

## **OHBM 2016 Educational Course: Electromagnetical Neuroimaging**

**This is one of many educational lectures.**

**This material is intended for the lecture entitled:**

**“Non-invasive imaging of cortical electric neuronal activity for the localization of function and for connectivity inference”**

**by:**

**Roberto D. Pascual-Marqui, PhD, PD**

**The KEY Institute for Brain-Mind Research, University of Zurich**

**Visiting Professor at Neuropsychiatry, Kansai Medical University, Osaka**

**[[www.keyinst.uzh.ch/loreta](http://www.keyinst.uzh.ch/loreta)] [[scholar.google.com/citations?user=pascualmarqui](https://scholar.google.com/citations?user=pascualmarqui)]**

**This material is a draft of the author’s version of the book chapter:**

**Pascual-Marqui RD. Theory of the EEG inverse problem. Quantitative eeg analysis: Methods and clinical applications. Boston: Artech House. 2009:121-40.**

### **Summary**

We deal here with the EEG neuroimaging problem: given measurements of scalp electric potential differences (EEG: electroencephalogram), find the 3D distribution of the generating electric neuronal activity. This problem has no unique solution. Particular solutions with optimal localization properties are of main interest, since neuroimaging is concerned with the correct localization of brain function. A brief historical outline of localization methods is given: from the single dipole, to multiple dipoles, to distributions. Technical details on the formulation and solution of this type of inverse problem are presented. Emphasis is placed on linear, discrete, 3D distributed EEG tomographies, that have a simple mathematical structure allowing a complete evaluation of their localization properties. One particular noteworthy member of this family is eLORETA (exact low resolution brain electromagnetic tomography [46]), which is a genuine inverse solution (not merely a linear imaging method, nor a collection of one-at-a-time single best fitting dipoles) with zero localization bias in the presence of measurement and structured biological noise.

### **5.1 Introduction**

Hans Berger [1] reported as early as 1929 on the human EEG (electroencephalogram), which consists of time varying measurements of scalp electric potential differences. At that time, using only one posterior scalp electrode with an anterior reference, he measured the alpha rhythm, an oscillatory activity in the range 8-12 Hz, that appears when the subject is awake, resting, with eyes closed. He observed that by simply opening the eyes, the alpha rhythm would disorganize and tend to disappear. Such observations led Berger to the belief that the EEG was a window into the brain. Through this “window”, one can “see” brain function, e.g., what posterior brain regions are doing when changing state from eyes open to closed.

The concept of “a window into the brain” already implies the localization of different brain regions, each one with certain characteristics and functions. From this point of view, Berger was already performing a very naïve type of low spatial resolution, low spatial sampling form of neuroimaging, by assuming that the electrical activity recorded at a scalp electrode was determined by the activity of the underlying brain structure. To this day, many published research papers still use the same technique, where brain localization inference is based on the scalp distribution of electric potentials (commonly known as topographic scalp maps).

It must be emphasized from the outset that this topographic-based method is in general not correct. In the case of EEG recordings, scalp electric potential differences are determined by electric neuronal activity from the entire cortex, and by the geometrical orientation of the cortex. The cortical orientation factor alone has a very dramatic effect: an electrode placed over an active gyrus or sulcus will be influenced in extremely different ways. The consequence is: ***a scalp electrode does not necessarily reflect activity of the underlying cortex.***

The route towards EEG-based neuroimaging must rely on the correct use of the physics laws that connect electric neuronal generators and scalp electric potentials. Formally, the EEG inverse problem can be stated as follows: given measurements of scalp electric potential differences, find the three-dimensional (3D) distribution of the generators, i.e. of the electric neuronal activity.

However, it turns out that in its most general form, this type of inverse problem has no unique solution, as was shown by Helmholtz in 1853 [2]. The curse of non-uniqueness [3] informally means that there is insufficient information in the scalp electric potential distribution to determine the actual generator distribution. Equivalently, given scalp potentials, there are infinitely different generator distributions that comply with the scalp measurements. The apparent consequence is that there is no way to determine the actual generators from scalp electric potentials.

This seemingly hopeless situation is not very true. The general statement of Helmholtz applies to arbitrary distributions of generators. However, the electric neuronal generators in the human brain are not arbitrary, and actually have properties that can be combined into the inverse problem statement, narrowing the possible solutions. In addition to endowing the possible inverse solutions with certain neuroanatomical and electrophysiology properties, we will be interested only in those solutions that have “good” localization properties, since that is what neuroimaging is all about: the localization of brain function.

Several solutions will be reviewed, with particular emphasis on the general family of linear imaging methods.

## 5.2 EEG generation

Details on the electrophysiology and physics of EEG/MEG generation can be found in Mitzdorf [4], Llinas [5], Martin [6], Hämmäläinen et al [7], Haalman and Vaadia [8], Sukov and Barth [9], Dale et al [10], and Baillet et al. [11]. The basic underlying physics can be studied in Sarvas [28].

### 5.2.1 The electrophysiological and neuroanatomical basis of the EEG

It is now widely accepted that scalp electric potential differences are generated by cortical pyramidal neurons undergoing post-synaptic potentials (PSPs). These neurons are oriented perpendicular to the cortical surface. The magnitude of experimentally recorded scalp electric potentials, at any given time instant, is due to the spatial summation of the impressed current density induced by highly synchronized PSPs occurring in large clusters of neurons. A typical cluster size must cover at least 40 to 200 mm<sup>2</sup> of cortical surface in order to produce a measurable scalp signal.

Summarizing, there are two essential properties:

1. The EEG sources are confined to the cortical surface, which is populated mainly by pyramidal neurons (constituting approximately 80% of the cortex), oriented perpendicular to the surface.

2. The frequent occurrence of highly synchronized PSPs in spatial clusters of cortical pyramidal neurons.

This information can be used to narrow significantly the non-uniqueness of the inverse solution, as will be explained later on.

It is important to keep in mind that there is a very strict limitation in the use of the equivalent terms “EEG generators” and “electric neuronal generators”. This is best illustrated with an example, such as the alpha rhythm. Cortical pyramidal neurons located mainly in occipital cortical areas are partly driven by thalamic neurons that make them beat synchronously at about 11 Hz (a thalamo-cortical loop). But the EEG does not “see” all parts of this electrophysiological mechanism. The EEG only sees the final electric consequence of this process, namely that the alpha rhythm is *electrically* generated in occipital cortical areas.

This raises the question: are scalp electric potentials only due to electrically active cortical pyramidal neurons? The answer is no. All active neurons contribute to the EEG. However, the contribution from the cortex is overwhelmingly large compared to all other structures, due to two factors:

1. The number of cortical neurons is much larger than subcortical neurons.
2. The distance from subcortical structures to the scalp electrodes is larger than from cortical structures to the electrodes.

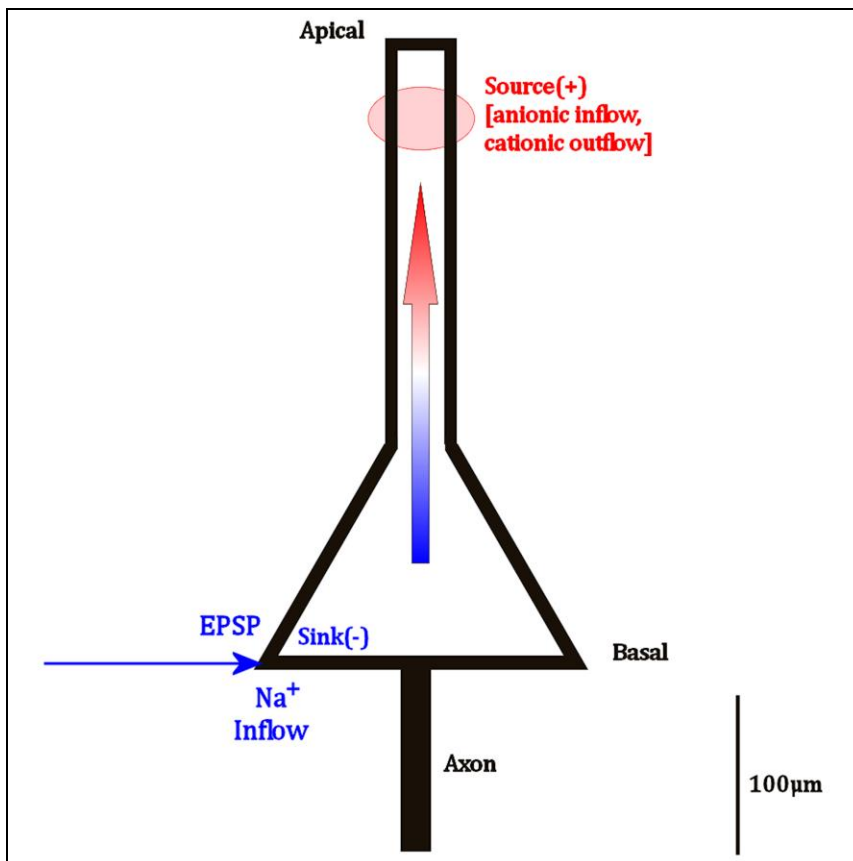
This is why EEG recordings are mainly generated by electrically active cortical pyramidal neurons.

It is possible to manipulate the measurements in order to enhance non-cortical generators. This can be achieved by averaging EEG measurements appropriately, such as is traditionally done in average event related potentials. Such an averaging manipulation usually reduces the amplitude of the background EEG activity, enhancing the brain response that is phase-locked to the stimulus. When the number of stimuli is very high, the average scalp potentials might be mostly due to non-cortical structures, as in a brain stem auditory evoked potential [13].

## 5.2.2 The equivalent current dipole

From the physics point of view, a cortical pyramidal neuron undergoing a PSP will behave as a current dipole, which consists of a current source and a current sink separated by a distance in the range of 100 to 500 micrometers. This means that both poles (the source and the sink) are always paired, and extremely close to each other, as seen from the macroscopic scalp electrodes. For this reason, the sources of the EEG can be modeled as a distribution of dipoles along the cortical surface.

Figure 5.1 illustrates the equivalent current dipole corresponding to a cortical pyramidal neuron undergoing an excitatory post-synaptic potential (EPSP) taking place at a basal dendrite. The cortical pyramidal neuron is outlined in black. Notice the approximate size scale (100 $\mu$ m bar in lower right). An incoming axon from a pre-synaptic neuron (blue arrow in lower left) terminates at a basal dendrite. The event taking place induces specific channels to open, allowing (typically) an inflow of Na<sup>+</sup>, which gives rise to a sink of current. Electrical neutrality must be conserved, and a source of current is produced at the apical regions, as shown in the transparent red ellipse. The corresponding current dipole vector is shown as the color coded arrow: blue (negative) to red (positive).



**Figure 5.1:** Schematic representation of the generators of the EEG: the equivalent current dipole corresponding to a cortical pyramidal neuron undergoing an excitatory post-synaptic potential (EPSP) taking place at a basal dendrite. The cortical pyramidal neuron is outlined in black. The incoming axon from a pre-synaptic neuron (blue arrow in lower left) terminates at a basal dendrite. Channels to open, allowing (typically) an inflow of  $\text{Na}^+$ , which gives rise to a sink of current. Due to the conservation of electrical neutrality, a source of current is produced at the apical regions, as shown in the transparent red ellipse. The corresponding current dipole vector is shown as the color coded arrow: blue (negative) to red (positive).

This implies that it would be very much against electrophysiology to model the sources as freely distributed, non-paired monopoles of current. An early attempt in this direction can be found in [12]. Those monopolar inverse solutions were not pursued any further because, as expected, they simply were incapable of correct localization when tested with real human data such as visual, auditory, and somatosensory event related potentials, for which the localization of the sensory cortices are well known.

It must be kept in mind that a single active neuron is not enough to produce measurable scalp electric potential differences. EEG measurements are possible due to the existence of relatively large spatial clusters of cortical pyramidal cells that are geometrically arranged parallel to each other, and that simultaneously undergo the same type of postsynaptic potential (synchronization). If these conditions are not met, then the total summed activity is too weak to produce non-negligible extracranial fields.

