Reference-Based Extraction
of Phase Synchronous Components

Jan-Hendrik Schleimer and Ricardo Vigário

Adaptive Informatics Research Centre
Helsinki University of Technology
P.O. Box 5400, FIN-02015 Espoo, Finland
`schleime@cis.hut.fi`, `rvigario@cis.hut.fi`

Abstract. Phase synchronisation is a phenomenon observed in measurements of dynamic systems, composed of several interacting oscillators. It can be quantified by the phase locking factor (PLF), which requires knowledge of the instantaneous phase of an observed signal. Linear sources separation methods treat scenarios in which measurements do not represent direct observations of the dynamics, but rather superpositions of underlying latent processes. Such a mixing process can cause spuriously high PLFs between the measurements, and camouflage the phase locking to a provided reference signal. The PLF between a linear projection of the data and a reference can be maximised as an optimisation criterion revealing the most synchronous source component present in the data, with its corresponding amplitude. This is possible despite the amplitude distributions being Gaussian, or the signals being statistically dependent, common assumptions in blind sources separation techniques without *a-priori* knowledge, *e.g.* in form of a reference signal.

1 Introduction

Interest in phase synchronisation phenomena has a long history, when studying the interaction of complex, natural or artificial, dynamic systems. A detailed documentation of the topic is given in Ref. [1]. Although not completely adopted, synchronisation was attributed a role in the interplay between different parts of the central nervous system (CNS) as well as across central and peripheral nervous systems. In that formulation, the elementary units are *self-sustained oscillators* $x_i(t)$, exhibiting stable limit cycles. If the coupling between the oscillators is of weak nature, any distortion that a mutual forcing would cause on the amplitudes, will be immediately compensated. Then the interactions of m self-sustained oscillators can be described with the Kuramoto model (*cf.* Ref. [2, 3] for a review), solely on the phases dynamics

$$\dot{\phi}_i(t) = \omega_i(t) + \frac{1}{m}\sum_{j=1}^{m} \kappa_{ij} \sin\big(\phi_j(t) - \phi_i(t)\big), \tag{1}$$

where $\phi_i(t)$ and $\omega_i(t)$ denote the oscillators' instantaneous phases and frequencies; and κ can either be the scalar-valued global coupling strength or a matrix in

which $[\boldsymbol{\kappa}]_{ij}$ describes the coupling between oscillators i and j. Postulated that the system in Eq. (1) is an adequate description of the dynamics of a phenomenon, it becomes meaningful to focus investigations of their interaction principles to phase synchronisation.

In the CNS, the basic unit — the *neuronal oscillator* — can be a single neuron, with an oscillating membrane potential, or a whole population of already synchronous neurons, that synchronises to another population at a different site of the brain. Examples of models for neuronal dynamics based on self-sustained oscillators can be found in Refs. [4, 5].

The phase synchronisation is commonly quantified by the phase locking factor (PLF, for definition see Sec. 2). In many applications, direct measurements of the individual sources $\boldsymbol{x}(t)$ are not available, but instead global multi-sensor measurements $\boldsymbol{y}(t)$ of the whole system, which represent mixtures of $\boldsymbol{x}(t)$. See [6] for a general treatment of such problems. Often, this mixing process can be described by a linear transformation $\boldsymbol{y}(t) = \boldsymbol{A}\boldsymbol{x}(t)$. If the PLF is evaluated w.r.t. $\boldsymbol{y}(t)$ two problems arise: (i) calculating the PLF between observations $y_i(t)$ will, due to the presence of individual oscillators in several sensors, lead to an erroneous detection of interactions between them; (ii) since each sensor measurement contains more than one of the oscillators the PLF of the $y_i(t)$ to a reference will be reduced, obscuring the true interactions.

