

Source Separation and Clustering of Phase-Locked Subspaces

Miguel Almeida, Jan-Hendrik Schleimer, José Mario Bioucas-Dias, *Member, IEEE*, and Ricardo Vigário

Abstract—It has been proven that there are synchrony (or phase-locking) phenomena present in multiple oscillating systems such as electrical circuits, lasers, chemical reactions, and human neurons. If the measurements of these systems cannot detect the individual oscillators but rather a superposition of them, as in brain electrophysiological signals (electro- and magnetoencephalogram), spurious phase locking will be detected. Current source-extraction techniques attempt to undo this superposition by assuming properties on the data, which are not valid when underlying sources are phase-locked. Statistical independence of the sources is one such invalid assumption, as phase-locked sources are dependent. In this paper, we introduce methods for source separation and clustering which make adequate assumptions for data where synchrony is present, and show with simulated data that they perform well even in cases where independent component analysis and other well-known source-separation methods fail. The results in this paper provide a proof of concept that synchrony-based techniques are useful for low-noise applications.

Index Terms—Clustering, phase locking, source separation, subspaces, synchrony.

I. INTRODUCTION

IN RECENT years, there has been an increase in scientific interest in the study of synchrony-related phenomena as well as in the amount of relevant results in this field, both theoretical and empirical. These phenomena are present in a multitude of physical substrata. The first detection of synchrony was made by Huygens in the 17th century, when he observed that two pendulum clocks interacting through a common supporting beam would always synchronize after a brief transient period [1]. Since then, other systems were found to exhibit similar behavior, such as organ pipes, electrical circuits, laser beams, astrophysical objects, some types of fireflies, and human neurons, among others. This behavior

Manuscript received September 13, 2010; revised June 22, 2011; accepted June 24, 2011. Date of publication July 25, 2011; date of current version August 31, 2011. This work was partially funded by Project P01075 of the Instituto de Telecomunicações, and by the Academy of Finland through its Centres of Excellence Program 2006–2011. The work of M. Almeida was supported in part by the Portuguese Science and Technology Foundation under Grant SFRH/BD/28834/2006 and in part by the Instituto de Telecomunicações under Grant BIM/15/2011.

M. Almeida and J. M. Bioucas-Dias are with the Instituto de Telecomunicações, Instituto Superior Técnico, Technical University of Lisbon, Lisbon 1169-047, Portugal (e-mail: malmeida@lx.it.pt; bioucas@lx.it.pt).

J.-H. Schleimer is with the Bernstein Center for Computational Neuroscience, Humboldt University, Berlin D-10099, Germany (e-mail: jan-hendrik.schleimer@bccn-berlin.de).

R. Vigário is with the Adaptive Informatics Research Centre, Aalto University, Aalto FI-00076, Finland (e-mail: ricardo.vigario@hut.fi).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TNN.2011.2161674

is often not caused by a strong interaction that forces the oscillators to oscillate in phase, but by weak interactions between the individual oscillators which, in time, drift their individual phases toward one another [1].

Our particular motivation for studying synchrony phenomena comes from the human brain. As an example, it has been shown that muscle activity measured with an electromyogram (EMG) and the activity of the motor cortex measured with an electroencephalogram (EEG) or magnetoencephalogram (MEG) have coherent oscillations when a person engages in a motor task. Because these coherent oscillations occur mostly in the beta-band¹ [2]–[5], the cortico-muscular coherence (CMC) phenomena are also denominated “beta-band synchronization,” although some studies have shown that they also occur in other frequency ranges [6], [7]. It was also found that, again during a motor task, several brain regions oscillate coherently with one another [5], [8]. In addition, there are multiple indications that several pathologies, including Alzheimer, Parkinson, and autism, are correlated with a disruption in the synchronization profile of the brain [9].

The typical formulation of a synchronization behavior is based on self-sustained oscillators as individual units. A self-sustained oscillator is a dynamical system, such as a pendulum clock, which has an intrinsic energy source and exhibits a periodic motion when isolated [1]. More rigorously, these oscillators have a limit cycle, which can be defined as a periodic trajectory in the oscillator’s phase space. The position along the limit cycle is the oscillator’s phase. Also, at least in a small neighborhood of this cycle, a self-sustained oscillator is stable, which means that, after being slightly perturbed, the trajectory always returns to the limit cycle after some time.

This return to the limit cycle has two distinct aspects. Perturbations along the limit cycle do not decay, and will result in a permanent change to the oscillator’s phase. On the other hand, perturbations orthogonal to this cycle decay exponentially.² For these reasons, a weak interaction can have a long-lasting effect on an oscillator’s phase, but the influence on its amplitude can be disregarded.

Assuming a weak, attractive, and time-invariant coupling, the interaction of N self-sustained oscillators affects their phases only, and can be described by the Kuramoto model [11]

$$\dot{\phi}_i(t) = \omega_i(t) + \frac{1}{N} \sum_{j=1}^N \kappa_{ij} \sin[\phi_j(t) - \phi_i(t)] \quad (1)$$

¹The beta-band range of brain signals is loosely defined as 15–30 Hz.

²The direction tangent to the limit cycle has a zero Lyapunov exponent, while the orthogonal directions have negative Lyapunov exponents [10].

where $\phi_i(t)$ and $\omega_i(t)$ are the instantaneous phase and intrinsic (natural) frequency of oscillator i , and $\kappa_{ij} > 0$ is called the coupling coefficient between oscillators i and j . Note that in this model, the coupling coefficients are constant in time, which implicitly assumes that the coupling between the oscillators is stationary in time. We do not address here the case of time-dependent coupling. The influence on ϕ_i is fully determined by the phase differences between oscillator i and each of the other oscillators. If oscillator j is slightly ahead of oscillator i (ϕ_j is slightly larger than ϕ_i), this will make oscillator i go slightly faster. Conversely, if oscillator j is slightly behind oscillator i , the latter will slow down. In both cases, this interaction tends to make the pairs of oscillators approach each other in phase, and synchrony can occur if this interaction is enough to compensate the difference in intrinsic frequencies $\omega_i - \omega_j$, called detuning. Naturally, if the detuning is large, a weak interaction cannot compensate it and the oscillators will never synchronize. These ideas are thoroughly explained in [1] and references therein.

Synchrony has been studied from a theoretical point of view since the original formulation by Kuramoto. Golomb *et al.* [12] have shown that, in a large and sparse neural network with weak coupling, there is a threshold value of the average number of synapses a neuron receives, above which synchrony is established and below which it is not, in a way reminiscent of Kuramoto's critical coupling coefficient [11]. The value of this threshold can be analytically found from the parameters of an "integrate-and-fire model." Integrate-and-fire oscillators, also known as relaxation oscillators, are described in more detail in the context of synchrony in [1], [13], and [14].

In real applications, it is often the case that one does not have direct access to the activity of the individual oscillators. Instead, one only has access to measurements of superpositions of the individual oscillators. The underlying oscillators are usually called "sources" in this context. When this is the case, synchrony between measurements is mostly due not to a true phase locking of the underlying sources, but instead to the mixture itself. We will discuss and illustrate this in Section II.

In this paper, we tackle the general problem of analyzing synchrony of the sources in the special case where the superposition is an instantaneous linear mixing. Such a problem is widely called "blind source separation" (BSS), because one generally only assumes rather generic properties of the sources, such as statistical independence or temporal structure. One important example of this problem is in brain electrophysiological signals (EEG and MEG), where instantaneous linear mixing is a valid assumption. This comes from the fact that most of the energy is in frequencies below 1 kHz, allowing the use of the quasistatic approximation in Maxwell's equations [15].

Independent component analysis (ICA) [16] is one of the most widely used BSS techniques. Traditional ICA approaches only use probability density function information (such as kurtosis) and therefore disregard the time dynamics of the sources. Thus, in applications where phase is relevant, ICA is not appropriate. A related approach, temporal decorrelation separation (TDSEP) [17], which is similar to second-order blind identification (SOBI) [18], is better suited for source

separation where dynamics are relevant. We will refer to these two approaches as "TD methods." TD methods extract sources using information on their autocorrelation function for several time lags simultaneously. Apart from the independence and temporal decorrelation criteria, approaches based on nonnegative matrix factorization have also been used (see [19], [20] for recent examples).

The main novelty of our paper is that it focuses exclusively on the phase information of the sources. By finding the sources in an appropriate way, and analyzing the synchrony between those, we avoid the problem of spurious synchronization mentioned above, allowing for more contrast between synchronized and nonsynchronized pairs of signals, thus permitting sharper detection. Two of the methods we propose are source-separation algorithms based on phase synchrony,³ applicable when one has a reference signal (Referenced Phase Analysis, RPA), or in a blind manner (Independent Phase Analysis, IPA). They are described in Section III-A and III-B, respectively.

We now review previous approaches on synchrony and coherence (which is a different but related concept), and their applications on brain signals. Vigário *et al.* and Meinecke *et al.* use TD methods to separate sources and observe that some of them are coherent [22] or synchronized [23]. This is a valid approach, but coherence and synchrony come as an epiphenomenon, rather than constituting the main searching criterion. We wish to focus on approaches that tackle synchrony directly, so that it is clear why the extracted sources are synchronous. We will show empirically in this paper that for certain types of sources TD methods are not adequate for source separation.

Nolte *et al.* [24] use the imaginary part of the Fourier coherence (IPC), noting that instantaneous mixing does not affect this measure. IPC does not work on the source space but only with the observed mixtures. Nolte *et al.* argue that this method should not replace other methods, but should instead be regarded as a "safe" method due to its low rate of false positives. In at least one experimental study, the IPC did not find any consistent results [25]. A drawback of this approach is that it has no time resolution. Therefore, it is only mostly applicable to stationary signals, and it will not be able to resolve nonstationarities. IPC should not be regarded as a BSS approach but rather as an approach that circumvents the problems that come from the superposition of the sources.

Allefeld and Kurths [26], [27] have proposed the synchronization cluster analysis (SCA) method, which detects synchrony through the eigendecomposition of the real-valued phase-locking factor (PLF) matrix, thus finding dominant phase-locked populations in the measured signals. This approach has the drawback of clustering together populations that are not interacting but that have similar frequencies. One of the methods we propose [phase SCA (pSCA)] is a generalization of such approach, which circumvents this drawback by using a complex-valued version of the PLF matrix (see Section III-C). Both SCA and pSCA are clustering methods and should not be considered source-separation techniques.

³Phase synchrony is a natural measure of dependence between signals. Other nonindependent source separation approaches have seen a recent growing interest, see [21].

The methods mentioned above vary on how the obtained phase is used. Apart from the methods we propose, only SCA focuses exclusively on the phase information. It has been shown that coupled oscillators can exhibit regimes with uncorrelated amplitudes but with bounded phase lag [28]. We therefore argue that there is a need for methods that focus only on the phase information. The techniques we propose attempt to fulfill this need: our methods focus on the phase information, unlike the traditional coherence-based approaches used in CMC studies. We believe that the proposed methods can contribute to a more precise characterization of synchrony in brain–brain or brain–muscle interactions. However, it is important to keep in mind that the algorithms we introduce here are applicable to any context where synchrony occurs, and do not assume anything specific of brain or muscle signals.

This paper is organized as follows. In Section II, we provide some background on synchrony and related concepts. The new algorithms we propose are detailed in Section III, along with results for simulated data. We discuss the algorithms' limitations and results in Section IV, and present concluding remarks in Section V. This paper summarizes and extends the results previously reported in [29]–[33].

II. SYNCHRONY

In this section, we provide some background on phase synchrony. We begin by noting the difference between synchrony and coherence. We then review the Hilbert transform and related concepts, introduce the PLF, which is central in the methods we propose, and define the synchronization matrix. We conclude with an illustration of the effect of linear mixing in synchrony.

A. Synchrony is Not Coherence

Terms such as “synchrony” and “coherence” are usually used to describe quantities related to frequency and phase. Here we make a formal distinction, to prevent confusion. Given two real signals $x(t)$ and $y(t)$, with power spectra $P_{xx}(\omega)$ and $P_{yy}(\omega)$, their spectral (or Fourier) coherence, also called coherence or coherency, is

$$C(\omega) \equiv \frac{|P_{xy}(\omega)|^2}{P_{xx}(\omega)P_{yy}(\omega)}$$

where $P_{xy}(\omega)$ is the cross-spectrum between the two signals. This quantity measures the similarity of two Fourier spectra. It can be understood as a correlation factor in frequency.