### 5.3 Localization of the electrically active neurons as a small number of “hot spots”

An early attempt towards the localization of the active brain region responsible for the scalp electric potential distribution was performed in a semi-quantitative manner by Brazier in 1949 [14]. It was suggested to make use of electric field theory to determine the location and orientation of the current dipole from the scalp potential map. This can be considered to be the starting point for what later developed into “dipole fitting”.

Immediately afterwards, using a spherical head model, the equations were derived that relate electric potential differences on the surface of a homogeneous conducting sphere due to a current dipole within [15, 16]. About a decade later, an improved, more realistic head model considered the different conductivities of neural tissue, skull, and scalp [17]. Use was made of these early techniques by Lehmann et al [18] to locate the generator of a visual evoked potential.

Note that in the single current dipole model, it is assumed that brain activity is due to a single small area of active cortex. In general, this model is very simplistic and non-realistic, since the whole cortex is never totally “quiet” except for a single small area. Nevertheless, the dipole model does produce reasonable results under some particular conditions. This was shown very convincingly by Henderson et al [19], both in an experimentally simulated head (a head phantom), and with real human EEG recordings. The conditions under which a dipole model makes sense are limited to cases where electric neuronal activity is dominated by a small brain area. Two examples where the model performs very well are in some epileptic spike events, and in the description of the early components of the average brain stem auditory evoked potential [13].

However, it would seem that the localization of higher cognitive functions could not be reliably modeled by dipole fitting.

#### 5.3.1 Single dipole fitting

Single dipole fitting can be seen as the localization of the electrically active neurons as a single “hot spot”.

Consider the case of a single current dipole located at position  $\mathbf{r}_v \in \mathbb{R}^{3 \times 1}$ , with dipole moment  $\mathbf{j}_v \in \mathbb{R}^{3 \times 1}$ , where:

$$\mathbf{r}_v = \begin{pmatrix} x_v & y_v & z_v \end{pmatrix}^T \quad (5.1)$$

denotes the position vector, with the superscript “ $T$ ” denoting vector/matrix transposition, and:

$$\mathbf{j}_v = \begin{pmatrix} j_x & j_y & j_z \end{pmatrix}^T \quad (5.2)$$

In order to introduce the basic form of the equations, consider the non-realistic, simple case of the current dipole in an infinite homogenous medium with conductivity  $\sigma$ . Then the electric potential at location  $\mathbf{r}_e \in \mathbb{R}^{3 \times 1}$ , for  $\mathbf{r}_e \neq \mathbf{r}_v$ , is:

$$\phi(\mathbf{r}_e, \mathbf{r}_v, \mathbf{j}_v) = \mathbf{k}_{e,v}^T \mathbf{j}_v + c \quad (5.3)$$

where:

$$\mathbf{k}_{e,v} = \frac{1}{4\pi\sigma} \frac{(\mathbf{r}_e - \mathbf{r}_v)}{\|\mathbf{r}_e - \mathbf{r}_v\|^3} \quad (5.4)$$

denotes what is commonly known as the lead field.

In equation (5. 3),  $c$  is a scalar accounting for the physics nature of electric potentials which are determined up to an arbitrary constant.

A slightly more realistic head model corresponds to a spherical homogeneous conductor in air. The lead field in this case is:

$$\mathbf{k}_{e,v} = \frac{1}{4\pi\sigma} \left[ 2 \frac{(\mathbf{r}_e - \mathbf{r}_v)}{\|\mathbf{r}_e - \mathbf{r}_v\|^3} + \frac{\mathbf{r}_e \|\mathbf{r}_e - \mathbf{r}_v\| + (\mathbf{r}_e - \mathbf{r}_v) \|\mathbf{r}_e\|}{\|\mathbf{r}_e\| \|\mathbf{r}_e - \mathbf{r}_v\| [\|\mathbf{r}_e\| \|\mathbf{r}_e - \mathbf{r}_v\| + \mathbf{r}_e^T (\mathbf{r}_e - \mathbf{r}_v)]} \right] \quad (5.5)$$

In the previous equation, the following notation was used:

$$\|\mathbf{X}\|^2 = \text{tr}(\mathbf{X}^T \mathbf{X}) = \text{tr}(\mathbf{X} \mathbf{X}^T) \quad (5.6)$$

where  $\text{tr}$  denotes the trace, and  $\mathbf{X}$  is any matrix or vector. If  $\mathbf{X}$  is a vector, then this is the squared Euclidean  $L_2$  norm; if  $\mathbf{X}$  is a matrix, then this is the squared Frobenius norm.

The equation for the lead field in a totally realistic head model (taking into account geometry and full conductivity profile) is not available in closed form, such equations (5. 4) and (5. 5). Numerical methods for computing the lead field can be found in [20].

Nevertheless, in general, the components of the lead field  $\mathbf{k}_{e,v} = (k_x \ k_y \ k_z)^T$  have a very simple interpretation:  $k_x$  corresponds to the electric potential at position  $\mathbf{r}_e$ , due to unit strength current dipole  $j_x = 1$  at position  $\mathbf{r}_v$ ; and similarly for the other two components.

Formally, we are now in a position to state the single dipole fitting problem. Let  $\hat{\phi}_e$ , for  $e=1\dots N_E$ , denote the scalp electric potential measurement at electrode “ $e$ ”, where  $N_E$  is the total number of cephalic electrodes. All measurements are made with using the same reference. Let  $\phi_e(\mathbf{r}_v, \mathbf{j}_v)$ , for  $e=1\dots N_E$ , denote the theoretical potential at electrode “ $e$ ”, due to a current dipole located at  $\mathbf{r}_v$  with moment  $\mathbf{j}_v$ . Then the problem consists in finding the unknown dipole position  $\mathbf{r}_v$  and moment  $\mathbf{j}_v$  that **best** explain the actual measurements. The simplest way to achieve this is to minimize the **distance** between theoretical and experimental potentials.

Consider the functional:

$$F = \sum_{e=1}^{N_E} [\phi_e(\mathbf{r}_v, \mathbf{j}_v) - \hat{\phi}_e]^2 \quad (5.7)$$

This expresses the distance between measurements and model, as a function of the two main dipole parameters: its location and its moment. The aim is to find the values of the parameters that minimize the functional, i.e. the **least squares** solution. There are many algorithms to find the parameters, as reviewed in [11, 13].

### 5.3.2 Multiple dipole fitting

A straightforward generalization of the previous case consists of attempting to explain the measured EEG as due to a small number of active brain spots. Based on the principle of superposition, the theoretical potential due to  $N_v$  dipoles is simply the sum of potentials due to each individual dipole. Therefore, the functional in equation (5. 7) generalizes to:

$$F = \sum_{e=1}^{N_E} \left[ \sum_{v=1}^{N_V} \phi_e(\mathbf{r}_v, \mathbf{j}_v) - \hat{\phi}_e \right]^2 \quad (5.8)$$

And the least squares problem for this multiple dipole fitting case consists of finding all dipole positions  $\mathbf{r}_v$  and moments  $\mathbf{j}_v$ , for  $v=1\dots N_V$ , that minimize  $F$ .

There are two major problems with multiple dipole fitting:

1. The number of dipoles  $N_V$  must be known beforehand. The dipole locations vary greatly for different values of  $N_V$ .
2. For realistic measurements (which includes measurement noise), and for a given fixed value of  $N_V > 1$ , the functional in equation (5. 8) has many local minima, with several of them very close in value to the absolute minimum, but all of them with very different locations for the dipoles. This makes it very difficult to choose objectively the correct solution.

## 5.4 Discrete, 3D distributed tomographic methods

The principles that will be used in this section are common to other tomographies, such as structural X-Ray (i.e. CAT scans), structural magnetic resonance imaging (MRI), and functional tomographies such as fMRI and positron emission tomography (PET).

For the EEG inverse problem, the solution space consists of a distribution of points in 3D space. A classical example is to construct a 3D uniform grid throughout the brain, and to retain the points that fall on the cortical surface (mainly populated by pyramidal neurons). At each such point, whose coordinates are known by construction, a current density vector with unknown moment components is placed. The current density vector (i.e. the equivalent current dipole) at a grid point represents the total electric neuronal activity of the volume immediately around the grid point, commonly called a voxel.

The scalp electric potential difference at a given electrode receives contributions, in an additive manner, from all voxels. The equation relating scalp potentials and current density can be conveniently expressed in vector/matrix notation as:

$$\Phi = \mathbf{KJ} + c\mathbf{1} \quad (5.9)$$

where the vector  $\Phi \in \mathbb{R}^{N_E \times 1}$  contains the instantaneous scalp electric potential differences measured at  $N_E$  electrodes with respect to a single common reference electrode (e.g., the reference can be linked earlobes, the toe, or one of the electrodes included in  $\Phi$ ); the matrix  $\mathbf{K} \in \mathbb{R}^{N_E \times (3N_V)}$  is the lead field matrix corresponding to  $N_V$  voxels;  $\mathbf{J} \in \mathbb{R}^{(3N_V) \times 1}$  is the current density;  $c$  is a scalar accounting for the physics nature of electric potentials which are determined up to an arbitrary constant; and  $\mathbf{1}$  denotes a vector of ones, in this case  $\mathbf{1} \in \mathbb{R}^{N_E \times 1}$ . Typically  $N_E \ll N_V$ , and  $N_E \geq 19$ .

In equation (5. 9), the structure of the lead field matrix  $\mathbf{K}$  is:

$$\mathbf{K} = \begin{pmatrix} \mathbf{k}_{11}^T & \mathbf{k}_{12}^T & \dots & \mathbf{k}_{1N_V}^T \\ \mathbf{k}_{21}^T & \mathbf{k}_{22}^T & \dots & \mathbf{k}_{2N_V}^T \\ \dots & & & \\ \mathbf{k}_{N_E 1}^T & \mathbf{k}_{N_E 2}^T & \dots & \mathbf{k}_{N_E N_V}^T \end{pmatrix} \quad (5.10)$$

where  $\mathbf{k}_{ev} \in \mathbb{R}^{3 \times 1}$ , for  $e=1 \dots N_E$  and for  $v=1 \dots N_V$ , corresponds to the scalp potentials at the  $e$ -th electrode due to three orthogonal unit strength dipoles at voxel  $v$ , each one oriented along the coordinate axes  $x$ ,  $y$ , and  $z$ . Equations (5. 4) and (5. 5) above were two examples of the lead field that can be written in closed form, although they correspond to head models that are too unrealistic.

Note that  $\mathbf{K}$  can also be conveniently written as:

$$\mathbf{K} = (\mathbf{K}_1, \mathbf{K}_2, \mathbf{K}_3, \dots, \mathbf{K}_{N_V}) \quad (5.11)$$

where  $\mathbf{K}_v \in \mathbb{R}^{N_E \times 3}$ , for  $v=1 \dots N_V$ , is defined as:

$$\mathbf{K}_v = \begin{pmatrix} \mathbf{k}_{1v}^T \\ \mathbf{k}_{2v}^T \\ \dots \\ \mathbf{k}_{N_E v}^T \end{pmatrix} \quad (5.12)$$

In equation (5. 9),  $\mathbf{J}$  is structured as:

$$\mathbf{J} = \begin{pmatrix} \mathbf{j}_1 \\ \mathbf{j}_2 \\ \dots \\ \mathbf{j}_{N_V} \end{pmatrix} \quad (5.13)$$

where  $\mathbf{j}_v \in \mathbb{R}^{3 \times 1}$  denotes the current density at the  $v$ -th voxel, as in equation (5. 2).

At this point, the basic EEG inverse problem for the discrete, 3D distributed case consists of solving equation (5. 9) for the unknown current density  $\mathbf{J}$  and constant  $c$ , given the lead field  $\mathbf{K}$  and measurements  $\Phi$ .