Here, an algorithm for the extraction of sources synchronised with a given reference is introduced (Sec. 2). The PLF is only evaluated in the source space, not for the observations, circumventing spurious synchronisations by cross-talk, and allowing the recovery of the true sources and their coupling strengths. The search for coupled oscillator networks is facilitated by the use of a reference signal, embodying existing information on the targeted networks. This can be a continuous stimulus to the complex system, an already extracted component of the system, or an external, more accessible part of the system. The algorithm is presented in a general gradient-based formulation, and can be applied to a variety of problems. Experimental results in a controlled simulated data set (Secs. 2.1,2.2), as well as in a preliminary investigation into cortico-muscular control are presented (Sec. 3).

2   Extraction of One Source Synchronised to a Reference

As stated above, assume that the observations result from a linear superposition of generative sources, $\boldsymbol{y}(t) = \boldsymbol{A}\boldsymbol{x}(t)$, with the restriction that $\boldsymbol{A}$ is invertible. The time index shall be discrete in the following and reside in a fixed interval $1 \leqslant t \leqslant T \in \mathbb{N}$. Further postulate that, for a given reference signal $u(t)$, a phase locking is taking place between the reference and at least one of the source signals, $x_i(t)$. Denote the analytic signals[1] as $\hat{y}_i(t) = Y_i(t)\,e^{\mathrm{i}\varphi_i(t)} = y_i(t) + \mathrm{i}\mathcal{H}[y_i](t)$, $\hat{u}(t) = U(t)\,e^{\mathrm{i}\psi(t)}$ and $\hat{s}(t) = S(t)\,e^{\mathrm{i}\phi(t)}$. $s(t)$ is the extracted source, an approximation of $x_i(t)$. For the phase difference between reference and the extracted source

---

[1] Here, $\mathcal{H}[x](t)$ is the Hilbert transform of a signal, $x(t)$.

Algo. 1. Extraction of a phase locked component.

1: input: $\boldsymbol{y}(t)$, $\hat{u}(t)$, η, n_{itr};
2: init: $\boldsymbol{w} \sim \mathcal{N}(0, \boldsymbol{I})$; $k = 1$;
3: **repeat**
4:   $s(t) \leftarrow \boldsymbol{w}^\mathsf{T} \boldsymbol{y}(t)$;
5:   $\hat{s}(t) \leftarrow s(t) + \mathrm{i}\mathcal{H}[s](t)$;
6:   $P \leftarrow \frac{1}{T} \sum_t \hat{s}(t)\hat{u}(t)/|\hat{s}(t)\hat{u}(t)|$;
7:   $\Psi \leftarrow \text{angle}(P)$; $\varrho \leftarrow |P|$;
8:   $\boldsymbol{\Delta} \leftarrow$ Eq. (9);
9:   $\boldsymbol{w} \leftarrow \boldsymbol{w} + \eta \boldsymbol{\Delta}$;
10:   $\boldsymbol{w} \leftarrow \boldsymbol{w}/\|\boldsymbol{w}\|$;
11:   $k \leftarrow k + 1$;
12: **until** $(\varrho > 1 - \delta) \wedge (\|\boldsymbol{\Delta}\| < \epsilon) \wedge (k > n_{\text{itr}})$

signal $\Delta\phi(t) = \phi(t) - \psi(t)$, define a function

$$\varrho\, e^{\mathrm{i}\Psi} = \frac{1}{T} \sum_{t=1}^{T} e^{\mathrm{i}\Delta\phi(t)} = \frac{1}{T} \sum_{t=1}^{T} \frac{\hat{s}(t)\hat{u}^*(t)}{|\hat{s}(t)\hat{u}(t)|}, \tag{2}$$