Since the Fourier transform forfeits time resolution, coherence measures are hard to interpret when computed from non-stationary signals, such as brain signals. However, coherence measures exhibiting time resolution have been proposed on the basis of the Morlet wavelet transform. This “wavelet transform coherence” is quite popular in geophysics [34], [35], but has also been used in the biomedical field [36].

B. Phase of a Real-Valued Signal

Typically, the signals under analysis are real-valued discrete signals. To obtain the phase of a real signal, popular choices

include using a complex Morlet (or Gabor) wavelet, which can be seen as a bank of band-pass filters [37], and using the Hilbert transform, which has to be applied to a locally narrowband signal or be preceded by appropriate filtering [38]. The Hilbert transform is then applied to multiple frequency bands of the original signal. The two approaches have been shown to be equivalent for the study of brain signals [39], but they may differ for other kinds of signals. In real applications, appropriate filtering must be used to ensure that the signals under analysis are narrowband. In this paper, the analyzed signals are narrowband by construction, thus we use the Hilbert transform.

C. Hilbert Transform, Analytic Signal, and Phase

The discrete Hilbert transform $x_h(t)$ of a band-limited discrete-time signal $x(t)$, $t \in \mathbb{Z}$, is given by a convolution [40]

$$x_h(t) \equiv x(t) * h(t), \text{ where } h(t) \equiv \begin{cases} 0, & \text{for } t = 0 \\ \frac{1-e^{i\pi t}}{\pi t}, & \text{for } t \neq 0. \end{cases}$$

The Hilbert filter $h(t)$ is not causal and has infinite duration, which makes direct implementation of the above formula impossible. In practice, the Hilbert transform is usually computed in the frequency domain, where the above convolution becomes a product of the discrete Fourier transforms of $x(t)$ and $h(t)$. A more thorough mathematical explanation of this transform is given in [38] and [40]. We used the Hilbert transform as implemented in MATLAB.

The analytic signal of $x(t)$, denoted by $\tilde{x}(t)$, is given by $\tilde{x}(t) \equiv x(t) + ix_h(t)$, where $i = \sqrt{-1}$ is the imaginary unit. The phase of $x(t)$ is defined as the angle of its analytic signal.

D. PLF

The common feature of our proposed methods is their use of the PLF, which plays a central role in this paper of synchrony phenomena [41]. For two oscillators with phases $\phi_1(t)$ and $\phi_2(t)$ for $t = 1, \dots, T$, the PLF is defined as⁴

$$\varrho \equiv \frac{1}{T} \sum_{t=1}^T e^{i[\phi_1(t)-\phi_2(t)]} = \langle e^{i(\phi_1-\phi_2)} \rangle$$

where $\langle \cdot \rangle$ denotes a time-average operation. The PLF obeys the constraints $0 \leq |\varrho| \leq 1$. The value of 1 is attained when the two oscillators are perfectly synchronized, i.e., they have a constant phase lag. The value of 0 is attained, e.g., if the phase difference $\phi_1 - \phi_2$ is uniformly distributed in $[0, 2\pi]$. Values between 0 and 1 indicate partial synchrony. Just as the coherence can be seen as a correlation factor in frequency, the PLF can be seen as a correlation factor in phase. Some authors (e.g., [27] and [41]) use the absolute value of ϱ as PLF. However, this complex version of the PLF can be used to obtain better results, as we will show in Section III-C.

⁴Recall that in 1 we assumed stationarity of the coupling coefficient, hence the use of the whole observation period in computing the PLF.

E. Synchronization Matrix

When one has multiple signals with phases $\phi_1(t), \dots, \phi_N(t)$, it is common to compute all the pairwise PLFs and store them in a complex synchronization matrix (or PLF matrix) \mathbf{Q}

$$\mathbf{Q}(j, k) \equiv \left\langle e^{i(\phi_j - \phi_k)} \right\rangle = \left\langle \mathbf{a}\mathbf{a}^H \right\rangle \quad (2)$$

where $\mathbf{a} \equiv [e^{i\phi_1(t)} \dots e^{i\phi_N(t)}]^T$ and $(\cdot)^H$ denotes the Hermitian. This matrix has all its diagonal values equal to 1, and it is Hermitian, which means that all its eigenvalues are real. Also, for any vector \mathbf{x} , $\mathbf{x}^H \mathbf{Q} \mathbf{x} = \mathbf{x}^H \langle \mathbf{a}\mathbf{a}^H \rangle \mathbf{x} = \langle \mathbf{x}^H \mathbf{a}\mathbf{a}^H \mathbf{x} \rangle = \langle \|\mathbf{x}^H \mathbf{a}\|^2 \rangle \geq 0$. Therefore \mathbf{Q} is positive semidefinite. Unlike the definition used in [27], in our case this matrix is complex.

F. Effect of Mixing in Synchrony

Two of the methods described in this paper aim at unmixing linear and instantaneous mixtures of sources, using the PLF as the main criterion. As motivated above, in EEG and MEG it is plausible to assume that the observed measurements are the result of a linear and instantaneous mixture of the underlying source signals. Thus, it is important to understand the effect of such mixing on the PLF. Such effect can be summarized as “tending toward partial synchrony”: if some sources have low synchrony (PLF ≈ 0), the mixed signals have a higher PLF, since each source is now present in both mixed signals. If some sources have high synchrony (PLF ≈ 1), the mixed signals have a lower PLF, because now each mixed signal has components from sources that were not phase-locked.

These statements are illustrated in Fig. 1. The top left subfigure shows a set of 12 sources. The pairwise PLFs of these sources are shown in the top middle subfigure, where larger squares represent higher PLF values. A random linear mixture of these 12 sources is depicted in the bottom left subfigure, and the pairwise PLFs of the mixed signals are shown in the bottom middle subfigure. Note that partial synchrony is now present in all pairs of measurements. This experiment illustrates the need for source separation methods that tackle synchrony problems directly. These methods should not be regarded as replacements or improvements of techniques such as ICA or TDSEP, but rather as related techniques that are applicable for different kinds of data.

III. PLF ALGORITHMS

In this section, we introduce three methods to analyze synchrony. We show their usefulness with simulated examples using instantaneous linear mixing.

- 1) RPA performs nonblind source separation when a reference signal is available, extracting the projection which has maximum PLF with the reference.
- 2) IPA performs blind unmixing of sources that are organized in subspaces, i.e., sources that have PLFs of 1 or 0 with one another. IPA works in two stages: it first unmixes the subspaces, and then unmixes the sources within each subspace.
- 3) pSCA is a nontrivial generalization of SCA [27]. In the original formulation of SCA, the essential procedure

is to perform an eigendecomposition of the real-valued synchronization matrix, i.e., the elementwise absolute value of the synchronization matrix defined in Section II. We will show that, instead of forfeiting the phase information, when taking the absolute value of the PLF, it is more useful to include it, to prevent multiple populations from ending in the same cluster, a problem which was present in the original SCA and will be illustrated here.

A. Referenced Phase Analysis

In electrophysiological recording signals from the brain, such as EEG and MEG, we often have not only the scalp measurements but also a reference signal that provides a hint of what to look for in the brain. This reference signal could be, for example, an EMG, collecting information of a muscle’s activity. The objective of RPA is to extract a source which is phase-locked to a reference signal [29]. In the context of brain signals, it would allow, for example, the identification of which cortical areas are phase-locked to the muscle activity, evidencing neuronal control of the muscle or sensory feedback from it.

As usual in linear separation techniques, we assume that the observations $\mathbf{x}(t)$ result from a linear and instantaneous superposition of the sources $\mathbf{s}(t)$, as in

$$\mathbf{x}(t) \equiv \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t) \quad (3)$$

where $\mathbf{n}(t)$ is noise. Throughout this paper, we will only consider the case where the noise $\mathbf{n}(t)$ is negligible.

Also usual in linear separation is a preprocessing step called “whitening,” or “sphering” [16], which usually results in a numerically better conditioned problem. We assume, with no loss of generality, that the data have zero mean. The whitening procedure starts with the computation of the empirical covariance matrix $\mathbf{C} = \langle \mathbf{x}\mathbf{x}^T \rangle$. Then one computes the eigenvector decomposition of \mathbf{C} . One can store the eigenvalues of \mathbf{C} in a diagonal matrix \mathbf{D} in descending order, and the corresponding eigenvectors as the columns of a matrix \mathbf{V} . Whitened data \mathbf{x}_w can be obtained through

$$\mathbf{x}_w(t) \equiv \mathbf{D}^{-\frac{1}{2}} \mathbf{V}^T \mathbf{x}(t). \quad (4)$$

We assume that we have a reference signal $u(t)$, with its phase denoted by $\psi(t)$. We define the estimated source as $y(t) \equiv \mathbf{w}^T \mathbf{x}(t)$, denoting its phase by $\phi(t)$, with the dephasing defined as $\Delta\phi(t) \equiv \phi(t) - \psi(t)$. We can now compute the absolute value of the PLF between the estimated source and the reference signal as

$$|\varrho| = \left| \frac{1}{T} \sum_{t=1}^T e^{i\Delta\phi(t)} \right| = \left| \frac{1}{T} \sum_{t=1}^T \frac{\tilde{y}(t)\tilde{u}^*(t)}{|\tilde{y}(t)\tilde{u}(t)|} \right| = \left| \left\langle \frac{\tilde{y}(t)\tilde{u}^*(t)}{|\tilde{y}(t)\tilde{u}(t)|} \right\rangle \right| \quad (5)$$

where \tilde{y} and \tilde{u} are the analytic signals of y and u , obtained through the Hilbert transform.

The idea of this algorithm is to maximize $|\varrho|$, or equivalently $|\varrho|^2$, with respect to \mathbf{w} . The global maximizer, denoted by \mathbf{w}_{opt} , will be the linear combination maximizing the synchrony between the estimated source $y(t) = \mathbf{w}_{\text{opt}}^T \mathbf{x}(t)$ and the

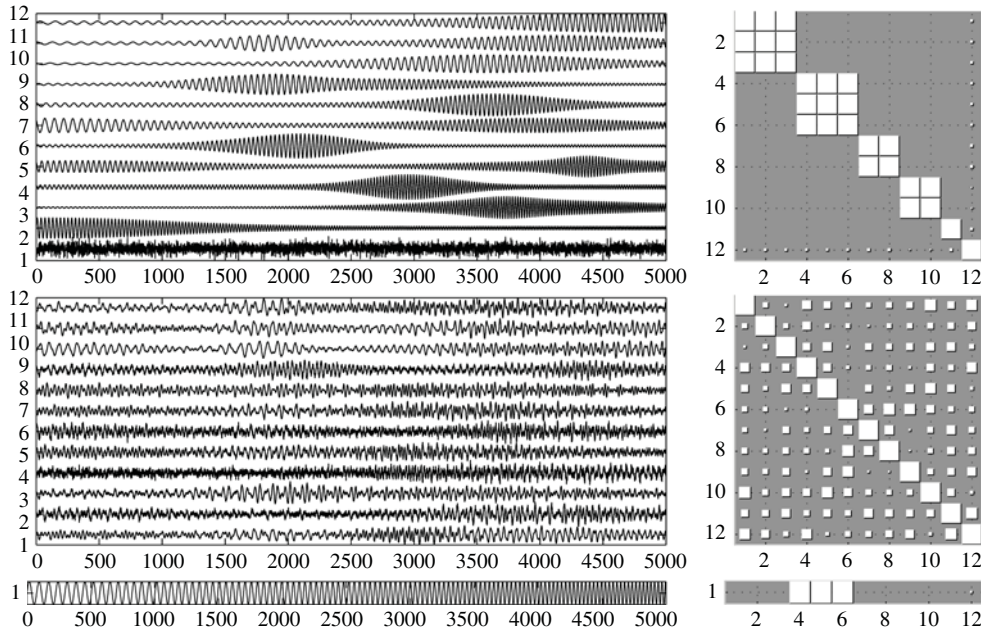


Fig. 1. Dataset used throughout this paper. (First row) Original sources (left) and PLFs between them (right). (Second row) Mixed signals and PLFs between them. (Third row) Reference signal (left), which is a sinusoid with varying frequency, and its PLF with the 12 sources (right). The reference is phase-locked with sources 4, 5, and 6. In the second column, the numbers denote the index of the sources, and the area of each square is proportional to the absolute value of their pairwise PLF.

reference signal. We want to find $y(t)$ only up to scale and sign, and therefore we constrain \mathbf{w} to have unit norm.