### 5.4.1 The reference electrode problem

As a first step, the reference electrode problem will be solved, by estimating “ $c$ ” in equation (5. 9). Given  $\Phi$  and  $\mathbf{KJ}$ , the reference electrode problem is:

$$\min_c \|\Phi - \mathbf{KJ} - c\mathbf{1}\|^2 \quad (5.14)$$

The solution is:

$$c = \frac{\mathbf{1}^T}{\mathbf{1}^T \mathbf{1}} (\Phi - \mathbf{KJ}) \quad (5.15)$$

Plugging equation (5. 15) into (5. 9) gives:

$$\mathbf{H}\Phi = \mathbf{HKJ} \quad (5.16)$$

where:

$$\mathbf{H} = \mathbf{I} - \frac{\mathbf{1}\mathbf{1}^T}{\mathbf{1}^T \mathbf{1}} \quad (5.17)$$

is the average reference operator, also known as the centering matrix, and  $\mathbf{I} \in \mathbb{R}^{N_E \times N_E}$  is the identity matrix.

This result establishes the fact that any inverse solution will not depend on the reference electrode. This applies to any form of the EEG inverse problem, including the inverse dipole fitting problems in equations (5. 7) and (5. 8).

Henceforth, it will be assumed that the EEG measurements and the lead field are average reference transformed, i.e.:

$$\begin{cases} \Phi \leftarrow \mathbf{H}\Phi \\ \mathbf{K} \leftarrow \mathbf{H}\mathbf{K} \end{cases} \quad (5.18)$$

and equation (5. 9) will be rewritten as:

$$\Phi = \mathbf{K}\mathbf{J} \quad (5.19)$$

Note that  $\mathbf{H}$  plays the role of the identity matrix for EEG data. It actually is the identity matrix, except for a null eigenvalue corresponding to an eigenvector of ones, accounting for the reference electrode constant.

### 5.4.2 The minimum norm inverse solution

Hämäläinen and Ilmoniemi [21] published in 1984 a technical report with a particular solution to the inverse problem corresponding to the forward equation of the type (5. 19). As the name of the method implies, this particular solution is the one that has minimum norm. The problem in its simplest form is stated as:

$$\begin{cases} \min_{\mathbf{J}} \mathbf{J}^T \mathbf{J} \\ \text{such that : } \Phi = \mathbf{K}\mathbf{J} \end{cases} \quad (5.20)$$

The solution is:

$$\hat{\mathbf{J}} = \mathbf{T}\Phi \quad (5.21)$$

with:

$$\mathbf{T} = \mathbf{K}^T (\mathbf{K}\mathbf{K}^T)^+ \quad (5.22)$$

The superscript “+” denotes the Moore-Penrose generalized inverse [22].

The minimum norm inverse solution (5. 21) and (5. 22) is a genuine solution to the system of equations (5. 19).

If the measurements are contaminated with noise, it is typically more convenient to change the statement of the inverse problem in such a way to avoid that the current density be influenced by errors. The new inverse problem now is:

$$\min_{\mathbf{J}} F \quad (5.23)$$

with:

$$F = \|\Phi - \mathbf{K}\mathbf{J}\|^2 + \alpha \mathbf{J}^T \mathbf{J} \quad (5.24)$$

In equation (5. 24), the parameter  $\alpha > 0$  controls the relative importance between the two terms on the right hand side: a penalty for being unfaithful to the measurements and a penalty for a large current density norm. This parameter is known as the Tikhonov regularization parameter [23]. The solution is:

$$\hat{\mathbf{J}} = \mathbf{T}\Phi \quad (5.25)$$

with:

$$\mathbf{T} = \mathbf{K}^T (\mathbf{K}\mathbf{K}^T + \alpha \mathbf{H})^+ \quad (5.26)$$

The current density estimator in (5. 25) and (5. 26) does not explain the measurements (5. 19) when  $\alpha > 0$ . In the limiting case  $\alpha \rightarrow 0$ , the solution is again the (non-regularized) minimum norm solution.

The main property of the original minimum norm method [21] was illustrated by showing correct, blurred localization of test point sources. The simulations corresponded to MEG sensors distributed on a plane, and with the cortex represented as a square grid of points on a plane located below the sensor plane. The test point source (i.e. the equivalent current dipole) was placed at a cortical voxel, and the theoretical MEG measurements were computed, which were then used in (5. 25) and (5. 26) to obtain the estimated minimum norm current density, which showed maximum activity at the correct location, but with some spatial dispersion. These first results were very encouraging. However, there was one essential omission: the method does not localize deep sources. In a 3D cortex, if the actual source is deep, the method misplaces it to the outermost cortex. The reason for this behavior was explained in Pascual-Marqui 1999 [24], where it was noted that the EEG/MEG minimum norm solution is a harmonic function [54] that can only attain extreme values (maximum activation) at the boundary of the solution space, i.e. at the outermost cortex.

### 5.4.3 Low resolution brain electromagnetic tomography (LORETA)

The discrete, 3D distributed, linear inverse solution that achieved low localization errors (in the sense defined above, by Hämäläinen and Ilmoniemi [21]) even for deep sources, was the method known as LORETA: low resolution electromagnetic tomography [25].

Informally, the basic property of this particular solution is that the current density at any given point on the cortex be maximally similar to the average current density of its neighbors. This “smoothness” property (see e.g. [26, 27]) must hold throughout the entire cortex. Note that the “smoothness” property approximates the electrophysiological constraint under which the EEG is generated: large spatial clusters of cortical pyramidal cells must undergo simultaneously and synchronously the same type of postsynaptic potentials.

The general inverse problem that includes LORETA as a particular case is stated as:

$$\min_{\mathbf{J}} F_w \quad (5.27)$$

with:

$$F_w = \|\Phi - \mathbf{KJ}\|^2 + \alpha \mathbf{J}^T \mathbf{WJ} \quad (5.28)$$

The solution is:

$$\hat{\mathbf{J}}_w = \mathbf{T}_w \Phi \quad (5.29)$$

with the pseudoinverse given by:

$$\mathbf{T}_w = \mathbf{W}^{-1} \mathbf{K}^T (\mathbf{K} \mathbf{W}^{-1} \mathbf{K}^T + \alpha \mathbf{H})^+ \quad (5.30)$$

where the matrix  $\mathbf{W} \in \mathbb{R}^{(3N_V) \times (3N_V)}$  can be tailored to endow the inverse solution with a particular property.

In the case of LORETA, the matrix  $\mathbf{W}$  implements the squared spatial Laplacian operator discrete. In this way, maximally synchronized PSPs at a relatively large macroscopic scale will be enforced. For the sake of simplicity, lead field normalization has not been mentioned in this description, although it is an integral part of the weight matrix used in LORETA. The technical details of the LORETA method can be found in [24, 25].

When LORETA is tested with point sources, low resolution images with very low localization errors are obtained. These results were shown in a non-peer reviewed publication [29] that included discussions with M.S. Hämäläinen, R.J. Ilmoniemi, and P.L. Nunez. The mean localization error of LORETA with EEG was, in the average, only one grid unit, which happened to be three times

smaller than that of the the minimum norm solution. These results were later reproduced and validated by an independent group [30].

It is important to take great care in implementing the Laplacian operator. For instance, Daunizeau and Friston [31] implement the Laplacian operator on a cortical surface consisting of 500 vertices, which are very irregularly sampled, as can be unambiguously appreciated from their Figure 2 in [31]. Obviously, the Laplacian operator is numerically worthless, and yet they conclude rather abusively that: “the LORETA method gave the worst results”. Because their Laplacian is numerically worthless, it is incapable of correctly implementing the smoothness requirement of LORETA. When this is done properly with a regularly sampled solutions space, as in [29, 30], LORETA localizes with very low localization error.

At the moment of this writing, LORETA has been extensively validated, such as in studies combining LORETA with functional MRI (fMRI) [32, 33], with structural MRI [34], and with PET [35]. Further LORETA validation has been based on accepting as ground truth localization findings obtained from invasive implanted depth electrodes, in which case there are several studies, in epilepsy [36, 37, 38,39], and cognitive event related responses [40].

#### 5.4.4 Dynamic statistical parametric maps (dSPM)

The inverse solutions previously described correspond to methods that estimate the electric neuronal activity directly as current density. An alternative approach within the family of “discrete, 3D distributed, linear imaging methods” is to estimate activity as statistically standardized current density.

This approach was introduced by Dale et al in 2000 [41], and is referred to as dynamic statistical parametric maps (dSPM) or noise-normalized current density. The method uses the ordinary minimum norm solution for estimating the current density, as given by equations (5. 25) and (5. 26). The standard deviation of the minimum norm current density is computed by assuming that its variability is exclusively due to noise in the measured EEG.

Let  $\mathbf{S}_{\Phi}^{Noise} \in \mathbb{R}^{N_E \times N_E}$  denote the EEG noise covariance matrix. Then the corresponding current density covariance is:

$$\mathbf{S}_{\mathbf{j}}^{Noise} = \mathbf{T} \mathbf{S}_{\Phi}^{Noise} \mathbf{T}^T \quad (5.31)$$

with  $\mathbf{T}$  given by (5. 26). This result is based on the quadratic nature of the covariance in (5. 31), as derived from the linear transform in (5. 19) (see e.g. Mardia et al 1979 [42]). From (5. 31), let  $[\mathbf{S}_{\mathbf{j}}^{Noise}]_v \in \mathbb{R}^{3 \times 3}$  denote the covariance matrix at voxel  $v$ . Note that this is the  $v$ -th  $3 \times 3$  diagonal block matrix in  $\mathbf{S}_{\mathbf{j}}^{Noise}$ , and it contains current density noise covariance information for all three components of the dipole moment. The noise-normalized imaging method of Dale et al (2000) then gives:

$$\mathbf{q}_v = \frac{\hat{\mathbf{j}}_v}{\sqrt{\text{tr}[\mathbf{S}_{\mathbf{j}}^{Noise}]_v}} \quad (5.32)$$

where  $\hat{\mathbf{j}}_v$  is the minimum norm current density at voxel  $v$ . The squared norm of  $\mathbf{q}_v$ :

$$q_v^t q_v = \frac{\hat{j}_v^T \hat{j}_v}{\text{tr}[\mathbf{S}_{\mathbf{j}}^{Noise}]_v} \quad (5.33)$$

is an  $F$ -distributed statistic.

Note that the noise-normalized method in (5. 32) is a linear imaging method in the case when it uses an estimated EEG noise covariance matrix based on a set measurements that are thought to contain no signal of interest (only noise), and that are independent from the measurements whose generators are sought.

Pascual-Marqui (2002) [43] and Sekihara et al (2005) [44] showed that this method has significant non-zero localization error, even under quasi-ideal conditions of negligible measurement noise.

#### 5.4.4 Standardized low resolution brain electromagnetic tomography (sLORETA)

Another discrete, 3D distributed, linear statistical imaging method is sLORETA: standardized low resolution brain electromagnetic tomography [43]. The basic assumption in this method is that the current density variance receives contributions from possible noise in the EEG measurements, but more importantly, from biological variance, i.e. variance in the actual electric neuronal activity. The biological variance is assumed to be due to electric neuronal activity that is independent and identically distributed all over the cortex, although any other a priori hypothesis can be accommodated. This implies that all the cortex is equally likely to be active. Under this hypothesis, sLORETA produces a linear imaging method that has exact, zero-error localization under ideal conditions, as shown empirically in [43], and theoretically in [44] and [45].