so that the amplitude ϱ measures the phase locking between the reference signal and the projection $s(t) = \boldsymbol{w}^\mathsf{T}\boldsymbol{y}(t)$. It is called the phase locking factor PLF and, since depending on the source signal, it is also a function of $\boldsymbol{w}$ and the data. Because the complex vector $\hat{s}(t)$ and $\hat{u}(t)$ in Eq. (2) are scaled to one, the PLF $\varrho \in \mathbb{R}$ lies in the interval $0 \leqslant \varrho \leqslant 1$. As the maximisation criterion for our algorithm we can use its square ϱ^2. The gradient w.r.t. $\boldsymbol{w}$ is given by the following expression

$$\nabla\varrho^2 = \frac{2\varrho}{T} \sum_{t=1}^{T} \frac{\sin(\Psi - \Delta\phi(t))}{S^2(t)} \boldsymbol{\Gamma}(t)\boldsymbol{w}, \tag{3}$$

with the amplitude ϱ and the mean phase Ψ as defined by Eq. (2), and a matrix $[\boldsymbol{\Gamma}(t)]_{ij} = Y_i(t)Y_j(t)\sin(\varphi_i(t) - \varphi_j(t))$, fully defined by the observations. The details of the derivation are shown in appendix A.

Eq. (9) can be used in a batch gradient ascent iteration to maximise ϱ. The learning rule reads

$$\Delta\boldsymbol{w} = \eta\, \frac{2\varrho}{T} \sum_{t=1}^{T} \frac{\sin(\Psi - \Delta\phi(t))}{S^2(t)} \boldsymbol{\Gamma}(t)\, \boldsymbol{w}. \tag{4}$$

For smoother convergence, the learning factor η can be chosen to decay in a variety of annealing strategies. Since $\varrho \leqslant 1$, a sufficient stopping criterion for the iteration, iff a phase locked component is present in the data, is $\varrho > 1 - \delta$ for $0 < \delta \ll 1$. A maximum number of iterations has to be specified, in case the reference signal has no phase locked component in the data, because then the objective function will not reach a high value. The batch algorithm is summarised in Algo. 1.
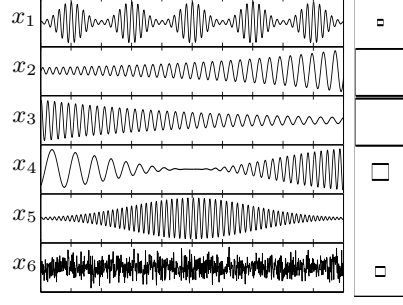
Fig. 1. Kurtosis values of the sources are $\text{kurt}(x_1) = -0.02$, $\text{kurt}(x_2) = 0.007$, $\text{kurt}(x_3) = 0.02$, $\text{kurt}(x_4) = 0.005$, $\text{kurt}(x_5) = 0.006$ and $\text{kurt}(x_6) = -0.18$.
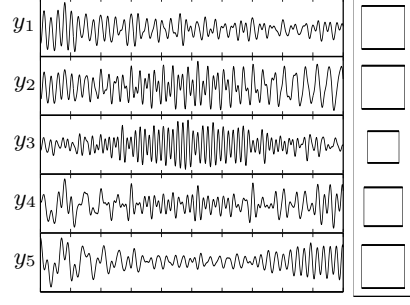
Fig. 2. The Linear mixtures $\boldsymbol{y}(t)$, with their PLF to the reference displayed as the area of opposite squares.

For larger data sets, with potential nonstationary phase locking behaviour, as can be produced by Eq. (1), the learning rule can be formulated in an online way comparable to stochastic gradient algorithms. Then, $\varrho_t\,e^{\mathrm{i}\Psi_t}$ will be evaluated in a time window and the update rule is

$$\Delta\boldsymbol{w}_t \propto \beta_t\,\frac{2\varrho_t\sin(\Psi_t - \Delta\phi(t))}{S^2(t)}\boldsymbol{\Gamma}(t)\,\boldsymbol{w}_t. \tag{5}$$

The evolution of the synchrony, or loss of it, for a component can be assessed by monitoring the quantity ϱ_t. The choice of forgetting factor β_t is then a critical element in the algorithm. A good choice will result in slowly varying component estimates.

