of the data covariance matrix (e.g. Hyvärinen *et al* (2001)). In the simplest case, the number of sources is taken to be equal to the number of dominant eigenvalues, where the latter is defined to be the number of eigenvalues which account for some (high) proportion of the total observed variance (e.g. 95% or 99%) or the number of eigenvalues whose individual contribution to the total variance is greater than some minimum amount (e.g. 1%). There are a number of problems with this approach:

(a) There is no *a priori* reason to suppose that the sources of interest are contained in the signal subspace spanned by the dominant principal components (e.g. if there is a lot of noise, or the sources of interest are relatively weak compared with other artifactual sources). This is also a main criticism of PCA-based dimension reduction and pre-whitening.
(b) The variance proportion threshold as an inclusion criterion is quite arbitrary, and will yield results that are dependent on the shape of the eigenvalue spectrum, especially the tail. They also do not give any indication of the confidence of the model order estimate (in the statistical sense).

(c) In the noise-free case, with fewer sources than sensors, an eigenvalue decomposition of the data covariance matrix can be inaccurate for numerical reasons.
(d) In the presence of noise (especially Gaussian), the number of sources that will be estimated using cumulative variance criterion varies with the number of sensors, since the proportion of total variance due to the sensor noise increases with the number of sensors. This results in overestimation of the number of sources, which in turn can lead to overfitting.

To address some of these issues, there are a number of source enumeration methods (e.g. Valaee and Kabal (2004), Green and Taylor (2002), and references therein) which use statistical or information theoretic measures to determine the number of dominant eigenvalues of the covariance matrix, and hence the number of sources. These criteria essentially determine how 'flat' the noise subspace part of the eigenvalue spectrum is. Yet this approach still relies on eigen-decomposition of the data covariance matrix (and is hence subject to problems associated therewith, i.e. overfitting when there are fewer sources and the noise levels are low) and the assumption that the measurement noise is the same in all channels. The latter may not always hold in biomedical electrophysiological recordings where, for example, electrical impedances (and therefore channel gain) can change due to sweat production of the skin, pressure of electrode contact on skin and related problems.

It is interesting to note, however, that despite advances in source enumeration approaches, the most commonly used criteria for selecting the number of sources for the ICA model are still based on cumulative and relative variance thresholds. Moreover, the advances in PCA based source enumeration methods notwithstanding, model order selection (or estimation) is still an issue in ICA research that has yet to be addressed satisfactorily. Especially with the advent of non-orthogonal diagonalization methods that do not require pre-whitening, it is highly desirable to remove PCA based pre-processing (dimension reduction and whitening) from ICA altogether. We are currently working towards a solution to model order selection that is based on stepwise estimation (i.e. sequential estimation of the components) of the ICA model and the use of statistical criteria based on the 'goodness-of-fit' of the ICA model to the data to determine a necessary and sufficient number of sources (Hesse and James 2004d, 2004e).

3. Bringing prior knowledge to bear: extensions to ICA

It is very often the case with biomedical measurements that we have some idea about the nature of some of the source signals we wish to extract from the recorded data. Many physiologically relevant signals or patterns have certain temporal, spectral or time–frequency characteristics, and in the case of multi-channel (body surface, volumetric) measurements also particular spatial projections. Examples include heart beat waveform morphology, rhythmic brain (EEG/MEG) activity such as alpha, seizures, or transients such as eye blinks, saccades and bursts of muscle activity. In the case of artifacts, most especially due to their good SNR, we can often obtain key information about temporal dynamics from a small subset of channels.

It is desirable and indeed possible to incorporate such prior information into the ICA model using only minor modifications of the estimation procedures, essentially by imposing constraints on the model, which can act on the spatial projections, or work on the temporal dynamics of the source waveforms. The idea is that we may be able to guide/bias the ICA solution to include an expected outcome. By including prior knowledge into the system and letting the ICA method of choice estimate the unknown portions (based on the assumptions as have already been covered) this helps us to interpret the results meaningfully.

In the following subsections we describe a number of ways in which the ICA solution can be constrained to take into account prior knowledge about temporal dynamics or spectral properties of the sources of interest, as well as their spatial projections—this is called constrained ICA (cICA). We show how ICA can be constrained to extract the spatial projections of individual components of interest with specified temporal dynamics and can be used to track components with a desired power spectrum. Subsequently we give a general outline of how constraints on the spatial projections of independent components (ICs) can be incorporated.

3.1. Temporal constraints with FastICA

In the HOS-based ICA technique of FastICA described previously, the algorithm would first (theoretically) converge to the single IC having the maximum negentropy of all the underlying ICs. When one desires a *specific* IC, this is of little use, unless the IC happened to carry the maximum negentropy (see section 2.2). Furthermore, the algorithm is not guaranteed to converge to the global maximum due to random initialization of the algorithm and other computational factors. The cICA algorithm described in Lu and Rajapakse (2001) brings in the use of a temporal constraint which is used to obtain an output which is statistically independent of other sources and is closest to some reference signal $r(t)$. This constraining signal need not be a perfect match (indeed, if it were, one would argue that there would be little point in performing the analysis at all) but it should be enough to point the algorithm in the direction of a particular IC spanning the measurement space. The closeness constraint can be written as

$$g(\mathbf{w}) = \varepsilon(\mathbf{w}) - \xi \leqslant 0, \tag{11}$$

where $\mathbf{w}$ denotes a single demixing weight vector, such that $\mathbf{y} = \mathbf{w}^{\mathrm{T}}\mathbf{x}$; $\varepsilon(\mathbf{w})$ represents the closeness between the estimated output $\mathbf{y}$ and the reference $\mathbf{r}$, and ξ some closeness threshold. The measure of closeness can take any form, such as mean-squared-error or correlation, or any other suitable closeness measure. In our implementation of the algorithm we used correlation as a measure of closeness such that $g(\mathbf{w})$ becomes

$$g(\mathbf{w}) = \xi - E\{\mathbf{r}(\mathbf{w}^{\mathrm{T}}x)\} \leqslant 0, \tag{12}$$