In this case, the covariance matrix for EEG measurements is:

$$\mathbf{S}_{\Phi} = \mathbf{K}\mathbf{S}_j\mathbf{K}^T + \mathbf{S}_{\Phi}^{Noise} \quad (5.34)$$

where  $\mathbf{S}_{\Phi}^{Noise}$  corresponds to noise in the measurements, and  $\mathbf{S}_j$  to the biological source of variability, i.e. the covariance for the current density. When  $\mathbf{S}_j$  is set to the identity matrix, it is equivalent to allowing equal contribution of all cortical neurons to the biological noise. Typically, the covariance of the noise in the measurements  $\mathbf{S}_{\Phi}^{Noise}$  is taken as being proportional to the identity matrix. Under these conditions, the current density covariance given by:

$$\mathbf{S}_j = \mathbf{T}\mathbf{S}_{\Phi}\mathbf{T}^T = \mathbf{T}(\mathbf{K}\mathbf{S}_j\mathbf{K}^T + \mathbf{S}_{\Phi}^{Noise})\mathbf{T}^T = \mathbf{T}(\mathbf{K}\mathbf{K}^T + \alpha\mathbf{H})\mathbf{T}^T = \mathbf{K}^T(\mathbf{K}\mathbf{K}^T + \alpha\mathbf{H})\mathbf{K} \quad (5.35)$$

The sLORETA linear imaging method then is:

$$\sigma_v = [\mathbf{S}_j]_v^{-1/2} \hat{\mathbf{j}}_v \quad (5.36)$$

where  $[\mathbf{S}_j]_v \in \mathbb{R}^{3 \times 3}$  denotes the  $v$ -th  $3 \times 3$  diagonal block matrix in  $\mathbf{S}_j$  (5. 35), and  $[\mathbf{S}_j]_v^{-1/2}$  is its symmetric square root inverse (as in the Mahalanobis transform, see e.g. Mardia et al 1979 [42]). The squared norm of  $\sigma_v$ , i.e.:

$$\sigma_v^T \sigma_v = \hat{j}_v^T [\mathbf{S}_j]_v^{-1} \hat{j}_v \quad (5.37)$$

can be interpreted as a pseudo-statistic with the form of an  $F$ -distribution.

It is worth emphasizing that Sekihara et al [44] and Greenblatt et al [45] showed that sLORETA has no localization bias in the absence of measurement noise; but in the presence of measurement noise, sLORETA has a localization bias. They did not consider the more realistic case where the brain in general is always active, as modeled here by the biological noise. A very recent result (Pascual-Marqui 2007) [46] presents proof that sLORETA has no localization bias under these arguably much more realistic conditions.

### 5.4.5 Exact low resolution brain electromagnetic tomography (eLORETA)

It is likely that the main reason for the development of EEG functional imaging methods in the form of standardized inverse solutions (e.g. dSPM and sLORETA) was that up to very recently all attempts to obtain an actual solution with no localization error have been fruitless. This has been a long-standing goal, as testified by the many publications that endlessly search for an appropriate weight matrix (refer to equations (5. 27) to (5. 30)). For instance, in order to correct for the large depth localization error of the minimum norm solution, one school of thought has been to give more importance (more weight) to deeper sources. A recent version of this method can be found in Lin et al 2006 [47]. That study showed that with the best depth-weighting, the average depth localization error was reduced from 12 mm to 7 mm.

The inverse solution denoted as exact low resolution brain electromagnetic tomography (eLORETA) achieves this goal [46, 48]. In [46], it is shown that eLORETA is a genuine inverse solution, not merely a linear imaging method, and endowed with the property of no localization bias in the presence of measurement and structured biological noise.

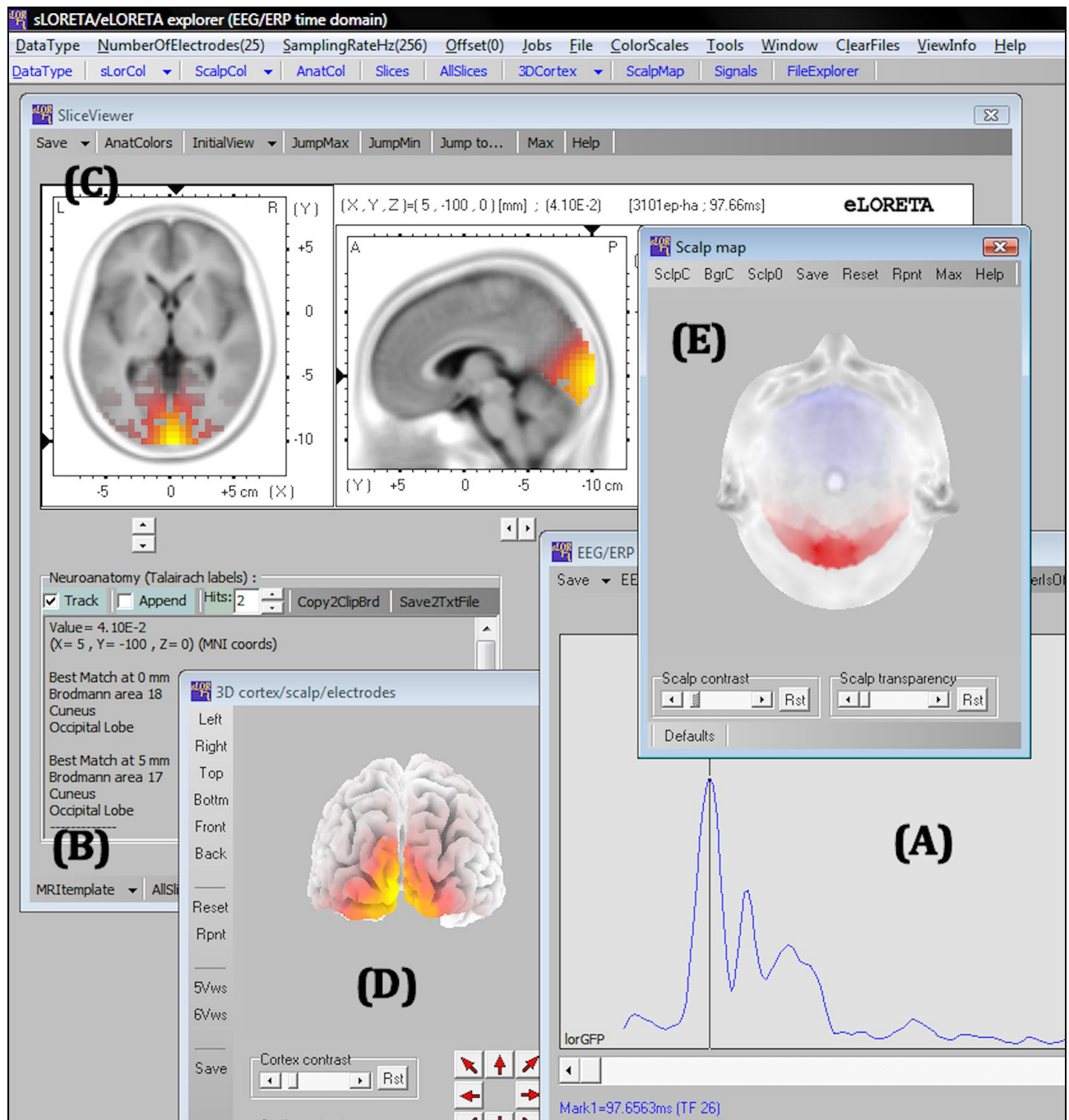
The eLORETA solution is of the weighted type, as given above by equations (5. 27) to (5. 30)). The weight matrix  $\mathbf{W}$  is block diagonal, with sub-blocks of dimension  $3 \times 3$  for each voxel. The eLORETA weights satisfy the system of equations:

$$\mathbf{W}_v = \left[ \mathbf{K}_v^T (\mathbf{K} \mathbf{W}^{-1} \mathbf{K}^T + \alpha \mathbf{H})^+ \mathbf{K}_v \right]^{1/2} \quad (5.38)$$

where  $\mathbf{W}_v \in \mathbb{R}^{3 \times 3}$  is the  $v$ -th diagonal sub-block of  $\mathbf{W}$ .

As shown in [46], eLORETA has no localization bias in the presence of measurement noise and biological noise with variance proportional to  $\mathbf{W}^{-1}$ .

The screenshot in Figure 5.2 shows a practical example for the eLORETA current density inverse solution corresponding to a single-subject visual evoked potential to pictures of flowers. The free academic eLORETA-KEY software and data is publicly available from the appropriate links at homepage of the KEY Institute for Brain-Mind Research, University of Zurich (<http://www.keyinst.uzh.ch>). Maximum total current density power occurs at about 100 ms after stimulus onset (shown in panel A). Current density is color coded, with maximum values represented in bright yellow. Maximum activation is found in Brodmann areas 17 and 18 (panel B). Panel C shows orthogonal slices through the point of maximum current density. Panel D shows a the posterior 3D cortex. Panel E shows the average reference scalp electric potential map, color coded as red for maximum potential, and blue for minimum potential.



**Figure 5.2:** Three-dimensional eLORETA inverse solution displaying estimated current density for a visual evoked potential to pictures of flowers (single subject data). Maximum current density occurs at about 100 ms after stimulus onset (Panel A). Current density is color coded with maximum shown in bright yellow. Maximum activation is found in Brodmann areas 17 and 18 (Panel B). Panel C shows orthogonal slices through the point of maximum activity. Panel D shows a the posterior 3D cortex. Panel E displays the average reference scalp map (positive potentials as red, negative as potentials blue).

#### 5.4.6 Other formulations and methods

There exists a great variety of very fruitful approaches and methods to the inverse EEG problem that lie outside the class of discrete, 3D distributed, linear imaging methods. In what follows, some noteworthy exemplary cases will be mentioned.

The beamformer methods [44, 45, 49, 50] have mostly been employed in MEG studies, but are readily applicable to EEG measurements. Beamformers can be seen as a spatial filtering approach to source localization. Mathematically, the beamformer estimate of activity is based on a weighted sum of the scalp potentials. This might appear to be a linear method, but the weights require and depend on the time-varying EEG measurements themselves, which implies that the method is not a linear one. The method is particularly well suited in the case that the EEG activity is generated by a small number of dipoles, whose time series have low correlation. The method tends to fail in the case of correlated sources. It must also be stressed that this method is an imaging technique that does not estimate the current density, which means that there is no control over how well does the image comply with the actual EEG measurements.

The functionals in equations (5. 24) and (5. 28) have a dual interpretation. On the one hand, they are conventional forms studied in mathematical functional analysis [23]. On the other hand, they can be derived from a Bayesian formulation of the inverse problem [51]. Recently, the Bayesian approach has been used in setting up very complicated and rich forms of the inverse problem, where many conditions can be imposed (in a soft or hard fashion) on the properties of the inverse solution at many levels. An interesting example with many layers of conditions on the solution and its properties can be studied in [52]. In general, this technique does not directly estimate the current density, but rather gives some probability measure of the current density. In addition, these methods are non-linear and are very computer intensive (a problem that is less important with the development of faster CPUs).

Another noteworthy approach to the inverse problem is to consider models that take into account the temporal properties of the current density. If the assumptions on dynamics are correct, the model will very likely perform better than the simple instantaneous models considered in the previous sections. One example of such an approach is [53].

## 5.5 Selecting the inverse solution

We are in a situation where there are many possible tomographies to choose from. The question of selecting the best solution is now essential. For instance:

1. Is there any way to know which method is correct?
2. If we cannot answer the first question, then at least is there any way to know which method is best?

The first question is the most important one, but it is so ill posed that it does not have an answer: there is no way to be certain of the validity of a given solution, unless it is validated by independent methods. This means that the best we can do is to validate the estimated localizations with some “ground truth”, if available.

The second question is also difficult to answer, since there are different criteria for judging the quality of a solution. In Pascual-Marqui [24, 29, 43, 46], the following arguments were used for selecting the “least worst” (as opposed to the possibly non-existent “best”) discrete, 3D distributed, linear tomography:

1. The “least worst” linear tomography is the one with minimum localization error.
2. In a linear tomography, the localization properties can be determined by using test-point sources, based on the principles of linearity and superposition.
3. If a linear tomography is incapable of zero error localization for test point sources that are active one at a time, then the tomography will certainly be incapable of zero error localization to two or more simultaneously active sources.

Based on these criteria, sLORETA and eLORETA are the only linear tomographies that have no localization bias, even under non-ideal conditions of measurement and biological noise.

These criteria are difficult to apply to non-linear methods, for the simple reason that in such a case the principles on linearity and superposition do not hold. Unlike the case of simple linear methods, in the case of non-linear methods, there will remain the uncertainty if the method localizes well in general.