If one suspects several components in the data to be synchronous with the reference signal, the algorithm can be applied several times, in a deflation manner. Each time a synchronous source $s(t)$ is found it needs to be removed from the data. The standard solution of projecting $s(t)$ back to the observation space and subtracting it from $\boldsymbol{y}(t)$, would require the data to be whitened. This can be achieved by an invertible linear transformation of $\boldsymbol{y}(t)$, prior to running Algo. 1. Since this presents just an additional linear mixing, Algo. 1 can, without any loss of generality, compensate for it. Let the whitened data be $\boldsymbol{z}(t)$, then each component can be subtracted, *e.g.* by $\boldsymbol{z}'(t) = \boldsymbol{z}(t) - (\boldsymbol{w}\boldsymbol{w}^{\mathsf{T}})^{-1}\boldsymbol{w}\,s(t)$, and the process continues. Even for two components with exactly the same phase evolution, *i.e.* identical PLFs, if their amplitudes vary, the algorithm would not converge to a mixture of those.

2.1 Simulation Examples

True blind source separation (BSS) algorithms use no explicit information about the sources to be extracted. The estimation relies typically on general assump-
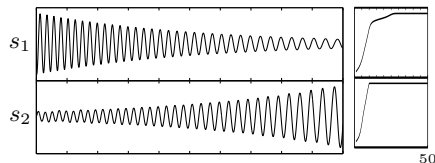
Fig. 3. Left: Examples of the two sources found by Algo.1. in the noiseless case ($\eta = 0.1$). Right: Corresponding objective function.
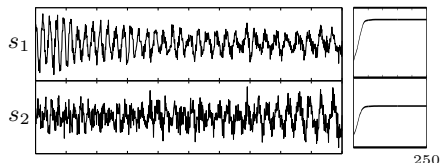
Fig. 4. Left: Examples of the two sources found by Algo.1. in the noisy case ($\eta = 0.1$). Right: Corresponding objective function.

tions such as statistical independence or non-Gaussianity of the sources' distributions. When in presence of oscillatory data, often a criterion based on temporal decorrelation can be employed (see [6] for an overview of various implementations of independent component analysis, one of the most used solutions to the BSS problem). No such requirements are necessary if knowledge of the source phase is available, up to an arbitrary constant phase lag.

To show the applicability of Algo. 1. to the search for components synchronous to a reference, we have generated a set of oscillatory signals $x_i(t) = A_i(t)\sin(\phi_i(t))$ (see Fig. 1). These can not be estimated from instantaneous linear mixtures by neither non-Gaussianity, nor temporal decorrelation criteria. This is because most of the sources have modulated amplitudes that insure histograms close to Gaussian. All have kurtosis close to that of $x_6(t)$, which corresponds to random Gaussian noise. Temporal decorrelation methods will fail also due to the varying frequency content of the sources.

The oscillators $x_i(t)|_{i=1,\cdots,6}$ are not phase coupled, thus the change of the instantaneous phase is proportional to their own natural frequency $\dot{\phi}_i = \omega_i(t)$. Only components 2 and 3 are coupled, such that $\phi_2(t) - \phi_3(t) = \text{const}$. Opposite to Fig. 1, is depicted the PLF of each source signal to the reference as the area of a square. The reference has the same phase dynamic as x_2 and x_3, but a different phase offset and an arbitrary amplitude, thus $\varrho_{x_2} = \varrho_{x_3} \approx 1$. This choice is just illustrative. Comparable results were reached using all other oscillators.

Figure 2 shows a set of linear mixtures of the signals in Fig. 1. Note that all mixtures have now a medium amount PLF to the reference, although clearly bellow those attained by the sources (no mixture has a PLF in excess of 0.75).