where ξ now becomes the threshold that defines the lower bound of the optimum correlation. With the constraint in place, the cICA problem is modelled as follows:

$$\begin{aligned} &\text{Maximize: } f(\mathbf{w}) = \rho[E\{G(\mathbf{w}^{\mathrm{T}}x)\} - E\{G(\nu)\}]^2, \\ &\text{subject to: } g(\mathbf{w}) \leqslant 0,\ h(\mathbf{w}) = E\{\mathbf{y}^2\} - 1 = 0 \text{ and } E\{\mathbf{r}^2\} - 1 = 0, \end{aligned} \tag{13}$$

where $f(\mathbf{w})$ denotes the contrast function described in (5); $g(\mathbf{w})$ is the closeness constraint; $h(\mathbf{w})$ constrains the output $\mathbf{y}$ to have unit variance; and the reference signal $\mathbf{r}$ is also constrained to have unit variance. In Lu and Rajapakse (2001), the problem of (13) is expressed as a constrained optimization problem which is solved through the use of an augmented Lagrangian function, where learning of the weights and Lagrange parameters is achieved through a Newton-like learning process.

3.2. Spectral constraints in temporal decorrelation

In the context of temporal decorrelation methods, prior knowledge of the spectral content of (some of) the sources to be extracted can be introduced into the model by means of reference channels. When the power spectrum of particular source activity is known, as is the case with rhythmic EEG signal components such as alpha activity or epileptic seizures, for example,

the reference channel(s) would consist of band-pass filtered noise with the desired power spectrum.

Specifically, a reference channel $\mathbf{r}_1(t)$, containing noise with a particular power spectrum, is included with the observed signal channels $\mathbf{x}(t)$ to form an augmented signal matrix

$$\widehat{\mathbf{x}}(t) = \begin{bmatrix} \mathbf{x}(t) \\ \mathbf{r}_1(t) \end{bmatrix}. \tag{14}$$

ICA using temporal decorrelation is then applied to this augmented signal matrix in the usual manner, i.e. using lagged covariance matrices. To eliminate the effects of phase differences between the reference channel and the source activity, cross-covariances can be computed in the frequency domain.

The ICA problem is now such that the extra row in the measurement space due to the reference vector results in an extra row in the IC space after the ICA step (as well as a corresponding extra column in the mixing matrix). For an m-channel system, the first m elements of the extra mixing matrix column depict the spatial distribution (topography) of the new IC given by the row vector $\hat{\mathbf{s}}_{m+1}(t)$. Furthermore, each of the elements of the $(m + 1)$th row of the mixing matrix reflects a weighting of each corresponding IC. This row vector, $\mathbf{a}_{m+1}$, can in fact be used to depict the contribution of each topography described by the columns of the mixing matrix, due to the reference channel $\mathbf{r}_1(t)$. In this way ICA provides a convenient spanning basis, which can be used to obtain the topography of interest, and which is extracted by summing the weighted contributions of each column of the mixing matrix. It can be seen that this can be readily extended to more than one reference.

3.3. Topographical maps as spatial constraints

When there is prior knowledge about the spatial projections of some of the sources to be estimated, this can be incorporated into the ICA model by means of constraints on (some of) the columns of the mixing matrix. The type of constraint can reflect the certainty (accuracy) that can be attached to prior knowledge.

Choosing a particular set of initial values for the columns of the mixing matrix constitutes a very simple and weak form of spatial constraint. Most ICA estimation methods involve numerical optimization and convergence can depend on initial conditions. The usual initialization for the mixing matrix is an identity matrix or a random orthogonal matrix. However, in EEG for example, approximate spatial topographies for eye-blink and saccade artefacts can be determined from the raw traces and included as an initial guess in the first two columns of the mixing matrix. While selective initialization of the mixing matrix does not guarantee that the constraint components form part of the solution, it can help to increase convergence when they are present in the data.

Staying with the example of ocular artefacts in EEG, when it is quite certain that particular spatial topographies form part of the ICA solution, they may be included in the mixing matrix and they can be kept fixed while the spatial projections of remaining components (brain activity) are estimated. This is preferable to simply applying the known topographies as a spatial filter to the data, since some of the other components may be spatially correlated with the former, and estimation of the remaining activity prevents data distortion. Using this type of 'hard' spatial constraint is only advisable when one can be absolutely certain of the accuracy of the *a priori* spatial projection.

In cases when there is only one approximate idea of the spatial projections that may form part of the ICA solution, it is possible to include these as *reference projections* and adapt the estimation algorithm in such a way that it seeks independent components whose spatial

projections lie 'near' the references, using some distance metric and penalty function. This type of 'soft' constraint is potentially a very useful and flexible way of taking into account prior knowledge that can be fairly vague.

Regardless of the type of spatial constraint used, when PCA based pre-whitening is used to pre-process the data, it is important to appropriately transform the spatial constraints using the whitening matrix.

4. Advanced applications of ICA: the neurophysiological domain

For this topical review we purposely did not set out to review the multitude of applications of ICA to biomedical signals, but rather to review some of the more prevalent approaches to ICA that are available in the literature and their potential benefits, with a specific intent of applying these to biomedical signals. For this reason, in this section we will show just a select number of examples of how we have applied ICA in our field of expertise—the neurophysiological domain and the application to EM brain signal analysis. The main idea is that of illustrating the variety of specializations that ICA allows, hopefully highlighting the many more potential uses of ICA in biomedicine than is currently presented in the literature. Few papers can be found in the literature that compare ICA with conventional methods for varied biomedical signals, because (we presume) that this is not always so easy to do. For the most part the use of ICA as a 'black-box' method may result in situations, such as the violation of some assumption (where another ICA method might have been better, or over/under fitting of the model occurs) which implies that ICA is inferior where, in fact, it could have been better had it been used appropriately.