The gradient of $|\varrho|^2$ w.r.t. \mathbf{w} is given by [42]

$$\nabla |\varrho|^2 = 2|\varrho| \left\langle \frac{\sin[\Phi - \Delta\phi(t)]}{Y^2(t)} \mathbf{\Gamma}_x(t) \right\rangle \mathbf{w} \quad (6)$$

where $Y(t) \equiv |\tilde{y}(t)|$ is the amplitude of the estimated source, $\Phi \equiv \text{angle}(\varrho)$ is the phase of the PLF, and $\mathbf{\Gamma}_x(t) = \mathbf{x}_h(t)\mathbf{x}_h^\top(t) - \mathbf{x}(t)\mathbf{x}_h^\top(t)$ [where $\mathbf{x}_h(t)$ is the Hilbert transform of $\mathbf{x}(t)$] is a matrix that can be precomputed, because it depends only on the data. Since $\mathbf{x}(t) = \text{Re}(\tilde{\mathbf{x}}(t))$ and $\mathbf{x}_h(t) = \text{Im}(\tilde{\mathbf{x}}(t))$, it can easily be seen that $\Gamma_{x_{ij}}(t) = X_i(t)X_j(t)\sin(\varphi_i(t) - \varphi_j(t))$, where $X_i(t)$ and $\varphi_i(t)$ are the amplitude and phase of $x_i(t)$.

There are many procedures to find the maximizer of a given objective function. We chose to use a gradient ascent algorithm with adaptive step sizes due to its simplicity. A step-by-step description of RPA is presented in Table I.

If the global maximizer is found, RPA outputs the linear combination of the data that is maximally phase-locked to the reference. If two sources are maximally phase-locked to the reference, there are two correct solutions for the problem, and the algorithm will output one or the other, depending on the initialization of the weight vector \mathbf{w} . However, the algorithm never outputs linear combinations of the two solutions, because such mixtures have a lower PLF with the reference.⁵

1) *Application to Simulated Data:* We applied the algorithm described above to a set of noiseless simulated data. The dataset's sources have unit variance and varying instantaneous

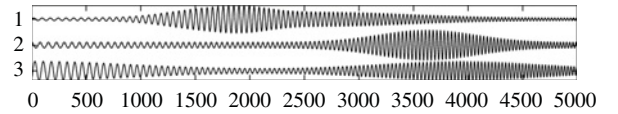


Fig. 2. Results of the RPA algorithm. The algorithm correctly finds the three sources phase-locked to the reference signal. PLF between the found sources and reference is 0.999 (chosen as stopping criterion). These three sources were found on three separate runs of the algorithm.

frequency (see the first row of Fig. 1). The reference signal, also shown in Fig. 1, is a sinusoid with the same instantaneous frequency as the fourth, fifth, and sixth source signals, with exactly zero phase lag with the fourth source.

These sources were mixed without noise, using a random mixing matrix, with entries uniformly distributed between -1 and 1 . The algorithm was then run on the mixed signals, which are shown on the second row of Fig. 1. Depending on the initial conditions, RPA will return one of the three solutions shown in Fig. 2 each time it is run. The results in Fig. 2 show that, despite having three sources phase-locked to the reference (sources 4–6), the algorithm returns one of them separately without mixing them.

These results show that RPA works as expected in the ideal noiseless condition.

B. Independent Phase Analysis

In this section, we introduce the IPA algorithm. This algorithm separates sources that are organized in subspaces, such that the intra-subspace PLFs are 1 and the inter-subspace PLFs are 0. The general idea is similar to independent subspace analysis [43]. IPA is a true BSS method, in contrast to RPA which is not blind. The original formulation of IPA was

⁵An exceptional case is when the two sources have a phase lag of exactly 0 or π between themselves, in which case linear combinations of those two sources have the same PLF with the reference.

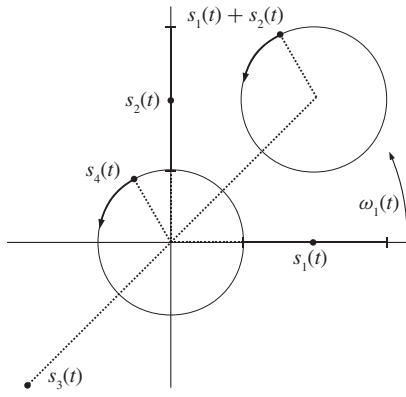


Fig. 3. Three signals, s_1 , s_2 , and s_3 , that are fully phase-locked (PLF = 1), and a linear combination of them, $s_4 = s_1 + s_2 + s_3$, that has zero PLF with all of them. The picture shows the four signals in a frame rotating with angular speed $\omega_1(t)$. The PLF between s_4 and any of the other three sources is zero since the average angle between s_4 and any other source is zero.

heavily motivated by ICA [30], but it has since been improved [32], [33].

We assume that a set of sources undergoes linear mixing as in (3), and also that the data \mathbf{x} has been whitened as in (4). Let the estimated sources be denoted by $\mathbf{y}(t) \equiv \mathbf{W}^T \mathbf{x}(t) = \mathbf{W}^T \mathbf{A} \mathbf{s}(t)$. Our goal is to find \mathbf{W} such that $\mathbf{W}^T \mathbf{A}$ is a permutation of a diagonal matrix, in which case the estimated sources $\mathbf{y}(t)$ are equal to the original sources $\mathbf{s}(t)$ up to permutation, scaling, and sign. We will limit ourselves to the case where the number of sources is equal to the number of measured signals. Therefore, \mathbf{W} is a square matrix.

Since the objective of this algorithm is to extract sources that have pairwise PLFs of 1 or 0, one might initially consider a cost function that is minimized at those values of PLF. That was the motivation behind our first attempts to solve this problem, which met with some success [30], [32]. Unfortunately, that approach gives poor results when the mixing matrix \mathbf{A} is far from orthogonal. This limitation is due to the following fact: given three phase-locked signals s_1 , s_2 , and s_3 , it is sometimes possible to construct a linear combination $s_4(t) = as_1 + bs_2 + cs_3$, such that the PLFs between s_4 and s_1 , s_2 and s_3 are zero. A simple example, illustrated in Fig. 3, is

$$\begin{aligned}\tilde{s}_1(t) &= (2 + \cos(\omega_2 t))e^{i\omega_1(t)t} \\ \tilde{s}_2(t) &= (2 + \sin(\omega_2 t))e^{i\omega_1(t)t + \frac{\pi}{2}} \\ \tilde{s}_3(t) &= 2\sqrt{2}e^{i\omega_1(t)t - \frac{3\pi}{4}} \\ \tilde{s}_4(t) &= \tilde{s}_1(t) + \tilde{s}_2(t) + \tilde{s}_3(t) \\ &= e^{i(\omega_1(t)t + \omega_2 t)}.\end{aligned}$$

The above-mentioned hypothetical cost function would be unable to distinguish between the correct subspace $\{s_1, s_2, s_3\}$ and an incorrect subspace such as $\{s_1, s_2, s_4\}$. In more formal terms, this problem formulation is ill posed.

To prevent this ill-posedness, we have recently proposed that the full problem be divided into two subproblems [33]: first, separate the subspaces from one another, even if some mixing remains within each subspace. Second, unmix the sources within each subspace, one subspace at a time. The key idea is

that, if the first subproblem is successful, the second part needs only separate sources that have PLFs equal to 1, therefore avoiding the above mentioned ill-posedness. We now discuss each of these subproblems in detail.

1) *Inter-Subspace Separation and Subspace Detection*: This first subproblem aims to find an unmixing matrix \mathbf{W} such that the estimated subspaces are correct. We assume that signals that are in distinct subspaces will have little interaction with each other, which should usually correspond to different dynamics in time. Techniques using temporal structure to perform separation should be adequate to this first subproblem. Therefore, we chose to use Ziehe *et. al.* implementation of TDSEP [17] for this first subproblem, but SOBI [18] can be used instead. Although we have no theoretical results that support TDSEPs adequacy to this task, we have repeatedly observed that it separates subspaces of various sizes quite well.

Detecting the subspace structure (which signals belong to which subspace) from the results of TDSEP is not trivial and warrants some discussion. From our experience, TDSEP can perform the inter-separation very well, but cannot adequately do the intra-subspace separation. This means that PLF values within each subspace will be underestimated. One (admittedly crude) solution for this step is to define a hard PLF threshold, above which signals are considered synchronized and form part of the same subspace, and below which they do not.

The matrix resulting from this hard thresholding should be block-diagonal, with each block having all elements equal to 1. If this is the case, no inconsistencies were found (i.e., no signal belongs to two subspaces), and we can move on to the second subproblem which separates the sources within each subspace (see Section III-B.2). Let the unmixing matrix estimated by TDSEP be denoted by \mathbf{W}_{tdsep} .

If the matrix resulting from the thresholding is not block-diagonal with all blocks filled with 1s, our algorithm considers that the subspaces were wrongly detected and returns the results of TDSEP. See Section IV for possible improvements.

2) *Intra-Subspace Separation*: In the second stage of IPA, we select the subset of columns of \mathbf{W}_{tdsep} that form the l th subspace, which is denoted by S_l , and concatenate them into a rectangular matrix $\mathbf{W}_{tdsep,l}$. Let N_l denote the number of signals in S_l and let $\mathbf{z}_l(t) = \mathbf{W}_{tdsep,l}^T \mathbf{x}(t)$ be the vector of sources in S_l estimated by TDSEP.

In this second stage, the goal is to separate the sources in S_l . This procedure is then repeated for all subspaces found by TDSEP. As explained above, each source in S_l should have a PLF of 1 with all other sources in S_l . Generally, TDSEP will severely underestimate this value (see top row of Fig. 4). We should therefore unmix the N_l sources found by TDSEP such that their PLFs are as high as possible. Mathematically, this corresponds to finding a N_l by N_l matrix \mathbf{W}_l such that the estimated sources in the l th subspace, $\mathbf{y}_l(t) = \mathbf{W}_l^T \mathbf{z}_l(t) = \mathbf{W}_l^T \mathbf{W}_{tdsep,l}^T \mathbf{x}(t)$, have the highest possible PLFs.