The perfect coupling between $x_2(t)$ and $x_3(t)$ suggests that any of the two can be found when the algorithm in Algo. 1 is used with a reference sharing their phase dynamics. Since any mixture of $x_2(t)$ and $x_3(t)$ is less synchronous to the reference, a single one is estimated at a time. The results depicted in Fig. 3 illustrate this fact. Note the correct estimation of the amplitude of the source signal, in addition to the phase recovery with the proper offset. In order to extract the second phase locked component, the first estimate should be removed by projecting it back to the observation space and subtracting it. This allows to extract the whole two dimensional subspace from the data, that is maximally phase locked to the reference.
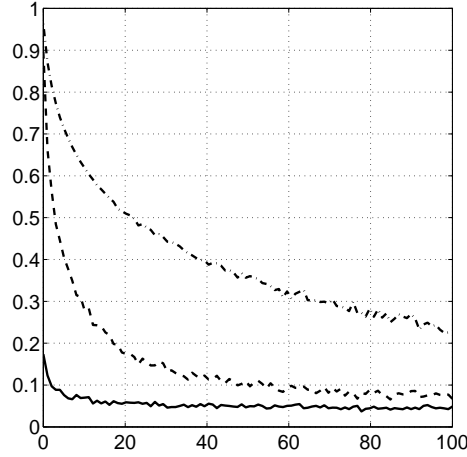
Fig. 5. *Ordinate:* PLF; *Abscissa:* σ^2. Maximal number of iterations and learning rate η are kept constant.

As in many source extraction algorithms the global amplitude scale and the sign of the sources will remain undetermined. For that reason, the projection vector is arbitrarily normalised to unit norm in step 10 of the algorithm.

The convergence speed for a particular run of the method can be inspected on the right part of Fig. 3. The exact values can vary, depending on the choice of η. If the phase of the reference signal is not present in the data, the algorithm will not reach a high PLF.

2.2   Sensitivity to Noise

The phase of a Gaussian white noise signal is typically mildly locked (PLF of ca. 0.1) to any other signal, including other Gaussian noise processes.

Let us assume that the observed mixtures $\boldsymbol{y}(t) = \boldsymbol{Ax}(t) + \sigma\,\boldsymbol{\varepsilon}(t)$, as well as the reference signal $v(t) = u(t) + \varsigma\,\nu(t)$, are corrupted with noise of variance σ^2 and ς^2 respectively. $\boldsymbol{\varepsilon}(t)$ and $\nu(t)$ are both drawn form a Gaussian distribution having zero mean and $\mathsf{Cov}[\varepsilon_i(s)\varepsilon_j(t)] = \delta_{ij}\delta_{st}$, $\mathsf{Cov}[\nu(s)\nu(t)] = \delta_{st}$.

Figure 4 shows a replication of the experiment reported in Fig. 3, for the case of added observational noise of the same unit variance $\sigma^2 = 1$ as the data. The estimation is not as perfect as in the noiseless environment, possibly due to a non-zero phase locking between the reference signal and the noise (see Fig. 1). The PLF serves as a quality measure for the extracted components. The obtained PLFs are $\varrho_{s_1} = 0.83$ and $\varrho_{s_2} = 0.59$, which are significantly beneath those of the true sources. A common problem of deflation schemes is that the estimation error accumulates with the number of extracted components. Also the convergence speed is, as should be expected, reduced slightly with the noise source present.

In Figure 5 the PLF is plotted as a function of the observation noise magnitude σ^2. The different graphs correspond to values of ς^2 (the steeper slops for lower ς^2). The maximally achieved objective function value (keeping the maximum number of iterations fixed) deteriorates with increasing noise variance in both observations and reference.