In our implementation of ICA for EM brain signal analysis we make assumptions that are in keeping with the general assumptions governing the application of ICA. In particular we assume that:

1. The measured EEG/MEG is a linear summation of the electrical/magnetic activity from various brain regions.
2. The EM field distribution is spatially fixed and only the electrical 'strength' is changing within these regions.
3. Any activity of interest is independent of the ongoing background EM brain activity. This certainly holds true for most artifacts and to activity such as seizure activity (at least early on in the evolution of a seizure).

Whilst ICA is not necessarily advocated for use in all problems in this domain, one of the biggest assets in favor of using ICA in EM brain signal analysis is the fact that multi-source activity can be naturally separated into neurophysiologically meaningful components. Standard signal processing techniques such as matched and/or adaptive filters *can* be used to detect and extract activity of interest, but these generally require much detailed *a priori* knowledge about the characteristics of each of the signals in question. Furthermore, such techniques are never as discrimantive as ICA can be, because there are usually residuals in performing unmixing in this way. ICA also unmixes signals by making very basic assumptions about the data (those of independence foremost) and it makes little difference if the signals are artifactual in origin or brain-signals, for example, for the technique to work—standard techniques are usually not so flexible.

4.1. Case study A: temporally constrained ICA for automatic artifact rejection

As shown in section 3.1, temporally constrained ICA can extract signals from measurements that are statistically independent, yet which are constrained to be similar to some temporal

reference signal, in this way incorporating *a priori* information. In James and Gibson (2003) we demonstrated this method on a synthetic dataset and on a number of artifactual waveforms identified in multi-channel recordings of EEG and MEG. cICA repeatedly converged to the desired component within a few iterations and subjective analysis showed the waveforms to be of the expected morphologies and with realistic spatial distributions. In this work cICA was applied with great success to EM brain signal analysis, in particular with a view to automating artifact extraction in EEG and MEG.

Figure 4 shows a 20 s epoch of multi-channel scalp EEG recorded from a long-term Epilepsy Monitoring Unit using the modified 10–20 electrode placement system (Lagerlund *et al* 1993) (with reference channel Fcz) with a sampling frequency of 200 Hz. Figure 4(a) depicts an epileptic seizure, the onset of which is around the 5–7 s mark. The seizure EEG is contaminated with ocular artifact throughout (see highlighted sections of EEG). In figure 4(b), a closer view is taken of channel Fp1 showing the seizure onset and the repeated ocular artifact. In order to obtain a reference signal for cICA a simple threshold is applied to Fp1 and a positive going pulse is recorded when the EEG at Fp1 exceeds that threshold, as shown in figure 4(c). After applying the cICA algorithm to the EEG data matrix, using the reference derived above, the ocular artifact depicted in figure 4(d) was extracted. The inset shows the topographic map of ocular artifact component derived from the de-mixing weight vector, which depicts a clear focus over the eyes. Finally, the extracted component is projected to the measurement space and subtracted from the recorded EEG matrix; a close-up view of channel Fp1 with the ocular artifact subtracted (from all channels of EEG) is shown in figure 4(e). This shows the seizure onset uncontaminated by ocular artifact.

In this case strong *a priori* information about the morphology of the desired signal was brought into play. This information allowed a FastICA based technique to rapidly isolate a statistically independent component from the measured data within a few iterations. As the desired end result was the automated detection of artifacts from ongoing EM brain signal recordings the underlying assumption of statistical independence and the use of a HOS based technique is probably quite justified.

4.2. Case study B: temporal decorrelation based constrained ICA to track changing brain states

In this study we use the temporal decorrelation based cICA of section 3.2 in tracking the changing scalp topographies of rhythmic brain activities (James and Hesse 2004a, 2004b). We demonstrate this method on a multi-channel recording of an epileptiform EEG, where we automate the repeated simultaneous extraction of both rhythmic seizure activity, as well as alpha-band ($\sim$10 Hz) activity, over an epoch of EEG. Subjective analysis of the results shows scalp topographies with realistic spatial distributions which conform to our neurophysiologic expectations.

It has already been stated that EEG recordings capture ongoing brain activity which can be interpreted as brain sources whose outputs vary over time. Some specific types of brain activity are associated with specific brain states. In general, rhythmic activity in the EEG is of interest (e.g. alpha-, beta-, delta- and gamma-band activities, or rhythmic seizure activity); furthermore, the spatial topographies of such activities showing the distribution over the scalp are also desirable. Automatically isolating, visualizing and tracking the scalp topographies of multiple neurophysiologically meaningful sources underlying the ongoing EEG recordings would be desirable. When tracking rhythmic brain activity in particular, it is generally the case that the frequency band of interest is known *a priori* and it is desirable to observe the changing power within that band and the topographies associated with that activity. Simple