It turns out that, if all the sources are in the same subspace and they are all fully synchronized with each other (PLF = 1), then the only vector of the form $\mathbf{y} = \mathbf{B} \mathbf{s}$ such that the estimated sources \mathbf{y} have a PLF of 1 with each other is $\mathbf{y} = \mathbf{s}$ up to permutation, scaling, and sign (see Appendix A). An immediate corollary is that, if the inter-subspace separation

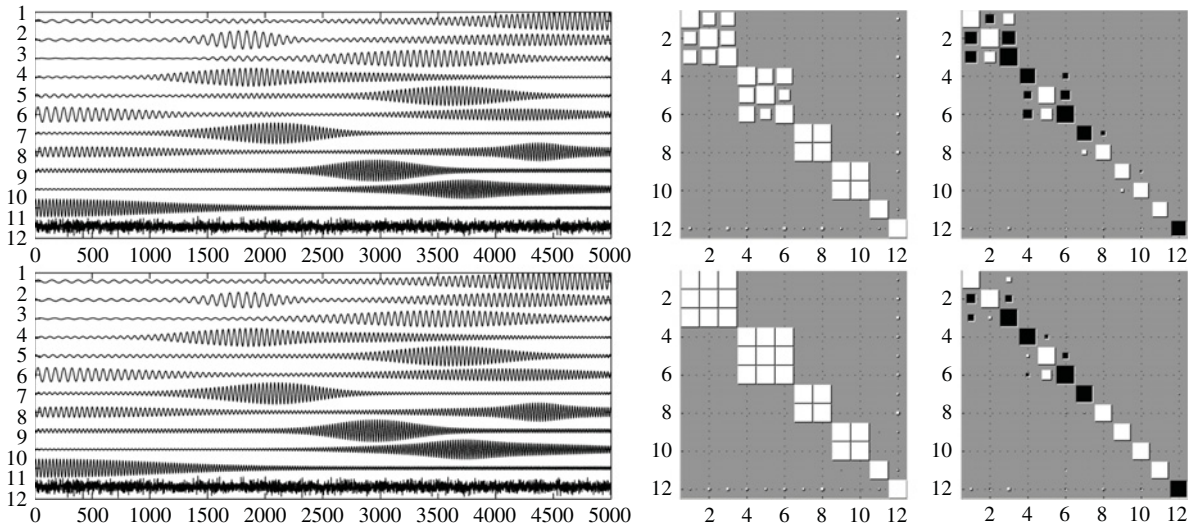


Fig. 4. (First row) Sources resulting from TDSEP (left). Note that the inter-subspace PLFs (middle) are very close to zero, but the intra-subspace PLFs are not all close to 1. Further, the intra-space separation is poor, as can be seen from inspection of the product $\mathbf{W}_{tdsep}^T \mathbf{A}$ (right). (Second row) Results found after the second stage of the algorithm. The estimated sources (left) are very similar to the original ones. This is corroborated by the PLFs between the estimated sources (middle) and the final unmixed matrix (right). The permutation of the sources was corrected manually. White squares represent positive values, while black squares represent negative values.

can fully separate the subspaces from one another, the intra-subspace separation has a unique solution.

For each subspace l , the objective function to maximize is

$$J_l = \frac{1 - \lambda}{N_l} \sum_{j,k \in S_l} |\varrho_{jk}|^2 + \lambda \log |\det \mathbf{W}_l| \quad (7)$$

where $\sum_{j,k \in S_l} |\varrho_{jk}|^2$ is the sum over all pairs of sources in subspace S_l of the square of the absolute PLF between those sources. We use the Hilbert transform [38] to obtain the phase values of the estimated sources (see Section II). The second term prevents the algorithm from finding solutions which trivially have $|\varrho_{jk}| = 1$, i.e., “finding the same source twice,” and $\lambda \in [0, 1]$ is a parameter controlling the relative weight of the two terms. Each column of \mathbf{W}_l is constrained to have unit norm to prevent trivial decreases of the penalty term.

Using this formulation, the problem is no longer ill posed as above, since we no longer consider a PLF of zero as a valid solution. Furthermore, only a subset of parameters need be optimized at a time, which can drastically reduce the time needed to separate a set of sources.

The gradient of J_l relative to a column \mathbf{w}_j of the matrix \mathbf{W}_l is given by [42] (we omit dependences on l for clarity)

$$\nabla_{\mathbf{w}_j} J_l = 4 \frac{1 - \lambda}{N^2} \sum_{k=1}^N |\varrho_{jk}| \left\langle \sin [\Psi_{jk} - \Delta\phi_{jk}(t)] \frac{\mathbf{\Gamma}_z(t)}{Y_j(t)^2} \right\rangle \mathbf{w}_j \quad (8)$$

where ϱ_{jk} is the PLF between two estimated sources j and k , $Y_j \equiv |\tilde{y}_j|$, where \tilde{y}_j is the analytic signal of the j th estimated source, $\phi_j \equiv \text{angle}(\tilde{y}_j)$ and $\Delta\phi_{jk}(t) \equiv \phi_j(t) - \phi_k(t)$ is the instantaneous phase difference between two estimated sources, $\Psi_{jk}(t) \equiv \langle \Delta\phi_{jk}(t) \rangle$ is the average phase difference between two estimated sources, and $\mathbf{\Gamma}_z(t) = \mathbf{z}_{hl}(t)\mathbf{z}_l^T(t) - \mathbf{z}_l(t)\mathbf{z}_{hl}^T(t)$,

as in RPA, is a matrix that can be precomputed, because it depends only on the data resulting from TDSEP.

3) *Application to Simulated Data:* We present results that show that IPA can successfully separate sets of sources with nontrivial subspace structure. After running TDSEP, we optimize the objective function in (7) using a gradient ascent algorithm with adaptive step sizes, which runs until a convergence criterion is met. The parameter λ was hand tuned for optimal performance. However, similar results are obtained for λ within a factor of 2 of the optimal value, which in this case is $\lambda = 0.1$.

We simulate the noiseless instantaneous linear mixture of the 12 sources depicted in the first row of Fig. 1. These sources are organized in six clusters of sizes 3, 2, and 1. We generate 300 mixing matrices with random entries taken from the Uniform($-1, 1$) distribution, and use those to generate 300 different datasets. We then run the algorithm once for each mixing matrix, for a total of 300 runs. Each run takes about 1 min on a modern laptop computer.

The second row of Fig. 1 shows the mixed signals that are the input to our algorithm. The first row of Fig. 4 shows the sources estimated by TDSEP. Visual inspection of the PLFs between these sources (second column) shows that some of the estimated sources do not have high PLFs. Inspection of the product $\mathbf{W}_{tdsep}^T \mathbf{A}$ (third column) reveals a very good inter-subspace separation, but a poor intra-subspace separation. Finally, the second row of Fig. 4 reveals that by maximizing the intra-subspace PLFs we can significantly improve the separation within each subspace. This is best noted through the product $\mathbf{W}^T \mathbf{A}$ depicted on the third column.

We analyze the results of IPA using the Amari performance index (API) [44], which measures the average relative contamination in each estimated source from all other sources. The API is nonnegative and decreases to zero as

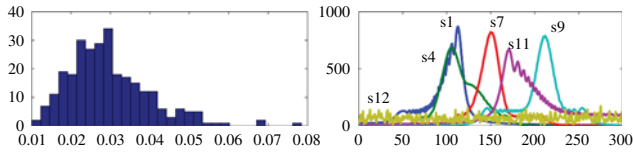


Fig. 5. Histogram of the Amari performance index for 300 runs, corresponding to 300 random mixing matrices for these sources (left), and discrete Fourier transform of the 1st, 4th, 7th, 9th, 11th, and 12th sources, one from each subspace (right). The latter illustrates that simple band-pass filters cannot separate the subspaces.

the separation quality improves. A histogram of the API for the 300 runs of IPA is shown on Fig. 5. The mode of this histogram is 0.03, which is also the API of the example in Fig. 5.

We also measured the performance of TDSEP and IPA using the well-known signal-to-noise (SNR) criterion. After matching the estimated sources to the true sources, we compute the SNR of the sources (in energy) estimated by TDSEP and by IPA.⁶ The average SNR is 16.78 dB for TDSEP and 24.18 dB for IPA, which shows that IPA yields an average increase of 7.4 dB of the estimated sources relative to TDSEP alone. Furthermore, the histograms in Fig. 6 show that the SNR distribution of TDSEP is skewed toward low values, while the values from IPA have much less skewness. In other words, IPA has a much lower probability of yielding very low SNR values (below, say, 10 dB).

We used a threshold of 0.1 on the squared PLF matrix for the detection of subspaces. In 7% of the runs, the matrix resulting from TDSEP and the thresholding is not block-diagonal with blocks full of ones, and therefore the algorithm stops. See Section IV for a discussion of this limitation.

C. pSCA

In pSCA, we are interested in clustering a population of oscillators into synchronous clusters. Contrary to RPA and IPA, here we assume that the sources are not mixed. It is important to remark that in SCA and pSCA no whitening is performed.

We begin by reviewing the original synchronization cluster analysis (SCA) recently proposed by Allefeld *et al.* [27] to automatically cluster a number of oscillators into subpopulations. We empirically show that, for signals that follow the Kuramoto model, SCA can sometimes cluster oscillators incorrectly. We will show that under this model the absolute value of the PLF is not enough to determine whether two signals are coupled and introduce pSCA as an improved method that can perform a correct clustering in some of those situations.

1) *SCA*: SCA can be seen as a clustering method that attempts to cluster the original set of oscillators into subpopulations such that the intra-population PLFs are high but the inter-population PLFs are low.

⁶See Section IV for more details on the SNR calculation.

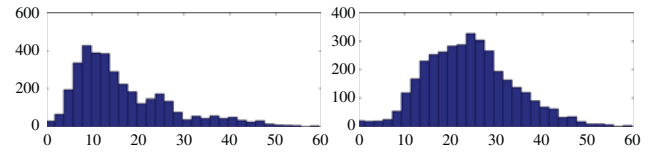


Fig. 6. (Left) Histogram of the signal-to-noise ratio (SNR) between the sources found by TDSEP and the original sources. (Right) Similar histogram for the sources found by IPA.

SCA begins with the construction of the real-valued synchronization matrix \mathbf{R}

$$\mathbf{R}_{ij} \equiv |\mathbf{Q}_{ij}| = \left| \left\langle e^{i(\phi_i - \phi_j)} \right\rangle \right|.$$

Note that this real-valued matrix is the elementwise absolute value of the complex synchronization matrix \mathbf{Q} defined in (2). The matrix \mathbf{R} is symmetric and its elements can only take real values between 0 and 1.

The main step in SCA consists of an eigendecomposition of \mathbf{R} . As shown in [27], the eigenvalues of this matrix are real and nonnegative, and their sum obeys $\text{tr}(\mathbf{R}) = N$, where N is the total number of oscillators. The largest eigenvalue of \mathbf{R} , λ_1 , is an estimate of the strength of the largest cluster, and the associated eigenvector \mathbf{v}_1 has in its j th component the relative participation of oscillator j in cluster 1. Similar properties hold for the remaining eigenvalues and eigenvectors. The participation index of oscillator j in cluster k is $\lambda_k v_{jk}^2$, where v_{jk} is the j th entry of the k th eigenvector \mathbf{v}_k [27]. Each oscillator is attributed to the cluster with which it has the highest participation index.

Although for many situations SCA works as expected [27], we now present an example showing that it can produce incorrect results. The dataset is depicted in Fig. 7. We simulate the phase of the oscillators using Kuramoto's model (1). The intrinsic frequencies were drawn randomly from a Gaussian with average value 0.003 and standard deviation 0.0005 (negative values are discarded and sampled again). The initial phase values are taken from a uniform distribution in $[0, 2\pi]$. Oscillators 1–5 are coupled in an “all-with-all” fashion, as are oscillators 6–8. There is zero coupling between these two subpopulations. The coupling strength is high enough to ensure that in each cluster all the oscillators are phase-locked to their respective clusters. The time series of all oscillators is shown in Fig. 7, along with their pairwise PLFs, and their relative phase values. When applied to this dataset, SCA clusters all eight oscillators into one cluster, despite finding two eigenvalues larger than 1 (6.35 and 1.64). Clearly, in this situation the real-valued PLF matrix \mathbf{R} does not contain enough information to perform the clustering. One simple way to aid in this distinction is to look at the relative phase values of the eight oscillators, which are shown in the right subfigure of Fig. 7. In that figure, the relative phase values are clearly clustered into two groups. This immediately suggests that phase values can be used to improve these results.

2) *Limitations of SCA*: Although SCA was not developed with the Kuramoto model in mind, there is nothing specific of this model preventing the application of SCA to oscillators that

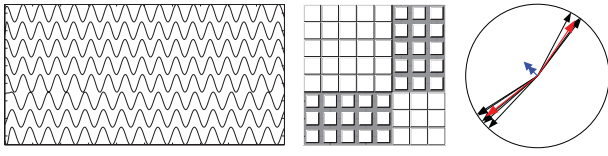


Fig. 7. (Left) Small dataset used to illustrate how pSCA works. Oscillators 1–5 are strongly coupled in an “all-with-all” fashion, as are oscillators 6–8. Note the transient behavior at the beginning of the observation period. (Center) Pairwise PLFs between the oscillators. Note that there is partial synchrony because of the limited time interval. (Right, black thin arrows) Phase values of the eight oscillators at the end of the observation period. (Thick red arrows) Mean field of each cluster. (Thick blue double arrow) Mean field of the whole population.