## References

- [1] Berger, H., Über das Elektroencephalogramm des Menschen. Archiv für Psychiatrie und Nervenkrankheit, Vol. 87, 1929, pp. 527-570.
- [2] Helmholtz, H., Ueber einige Gesetze der Vertheilung elektrischer Ströme in körperlichen Leitern, mit Anwendung auf die thierisch-elektrischen Versuche, Ann. Phys. Chem., Vol. 89, 1853, pp. 211-233, 353-377.
- [3] Pascual-Marqui, R.D., and Biscay-Lirio, R., Spatial resolution of neuronal generators based on EEG and MEG measurements, International Journal of Neuroscience, Vol. 68, 1993, pp. 93-105.
- [4] Mitzdorf, U., Current source-density method and application in cat cerebral cortex: investigation of evoked potentials and EEG phenomena. Physiol Rev., Vol. 65, 1985, pp. 37-100.
- [5] Llinas, R.R., The intrinsic electrophysiological properties of mammalian neurons: insights into central nervous system function. Science, Vol. 242, 1988, pp. 1654-1664.
- [6] Martin, J.H., "The collective electrical behavior of cortical neurons: The electroencephalogram and the mechanisms of epilepsy." In Principles of Neural Science, pp 777-791, E.R.Kandel, J.H. Schwartz, and T.M. Jessell (eds.), London: Prentice Hall International, 1991.
- [7] Hämäläinen, M.S., Hari, R., Ilmoniemi, R.J., Knuutila, J., and Lounasmaa, O.V., Magnetoencephalography - theory, instrumentation, and applications to noninvasive studies of the working human brain, Rev. Mod. Phys. Vol. 65, 1993, pp. 413-497.
- [8] Haalman, I., and Vaadia, E., Dynamics of neuronal interactions: relation to behavior, firing rates, and distance between neurons. Human Brain Mapping, Vol. 5, 1997, pp. 249-253.
- [9] Sukov, W., and Barth, D.S., Three-dimensional analysis of spontaneous and thalamically evoked gamma oscillations in auditory cortex, J. Neurophysiol., Vol. 79, 1998, pp: 2875-2884.
- [10] Dale, A.M., Liu, A.K., Fischl, B.R., Buckner, R.L., Belliveau, J.W., Lewine, J.D., and Halgren, E., Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity, Neuron, Vol. 26, 2000, pp. 55-67.
- [11] Baillet, S., Mosher, J.C., and Leahy, R.M., Electromagnetic Brain Mapping. IEEE Signal Processing Magazine, Vol. 18, 2001, pp. 14-30.
- [12] Pascual-Marqui, R.D., Biscay-Lirio, R., and Valdes-Sosa, P.A., "Physical basis of electrophysiological brain imaging: exploratory techniques for source localization and waveshape

- analysis of functional components of electrical brain activity.” In *Machinery of the Mind*, pp. 209-224, E.R. John (ed.), Boston: Birkhäuser, 1990.
- [13] Scherg, M., and von Cramon, D., A new interpretation of the generators of BAEP waves I-V: Results of a spatio-temporal dipole model, *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, Vol. 62, 1985, pp. 290-299.
- [14] Brazier, M. A. B., A study of the electrical fields at the surface of the head, *Electroencephalogr. Clin. Neurophysiol.*, Suppl. 2, 1949, pp. 38-52.
- [15] Wilson, F. N., and Bayley, R. H., The electric field of an eccentric dipole in a homogeneous spherical conducting medium, *Circulation*, 1950, pp: 84-92.
- [16] Frank, E., Electric potential produced by two point sources in a homogeneous conducting sphere, *J. appl. Phys.*, Vol. 23, 1952, pp: 1225-1228.
- [17] Geisler, C.D., and Gerstein, G.L., The surface EEG in relation to its sources, *Electroenceph. Clin. Neurophysiol.* Vol. 13, 1961, pp. 927-934.
- [18] Lehmann, D., Kavanagh, R.N., and Fender, D.H., Field studies of averaged visually evoked EEG potentials in a patient with a split chiasm, *Electroenceph. Clin. Neurophysiol.* Vol. 26, 1969, pp. 193-199.
- [19] Henderson, C. J., Butler, S. R., and Glass, A., The localisation of equivalent dipoles of EEG sources by the application of electrical field theory, *Electroencephalography and Clinical Neurophysiology*, Vol. 39, 1975, pp. 117-130.
- [20] Fuchs, M., Wagner, M., and Kastner, J., Development of volume conductor and source models to localize epileptic foci, *J Clin Neurophysiol* , Vol. 24, 2007, pp. 101-119.
- [21] Härmäläinen, M.S., and Ilmoniemi, R.J., Interpreting measured magnetic fields of the brain: estimates of current distributions, Tech. Rep. TKK-F-A559, Helsinki University of Technology, Espoo, 1984.
- [22] Rao, C.R., and Mitra, S.K., Theory and application of constrained inverse of matrices, *SIAM J. Appl. Math.*, Vol. 24, 1973, pp. 473-488.
- [23] Tikhonov, A., and Arsenin, V., *Solutions to Ill-Posed Problems*, Washington D.C.: Winston, 1977.
- [24] Pascual-Marqui, R.D., Review of methods for solving the EEG inverse problem, *International Journal of Bioelectromagnetism*, Volume 1, 1999, pp. 75-86.
- [25] Pascual-Marqui, R.D., Michel, C.M., and Lehmann, D., Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain, *International Journal of Psychophysiology*, Vol. 18, 1994, pp. 49-65.
- [26] Titterton, D.M., Common structure of smoothing techniques in statistics, *International statistical review*, Vol. 53, 1985, pp. 141-170.
- [27] Wahba, G., *Spline models for observational data*, Philadelphia: SIAM, 1990.

- [28] Sarvas, J., Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem, *Phys. Med. Biol.*, Vol. 32, 1987, pp. 11-22.
- [29] Pascual-Marqui, R.D., "Reply to comments by Hämäläinen, Ilmoniemi and Nunez." In *Source Localization: Continuing Discussion of the Inverse Problem*, pp. 16-28, W. Skrandies (ed.), ISBET Newsletter No.6 (ISSN 0947-5133), 1995.  
(<http://www.uzh.ch/keyinst/NewLORETA/BriefHistory/LORETA-NewsLett2b.pdf>).
- [30] Grave de Pealta, R., Gonsalez, S., Lantz, G., Michels, C.N., and Landis, T., Noninvasive localization of electromagnetic epileptic activity. I. Method descriptions and simulations, *Brain Topography*, Vol. 14, 2001, pp. 131-137.
- [31] Daunizeau, J., and Friston, K.J., A mesostate-space model for EEG and MEG, *NeuroImage*, Vol. 38, 2007, pp. 67-81
- [32] Mulert, C., Jager, L., Schmitt, R., Bussfeld, P., Pogarell, O., Moller, H.J., Juckel, G., and Hegerl, U., Integration of fMRI and simultaneous EEG: towards a comprehensive understanding of localization and time-course of brain activity in target detection, *Neuroimage*, Vol. 22, 2004, pp. 83-94.
- [33] Vitacco, D., Brandeis, D., Pascual-Marqui, R.D., and Martin, E., Correspondence of event-related potential tomography and functional magnetic resonance imaging during language processing, *Hum Brain Mapp*, Vol. 17, 2002, pp. 4-12.
- [34] Worrell, G.A., Lagerlund, T.D., Sharbrough, F.W., Brinkmann, B.H., Busacker, N.E., Cicora, K.M., and O'Brien, T.J., Localization of the epileptic focus by low-resolution electromagnetic tomography in patients with a lesion demonstrated by MRI, *Brain Topography*, Vol. 12, 2000, pp. 273-282.
- [35] Pizzagalli, D.A., Oakes, T.R., Fox, A.S., Chung, M.K., Larson, C.L., Abercrombie, H.C., Schaefer, S.M., Benca, R.M., and Davidson, R.J., Functional but not structural subgenual prefrontal cortex abnormalities in melancholia, *Mol Psychiatry*, Vol. 9, 2004, pp. 393-405.
- [36] Zumsteg, D., Wennberg, R.A., Treyer, V., Buck, A., and Wieser, H.G., H<sub>2</sub>(15)O or 13NH<sub>3</sub> PET and electromagnetic tomography (LORETA) during partial status epilepticus, *Neurology*, Vol. 65, 2005, pp. 1657-1660.
- [37] Zumsteg, D., Lozano, A.M., Wieser, H.G., and Wennberg, R.A., Cortical activation with deep brain stimulation of the anterior thalamus for epilepsy, *Clin Neurophysiol*, Vol. 117, 2006, pp. 192-207.
- [38] Zumsteg, D., Lozano, A.M., and Wennberg, R.A., Depth electrode recorded cerebral responses with deep brain stimulation of the anterior thalamus for epilepsy, *Clin Neurophysiol*, Vol. 117, 2006, pp. 1602-1609.
- [39] Zumsteg, D., Friedman, A., Wieser, H.G., and Wennberg, R.A., Propagation of interictal discharges in temporal lobe epilepsy: correlation of spatiotemporal mapping with intracranial foramenovale electrode recordings, *Clin Neurophysiol*, Vol. 117, 2006, pp. 2615-2626.
- [40] Volpe, U., Mucci, A., Bucci, P., Merlotti, E., Galderisi, S., and Maj, M., The cortical generators of P3a and P3b: a LORETA study, *Brain Res Bull*, Vol. 73, 2007, pp. 220-230.

- [41] Dale, A.M., Liu, A.K., Fischl, B.R., Buckner, R.L., Belliveau, J.W., Lewine, J.D., and Halgren, E., Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity, *Neuron*, Vol. 26, 2000, pp. 55-67.
- [42] Mardia, K.V., Kent, J.T., and Bibby J.M., *Multivariate Analysis*, London: Academic Press, 1979.
- [43] Pascual-Marqui, R.D., Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. *Methods Find. Exp. Clin. Pharmacol.*, Vol. 24 (Suppl D.), 2002, pp. 5-12.
- [44] Sekihara, K., Sahani, M., and Nagarajan, S. S., Localization bias and spatial resolution of adaptive and nonadaptive spatial filters for MEG source reconstruction, *Neuroimage*, Vol. 25, 2005, pp. 1056-1067.
- [45] Greenblatt, R.E., Ossadtchi, A., and Pflieger, M.E., Local Linear Estimators for the Bioelectromagnetic Inverse Problem, *IEEE transactions on signal processing*, Vol. 53, 2005, pp. 3403-3412.
- [46] Pascual-Marqui, R.D., Discrete, 3D distributed, linear imaging methods of electric neuronal activity. Part 1: exact, zero error localization. arXiv:0710.3341 [math-ph], 2007-October-17; <http://arxiv.org/pdf/0710.3341>
- [47] Lin, F.H., Witzel, T., Ahlfors, S.P., Stufflebeam, S.M., Belliveau, J.W., and Hämäläinen, M.S., Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates, *NeuroImage*, Vol. 31, 2006, pp. 160-171.
- [48] Pascual-Marqui, R.D., Pascual-Montano, A.D., Lehmann, D., Kochi, K., Esslen, M., Jancke, L., Anderer, P., Saletu, B., Tanaka, H., Hirata, K., John, E.R., Prichep, L., Exact low resolution brain electromagnetic tomography (eLORETA), *Neuroimage*, Vol. 31, Suppl. 1, 2006, pp:S86.
- [49] Brookes, M.J., Vrba, J., Robinson, S.E., Stevenson, C.M., Peters, A.M., Barnes, G.R., Hillebrand, A., and Morris, P.G., Optimising experimental design for MEG beamformer imaging, *NeuroImage*, Vol. 39, 2008, pp. 1788-1802.
- [50] Van Veen, B.D., Van Drongelen, W., Yuchtman, M., and Suzuki, A., Localisation of brain electrical activity via linearly constrained minimum variance spatial filtering, *IEEE Transa. Biomed. Eng.*, Vol. 44, 1997, pp. 867-880.
- [51] Tarantola, A., *Inverse problem theory and methods for model parameter estimation*, Philadelphia: SIAM, 2005.
- [52] Nummenmaa, A., Auranen, T., Hämäläinen, M.S., Jääskeläinen, I.P., Lampinen, J., Sams, M., and Vehtaria, A., Hierarchical Bayesian estimates of distributed MEG sources: Theoretical aspects and comparison of variational and MCMC methods, *NeuroImage*, Vol. 35, 2007, pp. 669-685.
- [53] Trujillo-Barreto, N.J., Aubert-Vázquez, E. and Penny, W.D., Bayesian M/EEG source reconstruction with spatio-temporal priors, *NeuroImage*, Vol. 39, 2008, pp. 318-335.
- [54] Axler, S., Bourdon, P., and Ramey, W., *Harmonic Function Theory*, New York: Springer-Verlag, 1992.