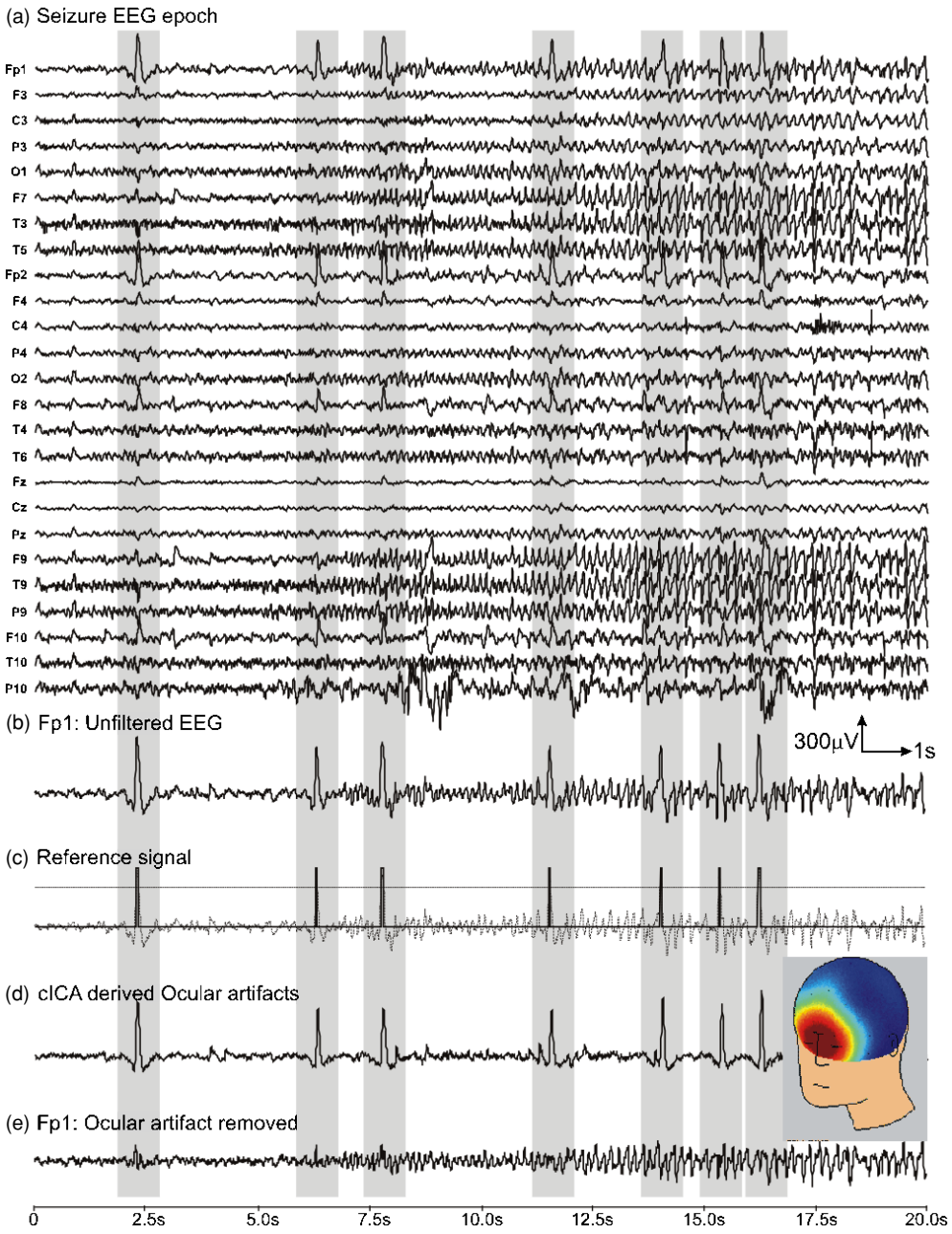

Figure 4. (a) A 20 s epoch of multi-channel scalp EEG recorded with modified 10–20 electrode system (ref channel Fcz) depicting seizure onset contaminated with ocular artifact. (b) A closer look at channel Fp1 showing seizure onset (∼7 s into recording) and ocular artifact. (c) Reference signal obtained by applying a threshold to Fp1. (d) cICA output after applying a cICA algorithm to EEG data using reference derived at (c) (inset shows topographic map of ocular artifact component). (e) Channel Fp1 with ocular artifact of (d) subtracted (from all channels of EEG).

band-pass (BP) filtering in the single-channel recordings of the EEG is generally of little use as the activities of interest are usually small in comparison to other ongoing activities, and

the move to multi-channel analysis to visualize scalp topographies further compounds the problems.

The method of section 3.2 was repeatedly applied to 20 s of multi-channel epileptiform EEG (figure 5(a)) data using two reference signals, each of which was derived as above. The first reference described the seizure activity and consisted of BP filtered white noise with lower and upper corner frequencies of 2 Hz and 6 Hz, respectively. The second reference represented alpha-band activity, with lower and upper corner frequencies set at 9 Hz and 11 Hz, respectively. A series of overlapping (each time shifted by 125 ms) 3 s windows of multi-channel EEG was analysed and in each case the resulting topographies of the two reference channels were recorded along with the relative power of each (relative to total power within the data matrix). Figure 5(b) shows the relative power of each set of consecutive scalp topographies (normalized) obtained for each reference and figure 5(c) shows some of the normalized topographies for each.

This study shows a different implementation of the concept of cICA—this time through temporal decorrelation based ICA and the constraint is applied as an extra channel in the measurement space. This has been shown to be useful in tracking the changing topographies of different rhythmic activities which are manifest as the changing columns of the mixing matrix **A**—each time the algorithm is applied to a new time-shifted dataset.

4.3. Case study C: model order selection/non-square mixing

In this example we present an approach to simultaneous ICA model order and source estimation (Hesse and James 2004d, 2004e), which uses a method for direct estimation of the mixing matrix adapted from Yeredor (2002). Starting with one component, the mixing matrix is constructed in a stepwise fashion, one column at a time, until a number of sources have been found which strikes an optimal balance between model parsimony and goodness-of-fit to the data (in the least-squares sense). In this case, the stopping criterion was based on a statistical test (an F-test) which determines whether the difference in goodness-of-fit between two consecutive models is significantly different (at a particular significance level, e.g. $\alpha = 0.01$). Thus, the test seeks to determine whether the addition of an extra source leads to an improvement in the ICA model goodness-of-fit that can be justified against the associated increase in the number of model parameters, i.e. whether the extra source is *statistically significant*. This approach is analogous to that of conventional stepwise regression analysis.

We applied the stepwise ICA model estimation method to a 20 s segment of 25-channel seizure EEG (see figure 6(a)). The electrodes were arranged on the scalp according to the international 10–20 system, and signals were recorded with Fpz as reference at a sampling rate of 200 Hz, and off-line re-referenced to an average reference. The EEG shows the onset of a seizure with a right temporal focus (T10, F10) about 6 s into the recording, as well as ocular artifacts (Fp1, Fp2) and bursts of EMG activity due to chewing.

For stepwise ICA estimation we captured the spatio-temporal dependency structure of the non-stationary EEG using a set of time delayed cross-covariance matrices with $\tau = [0, 1, \ldots, 5]$ from 4 s non-overlapping windows. For comparison we also applied conventional FastICA (kurtosis based) where the number of sources was determined by the index of the last eigenvalue of the data covariance matrix that accounted for at least 1% of the overall variance, which in this case yielded an estimate of 10 sources.