The presence of noise has a profound influence in many real world applications. Furthermore, is it possible for real signals to exhibit very broad spectra, with oscillatory dynamics in different frequency ranges. The Hilbert transform is not able to estimate a meaningful phase for such broad band signals. In conclusion, it is therefore advisable to remove, or reduce the noise and filter the signal in a targeted frequency band of interest, prior to the phase analysis. A way to combine filtering and phase estimation, that was reported to perform reliably on biological signals, is the convolution with complex Morlet wavelets [7]. Another valuable preprocessing approach is singular spectrum analysis (SSA), since it allows to decompose a signal into trends, oscillators and noise components, *cf.* Ref. [8].

3   Cortico-muscular Phase Locking in MEG Revisited

Strong coherence, *i.e.*, spectral cross-correlation (see [9] and references therein), and synchronisation have been observed between electrophysiological measurements from the brain and muscles (using electroencephalograms, EEG; magnetoencephalograms, MEG; and electromyograms, EMG). *Cortico-muscular* and *cortico-cortical* interactions were found in frequency bands centred around 15Hz, 20Hz and 40Hz. These have been supported by physiological consideration upon the biological processes involved.

An obstacle in these studies, *e.g.*, addressed in [10], is that the synchronisation among EEG or MEG channels is likely to result partially from cross-talk and volume conduction, *i.e.* the same oscillator being present in different measurements, because of a natural mixing process. Synchronisation between EEG/MEG and EMG, on the contrary, will be decreased as a result of the same process, since there is no single EEG/MEG channel that presents directly the underlying oscillator that is synchronous to the EMG.

In [11] the imaginary part of coherence has been introduced as a promising measure for brain interactions. It has the appealing property of not being sensitive to volume conduction, though it could possibly oversee interactions with very small phase lags. Such zero phase lag synchronisation could arise if the neuronal coupling between the two subsystems is strong and symmetric. On the other hand, the amount of phase lag between the compared signals does not affect the estimate of the PLF, on which Algo. 1. is based. Since it only measures synchrony between the *source* signals and the reference, the volume conduction and cross-talk should also be decreased. The algorithm's assumptions of a linear and instantaneous nature of the mixing process, are both substantiated from a theoretical viewpoint (see, *e.g.*, [12, 13]).
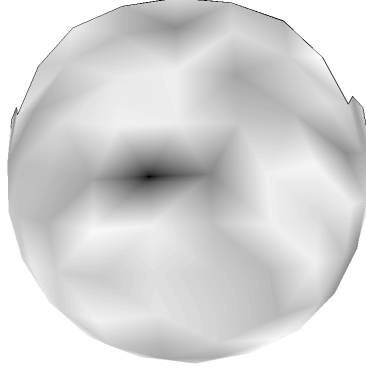
Fig. 6. Topographic map of the Neuromag MEG helmet at 18–20Hz.

We tested Algo. 1 on real measurements, using the data set described in [9]. It consists of simultaneous MEG recordings, with a 306-sensor Vectorview neuromagnetometer (Neuromag Ltd; 204 planar gradiometers and 102 magnetometers), together with left and right hand EMG's. The subject was instructed to simultaneously keep isometric contraction in left and right hand muscles, using a special squeezing device. Only the measurements of the planar gradiometers were analysed. The sampling rate is 600 Hz for a duration of 3 minutes.

Based on physiological considerations, we have scanned a series of frequency ranges for targeting the algorithm. The results attained for the range 18–20Hz are shown in Fig. 6. This corresponds to the estimated projection between the extracted source and the measurements. This topographic map is conventionally called the component's field map. The view is taken from above, preserving right and left orientations, and with front facing up.

A comparison between the results shown, and the ones presented in [9], suggests the phase locked component to represent activity originating from the primary motor cortex. The variance of the extracted source is of the same magnitude as the measurements, indicating that the component has a significant presence in the recordings.

4   Concluding Remarks

Synchronisation plays a capital role in interacting oscillatory systems. It has been proposed in the literature that brain communication is implemented through synchronisation. We introduced a gradient based algorithm for the extraction of components, from measurements that are phase synchronous to a given reference signal. This can potentially elicit information about neuronal oscillator interactions from brain signals. The problem of noise was addressed in a controlled simulated environment. A preliminary study of its usage in cortico-muscular interactions was also presented.