# Discrete, 3D distributed linear imaging methods of electric neuronal activity. Part 1: exact, zero error localization

Roberto D. Pascual-Marqui  
The KEY Institute for Brain-Mind Research  
University Hospital of Psychiatry  
Lenggstr. 31, CH-8032 Zurich, Switzerland  
pascualm at key.uzh.ch  
www.keyinst.uzh.ch/loreta

## 1. Abstract

This paper deals with the EEG/MEG neuroimaging problem: given measurements of scalp electric potential differences (EEG: electroencephalogram) and extracranial magnetic fields (MEG: magnetoencephalogram), find the 3D distribution of the generating electric neuronal activity. This problem has no unique solution. Only particular solutions with "good" localization properties are of interest, since neuroimaging is concerned with the localization of brain function. In this paper, a general family of linear imaging methods with exact, zero error localization to point-test sources is presented. One particular member of this family is sLORETA (standardized low resolution brain electromagnetic tomography; Pascual-Marqui, Methods Find. Exp. Clin. Pharmacol. 2002, 24D:5-12; <http://www.unizh.ch/keyinst/NewLORETA/sLORETA/sLORETA-Math01.pdf>). It is shown here that sLORETA has no localization bias in the presence of measurement and biological noise. Another member of this family, denoted as eLORETA (exact low resolution brain electromagnetic tomography; Pascual-Marqui 2005: <http://www.research-projects.unizh.ch/p6990.htm>), is a genuine inverse solution (not merely a linear imaging method) with exact, zero error localization in the presence of measurement and structured biological noise. The general family of imaging methods is further extended to include data-dependent (adaptive) quasi-linear imaging methods, also with the exact, zero error localization property.

## 2. The forward equation

Details on the electrophysiology and physics of EEG/MEG generation can be found in Mitzdorf (1985), Llinas (1988), Martin (1991), Hämmäläinen et al (1993), Haalman and Vaadia (1997), Sukov and Barth (1998), Dale et al (2000), Baillet et al. (2001). The basic underlying physics can be studied in Sarvas (1987).

Consider the forward EEG equation:

$$\text{Eq. 1:} \quad \Phi = \mathbf{K}\mathbf{J} + \mathbf{c}\mathbf{1}$$

where the vector  $\Phi \in \mathbb{R}^{N_E \times 1}$  contains instantaneous scalp electric potential differences measured at  $N_E$  electrodes with respect to a single common reference electrode (e.g., the reference can be linked earlobes, the toe, or one of the electrodes included in  $\Phi$ ); the matrix  $\mathbf{K} \in \mathbb{R}^{N_E \times (3N_V)}$  is the lead field matrix corresponding to  $N_V$  voxels;  $\mathbf{J} \in \mathbb{R}^{(3N_V) \times 1}$  is the current

density;  $c$  is a scalar accounting for the physics nature of electric potentials which are determined up to an arbitrary constant; and  $\mathbf{1}$  denotes a vector of ones, in this case  $\mathbf{1} \in \mathbb{R}^{N_E \times 1}$ . Typically  $N_E \ll N_V$ , and  $N_E \geq 19$ .

In Eq. 1, the structure of  $\mathbf{K}$  is:

$$\text{Eq. 2:} \quad \mathbf{K} = \begin{pmatrix} \mathbf{k}_{11}^T & \mathbf{k}_{12}^T & \dots & \mathbf{k}_{1N_V}^T \\ \mathbf{k}_{21}^T & \mathbf{k}_{22}^T & \dots & \mathbf{k}_{2N_V}^T \\ \dots & & & \\ \mathbf{k}_{N_E 1}^T & \mathbf{k}_{N_E 2}^T & \dots & \mathbf{k}_{N_E N_V}^T \end{pmatrix}$$

where the superscript " $T$ " denotes transposition; and  $\mathbf{k}_{ij} \in \mathbb{R}^{3 \times 1}$ , for  $i=1 \dots N_E$  and for  $j=1 \dots N_V$ , corresponds to the scalp potentials at the  $i$ -th electrode due to three orthogonal unit strength dipoles at voxel  $j$ , each one oriented along the coordinate axes  $x$ ,  $y$ , and  $z$ . For instance, in infinite homogeneous medium with conductivity  $\sigma$ :

$$\text{Eq. 3:} \quad \mathbf{k}_{ij} = \frac{1}{4\pi\sigma} \frac{(\mathbf{r}_{Ei} - \mathbf{r}_{Vj})}{\|\mathbf{r}_{Ei} - \mathbf{r}_{Vj}\|^3}$$

where  $\mathbf{r}_{Ei}, \mathbf{r}_{Vj} \in \mathbb{R}^{3 \times 1}$  are position vectors for the  $i$ -th scalp electrode and for the  $j$ -th voxel, respectively. As another example, for the case of a homogeneous conducting sphere in air, the lead field is:

$$\text{Eq. 4:} \quad \mathbf{k}_{ij} = \frac{1}{4\pi\sigma} \left[ 2 \frac{(\mathbf{r}_{Ei} - \mathbf{r}_{Vj})}{\|\mathbf{r}_{Ei} - \mathbf{r}_{Vj}\|^3} + \frac{\mathbf{r}_{Ei} \|\mathbf{r}_{Ei} - \mathbf{r}_{Vj}\| + (\mathbf{r}_{Ei} - \mathbf{r}_{Vj}) \|\mathbf{r}_{Ei}\|}{\|\mathbf{r}_{Ei}\| \|\mathbf{r}_{Ei} - \mathbf{r}_{Vj}\| [\|\mathbf{r}_{Ei}\| \|\mathbf{r}_{Ei} - \mathbf{r}_{Vj}\| + \mathbf{r}_{Ei}^T (\mathbf{r}_{Ei} - \mathbf{r}_{Vj})]} \right]$$

In the previous equations, the following notation was used:

$$\text{Eq. 5:} \quad \|\mathbf{X}\|^2 = \text{tr}(\mathbf{X}^T \mathbf{X}) = \text{tr}(\mathbf{X} \mathbf{X}^T)$$

where  $\text{tr}$  denotes the trace, and  $\mathbf{X}$  is any matrix or vector. If  $\mathbf{X}$  is a vector, then this is the squared Euclidean  $L_2$  norm; if  $\mathbf{X}$  is a matrix, then this is the squared Frobenius norm.

Note that  $\mathbf{K}$  can also be conveniently written as:

$$\text{Eq. 6:} \quad \mathbf{K} = (\mathbf{K}_1, \mathbf{K}_2, \mathbf{K}_3, \dots, \mathbf{K}_{N_V})$$

where  $\mathbf{K}_j \in \mathbb{R}^{N_E \times 3}$ , for  $j=1 \dots N_V$ , is defined as:

$$\text{Eq. 7:} \quad \mathbf{K}_j = \begin{pmatrix} \mathbf{k}_{1j}^T \\ \mathbf{k}_{2j}^T \\ \dots \\ \mathbf{k}_{N_E j}^T \end{pmatrix}$$

Ideally, the lead field should correspond to the real head (with realistic geometry and conductivities). For the EEG problem, the voxels should correspond to cortical grey matter. For other situations (e.g. EKG), appropriate volume conductor models and solution spaces should be used.

In Eq. 1,  $\mathbf{J}$  is structured as:

$$\text{Eq. 8:} \quad \mathbf{J} = \begin{pmatrix} \mathbf{j}_1 \\ \mathbf{j}_2 \\ \dots \\ \mathbf{j}_{N_V} \end{pmatrix}$$

where  $\mathbf{j}_i \in \mathbb{R}^{3 \times 1}$  denotes the current density at the  $i$ -th voxel.

### 3. The reference electrode problem

As a first step, before even stating the inverse problem, the reference electrode problem will be solved, by estimating "c" in Eq. 1. Given  $\Phi$  and  $\mathbf{KJ}$ , the reference electrode problem is:

$$\text{Eq. 9:} \quad \min_c \|\Phi - \mathbf{KJ} - c\mathbf{1}\|^2$$

The solution is:

$$\text{Eq. 10:} \quad c = \frac{\mathbf{1}^T}{\mathbf{1}^T \mathbf{1}} (\Phi - \mathbf{KJ})$$

Plugging Eq. 10 into Eq. 1 gives:

$$\text{Eq. 11:} \quad \mathbf{H}\Phi = \mathbf{H}\mathbf{KJ}$$

where:

$$\text{Eq. 12:} \quad \mathbf{H} = \mathbf{I} - \frac{\mathbf{1}\mathbf{1}^T}{\mathbf{1}^T \mathbf{1}}$$

is the average reference operator, also known as the centering matrix, and  $\mathbf{I} \in \mathbb{R}^{N_E \times N_E}$  is the identity matrix.

This result establishes the fact that any inverse solution (of any form, not necessarily linear) will not depend on the reference electrode.

Henceforth, it will be assumed that the EEG measurements and the lead field are average reference transformed, i.e.:

$$\text{Eq. 13:} \quad \begin{cases} \Phi \leftarrow \mathbf{H}\Phi \\ \mathbf{K} \leftarrow \mathbf{H}\mathbf{K} \end{cases}$$

and Eq. 1 will be rewritten as:

$$\text{Eq. 14:} \quad \Phi = \mathbf{KJ}$$

Note that  $\mathbf{H}$  plays the role of the identity matrix for EEG data. It actually is the identity matrix, except for a null eigenvalue corresponding to an eigenvector of ones (see Eq. 12), accounting for the reference electrode constant.

#### 4. *A family of discrete, 3D distributed linear imaging methods with exact, zero error localization*

The family of linear imaging methods considered here is parameterized by a symmetric matrix  $\mathbf{C} \in \mathbb{R}^{N_E \times N_E}$ , such that:

$$\text{Eq. 15:} \quad \hat{\mathbf{j}}_i = \left[ (\mathbf{K}_i^T \mathbf{C} \mathbf{K}_i)^{-1/2} \mathbf{K}_i^T \mathbf{C} \right] \Phi$$

where  $\hat{\mathbf{j}}_i \in \mathbb{R}^{3 \times 1}$  is any estimator for the electric neuronal activity at the  $i$ -th voxel, not necessarily current density (e.g. it can be standardized current density, as in Pascual-Marqui 2002).

Note that in the case of MEG,  $\mathbf{C}$  must be non-singular. In the case of EEG,  $\mathbf{C}$  must be of rank  $(N_E - 1)$ , with its null eigenvector equal to a vector of ones (accounting for the reference constant).

Note that in Eq. 15, the symmetric matrix  $(\mathbf{K}_i^T \mathbf{C} \mathbf{K}_i)$  is of dimension  $3 \times 3$ , and the notation  $\left[ (\mathbf{K}_i^T \mathbf{C} \mathbf{K}_i)^{-1/2} \right]$  indicates the symmetric square root inverse. In the particular case of MEG in a spherical head model, the matrix  $(\mathbf{K}_i^T \mathbf{C} \mathbf{K}_i)$  is of rank two, and its symmetric square root pseudo-inverse must be used.

Localization inference in neuroimaging is typically based on the search for large values of the power (squared amplitude) of the estimator for electric neuronal activity, i.e.  $\|\hat{\mathbf{j}}_i\|^2$ .