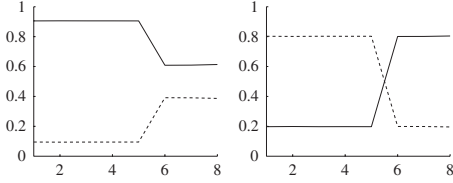


Fig. 8. (Left) Participation indices for each oscillator for SCA. (Right) Participation indices for each oscillator for pSCA. The horizontal axis is the oscillator number. Each oscillator is assigned to the cluster with which it has the highest participation index. SCA fails to find the two clusters.

follow that model. We now demonstrate that high PLF values, which are independent of the relative phase difference between two oscillators, are a necessary but not sufficient condition for two oscillators to be coupled under the Kuramoto model.

If the intra-cluster coupling within cluster c_j is strong enough and the inter-cluster couplings are weak enough to be disregarded, and if all intra-cluster interactions have the same coupling strength κ , a mean field can be defined for that cluster as

$$\varrho_{c_j} e^{i\Phi_{c_j}} \equiv \frac{1}{N_j} \sum_{k \in c_j} e^{i\phi_k}$$

where N_j is the number of oscillators in cluster c_j . The mean fields for the above example are shown as thick red arrows in Fig. 7. The mean field of the whole population of eight oscillators is represented by the thick blue double arrow. In this case, the original Kuramoto model (1) can be written as

$$\dot{\phi}_i(t) = \omega_i + 2N_j \kappa \varrho_{c_j} \sin[\Phi_{c_j}(t) - \phi_i(t)] \quad (9)$$

for oscillators $i \in c_j$ [42]. Such formulation allows for an interpretation of this case as an interaction between the oscillator and the cluster to which it belongs, instead of the pairwise interactions. This idea was introduced in [27].

In the equilibrium state, we have

$$\sin(\phi_i - \Phi_{c_j}) = \left(\frac{\omega_i}{2N_j \kappa \varrho_{c_j}} \right).$$

Equation (9) thus has two equilibrium solutions, one which has $|\phi_i - \Phi_{c_j}| < \pi/2$ and one with $|\phi_i - \Phi_{c_j}| > \pi/2$. The latter is an unstable equilibrium point, because, if $|\phi_i - \Phi_{c_j}|$ increases, then $|\sin(\phi_i - \Phi_{c_j})|$ will decrease. Therefore, the interaction term in (9) will be smaller and the oscillator’s phase will move further away from the mean field.

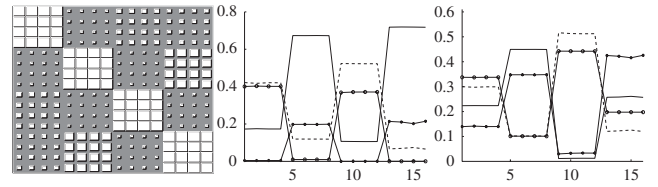


Fig. 9. (Left) Pairwise PLFs between 16 oscillators in a second pSCA dataset. (Center) Results of the original SCA algorithm. SCA detects only two clusters, with eight oscillators each. (Right) Results of the pSCA algorithm. All the clusters are correctly detected, and with a larger discriminating power.

A similar reasoning shows that the first solution, with $|\phi_i - \Phi_{c_j}| < \pi/2$, is stable. We can write, for this solution

$$\phi_i - \Phi_{c_j} = \arcsin\left(\frac{\omega_i}{2N_j \kappa \varrho_{c_j}}\right).$$

This result shows that oscillators for which $|\phi_i - \Phi_{c_j}| > \pi/2$ cannot physically belong to cluster c_j under the Kuramoto model. Graphically, all the oscillators should be contained in a half circle around the mean field direction. Therefore, a common frequency is a necessary but not sufficient condition for two oscillators to be considered phase-synchronous.

To surpass this limitation, we have developed a more complete approach. Instead of keeping only the real-valued PLF matrix \mathbf{R} , we use the complex PLF matrix \mathbf{Q} to cluster the oscillators. We now describe the procedure in detail.

3) *pSCA*: We present a small theoretical introduction to pSCA, illustrated by two small examples. After this, we present the general algorithm and show results in simulated datasets.

Example 1: Find the maximum of the average energy $\langle |y|^2 \rangle = \langle yy^* \rangle$ of a signal given by the superposition of two sources with the same frequency: $y = ae^{i\omega t} + \beta e^{i(\omega t + \phi)}$, where $a, \beta \in \mathbb{C}$ are the problem variables. To prevent unboundedness, we add the constraint $|a|^2 + |\beta|^2 = 1$.

The Lagrangian for this problem is $L = \langle |y|^2 \rangle - \lambda(ae^{i\omega t} + \beta e^{i(\omega t + \phi)} - 1)$; the conditions $\partial L / \partial a^* = \partial L / \partial \beta^* = \partial L / \partial \lambda^* = 0$ yield⁷ $(\lambda - 1)a = \beta e^{i\phi}$, $(\lambda - 1)\beta = a e^{-i\phi}$, $|a|^2 + |\beta|^2 = 1$.

The solutions of these equations are of the form $|a| = |\beta| = 1/\sqrt{2}$ and $\beta = a e^{-i\phi}$, such that $y = a e^{i\omega t} + \beta e^{i(\omega t + \phi)}$ becomes the sum of two signals exactly in phase. Intuitively, we can say that this maximization problem finds the best coefficients to compensate the dephasing of the sources.

Example 2: We now consider three oscillators: the first two with a frequency ω and the third with a frequency $\omega + \Delta\omega$, where $\Delta\omega > 0$, all with the same amplitude and initial phase. Suppose the signals are observed from $t = 0$ to $t = T$.

We again want to maximize the average energy of $y = a e^{i\omega t} + \beta e^{i\omega t} + \gamma e^{i(\omega + \Delta\omega)t}$ subject to the constraint $|a|^2 + |\beta|^2 + |\gamma|^2 = 1$. We can obtain the following two relations

$$|\alpha| = |\beta|, \quad |\gamma| \leq \frac{4}{\Delta\omega T} |\alpha|$$

(see Appendix B). This shows that, for nonzero a and β , the magnitude of γ will be close to zero if either the observation

⁷We use the complex gradient as defined in [45].

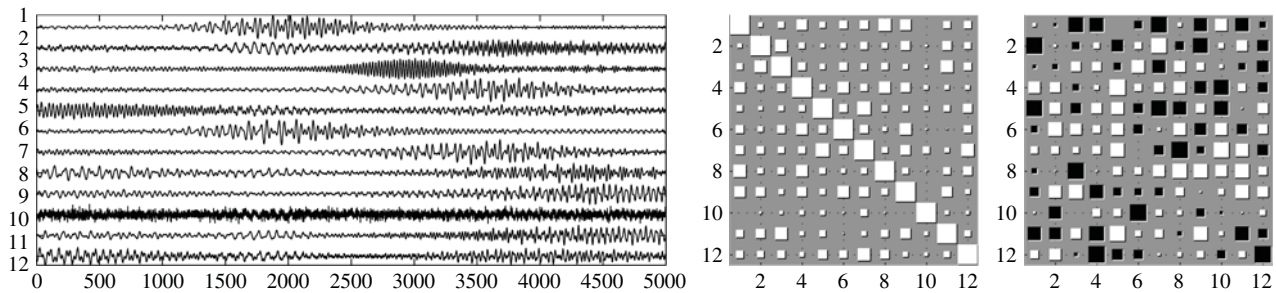


Fig. 10. Result of FastICA applied to the dataset of Fig. 1. (Left) Sources estimated by FastICA. (Center) PLFs between the estimated sources. (Right) Estimated unmixing matrix. It is clear that FastICA is not adequate to the problem discussed in this paper.

time T or the frequency detuning $\Delta\omega$ (or both) is large. In particular, it shows that, if we have a large enough observation time, this maximization problem “detects” that there are two phase-locked sources and a third nonlocked source.

Example 1 shows that this maximization procedure can compensate relative phase offsets between oscillators, while example 2 shows that it can select synchronized oscillators from a population that also includes nonsynchronized oscillators. We now formulate a general algorithm based on the intuitive ideas provided by these examples. This algorithm has two steps: an eigendecomposition and an additional rotation.

We assume that we have N measurements. We extract their phases $\phi_i(t)$ using the Hilbert transform and construct the normalized signals $a_i(t) = e^{i\phi_i(t)}$, which we use to form the vector $\mathbf{a}(t)$. Then we search for a complex superposition of the form $y(t) = \mathbf{c}^H \mathbf{a}(t)$ that maximizes the average energy $E = \langle |y|^2 \rangle$. To prevent unboundedness, we constrain the projection vector \mathbf{c} to have unit norm. This maximization problem is the general case of Examples 1 and 2.

We can rewrite the energy as $E = \langle |y|^2 \rangle = \langle yy^* \rangle = \mathbf{c}^H \langle \mathbf{a} \mathbf{a}^H \rangle \mathbf{c} = \mathbf{c}^H \mathbf{Q} \mathbf{c}$, where $\mathbf{Q} \equiv \langle \mathbf{a} \mathbf{a}^H \rangle$. Note that \mathbf{Q} is exactly the synchronization matrix defined in (2). The Lagrangian for the maximization of E is thus $L = \mathbf{c}^H \mathbf{Q} \mathbf{c} - \lambda (\mathbf{c}^H \mathbf{c} - 1)$, and the solutions of the problem must obey $\nabla_{\mathbf{c}} L = 0$. If \mathbf{v} is a solution of this problem, it must obey the eigenvalue condition $\mathbf{Q} \mathbf{v} = \lambda \mathbf{v}$.

This reasoning presents additional motivation to the use of the eigendecomposition originally proposed in [27]. We now assume, without loss of generality, that the eigenvalues of \mathbf{Q} are ordered in descending order: $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$. Let \mathbf{v}_k denote the eigenvector associated with λ_k , and let v_{jk} denote the j th coordinate of \mathbf{v}_k . The largest eigenvalue of \mathbf{Q} is the maximum of L , and its corresponding eigenvector \mathbf{v}_1 is thus the solution to the maximization problem above. Since this is a global maximum, the Hessian of L , $\nabla \nabla L = \mathbf{Q} - \lambda \mathbf{I}$, is negative (semi)definite for $\lambda = \lambda_1$.

The value of λ_i is for noiseless data and infinite observation time, the number of oscillators in the i th largest population [this is easy to show from (2)]. Thus, $\text{tr}(\mathbf{Q}) = N$, the total number of oscillators. As in [27], we use the number of eigenvalues of \mathbf{Q} larger than 1 as the number of clusters.

Note that the main difference between the eigendecomposition in pSCA and the one done in SCA is that in pSCA we retain the value of the phase lag between any oscillator i and

cluster j , which is the angle of the i th component of \mathbf{v}_j . We now describe how these phase values can be used to overcome the limitations of SCA described in Section III-C.2.

Assume that the population has N oscillators in d clusters. As mentioned before, we choose d as the number of eigenvalues of \mathbf{Q} larger than 1. Our goal is to find a new set of complex-valued coefficients u_{ij} , which contain the participation of oscillator i in cluster j , such that the phase compactness values of all clusters j , defined as $|\sum_{i=1}^N u_{ij}| = |\sum_{i=1}^N |u_{ij}| e^{i\phi_i}|$, are as high as possible. The phase compactness measures the compactness of the phase angles ϕ_i weighted by their participation in cluster j . Note that the sum is over all oscillators, regardless of which cluster they belong to (if oscillator i does not belong to cluster j , then $u_{ij} = 0$ and it will not influence the sums). We define the vectors \mathbf{u}_j , $j = 1, \dots, d$, as column vectors containing the u_{ij} coefficients, similar to the definition of \mathbf{v}_j . Our objective will be to maximize the sum of the phase compactnesses of all the d clusters

$$J = \sum_{j=1}^d \left| \sum_{i=1}^N u_{ij} \right|.$$

This criterion can be expressed as “assign these N oscillators to d clusters in such a way that the phase compactness of the clusters is as high as possible.” Regarding the right subfigure of Fig. 7, the goal is to assign the eight oscillators into two clusters such that the red arrows are as big as possible.