Figures 6(b) and (c) show the waveforms and scalp topographies of the ICs estimated using FastICA and the stepwise temporal decorrelation method, respectively. The ICs extracted using FastICA recognizably reflect seizure, eye blink and eye movement activity. However, the decomposition is not particularly 'clean' and appears to be dominated by muscle activity

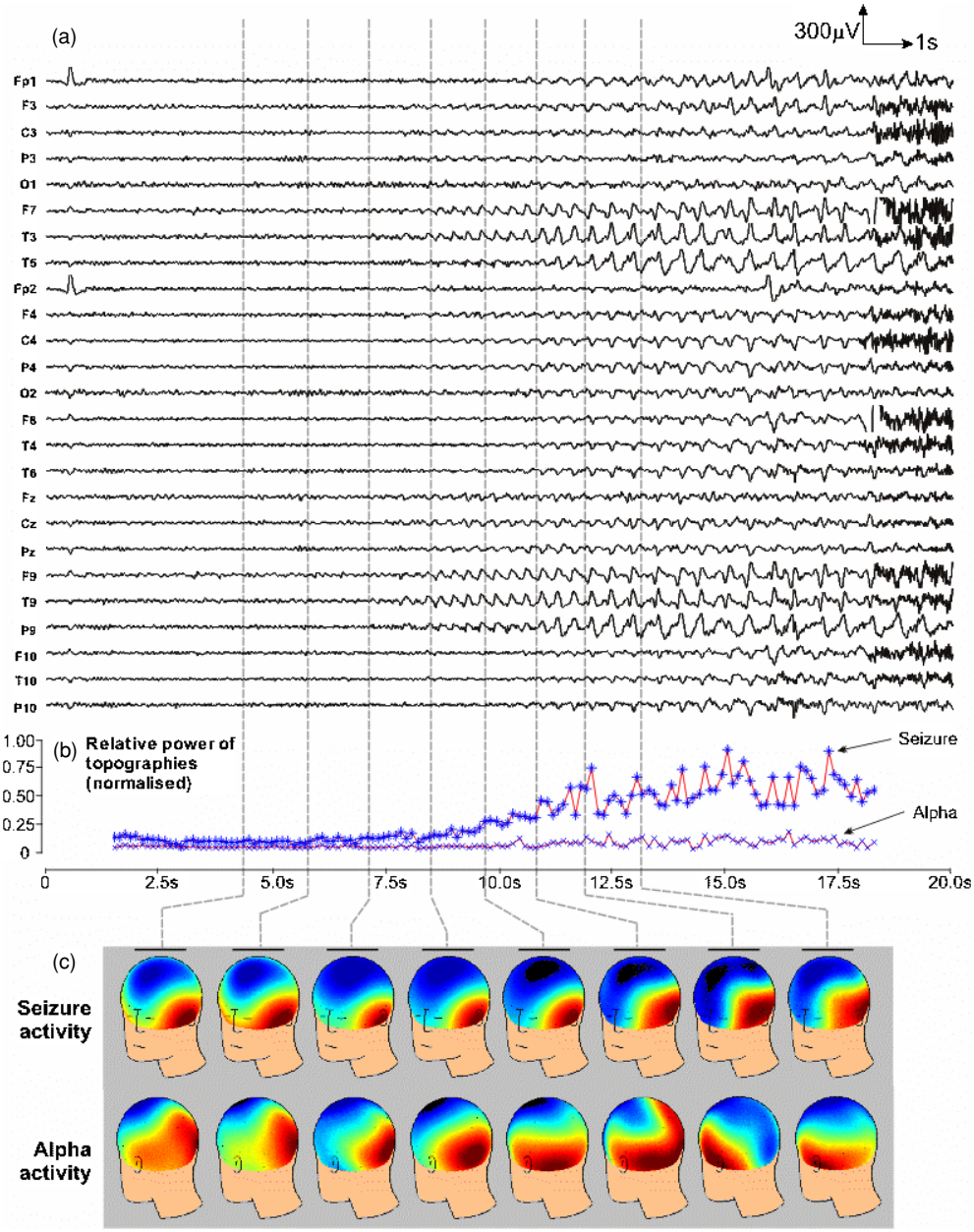

Figure 5. (a) A 20 s epoch of multi-channel seizure EEG to which the cICA method was repeatedly applied using a sliding window of 3 s width and moving by 125 ms. Rhythmic ∼5 Hz seizure activity appears 7–8 s into the recording with a left-temporal focus. (b) The relative power of the consecutive topographies obtained from cICA for each of the two references (normalized). (c) A number of topographies obtained for each reference channel.

(which in this case has high kurtosis). In contrast, the stepwise temporal decorrelation method yields only five sources (figure 6(c)), two of which relate to the seizure (S1, S2). The muscle