In order to test the localization properties of a linear imaging method, consider the case when the actual source is an arbitrary point-test source at the  $j$ -th voxel. This means that:

$$\text{Eq. 16:} \quad \Phi = \mathbf{K}_j \mathbf{A}$$

where  $\mathbf{K}_j$  is defined in Eq. 7 above, and  $\mathbf{A} \in \mathbb{R}^{3 \times 1}$  is an arbitrary non-zero vector (containing the dipole moments).

Plugging Eq. 16 into Eq. 15 and taking the squared amplitude gives:

$$\text{Eq. 17:} \quad \|\hat{\mathbf{j}}_i\|^2 = \mathbf{A}^T \mathbf{K}_j^T \mathbf{C} \mathbf{K}_i (\mathbf{K}_i^T \mathbf{C} \mathbf{K}_i)^+ \mathbf{K}_i^T \mathbf{C} \mathbf{K}_j \mathbf{A}$$

where the superscript "+" denotes the Moore-Penrose pseudoinverse (which is equal to the common inverse if the matrix is non-singular).

Following the same type of derivations as in Greenblatt et al (2005), the derivative of  $\|\hat{\mathbf{j}}_i\|^2$  in Eq. 17 with respect to  $\mathbf{K}_i$  is:

$$\text{Eq. 18: } \left\{ \begin{aligned} \frac{\partial \|\hat{\mathbf{j}}_i\|^2}{\partial \mathbf{K}_i} &= 2\mathbf{C}\mathbf{K}_j\mathbf{A}\mathbf{A}^T\mathbf{K}_j^T\mathbf{C}\mathbf{K}_i(\mathbf{K}_i^T\mathbf{C}\mathbf{K}_i)^+ \\ &\quad - 2\mathbf{C}\mathbf{K}_i(\mathbf{K}_i^T\mathbf{C}\mathbf{K}_i)^+ \mathbf{K}_i^T\mathbf{C}\mathbf{K}_j\mathbf{A}\mathbf{A}^T\mathbf{K}_j^T\mathbf{C}\mathbf{K}_i(\mathbf{K}_i^T\mathbf{C}\mathbf{K}_i)^+ \end{aligned} \right\}$$

It can be easily shown that this derivative is zero when  $\mathbf{K}_i$  is equal to  $\mathbf{K}_j$ , demonstrating that this family of methods produces exactly localized maxima to point-test sources anywhere in the brain, i.e. this family of linear imaging methods attains exact, zero error localization.

Note that the choice:

$$\text{Eq. 19: } \mathbf{C} = (\mathbf{K}\mathbf{K}^T + \alpha\mathbf{H})^+$$

gives the sLORETA method (Pascual-Marqui 2002), where  $\alpha \geq 0$  is the regularization parameter.

Note that these results can be applied in a straightforward manner to the case where the current density orientation is known (i.e. known cortical geometry), but with unknown current density amplitude.

## 5. Unbiased localization for sLORETA

As in the previous section, consider the case when the actual source is any arbitrary point-test source at the  $j$ -th voxel, but now the measurements are contaminated with measurement and biological noise. This means that:

$$\text{Eq. 20: } \Phi = \mathbf{K}_j\mathbf{A} + \epsilon_\Phi + \mathbf{K}\epsilon_j$$

where  $\epsilon_\Phi$  represents the measurement noise and  $\epsilon_j$  the biological noise. It will be assumed that both noise sources are zero mean and independent, with covariance matrices:

$$\text{Eq. 21: } \text{cov}(\epsilon_\Phi) = \sigma_\Phi\mathbf{H}$$

$$\text{Eq. 22: } \text{cov}(\epsilon_j) = \sigma_j\mathbf{I}$$

This gives the following expected covariance matrix for the measurements:

$$\text{Eq. 23: } \text{cov}(\Phi) = \Sigma_\Phi = \mathbf{K}_j\mathbf{A}\mathbf{A}^T\mathbf{K}_j^T + \sigma_\Phi\mathbf{H} + \sigma_j\mathbf{K}\mathbf{K}^T$$

The corresponding expected square amplitude then is:

$$\text{Eq. 24: } \left\{ \begin{aligned} E\left(\|\hat{\mathbf{j}}_i\|^2\right) &= \text{tr}\left[\left(\mathbf{K}_i^T\mathbf{C}\mathbf{K}_i\right)^{-1/2} \mathbf{K}_i^T\mathbf{C}\left(\mathbf{K}_j\mathbf{A}\mathbf{A}^T\mathbf{K}_j^T + \sigma_\Phi\mathbf{H} + \sigma_j\mathbf{K}\mathbf{K}^T\right)\mathbf{C}\mathbf{K}_i\left(\mathbf{K}_i^T\mathbf{C}\mathbf{K}_i\right)^{-1/2}\right] \\ &= \text{tr}\left[\left(\mathbf{K}_j\mathbf{A}\mathbf{A}^T\mathbf{K}_j^T + \sigma_\Phi\mathbf{H} + \sigma_j\mathbf{K}\mathbf{K}^T\right)\mathbf{C}\mathbf{K}_i\left(\mathbf{K}_i^T\mathbf{C}\mathbf{K}_i\right)^+ \mathbf{K}_i^T\mathbf{C}\right] \end{aligned} \right\}$$

The derivative of  $E\left(\left\|\hat{\mathbf{j}}_i\right\|^2\right)$  in Eq. 24 with respect to  $\mathbf{K}_i$  is:

$$\text{Eq. 25: } \left\{ \begin{aligned} \frac{\partial E\left(\left\|\hat{\mathbf{j}}_i\right\|^2\right)}{\partial \mathbf{K}_i} &= 2\mathbf{C}\left(\mathbf{K}_j\mathbf{A}\mathbf{A}^T\mathbf{K}_j^T + \sigma_{\Phi}\mathbf{H} + \sigma_j\mathbf{K}\mathbf{K}^T\right)\mathbf{C}\mathbf{K}_i\left(\mathbf{K}_i^T\mathbf{C}\mathbf{K}_i\right)^+ \\ &\quad - 2\mathbf{C}\mathbf{K}_i\left(\mathbf{K}_i^T\mathbf{C}\mathbf{K}_i\right)^+\mathbf{K}_i^T\mathbf{C}\left(\mathbf{K}_j\mathbf{A}\mathbf{A}^T\mathbf{K}_j^T + \sigma_{\Phi}\mathbf{H} + \sigma_j\mathbf{K}\mathbf{K}^T\right)\mathbf{C}\mathbf{K}_i\left(\mathbf{K}_i^T\mathbf{C}\mathbf{K}_i\right)^+ \end{aligned} \right\}$$

It can be easily shown that the derivative in Eq. 25 is zero for the sLORETA case, when the parameter matrix is:

$$\text{Eq. 26: } \mathbf{C} = \left( \mathbf{K}\mathbf{K}^T + \frac{\sigma_{\Phi}}{\sigma_j}\mathbf{H} \right)^+$$

and when  $\mathbf{K}_i$  is equal to  $\mathbf{K}_j$ , thus demonstrating that sLORETA produces exactly localized maxima to point-test sources anywhere in the brain, even in the presence of noise, i.e. sLORETA is unbiased.

This new result is to be contrasted with those published by Sekihara et al (2005) and Greenblatt et al (2005). They showed that under pure measurement noise, sLORETA is biased, and only attains exact localization under ideal conditions of no noise. They did not consider the more realistic case where the brain in general is always active, as modeled here by the biological noise. Under these arguably much more realistic conditions, sLORETA is unbiased.

## 6. eLORETA: exact low resolution brain electromagnetic tomography

The eLORETA method was developed and officially recorded as a working project at the University of Zurich in March 2005. A description (including the official registration date) can be obtained from the University of Zurich server at: <http://www.research-projects.unizh.ch/p6990.htm>

An additional reference to eLORETA is:

Roberto D. Pascual-Marqui, Alberto D. Pascual-Montano, Dietrich Lehmann, Kieko Kochi, Michaela Esslen, Lutz Jancke, Peter Anderer, Bernd Saletu, Hideaki Tanaka, Koichi Hirata, E. Roy John, Leslie Prichep. Exact low resolution brain electromagnetic tomography (eLORETA). Neuroimage 2006, Vol 31, Suppl. 1, page:S86

Consider the general weighted minimum norm solution (see, e.g. Pascual-Marqui 1999):

$$\text{Eq. 27: } \hat{\mathbf{j}} = \mathbf{T}\Phi$$

with:

$$\text{Eq. 28: } \mathbf{T} = \mathbf{W}^{-1}\mathbf{K}^T\left(\mathbf{K}\mathbf{W}^{-1}\mathbf{K}^T + \alpha\mathbf{H}\right)^+$$

where  $\mathbf{W} \in \mathbb{R}^{(3N_V) \times (3N_V)}$  denotes the symmetric weight matrix, and  $\alpha \geq 0$  denotes the regularization parameter.

The particular case of interest here will only consider a structured block-diagonal weight matrix  $\mathbf{W}$ , where all matrix elements are zero except for the diagonal sub-blocks denoted as  $\mathbf{W}_i \in \mathbb{R}^{3 \times 3}$ , the  $i$ -th voxel, with  $i = 1 \dots N_V$ .

Note that for  $\alpha = 0$ , this is a genuine solution, in the sense that  $\hat{\mathbf{J}}$  is a direct estimator for the current density, and it reproduces exactly the measurements. In other words, for  $\alpha = 0$ :

$$\text{Eq. 29:} \quad \begin{cases} \Phi = \mathbf{K}\hat{\mathbf{J}} = \mathbf{K}\mathbf{T}\Phi \\ \mathbf{H} = \mathbf{K}\mathbf{T} \end{cases}$$

The current density estimator at the  $i$ -th voxel then is:

$$\text{Eq. 30:} \quad \hat{\mathbf{j}}_i = \mathbf{W}_i^{-1} \mathbf{K}_i^T (\mathbf{K}\mathbf{W}^{-1} \mathbf{K}^T + \alpha \mathbf{H})^+ \Phi$$

Based on the results of the previous section (entitled "A family of discrete, 3D distributed linear imaging methods with exact, zero error localization"), by comparing Eq. 30 with Eq. 15, exact, zero error localization is attained with weights satisfying:

$$\text{Eq. 31:} \quad \mathbf{W}_i = \left[ \mathbf{K}_i^T (\mathbf{K}\mathbf{W}^{-1} \mathbf{K}^T + \alpha \mathbf{H})^+ \mathbf{K}_i \right]^{1/2}$$

This result is easily derived by noting that Eq. 30 matches Eq. 15 when:

$$\text{Eq. 32:} \quad \mathbf{C} \leftarrow (\mathbf{K}\mathbf{W}^{-1} \mathbf{K}^T + \alpha \mathbf{H})^+$$

and:

$$\text{Eq. 33:} \quad \mathbf{W}_i^{-1} \leftarrow (\mathbf{K}_i^T \mathbf{C} \mathbf{K}_i)^{-1/2}$$

The weights satisfying the system of equations given by Eq. 31 define the eLORETA method, which is a genuine solution to the inverse problem (not merely a linear imaging method), and attains exact, zero error localization. Additionally, eLORETA is standardized by definition, meaning that its theoretical expected variance is unity.

Furthermore, following the derivations as in the previous section entitled "Unbiased localization for sLORETA", it can easily be shown that eLORETA is unbiased in the presence of measurement and structured biological noise of the form:

$$\text{Eq. 34:} \quad \text{cov}(\varepsilon_J) = \sigma_J \mathbf{W}^{-1}$$

Unfortunately, such a structure on background brain activity (the so-called biological noise) is determined by the physics properties of the head model and the laws of electrodynamics, and might have little relation to electrophysiological reality. This might be seen as a disadvantage of eLORETA as compared to sLORETA.

## 7. An alternative theoretical approach to eLORETA, including numerical methods

### 7.1. The classical weighted minimum norm tomography

Consider the regularized, weighted minimum norm problem:

$$\text{Eq. 35:} \quad \min_{\mathbf{J}} \left[ \|\Phi - \mathbf{KJ}\|^2 + \alpha \mathbf{J}^T \mathbf{WJ} \right]$$

where  $\mathbf{W} \in \mathbb{R}^{(3N_V) \times (3N_V)}$  denotes a given symmetric weight matrix, and  $\alpha \geq 0$  denotes the regularization parameter.