To obtain the u_{ij} coefficients, we perform a linear transformation of the v_{ij} coefficients obtained from the eigendecomposition, as motivated by Examples 1 and 2. Thus, the new coefficients are given by $|u_{ij}| \equiv |\sum_k v_{ik} w_{kj}|$, where w_{kj} are real coefficients to be optimized through the maximization of the objective function J . If we concatenate the w_{kj} in a matrix $\mathbf{W} \in \mathbb{R}^{d \times d}$, the v_{ij} in $\mathbf{V} \in \mathbb{C}^{N \times d}$, and the u_{ij} in $\mathbf{U} \in \mathbb{C}^{N \times d}$, this linear transformation can be written as $\mathbf{U} = \mathbf{V} \mathbf{W}$.

There are some restrictions on \mathbf{W} . Since we want to measure the compactness of the phase values in the clusters, the coefficients in \mathbf{W} should not compensate for the different phase values (as in Example 1) and should therefore be real. Also, we force \mathbf{W} to be orthonormal so that the total number of oscillators remains equal to N . Mathematically, if $\mathbf{W}^T \mathbf{W} = \mathbf{I}$, then $\text{tr}(\mathbf{U}^H \mathbf{Q} \mathbf{U}) = \text{tr}(\mathbf{W}^T \mathbf{V}^H \mathbf{D} \mathbf{V} \mathbf{W}) = \text{tr}(\mathbf{W}^T \mathbf{Q} \mathbf{W}) = N$, where \mathbf{D} is a diagonal $d \times d$ matrix with only the d largest eigenvalues of \mathbf{Q} in descending order.

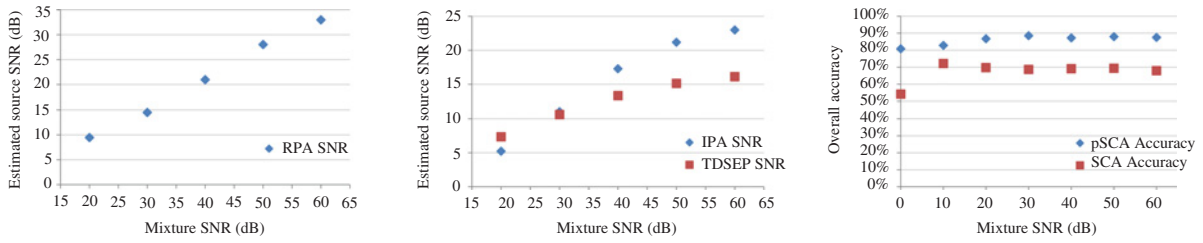


Fig. 11. Results with noise for RPA (left), IPA and TDSEP (middle), and pSCA and SCA (right).

The gradient of the objective J with respect to w_{ij} is [42]

$$\mathbf{G}_{ij} = \frac{\partial J}{\partial w_{ij}} = \frac{1}{|\bar{u}_j|} \left[\text{Re}(\bar{v}_i) \times \text{Re}(\bar{u}_j) + \text{Im}(\bar{v}_i) \times \text{Im}(\bar{u}_j) \right] \quad (10)$$

where $\text{Re}(\cdot)$ and $\text{Im}(\cdot)$ are real and imaginary parts, and $\bar{u}_j \equiv \sum_i u_{ij}$ and $\bar{v}_i \equiv \sum_k v_{ki}$ are column sums of \mathbf{U} and \mathbf{V} .

Since \mathbf{W} is constrained to be orthonormal, a simple gradient method is not appropriate to maximize J . The manifold of orthonormal \mathbf{W} matrices is called the ‘‘Stiefel manifold’’ \mathcal{S} [46]. To maximize J subject to the constraint $\mathbf{W} \in \mathcal{S}$, one possibility is to project the gradient in (10) onto the tangent space of \mathcal{S} at \mathbf{W} by $\mathbf{G}_{rang} = \mathbf{W} \text{skew}(\mathbf{W}^T \mathbf{G}) + (\mathbf{I} - \mathbf{W} \mathbf{W}^T) \mathbf{G}$ [46], where \mathbf{G} is a matrix whose (i, j) element is given by (10), and $\text{skew}(\cdot)$ is the antisymmetric part of a matrix. Another possibility is to take directly into account the geometry of the Stiefel manifold when computing the gradient. This yields $\mathbf{G}_{geom} = \mathbf{G} - \mathbf{W} \mathbf{G}^T \mathbf{W}$ [46]. We used \mathbf{G}_{rang} in all the results presented here, due to its simpler interpretation.

The final result of pSCA is a clustering of the oscillators into clusters with tightly packed phase angles. We define the unnormalized participation index of oscillator i in cluster c_j as $|u_{ij}|$. Each oscillator is assigned to the cluster with which it has the largest unnormalized participation index

$$i \in c_j \Leftrightarrow j = \arg \max_k \{|u_{ik}|\}. \quad (11)$$

Because of the rotation mentioned above, the eigenvalues of \mathbf{Q} are not necessarily the correct cluster sizes after the optimization of J . Therefore, in pSCA, the unnormalized participation indices for each oscillator do not necessarily sum to 1 as in SCA. Thus we force this normalization by scaling the participation indices. For each oscillator i , we define the normalized participation indices as

$$p_{ij} = \frac{u_{ij}}{\sum_k |u_{ik}|}. \quad (12)$$

Note that, for each oscillator, this normalization is a scaling of its participation in all clusters such that $\sum_j |p_{ij}| = 1$. This scaling does not alter the assignment rule in (11), i.e., the same rule applied to the normalized participation indices in (12) yields the same clustering.

The complete pSCA procedure is summarized in Table I.

4) *Application to Simulated Data:* We now show the result of some simulated experiments. We begin by considering the small example with eight oscillators already mentioned above. We applied SCA and pSCA to this toy dataset for comparison. The only two eigenvalues larger than 1 are 6.35 and 1.64, indicating the presence of two clusters. Note that the

eigenvalues are equal for the eigendecomposition of the real PLF matrix and the complex PLF matrix. The participation indices are shown in Fig. 8 and show that in this example pSCA is capable of distinguishing the two clusters, while SCA is not. One interesting remark is that, although there are two eigenvalues larger than 1, SCA assigns all eight oscillators to one cluster. This example shows that pSCA yields better results than SCA for this small dataset.

We also considered a larger example, with 16 oscillators organized into four clusters. The oscillators are simulated in the same way: we use Euler’s method (1) to simulate the dynamics of the population, and all oscillators start with a random phase value between 0 and 2π . The only eigenvalues larger than 1 are 6.55, 4.78, 3.07, and 1.59. The pairwise PLFs and participation indices are shown in Fig. 9.

These results show that pSCA can use the additional phase information to improve the results of the original SCA algorithm. Despite finding four eigenvalues larger than 1 (the criterion mentioned in [27] for presence of a cluster), the original SCA fails to find four distinct clusters, instead finding two clusters with double the correct number of oscillators.

IV. DISCUSSION

As mentioned earlier, observed signals are often mixtures of underlying sources. This mixing process has a serious influence on phase-locking relations, and should be inverted as much as possible when one analyzes the synchrony of such signals. In the absence of prior knowledge of the sources, often one uses BSS approaches. One of the most widely used BSS techniques is ICA, which works by finding linear combinations of the measurements that are as independent as possible. ICA uses information from the probability density function of the data, and therefore discards all temporal dynamics information, rendering it useless for synchrony studies (see Fig. 10).

Other BSS approaches such as TDSEP are based on the temporal structure of the sources, and therefore take into account their temporal dynamics. The first row of Fig. 4 already shows that TDSEP can separate subspaces from each other but fails to separate sources within each subspace.

One important aspect to study is the robustness of the proposed algorithms to small levels of noise, especially since the influence of, e.g., Gaussian noise on the phase values is quite complex.⁸ We present in Fig. 11 results that show

⁸This noise would be projected into a phase component and an amplitude component, each of which has different impacts, see [47].

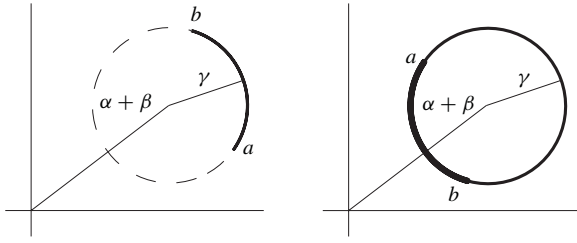


Fig. 12. Integration path in (15) $a = -T/2$ and $b = +T/2$ represent the end points of the integration. (Left) Case with $\Delta\omega T < \pi/2$. (Right) Case with $\pi < \Delta\omega T < 3\pi/2$. Note that the arc between a and b is counted twice in the integration for this second case.

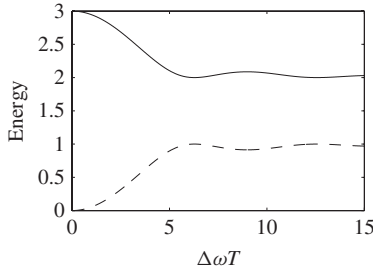


Fig. 13. Two solutions for E given by (21) plotted against $\Delta\omega T$. (Solid line) Maximum energy, obtained by choosing the plus sign in (21). (Dashed line) Minimum energy, obtained by choosing the minus sign in (21).

that all the algorithms are robust to small levels of noise. We define the SNR of the i th mixture signal $x_i = (\mathbf{A}\mathbf{s})_i + n_i$ as $10 \log_{10}(E[(\mathbf{A}\mathbf{s})_i^2]/E[n_i^2])$. To produce the noisy mixture, we first generate the noiseless mixture and then add noise to each mixture signal to ensure that all mixture signals have the same SNR. We generated 300 datasets for RPA and IPA similar to the one shown in Fig. 1, and 300 datasets for pSCA similar to the one in Fig. 9, but with amplitude bursts of random position and width as in RPA and IPA.⁹

After matching the estimated sources with the original sources, the SNR of the estimated source y_i is defined as $10 \log_{10}(E[(\alpha y_i)^2]/E[(\alpha y_i \pm s_i^2)])$, where the real scalar α and the \pm sign are chosen to maximize the SNR value. This ensures that these SNR values are independent of permutation, scaling, and sign, as is common in source separation contexts.

Much in line with other methods dealing with analysis of phase-locking relations, one limitation of the methods presented in this paper is the use of the Hilbert transform, which only has a clear interpretation when applied to locally narrowband signals. Note that since a wavelet decomposition can be interpreted as a narrowband filter bank [37], using wavelets does not solve this problem. All the signals generated for our simulated examples are narrowband, therefore the use of the Hilbert transform is justified.

One could consider using a deflationary version of RPA to extract multiple sources with a single run of the algorithm. This cannot be done simply by removing the orthogonal projection of the first estimated source from the data and rerunning the algorithm. Since phase-locked sources may be correlated, removing the orthogonal projection of one

estimated source from the data will also remove parts of other sources. Therefore, a deflationary approach is currently not possible, and if multiple sources are desired RPA must be run multiple times with different initial conditions.

IPA is quite robust to the choice of λ : values within a factor of two from the optimal value yield similar results.

We observed that pSCA does not always find the correct solution, especially in cases with multiple populations with equal sizes, where the numerical computation of the eigenvalues is less accurate. Yet, we found that not only does pSCA consistently yield better clustering results than SCA, it usually has a larger distance between the two largest participation factors (see, oscillators 1–4 in Fig. 9). This suggests that pSCA has a larger “discriminant power” than SCA, which explains pSCA’s higher tolerance to noise.

Another aspect to be analyzed further is the performance of these algorithms in situations of partial synchrony. Preliminary tests suggest that the algorithms behave differently in this situation. RPA performs quite well when the reference is only partially phase-locked to some of the sources, as long as the stopping criteria are well chosen. IPA performs well for phase-locking values above approximately 0.9 or below approximately 0.1. pSCA seems to work well for intermediate values of the PLF, as suggested by the results obtained on the small dataset in Fig. 8. A related aspect to be investigated in the future is how to tackle a time-dependent PLF, which is relevant in some real situations. In this case, using time windows allows the algorithms to look only at parts of the data, but further work would be needed to implement a change detection technique to track these time variations.