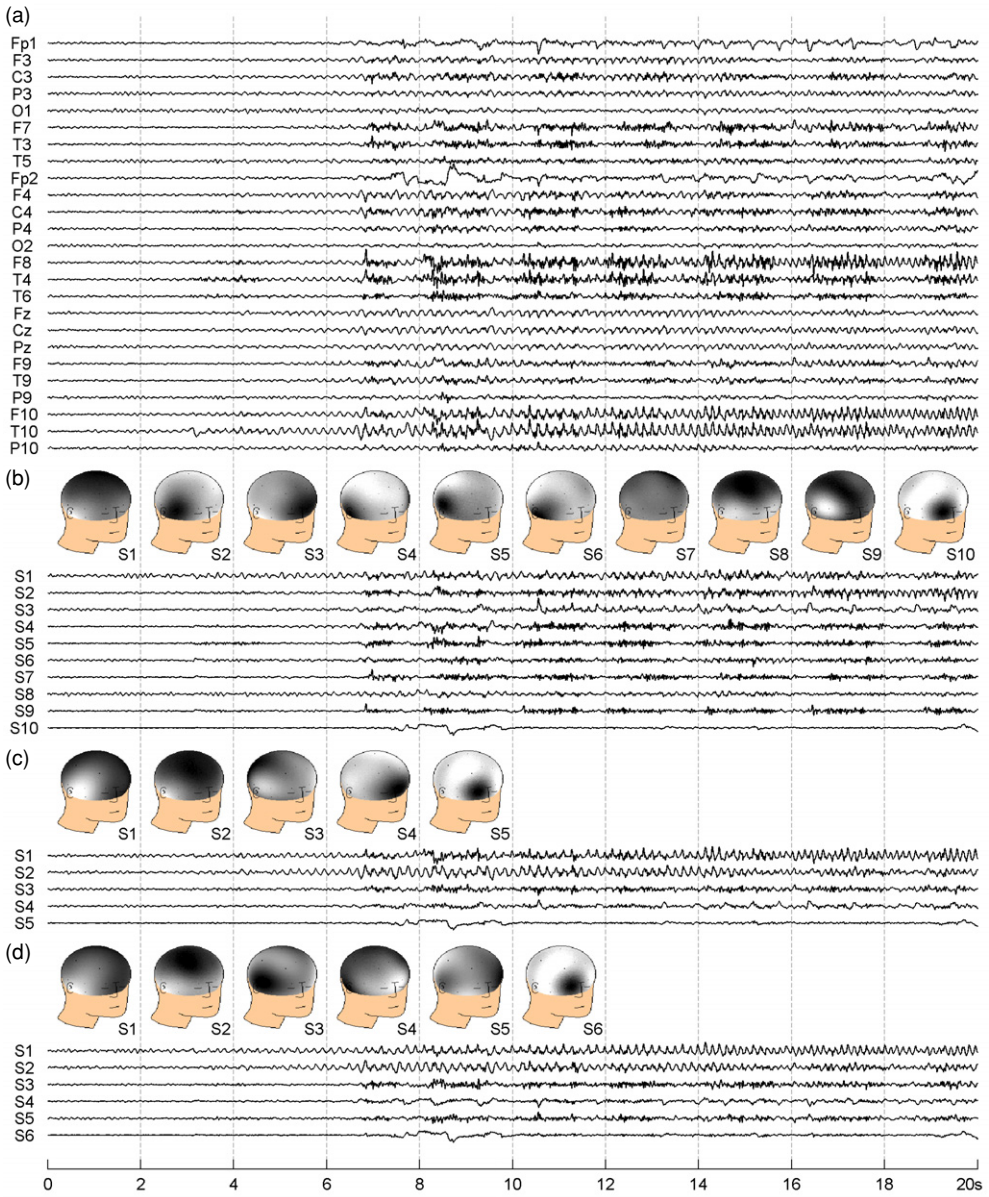

Figure 6. (a) A 20 s segment of 25-channel seizure EEG. The seizure has a right temporal focus and begins about 6 s into the segment. Waveforms and scalp topographies of sources extracted using FastICA with PCA based model order estimation are shown in (b), and stepwise estimation of the mixing matrix using time-lagged cross-covariances over a series of 4 s windows in (c), and wavelet transform sub-band cross-covariances in (d).

activity that dominated the FastICA solution is here only reflected in one component (S3). The ocular artifacts are clearly separated into blinks (S4) and saccades (S5).

This example shows how a small number of statistically independent, statistically significant and neurophysiologically meaningful components can be extracted by direct,

stepwise estimation of the mixing matrix, without the need for pre-whitening. Moreover, in this particular case, conventional PCA based model order estimation causes a HOS based ICA method such as FastICA to converge to an unacceptable, 'messy' solution.

4.4. Case study D: time–frequency based techniques

In this study we illustrate the use of ICA by multispectral decorrelation based on wavelets as a robust and efficient alternative to (windowed) temporal decorrelation (Hesse and James 2004a, 2004b, 2004c). Here, the data dependency structure is represented by means of cross-covariance matrices computed for each of the sub-bands of the discrete wavelet transform (DWT) of each channel. These are then used for stepwise estimation of the mixing matrix as described in the previous section.

The same set of EEG data as in case study C was used, and wavelet decomposition of each channel was based on a fourth-order Daubechies wavelet with decomposition level 6 (e.g. Mallat (1999)). Sub-band cross-covariances from all bands were used to estimate the mixing matrix. Figure 6(d) shows the waveforms and scalp topographies of the resultant six sources, which clearly reflect seizure activity (S1, S2), bilateral muscle activity (S3, S5) and ocular artefacts (S4, S6). Despite the presence of an additional source, these results are comparable with the windowed temporal decorrelation method.

One of the advantages of using the DWT is that the computation of all of the sub-band cross-covariances has the same computational cost as computing the (instantaneous) data covariance. Conversely, the cost associated with computing time-lagged cross-covariances for temporal correlation is proportional to the cost of computing the data covariance times the number of lags. Moreover, the number of cross-covariance matrices used to estimate the mixing matrix can be much smaller when compared with (windowed) temporal decorrelation methods, in this case there were only seven wavelet sub-band cross-covariance matrices compared with 30 time lagged cross-covariance matrices.

This example shows that a small number of neurophysiologically meaningful sources with non-stationary waveforms can be extracted from multi-channel seizure EEG using ICA based on multispectral decorrelation in the wavelet domain. Compared with ICA based on temporal decorrelation, this *Wavelet ICA* method is computationally more efficient.

5. Conclusions

This topical review sets out to describe the technique of ICA as a method for performing BSS in the context of biomedical signal processing. The generic technique of ICA is first discussed and the fundamental assumptions that are generally made in order to make the problem a more tractable one are introduced. Essentially ICA techniques make assumptions based on the mixing of the independent sources and (most importantly) based on the statistical independence of those sources. As already stated, the mixing assumptions such as those of linearity, stationarity and square mixing are made in order to allow specific embodiments of ICA to be easily formulated and may be relaxed at will—depending on the algorithm in use. The same holds, for example, for the assumption of noiseless mixing. It has been shown that although these assumptions might make ICA seem like a technique which is quite restricted in its potential applications, it has found many applications in the biomedical signal processing field in the literature.