In the future the robustness and convergence behaviour of the algorithm shall be determined in more detail. On a practical side, one should investigate which preprocessing techniques are useful when applying the algorithm to real world problems. Beyond the cortico-muscular example, we intend to investigate communication inside the central nervous system. This will require an extension of the algorithm in which the reference signal is estimated also from the measured signals in an unsupervised manner.

Acknowledgements

References

1. Pikovsky, A., Rosenblum, M., Kurths, J.: Synchronization – A universal concept in nonlinear sciences. Volume 12 of Cambridge Nonlinear Science Series. Cambridge University Press, UK (2001)
2. Kuramoto, Y.: Chemical Oscillations, Waves and Turbulences. Springer Berlin (1984)
3. Strogatz, S.H.: From Kuramoto to Crawford: Exploring the Onset of Synchronization in Populations of Coupled Oscillators. Physica D **143** (2000) 1–20
4. Hindmarsh, J.L., Rose, R.M.: A model of neuronal bursting using three coupled first–order differential equations. Proceedings of the Royal Society of London **221** (1984) 87–102
5. Frank, T.D., Daffertshofer, A., Pepper, C.E., Beek, P.J., Haken, H.: Towards a comprehensive theory of brain activity: Coupled oscillator systems under external forces. Physica D **144** (2000) 62–86
6. Hyvärinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. Adaptive and Learning Systems for Signal Processing, Communications, and Control. John Wiley & Sons, Inc. (2001)
7. Lachaux, J.P., Rodriguez, E., Martinerie, J., Varela, F.J.: Measuring phase synchrony in brain signals. Human Brain Mapping **8**(4) (1999) 194–208
8. Ghil, M., Allen, M.R., Dettinger, M.D., Ide, K., Kondrashov, D., Mann, M.E., Robertson, A., Saunders, A., Tian, Y., Yiou, P.: Advanced Spectral Methods for Climatic Time Series. Reviews of Geophysics **40**(1) (2001)
9. Vigário, R., Jensen, O.: Identifying Cortical Sources of Corticomuscle Coherence During Bimanual Muscle Contraction by Temporal Decorrelation. In: Proceedings of IEEE International Symposium on Signal Processing and Its Applications. (2003)
10. Meinecke, F.C., Ziehe, A., Kurths, J., Müller, K.R.: Measuring Phase Synchronization of Superimposed Signals. Physical Review Letters **94**(8) (2005)
11. Nolte, G., Bai, O., Wheaton, L., Mari, Z., Vorbach, S., Hallett, M.: Identifying true brain interaction from EEG data using the imaginary part of coherency. Clinical Neurphysiology **115** (2004) 2292–2307
12. Hämäläinen, M., Hari, R., Ilmoniemi, R., Knuutila, J., Lounasmaa, O.V.: Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. Reviews of Modern Physics **65**(2) (1993) 413–497

13. Vigário, R., Särelä, J., Jousmäki, V., Hämäläinen, M., Oja, E.: Independent component approach to the analysis of EEG and MEG recordings. IEEE Transactions on Biomedical Engineering **47**(5) (2000) 589–593

A   Gradient of ϱ^2

The gradient can be written out as $\nabla\varrho^2 = \nabla\big(\varrho\cos(\Psi)\big)^2 + \nabla\big(\varrho\sin(\Psi)\big)^2$, which equals

$$2\varrho\big(\cos(\Psi)[\nabla\varrho\cos(\Psi)] + \sin(\Psi)[\nabla\varrho\sin(\Psi)]\big). \qquad (6)$$