Apart from the results with simulated data, we have previously applied RPA to real data from cortico-muscle control using EMG as a reference for MEG recordings [29]. Those results agreed with established knowledge on relations between cortical and muscle activity, suggesting that the use of the methods we propose here in real applications is possible.

V. CONCLUSION

Synchrony phenomena are present in many scientific domains, including astrophysics, biology, chemistry, and neuroscience. Even though many different systems exhibit synchrony, a common framework, such as Kuramoto’s model, can be used for all of them. Whichever the underlying physical system may be, in any synchrony phenomenon it is important to have direct access to the underlying oscillators. If that is not possible or feasible, as in brain electrophysiological recordings, it is crucial that one looks at the sources and not at the indirect measurements, as these will have spurious synchrony. Unfortunately, few researchers in neuroscience perform synchrony (or coherence) analysis on the underlying sources instead of on the measurements.

There is a need for source separation methods that tackle this type of problems. We showed empirically that current state-of-the-art techniques such as ICA and TD methods are not adequate. Since these methods make assumptions that are not valid when the underlying sources are phase-locked, synchrony between sources extracted with these methods should be regarded as an epiphenomenon.

⁹The position and width of the amplitude bursts, as well as the mixing matrices (for RPA and IPA), are chosen randomly for each of the 300 runs.

We have presented three methods deeply rooted in the concept of phase synchrony, which can be used in different circumstances. The methods are inspired by concepts from BSS and attempt to circumvent the disruption of synchrony values in the event of a linear mixing. RPA is used to extract an underlying source phase-locked to a reference signal. IPA performs BSS of phase-locked sources. pSCA clusters a population of oscillators into synchronous subpopulations. The three methods work for simulated data and are robust to small levels of noise. We believe these algorithms provide the community with methods applicable in situations which were not addressable previously.

APPENDIX A

UNICITY OF SOLUTION IN IPA

A. Assumptions and Definitions

Consider a set of N sources, which are defined as complex time series: $s_k(t) \equiv A_k(t)e^{i(\omega(t)+\phi_k)}$, with k from 1 to N . Also, consider N measured signals, obtained by linear instantaneous mixture of the sources

$$y_j(t) \equiv b_{j1}s_1(t) + b_{j2}s_2(t) + \dots + b_{jN}s_N(t). \quad (13)$$

Our goal is to prove that to put y_1, y_2, \dots in the form¹⁰

$$y_j(t) = C_j(t)e^{i(\omega(t)+\alpha_j)} \quad (14)$$

with $C_i(t)$ nonnegative functions and α_i scalar values, one must always have $y_i = Ks_j$ for some i and j . In other words, for each $i = 1, 2, \dots$, exactly one of the coefficients b_{i1}, b_{i2}, \dots must be nonzero, with all others being zero. One also demands, as is natural in source separation, that $y_i \neq y_j$. The matrix \mathbf{B} , whose (i, j) entry is b_{ij} , must therefore be a permutation of a diagonal matrix with nonzero diagonal.

The proposition we are trying to prove is not generally true. However, it is true under mild assumptions, which we list here.

- 1) For all i , $s_i(t), y_i(t) \neq 0$ for some t .
- 2) For all $i \neq j$, $y_i(t) \neq y_j(t)$.
- 3) ϕ_1, ϕ_2, \dots are all distinct modulo π .
- 4) The amplitudes $A_i(t)$ are linearly independent. This means that if one has two linear combinations that obey $\sum_i c_i A_i(t) = \sum_i d_i A_i(t)$, then necessarily one has $c_i = d_i$ for all i .

Assumption 1 immediately rules out the possibility that for some i , all b_{i1}, b_{i2}, \dots are all zero.

B. Proof

We start by proving that y_1 is equal to one of the sources up to scaling and sign. If we define a matrix \mathbf{A} where the (i, t) -th entry is $A_i(t)$, $\mathbf{C} \equiv [C_1(1) \dots C_1(T)]$, $\Phi \equiv [\phi_1 \dots \phi_N]^T$, $\Phi - \alpha_1 \equiv [(\phi_1 - \alpha_1) \dots (\phi_N - \alpha_1)]^T$, and $\mathbf{Z}(\Phi) = [b_{11}e^{i\phi_1} \dots b_{1N}e^{i\phi_N}]$, then by equating (13) and (14), and by eliminating the common term $e^{i\omega(t)}$, the equation for y_1 yields $\mathbf{C}^T e^{i\alpha_1} = \mathbf{A}^T \mathbf{Z}(\Phi) \Leftrightarrow \mathbf{C}^T = \mathbf{A}^T \mathbf{Z}(\Phi - \alpha_1) \Leftrightarrow \mathbf{C}^T = \mathbf{A}^T \mathbf{Z}(\alpha_1 - \Phi)$, where the last equation was obtained by taking

¹⁰The y_j 's must be of this form to have a PLF of 1 with the sources: the PLF between two signals is 1 if and only if their phase difference is constant.

the complex conjugate of both sides (note that \mathbf{C} and \mathbf{A} are real). Because of Assumption 4, \mathbf{A} is a full rank matrix, so we can conclude that $\mathbf{Z}(\Phi - \alpha_1) = \mathbf{Z}(\alpha_1 - \Phi)$, which is equivalent to either $\sin(\phi_j - \alpha_1) = 0$ or $b_{1j} = 0$, for all j . Because of Assumption 1, at least one of the b_{1j} must be nonzero. On the other hand, because of Assumption 3, $\sin(\phi_j - \alpha_1) = 0$ is verified for at most one j . Combining these two statements, one can conclude that exactly one of the b_{1j} coefficients is non-zero.

The demonstration for y_2, \dots, y_N is similar: if we force $y_i \neq y_j$ if $i \neq j$ (Assumption 2), this immediately forces that if $b_{1k} \neq 0$, then $b_{mk} = 0$ for all $m \neq 1$, etc.

APPENDIX B

INEQUALITY IN EXAMPLE 2 OF PSCA

We can begin by choosing the observation period as $[-T/2, T/2]$, with no loss of generality. In this case, the average energy of $y(t)$ is given by the integral

$$\begin{aligned} E &= \langle |y(t)|^2 \rangle = \left\langle \left| y(t) \cdot e^{-i\omega t} \right|^2 \right\rangle \\ &= \frac{1}{T} \int_{-T/2}^{T/2} \left(\alpha + \beta + \gamma e^{i\Delta\omega t} \right) \left(\alpha^* + \beta^* + \gamma^* e^{-i\Delta\omega t} \right) dt \\ &= |\alpha + \beta|^2 + |\gamma|^2 + (\Delta\omega T)^{-1} 2 \sin \frac{\Delta\omega T}{2} (\alpha^* + \beta^*) \gamma \\ &\quad + (\Delta\omega T)^{-1} 2 \sin \frac{\Delta\omega T}{2} (\alpha + \beta) \gamma^*. \end{aligned} \quad (15)$$

$$\begin{aligned} &= |\alpha + \beta|^2 + |\gamma|^2 + (\Delta\omega T)^{-1} 2 \sin \frac{\Delta\omega T}{2} (\alpha^* + \beta^*) \gamma \\ &\quad + (\Delta\omega T)^{-1} 2 \sin \frac{\Delta\omega T}{2} (\alpha + \beta) \gamma^*. \end{aligned} \quad (16)$$

This integral is depicted geometrically in Fig. 12.

From Fig. 12, it is clear that the maximum of E will occur when α and β are parallel and γ is either parallel or antiparallel to α and β . With no loss of generality, we can choose α, β , and γ to be real numbers. In that case, we can rewrite the constraint $|\alpha|^2 + |\beta|^2 + |\gamma|^2 = 1$ as $\alpha^2 + \beta^2 + \gamma^2 = 1$, and the energy of y as

$$\begin{aligned} &(\alpha + \beta)^2 + \gamma^2 + (\Delta\omega T)^{-1} 4 \sin \frac{\Delta\omega T}{2} (\alpha + \beta) \gamma \\ &= 1 + 2\alpha\beta + (\Delta\omega T)^{-1} 4 \sin \frac{\Delta\omega T}{2} (\alpha + \beta) \gamma. \end{aligned}$$

We can now write the Lagrangean of this problem as

$$L = 1 + 2\alpha\beta + \frac{4 \sin \frac{\Delta\omega T}{2}}{\Delta\omega T} (\alpha + \beta) \gamma - \lambda (\alpha^2 + \beta^2 + \gamma^2 - 1)$$

and force the derivatives with respect to α, β, γ , and λ to be zero

$$\begin{cases} \frac{\partial L}{\partial \alpha} = 2\beta + \frac{4 \sin \frac{\Delta\omega T}{2}}{\Delta\omega T} \gamma - 2\lambda\alpha = 0 \\ \frac{\partial L}{\partial \beta} = 2\alpha + \frac{4 \sin \frac{\Delta\omega T}{2}}{\Delta\omega T} \gamma - 2\lambda\beta = 0 \\ \frac{\partial L}{\partial \gamma} = \frac{4 \sin \frac{\Delta\omega T}{2}}{\Delta\omega T} (\alpha + \beta) - 2\lambda\gamma = 0 \\ \frac{\partial L}{\partial \lambda} = \alpha^2 + \beta^2 + \gamma^2 - 1 = 0. \end{cases} \quad (17)$$

The two first equations in the set 17 yield $\lambda = -1 \vee \alpha = \beta$. Let us start with the choice $\lambda = -1$. Substitution in the first or second equation in (17) shows that $\gamma = 0$ or $\Delta\omega T = 0$. The latter is impossible because we assumed positive detuning and observation time. On the other hand, substituting

(see Fig. 13). Therefore, the plus sign is clearly the maximum of E that we were looking for.

Let us recover (20) with the plus sign. By taking the absolute value on both sides we obtain

$$|\gamma| = \left| \frac{4A}{1 + \sqrt{1 + 8A^2}} \right| |\alpha|. \quad (22)$$

Since $|A| < 1$, the numerator in (22) is not greater than $8/\Delta\omega T$ in absolute value, and the absolute value of the denominator is not smaller than 2. We can conclude that

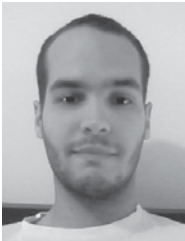
$$|\gamma| \leq \frac{4}{\Delta\omega T} |\alpha|$$

which is the inequality in Example 2.