Of the many possible algorithms devised towards solving the BSS problem, ICA is popularly solved through the use of HOS techniques—basically trying to separate statistically independent sources based on their non-Gaussianity. We show that another, more appealing

(from the viewpoint of biomedical signal analysis) viewpoint is that of using spatio-temporal and spatial-time frequency based ICA techniques. The main difference between the two being that in the latter technique the information inherent in the time-sequence of the measured data points is made use of—whereas in the former it is not. It can be seen that in the biomedical signal processing field where the analysis of information is generally based on the frequency content of recordings and on waveform morphology, such ICA techniques are invaluable.

We also show that ICA becomes much more informative as a technique when prior information is used to enhance the performance of the standard ICA models. We show this can be done by using only minor modifications of the estimation procedures, by essentially imposing constraints on the model. By including prior knowledge into the system and letting the ICA method of choice estimate the unknown portions this helps to interpret the results meaningfully. Whilst this can be simply seen as just making more rigid assumptions on the BSS model, as before, if the assumptions made are meaningful in the settings which they are made, then ICA will generally be more informative.

Finally it can be seen that the use of spatial, time and time–frequency dependency based techniques coupled with prior information make for a more flexible set of tools available to extract information from the underlying set of measurements. The technique of ICA can be in fact used to extract a lot of meaningful information from a set of measurements through just a few assumptions about the underlying sources and their mixing in the measurement space.

References

Amin M G, Zhang Y, Frazer G J and Lindsey A R 2003 Spatial time-frequency distributions: theory and applications *Wavelets and Signal Processing* ed L Debnath (Basle: Birkhäuser)

Bell A J and Sejnowski T J 1995 An information-maximization approach to blind separation and blind deconvolution *Neural Comput.* **7** 1129–59

Belouchrani A and Amin M G 1998 Blind source separation based on time-frequency signal representations *IEEE Trans. Signal Process.* **46** 2888–97

Belouchrani A, Abed-Meraim K, Cardoso J F and Moulines E 1997 A blind source separation technique using second order statistics *IEEE Trans. Signal Process.* **45** 434–44

Cardoso J F and Souloumiac A 1993 Blind beamforming for non Gaussian signals *IEE Proceedings* F **140** 362–70

Cardoso J F and Souloumiac A 1996 Jacobi angles for simultaneous diagonalization. *SIAM J. Matrix Anal. Appl.* **17** 161–4

Cardoso J F, Jutten C and Loubaton P (ed) 1999 *Proc. 1st Int. Workshop on Independent Component Analysis and Signal Separation (Aussois, France)*

Chatfield C 1996 *The Analysis of Time Series: An Introduction*. 5th edn (London: Chapman and Hall/CRC)

Choi S, Cichocki A and Belouchrani A 2002 Second order nonstationary source separation. *J. VLSI Signal Process.* **32** 93–104

Cichocki A and Amari S 2002 *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications* (New York: Wiley)

Cichocki A and Belouchrani A 2001 Source separation of temporally correlated source using bank of band pass filters *Proc. Int. Conf. on Independent Component Analysis and Blind Signal Separation (ICA2001) (San Diego, CA)* pp 173–8

Cohen L 1995 *Time-Frequency Analysis* (New York: Prentice-Hall)

Comon P 1994 Independent component analysis, a new concept? *Signal Process.* **36** 287–314

De Lathauwer L, De Moor B and Vandewalle J 2000 Fetal electrocardiogram extraction by blind source subspace separation *IEEE Trans. Biomed. Eng.* **47** 567–72

Donoho D L 1981 On minimum entropy deconvolution *Applied Time Series Analysis II* (New York: Academic) pp 565–608

EEGLAB MATLAB Package online at http://www.sccn.ucsd.edu/~scott/ica.html

FastICA MATLAB Package online at http://www.cis.hut.fi/projects/ica/fastica

Girolami M 2000 *Advances in Independent Component Analysis* (London: Springer)

Green P J and Taylor D P 2002 Dynamic signal enumeration algorithm for smart antennas *IEEE Trans. Signal Process.* **50** 1307–14

Haykin S 2000 *Unsupervised Adaptive Filtering, Vol 1: Blind Source Separation* (New York: Wiley)

Herault J and Jutten C 1986 Space or time processing by neural network models *Proc. AIP Conference: Neural Networks for Computing* vol 151 (New York: AIP)

Hesse C W and James C J 2004a Blind source separation through multi-spectral signal decorrelation applied to multi-channel ictal EEG *Proc. Med. Conf. on Medical and Biological Engineering (MEDICON2004, 31 July–5 August 2004, Ischia, Italy)*

Hesse C W and James C J 2004b An efficient time-frequency approach to blind source separation based on wavelets *Proc. Int. Conf. on Independent Component Analysis and Blind Signal Separation (ICA2004, 22–24 September 2004, Granada, Spain)*

Hesse C W and James C J 2004c A time-frequency approach to blind source separation using statistically optimal wavelet packets applied to ictal EEG *Proc. IEE Medical Signal and Information Processing Conference (MEDSIP2004, 5–8 September, Malta)*

Hesse C W and James C J 2004d Stepwise model order estimation in blind source separation applied to ictal EEG *Proc. Int. Conf. IEEE Engineering in Medicine and Biology Society (San Francisco, CA 1–5 September 2004)*

Hesse C W and James C J 2004e A stepwise statistical approach to model order estimation in blind source separation *Proc. IMA Mathematics in Signal Processing Conference (14–16 December 2004, Cirencester, UK)* pp 99–102

Hyvärinen A and Oja E 1997 A fast fixed-point algorithm for independent component analysis *Neural Comput.* **9** 483–1492

Hyvärinen A, Karhunen J and Oja E 2001 *Independent Component Analysis* (New York: Wiley)

ICA 1999 *Proc. 1st Int. Conf. on Independent Component Analysis and Blind Signal Separation (January 11–15, Aussois, France)*

ICA 2000 *Proc. 2nd Int. Conf. on Independent Component Analysis and Blind Signal Separation (June 19–22, Helsinki, Finland)*