From the definition in Eq. (2) it follows that $\varrho\cos(\Psi) = \frac{1}{T}\sum_{t=1}^{T}\cos(\Delta\phi(t))$. Inserting this and the equivalent identity for $\varrho\sin(\Psi)$ into Eq. (6) and further evaluating the gradient yields

$$\frac{2\varrho}{T}\sum_{t=1}^{T}\Big[\sin(\Psi)\cos\big(\Delta\phi(t)\big) - \cos(\Psi)\sin\big(\Delta\phi(t)\big)\Big]\nabla\phi(t). \qquad (7)$$

The phase $\phi(t)$ is the angle of $\hat{s}(t)$ in the complex plane. This is given as $\phi(t) = \text{angle}\,\hat{s}(t) = \arctan2(\mathcal{H}[s](t), s(t))$, where the two arguments arctan maps the angle into the correct quadrant. Let the Hilbert transform $\mathcal{H}[\cdot]$ operate on the coordinates of a vector. For the gradient of $\phi(t) = \arctan2(\boldsymbol{w}^\mathsf{T}\boldsymbol{y}(t), \boldsymbol{w}^\mathsf{T}\mathcal{H}[\boldsymbol{y}](t))$, the arctan2-function can be substituted by the normal arctan, so that

$$\nabla\phi(t) = \nabla\arctan\Big(\frac{\boldsymbol{w}^\mathsf{T}\mathcal{H}[\boldsymbol{y}](t)}{\boldsymbol{w}^\mathsf{T}\boldsymbol{y}(t)}\Big) = \frac{(\boldsymbol{w}^\mathsf{T}\boldsymbol{y}(t))\mathcal{H}[\boldsymbol{y}](t) - (\boldsymbol{w}^\mathsf{T}\mathcal{H}[\boldsymbol{y}](t))\boldsymbol{y}(t)}{\Big(1 + \big(\frac{\boldsymbol{w}^\mathsf{T}\mathcal{H}[\boldsymbol{y}](t)}{\boldsymbol{w}^\mathsf{T}\boldsymbol{y}(t)}\big)^2\Big)(\boldsymbol{w}^\mathsf{T}\boldsymbol{y}(t))^2}.$$

The first factor in the denominator is the derivative of arctan and the second is the result of the quotient rule of differentiation. This can be rearranged to

$$\nabla\phi(t) = \frac{(\mathcal{H}[\boldsymbol{y}](t)\boldsymbol{y}^\mathsf{T}(t) - \boldsymbol{y}(t)\mathcal{H}[\boldsymbol{y}]^\mathsf{T}(t))\,\boldsymbol{w}}{(\boldsymbol{w}^\mathsf{T}\boldsymbol{y}(t))^2 + (\boldsymbol{w}^\mathsf{T}\mathcal{H}[\boldsymbol{y}](t))^2}, \qquad (8)$$

reviling the denominator to be the square magnitude of extracted source $s^2(t) + \mathcal{H}[s]^2(t) = S^2(t)$. The matrix $\boldsymbol{\Gamma}(t) = (\mathcal{H}[\boldsymbol{y}](t)\boldsymbol{y}^\mathsf{T}(t) - \boldsymbol{y}(t)\mathcal{H}[\boldsymbol{y}]^\mathsf{T}(t))$ in the numerator can also be written in terms of the phase of the observation signal as

$$[\boldsymbol{\Gamma}(t)]_{ij} = Y_i(t)Y_j(t)\sin(\varphi_i(t))\cos(\varphi_j(t)) - Y_i(t)Y_j(t)\cos(\varphi_i(t))\sin(\varphi_j(t))$$
$$= Y_i(t)Y_j(t)\sin(\varphi_i(t) - \varphi_j(t)).$$

The same simplification can be applied to Eq. (7), to finally arrive at

$$\nabla\varrho^2 = \frac{2\varrho}{T}\sum_{t=1}^{T}\frac{\sin(\Psi - \Delta\phi(t))}{S^2(t)}\boldsymbol{\Gamma}(t)\,\boldsymbol{w}. \qquad (9)$$