REFERENCES

- [1] A. Pikovsky, M. Rosenblum, and J. Kurths, *Synchronization: A Universal Concept in Nonlinear Sciences* (Nonlinear Science). Cambridge, U.K.: Cambridge Univ. Press, 2001.
- [2] B. A. Conway, D. M. Halliday, S. F. Farmer, U. Shahani, P. Maas, A. I. Weir, and J. R. Rosenberg, "Synchronization between motor cortex and spinal motoneuronal pool during the performance of a maintained motor task in man," *J. Physiol.*, vol. 489, pp. 917–924, Dec. 1995.
- [3] S. Salenius, K. Portin, M. Kajola, R. Salmelin, and R. Hari, "Cortical control of human motoneuron firing during isometric contraction," *J. Neurophysiol.*, vol. 77, no. 6, pp. 3401–3405, Jun. 1997.
- [4] J. M. Kilner, S. N. Baker, S. Salenius, V. Jousmäki, R. Hari, and R. N. Lemon, "Task-dependent modulation of 15–30 Hz coherence between rectified EMGs from human hand and forearm muscles," *J. Physiol.*, vol. 516, pp. 559–570, Apr. 1999.
- [5] J.-M. Schoffelen, R. Oostenveld, and P. Fries, "Imaging the human motor system's beta-band synchronization during isometric contraction," *NeuroImage*, vol. 41, no. 2, pp. 437–447, Jun. 2008.
- [6] W. Omlor, L. Patino, M.-C. Hepp-Raymond, and R. Kristeva, "Gamma-range corticomuscular coherence during dynamic force output," *NeuroImage*, vol. 34, no. 3, pp. 1191–1198, Feb. 2007.
- [7] A. Andrykiewicz, L. Patino, J. R. Naranjo, M. Witte, M. Hepp-Raymond, and R. Kristeva, "Corticomuscular synchronization with small and large dynamic force output," *BMC Neurosci.*, vol. 8, no. 101, Nov. 2007.
- [8] J. M. Palva, S. Palva, and K. Kaila, "Phase synchrony among neuronal oscillations in the human cortex," *J. Neurosci.*, vol. 25, no. 15, pp. 3962–3972, Apr. 2005.
- [9] P. J. Uhlhaas and W. Singer, "Neural synchrony in brain disorders: Relevance for cognitive dysfunctions and pathophysiology," *Neuron*, vol. 52, no. 1, pp. 155–168, Oct. 2006.
- [10] S. Strogatz, *Nonlinear Dynamics and Chaos*. Boulder, CO: Westview, 2000.
- [11] Y. Kuramoto, *Chemical Oscillations, Waves and Turbulences*. Berlin, Germany: Springer-Verlag, 1984.
- [12] D. Golomb and D. Hansel, "The number of synaptic inputs and the synchrony of large, sparse neuronal networks," *Neural Comput.*, vol. 12, no. 5, pp. 1095–1139, May 2000.
- [13] A. N. Burkitt, "A review of the integrate-and-fire neuron model: I. Homogeneous synaptic input," *Biol. Cybern.*, vol. 95, no. 1, pp. 1–19, 2006.
- [14] A. N. Burkitt, "A review of the integrate-and-fire neuron model: II. Inhomogeneous synaptic input and network properties," *Biol. Cybern.*, vol. 95, no. 2, pp. 97–112, 2006.
- [15] R. Vigiário, J. Särelä, V. Jousmäki, M. Hämäläinen, and E. Oja, "Independent component approach to the analysis of EEG and MEG recordings," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 5, pp. 589–593, May 2000.
- [16] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: Wiley, 2001.
- [17] A. Ziehe and K.-R. Müller, "TDSEP - an efficient algorithm for blind separation using time structure," in *Proc. Int. Conf. Artif. Neural Netw.*, Sep. 1998, pp. 675–680.
- [18] A. Belouchrani, K. A. Meraim, J.-F. Cardoso, and E. Moulines, "Second-order blind separation of temporally correlated sources," in *Proc. Int. Conf. Digital Signal Process.*, 1993, pp. 346–351.
- [19] G. Zhou, Z. Yang, S. Xie, and J.-M. Yang, "Online blind source separation using incremental nonnegative matrix factorization with volume constraint," *IEEE Trans. Neural Netw.*, vol. 22, no. 4, pp. 550–560, Apr. 2011.
- [20] M. Almeida, R. Vigiário, and J. Bioucas-Dias, "Phase locked matrix factorization," in *Proc. EUSIPCO Conf.*, Aug.–Sep. 2011.
- [21] Y. Xiang, S. K. Ng, and V. K. Nguyen, "Blind separation of mutually correlated sources using precoders," *IEEE Trans. Neural Netw.*, vol. 21, no. 1, pp. 82–90, Jan. 2010.
- [22] R. Vigiário and O. Jensen, "Identifying cortical sources of corticomuscle coherence during bimanual muscle contraction by temporal decorrelation," in *Proc. IEEE 7th Int. Symp. Signal Process. Appl.*, vol. 2, Jul. 2003, pp. 109–112.
- [23] C. Meinecke, A. Ziehe, J. Kurths, and K.-R. Müller, "Measuring phase synchronization of superimposed signals," *Phys. Rev. Lett.*, vol. 94, no. 8, pp. 084102-1–084102-4, Mar. 2005.
- [24] G. Nolte, O. Bai, L. Wheaton, Z. Mari, S. Vorbach, and M. Hallett, "Identifying true brain interaction from EEG data using the imaginary part of coherency," *Clin. Neurophysiol.*, vol. 115, no. 10, pp. 2292–2307, Oct. 2004.
- [25] L. A. Wheaton, G. Nolte, S. Bohlhalter, E. Fridman, and M. Hallett, "Synchronization of parietal and premotor areas during preparation and execution of praxis hand movements," *Clin. Neurophysiol.*, vol. 116, no. 6, pp. 1382–1390, 2005.
- [26] C. Allefeld and J. Kurths, "An approach to multivariate phase synchronization analysis and its application to event-related potentials," *Int. J. Bifur. Chaos*, vol. 14, no. 2, pp. 417–426, 2004.
- [27] C. Allefeld, M. Müller, and J. Kurths, "Eigenvalue decomposition as a generalized synchronization cluster analysis," *Int. J. Bifur. Chaos*, vol. 17, no. 10, pp. 3493–3497, Jun. 2007.
- [28] M. G. Rosenblum, A. S. Pikovsky, and J. Kurths, "Phase synchronization of chaotic oscillators," *Phys. Rev. Lett.*, vol. 76, no. 11, pp. 1804–1807, Mar. 1996.
- [29] J.-H. Schleimer and R. Vigiário, "Reference-based extraction of phase synchronous components," in *Proc. Int. Conf. Artif. Neural Netw.*, Athens, Greece, 2006, pp. 230–239.
- [30] J.-H. Schleimer and R. Vigiário, "Order in complex systems of nonlinear oscillators: Phase locked subspaces," in *Proc. Eur. Symp. Neural Netw.*, 2007, pp. 1–6.
- [31] J.-H. Schleimer and R. Vigiário, "Clustering limit cycle oscillators by spectral analysis of the synchronisation matrix with an additional phase sensitive rotation," in *Proc. Int. Conf. Artif. Neural Netw.*, 2007, pp. 1–10.
- [32] M. Almeida and R. Vigiário, "Source separation of phase-locked subspaces," in *Proc. Independ. Comp. Anal. Conf.*, 2009, pp. 1–8.
- [33] M. Almeida, J. Bioucas-Dias, and R. Vigiário, "Independent phase analysis: Separating phase-locked subspaces," in *Proc. Latent Variable Anal. Conf.*, 2010, pp. 189–196.
- [34] A. Grinsted, J. C. Moore, and S. Jevrejeva, "Application of the cross wavelet transform and wavelet coherence to geophysical time series," *Nonlin. Processes Geophys.*, vol. 11, nos. 5–6, pp. 561–566, 2004.
- [35] S. Bloomfield, J. McAteer, B. W. Lites, P. G. Judge, M. Mathioudakis, and F. P. Keenan, "Wavelet phase coherence analysis: Application to a quiet-sun magnetic element," *Astrophys. J.*, vol. 617, no. 1, pp. 623–632, 2004.
- [36] A. Klein, T. Sauer, A. Jedynek, and W. Skrandies, "Conventional and wavelet coherence applied to sensory-evoked electrical brain activity," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 2, pp. 266–272, Feb. 2006.
- [37] C. Torrence and G. P. Compo, "A practical guide to wavelet analysis," *Bull. Amer. Meteorol. Soc.*, vol. 79, no. 1, pp. 61–78, Jan. 1998.
- [38] A. V. Oppenheim, R. W. Schaffer, and J. R. Buck, *Discrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1999.
- [39] M. L. V. Quyen, J. Foucher, J.-P. Lachaux, E. Rodriguez, A. Lutz, J. Martinerie, and F. J. Varela, "Comparison of Hilbert transform and wavelet methods for the analysis of neuronal synchrony," *J. Neurosci. Methods*, vol. 111, no. 2, pp. 83–98, Sep. 2001.
- [40] B. Gold, A. V. Oppenheim, and C. M. Rader, "Theory and implementation of the discrete Hilbert transform," in *Discrete Signal Processing*, L. R. Rabiner and C. M. Rader, Eds. Cambridge, MA: MIT Press, 1973.
- [41] J.-P. Lachaux, E. Rodriguez, J. Martinerie, and F. J. Varela, "Measuring phase synchrony in brain signals," *Human Brain Mapp.*, vol. 8, no. 4, pp. 194–208, 1999.
- [42] M. Almeida, J.-H. Schleimer, J. Bioucas-Dias, and R. Vigiário. (2011). *Source Separation and Clustering of Phase-Locked Subspaces: Derivations and Proofs* [Online]. Available: <http://arxiv.org/abs/1106.2474>
- [43] A. Hyvärinen and P. Hoyer, "Emergence of phase- and shift-invariant features by decomposition of natural images into independent feature subspaces," *Neural Comput.*, vol. 12, no. 7, pp. 1705–1720, Mar. 2000.

- [44] S. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 8, 1996, pp. 757–763.
- [45] D. H. Brandwood, "A complex gradient operator and its application in adaptive array theory," *IEE Proc. F Commun., Radar Signal Process.*, vol. 130, no. 1, pp. 11–16, Feb. 1983.
- [46] A. Edelman, T. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM J. Matrix Anal. Appl.*, vol. 20, no. 2, pp. 303–353, 1998.
- [47] P. Celka, "Statistical analysis of the phase-locking value," *IEEE Signal Process. Lett.*, vol. 14, no. 9, pp. 577–580, Sep. 2007.



Miguel Almeida received the M.Sc. degree in physics and technology engineering from the Instituto Superior Técnico, Technical University of Lisbon, Lisbon, Portugal, in 2006, and Advanced Training in biophysics from the Instituto de Biofísica e Engenharia Biomédica, University of Lisbon, in 2007. Since 2010, he has been pursuing the Doctoral degree in electrical and computer engineering with the Pattern and Image Analysis Group, Instituto de Telecomunicações, Technical University of Lisbon.

He was with the Neuroinformatics Group, Helsinki University of Technology, (currently Aalto University), Aalto, Finland in 2008. His current research interests include statistical machine learning, optimization, and pattern recognition, and application of these topics to brain research using electroencephalography, magnetoencephalography, or functional magnetic resonance imaging data, in particular to study sleep and synchrony between distant brain regions.



Jan-Hendrik Schleimer received the B.Sc. degree in bioinformatics from the University of Tuebingen, Tuebingen, Germany, and the M.Sc. degree in computer and information science from the Helsinki University of Technology, Aalto, Finland, in 2003 and 2007, respectively. Since 2007, he has been carrying out doctoral research at the Institute of Theoretical Biology, Humbolt University, Berlin, Germany, and the Bernstein Center for Computational Neuroscience, Berlin.

His current research interests include addressing the links between neural coding and the dynamics of single neurons using techniques from information and dynamical systems theory.



José Mário Bioucas-Dias (S'87–M'95) received the E.E., M.Sc., Ph.D., and Agregado degrees from the Technical University of Lisbon, Lisbon, Portugal, in 1985, 1991, 1995, and 2007, respectively, all in electrical and computer engineering.

He has been with the Department of Electrical and Computer Engineering, Instituto Superior Técnico, Lisbon, Portugal, since 1995. He is also a Senior Researcher with the Pattern and Image Analysis group, Instituto de Telecomunicações, Lisbon, Portugal, which is a private nonprofit research institution.

He is involved in several national and international research projects and networks. His current research interests include signal and image processing, pattern recognition, optimization, and remote sensing.

Dr. Bioucas-Dias is an Associate Editor of the IEEE Transactions on Image Processing, and he was an Associate Editor of the IEEE Transactions on Circuits and Systems. He is a Guest Editor of a special issue of the IEEE Transactions on Geoscience and Remote Sensing and of a special issue of the IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. He has been a member of the program/technical committees of several international conferences, and General Co-Chair of the 3rd IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing-Whispers in 2011.



Ricardo Vigário received the Licenciatura and Mestrado degrees in applied physics and biomedical engineering from the University of Lisbon, Lisbon, Portugal, in 1992 and 1994, respectively, and the Dr.Sc. degree from the Helsinki University of Technology (currently Aalto University, since 2010), Aalto, Finland, in 1999. In 2004, he became a Docent of statistical machine learning in biomedical systems in that university.

He was a Visiting Lecturer at the Technical University of Graz, Graz, Austria, in 2000 and 2001, respectively and at Zaragoza University, Zaragoza, Spain, in 2003. He has also been a Visiting Associate Professor at the Institut National Polytechnique de Grenoble, Grenoble, France, since 2007. He is currently a Senior Researcher with the Department of Information and Computer Science, Aalto University School of Science, where he leads a research team in Neuroinformatics. His current research interests include adaptive informatics and statistical machine learning, as well as their application to neuroinformatics, functional human brain mapping, and biomedical signal processing.