ICA 2001 *Proc. 3rd Int. Conf. on Independent Component Analysis and Blind Signal Separation (December 9–12, San Diego, CA)*

ICA 2003 *Proc. 4th Int. Conf. on Independent Component Analysis and Blind Signal Separation (April 1–4, Nara, Japan)*

ICA 2004 *Proc. 5th Int. Conf. on Independent Component Analysis and Blind Signal Separation (September 22–24, Granada, Spain)*

James C J and D'Alimonte D 2004 Tracking multisource brain activity in multi-channel EEG with a probabilistic model through ICA *Proc. Mediterranean Conf. on Medical and Biological Engineering—MEDICON2004 (Ischia, Italy, 31 July–5 Aug.)* (CDROM)

James C J and Gibson O J 2003 Temporally constrained ICA: an application to artifact rejection in electromagnetic brain signal analysis *IEEE Trans. Biomed. Eng.* **50** 1108–16

James C J and Hesse C W 2004a Mapping Scalp topographies of rhythmic EEG activity using temporal decorrelation based constrained ICA *Proc. 26th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (1–5 September 2004, San Francisco, CA)* (CDROM)

James C J and Hesse C W 2004b Observing and tracking brain dynamics in electromagnetic brain signals through blind source separation *Proc. IEE Medical Signal and Information Processing Conference—MEDSIP2004 (Malta, 5–8 September 2004)* (CDROM)

James C J and Hesse C W 2004c A comparison of time structure and statistically based BSS methods in the context of long-term epileptiform EEG recordings *Independent Component Analysis and Blind Signal Separation (Proc. 5th Int. Conf., ICA2004, Granada, Spain, 22–24 September 2004) Lecture Notes in Computer Science 3195* ed C G Puntonet and A Prieto (Berlin: Springer) pp 1025–32

James C J and Lowe D 2000a Using dynamical embedding to isolate seizure components in the Ictal EEG *IEE Medical Signal and Information Processing Conference—MEDSIP 2000 (4–6 September 2000, Bristol)* pp 158–65

James C J and Lowe D 2000b Clustering Epileptiform discharges with an adaptive subspace self-organizing feature map: a simulation study *IEE Medical Signal and Information Processing Conference—MEDSIP 2000 (4–6 September 2000, Bristol)* pp 238–43

James C J and Lowe D 2001 ICA in electromagnetic brain signal analysis *Proc. 4th Int. Conf. Neural Networks and Expert Systems in Medicine & Healthcare—NNESMED 2001 (Greece)* pp 197–202

James C J and Lowe D 2003 Extracting multisource brain activity from a single electromagnetic channel *Artific. Intelligence Med. (Special issue)* **28** 89–104

Jung T P, Humphries C, Lee T W, Makeig S, McKeown M J, Iragui V and Sejnowski T J 1998 Extended ICA removes artifacts from electroencephalographic recordings *Adv. Neural Inf. Process. Syst.* **10** 894–900

Koehler B U and Orglmeister R 1998 Independent component analysis of electroencephalographic data using wavelet decomposition *Proc. Med. Conf. on Medical and Biological Engineering and Computing (Lemesos, Cyprus)*

Lagerlund T D, Sharbrough F W, Jack C R, Erickson B J, Strelow D C, Cicora K M and Busacker NE 1993 Determination of 10–20 system electrode locations using magnetic resonance image scanning with markers *Electroenceph. Clin. Neurophysiol.* **86** 7–14

Lee T W 1998 *Independent Component Analysis: Theory and Applications* (Boston: Kluwer)

Lu W and Rajapakse J C 2001 ICA with reference *Proc. 3rd Int. Conf. on Independent Component Analysis and Blind Signal Separation: ICA2001* pp 120–5

Makeig S, Jung T P, Bell A J, Ghahremani A and Sejnowski T J 1997 Blind separation of auditory event-related brain responses into independent components *Proc. Natl. Acad. Sci. USA,* **94** 10979–84

Mallat S 1999 *A Wavelet Tour of Signal Processing* 2nd edn (San Diego, CA: Academic)

McKeown M J, Makeig S, Brown G G, Jung T-P, Kindermann S S and Sejnowski T J 1998 Analysis of fMRI by blind separation into independent spatial components *Hum. Brain Mapping* **6** 160–88

Molgedey L and Schuster H G 1994 Separation of a mixture of independent signals using time delayed correlations *Phys. Rev. Lett.* **72** 3634–6

Nandi A (ed) 1999 *Blind Estimation Using Higher-Order Statistics* (Dordrecht: Kluwer)

Pajunen P and Karhunen J (ed) 2000 *Proc of the 2nd Int. Workshop on Independent Component Analysis and Blind Signal Separation (Helsinki, Finland)*

Pham D H 2001 Joint approximate diagonalization of positive definite matrices *SIAM J. Matrix Anal. Appl.* **22** 1136–52

Roberts S and Everson E 2001 *Independent Component Analysis: Principles and Practice* (Cambridge: Cambridge University Press)

Shalvi O and Weinstein E 1993 New criteria for blind deconvolution of nonminimum phase systems (channels) *IEEE Trans. Inf. Theory* **36** 312–21

Tong L, Liu R W, Soon V C and Huang Y F 1991 Indeterminacy and identifiability in blind identification *IEEE Trans. Circuits Syst.* **38** 499–509

Valaee S and Kabal P 2004 An information theoretic approach to source enumeration in array signal processing *IEEE Trans. Signal Process.* **52** 1171–8

Yeredor A 2002 Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation *IEEE Trans. Signal Process.* **50** 1545–53

Ziehe A, Laskov P, Müller K R and Nolte G 2003 A linear least-squares algorithm for joint diagonalization *Proc. Int. Conf. on Independent Component Analysis and Blind Signal Separation (ICA2003) (Nara, Japan)* pp 469–74

Ziehe A and Müller K R 1998 TDSEP—an efficient algorithm for blind separation using time structure *Proc. Int. Conf. on Artificial Neural Networks (ICANN'98, Skövde, Sweden)* pp 675–80