The solution is linear:

$$\text{Eq. 36:} \quad \hat{\mathbf{J}} = \mathbf{T}\Phi$$

with:

$$\text{Eq. 37:} \quad \mathbf{T} = \mathbf{W}^{-1} \mathbf{K}^T (\mathbf{KW}^{-1} \mathbf{K}^T + \alpha \mathbf{H})^+$$

where the superscript "+" denotes the Moore-Penrose pseudoinverse (which is equal to the common inverse if the matrix is non-singular).

The choice  $\mathbf{W} = \mathbf{I}$  gives the classical minimum norm solution. This was the first 2D distributed linear solution introduced in MEG by Hämäläinen and Ilmoniemi (1984). Some of the images in that publication show that when the solution space is parallel to the measurement space, point-test sources are correctly localized, albeit with low resolution.

However, when the solution space is extended to 3D, the minimum norm solution is utterly incapable of correct localization of depth. This was clarified in Pascual-Marqui (1999), where it was shown that the minimum norm solution is harmonic, and harmonic functions attain their extreme values on the boundary of their domain of definition. This means that deep sources are always incorrectly localized to the outermost cortex.

Another popular choice is depth weighting for the 3D solution space, i.e. larger weights are assigned to deeper sources, with the hope of correcting depth localization error. These solutions achieve lower localization error than the classical minimum norm, but their errors are still significant, no matter what inverse power for depth weighting is used.

The weighted minimum norm method that uses combined depth weighting and Laplacian smoothing, known as LORETA (low resolution brain electromagnetic tomography; Pascual-Marqui et al 1994), achieved the lowest localization error up to the present, among linear solutions. Yet, the method has non-zero error, but quite lower than the two previous methods.

## 7.2. eLORETA: optimal weights that produce exact localization

The regularized problem in Eq. 35 was presented from a "functional analysis" point of view. Alternatively, a Bayesian point of view renders the same formulation, where the quadratic functional in Eq. 35 is part of the posterior density, with:

Eq. 38:  $\Sigma_{\Phi}^{noise} = \alpha H$

being the covariance matrix for the noise in the measurements, and:

Eq. 39:  $\Sigma_J = W^{-1}$

being the "a priori" covariance matrix for the current density  $J$ .

Based on the linear relation in Eq. 14, extending it to include possible additive noise in the measurements, making use of Eq. 38 and Eq. 39, and assuming independence of neuronal activity and measurement noise, the covariance matrix for the electric potential is:

Eq. 40:  $\Sigma_{\Phi} = K \Sigma_J K^T + \Sigma_{\Phi}^{noise} = K W^{-1} K^T + \alpha H$

Based on the linear relation in Eq. 36, and making use of Eq. 40, the covariance matrix for the estimated current density is:

Eq. 41: 
$$\left\{ \begin{array}{l} \Sigma_J = T \Sigma_{\Phi} T^T \\ = T (K W^{-1} K^T + \alpha H) T^T \\ = W^{-1} K^T (K W^{-1} K^T + \alpha H)^+ (K W^{-1} K^T + \alpha H) (K W^{-1} K^T + \alpha H)^+ K W^{-1} \\ = W^{-1} K^T (K W^{-1} K^T + \alpha H)^+ K W^{-1} \end{array} \right\}$$

When  $W$  is restricted to be a block-diagonal matrix, with the  $j$ -th block denoted as  $W_j \in \mathbb{R}^{3 \times 3}$ , for  $j = 1 \dots N_V$ , then the solution to the problem:

Eq. 42:  $\min_W \|I - \Sigma_J\|^2 = \min_W \|I - W^{-1} K^T (K W^{-1} K^T + \alpha H)^+ K W^{-1}\|^2$

produces an inverse solution (Eq. 36 and Eq. 37) with zero localization error.

Zero localization error is defined in this study as follows: For a given point-test source anywhere in the solution space, with arbitrary orientation, compute the extracranial EEG/MEG measurements, give them to the linear inverse solution, threshold the inverse solution to the absolute maximum of the amplitude of the current density vector field, and compute as localization error the distance between the actual point-test source and the position of the absolute maximum.

This property has not been achieved by any previously published discrete 3D distributed linear solution.

Note that the covariance matrix for the estimated current density (Eq. 41) is not the resolution matrix of Backus and Gilbert.

The solution to the problem in Eq. 42 satisfies the following set of matrix equations:

Eq. 43:  $\mathbf{W}_j^2 = \mathbf{K}_j^T (\mathbf{K} \mathbf{W}^{-1} \mathbf{K}^T + \alpha \mathbf{H})^+ \mathbf{K}_j$ , for  $j = 1 \dots N_V$

where the matrix  $\mathbf{K}_j$  is defined in Eq. 7.

The following simple iterative algorithm (in pseudo-code) converges to the block-diagonal weights  $\mathbf{W}$  that solve the problem in Eq. 42 and equivalently satisfies Eq. 43:

1. Given the average reference lead field  $\mathbf{K}$  and a regularization parameter  $\alpha \geq 0$ , initialize the block-diagonal weight matrix  $\mathbf{W}$  as the identity matrix.
2. Set:  
Eq. 44:  $\mathbf{M} = (\mathbf{K} \mathbf{W}^{-1} \mathbf{K}^T + \alpha \mathbf{H})^+$
3. For  $j = 1 \dots N_V$  do:  
Eq. 45:  $\mathbf{W}_j = [\mathbf{K}_j^T \mathbf{M} \mathbf{K}_j]^{SymmSqrt}$   
Comment:  $[\mathbf{K}_j^T \mathbf{M} \mathbf{K}_j]^{SymmSqrt}$  denotes the symmetric square root of the matrix  $[\mathbf{K}_j^T \mathbf{M} \mathbf{K}_j]$ .
4. Go to step 2 until convergence (negligible changes in  $\mathbf{W}$ ).

Finally, the block-diagonal matrix  $\mathbf{W}$  produced by this algorithm should be plugged into the pseudoinverse matrix  $\mathbf{T}$  (in Eq. 37). This is denoted as the eLORETA inverse solution.

### 7.3. eLORETA for EEG with known current density vector orientation, unknown amplitude

The average reference forward EEG equation (Eq. 14) is now written as:

Eq. 46:  $\Phi = \mathbf{K} \mathbf{J} = \mathbf{K} \mathbf{N} \mathbf{L}$

with:

Eq. 47:  $\mathbf{J} = \mathbf{N} \mathbf{L}$

where  $\mathbf{L} \in \mathbb{R}^{N_V \times 1}$  contains the current density amplitudes at each voxel, and  $\mathbf{N} \in \mathbb{R}^{(3N_V) \times N_V}$  contains the outward normal vectors to the cortical surface at each voxel. Note that the columns of  $\mathbf{N}$ , denoted as  $\mathbf{N}_j \in \mathbb{R}^{(3N_V) \times 1}$  for  $j = 1 \dots N_V$  are:

Eq. 48:  $\mathbf{N}_1 = \begin{pmatrix} \mathbf{n}_1 \\ \mathbf{o} \\ \mathbf{o} \\ \mathbf{o} \\ \dots \\ \mathbf{o} \end{pmatrix}; \mathbf{N}_2 = \begin{pmatrix} \mathbf{o} \\ \mathbf{n}_2 \\ \mathbf{o} \\ \mathbf{o} \\ \dots \\ \mathbf{o} \end{pmatrix}; \mathbf{N}_3 = \begin{pmatrix} \mathbf{o} \\ \mathbf{o} \\ \mathbf{n}_3 \\ \mathbf{o} \\ \dots \\ \mathbf{o} \end{pmatrix}; \dots; \mathbf{N}_{N_V} = \begin{pmatrix} \mathbf{o} \\ \mathbf{o} \\ \mathbf{o} \\ \mathbf{o} \\ \dots \\ \mathbf{n}_{N_V} \end{pmatrix}$

where  $\mathbf{o} \in \mathbb{R}^{3 \times 1}$  is a vector of zeros, and  $\mathbf{n}_j \in \mathbb{R}^{3 \times 1}$  is the normal vector at the  $j$ -th voxel, i.e.:

Eq. 49:  $\mathbf{n}_j^T \mathbf{n}_j = 1$

In this section,  $\mathbf{N}$  is assumed to be known.

The regularized, weighted minimum norm problem is:

$$\text{Eq. 50:} \quad \min_{\mathbf{L}} \left[ \|\Phi - \mathbf{KNL}\|^2 + \alpha \mathbf{L}^T \mathbf{W} \mathbf{L} \right]$$

where  $\mathbf{W} \in \mathbb{R}^{N_V \times N_V}$  in this case denotes a given symmetric weight matrix, and  $\alpha > 0$  denotes the regularization parameter.

The solution is linear:

$$\text{Eq. 51:} \quad \hat{\mathbf{L}} = \mathbf{T} \Phi$$

with:

$$\text{Eq. 52:} \quad \mathbf{T} = \mathbf{W}^{-1} (\mathbf{KN})^T \left( (\mathbf{KN}) \mathbf{W}^{-1} (\mathbf{KN})^T + \alpha \mathbf{H} \right)^+$$

Following similar lines of reasoning as in the previous section, the covariance matrix for the electric potential is:

$$\text{Eq. 53:} \quad \Sigma_{\Phi} = (\mathbf{KN}) \Sigma_{\mathbf{L}} (\mathbf{KN})^T + \Sigma_{\Phi}^{\text{noise}} = (\mathbf{KN}) \mathbf{W}^{-1} (\mathbf{KN})^T + \alpha \mathbf{H}$$

where:

$$\text{Eq. 54:} \quad \Sigma_{\mathbf{L}} = \mathbf{W}^{-1}$$

is the "a priori" covariance matrix for the current density amplitudes  $\mathbf{L}$ . In addition, the covariance matrix for the estimated current density is:

$$\text{Eq. 55:} \quad \Sigma_{\hat{\mathbf{L}}} = \mathbf{W}^{-1} (\mathbf{KN})^T \left( (\mathbf{KN}) \mathbf{W}^{-1} (\mathbf{KN})^T + \alpha \mathbf{H} \right)^+ (\mathbf{KN}) \mathbf{W}^{-1}$$

When  $\mathbf{W}$  is restricted to be a diagonal matrix, with the  $j$ -th element denoted as  $W_j$ , for  $j = 1 \dots N_V$ , then the solution to the problem:

$$\text{Eq. 56:} \quad \min_{\mathbf{W}} \|\mathbf{I} - \Sigma_{\hat{\mathbf{L}}}\|^2 = \min_{\mathbf{W}} \left\| \mathbf{I} - \mathbf{W}^{-1} (\mathbf{KN})^T \left( (\mathbf{KN}) \mathbf{W}^{-1} (\mathbf{KN})^T + \alpha \mathbf{H} \right)^+ (\mathbf{KN}) \mathbf{W}^{-1} \right\|^2$$

produces an inverse solution (Eq. 51 and Eq. 52) with zero localization error.

The solution to the problem in Eq. 56 satisfies the following set of equations:

$$\text{Eq. 57:} \quad W_j = \sqrt{(\mathbf{KN})_j^T (\mathbf{KW}^{-1} \mathbf{K}^T + \alpha \mathbf{H})^+ (\mathbf{KN})_j}, \text{ for } j = 1 \dots N_V$$

where the vector  $(\mathbf{KN})_j \in \mathbb{R}^{N_E \times 1}$  corresponds to the  $j$ -th column of  $(\mathbf{KN})$ .

The following simple iterative algorithm (in pseudo-code) converges to the diagonal weights  $\mathbf{W}$  that solve the problem in Eq. 56 and equivalently satisfies Eq. 57:

1. Given the average reference lead field  $\mathbf{K}$ , the cortical normal vectors  $\mathbf{N}$ , and a regularization parameter  $\alpha \geq 0$ , initialize the diagonal weight matrix  $\mathbf{W}$  as the identity matrix.
2. Set:
 
$$\text{Eq. 58:} \quad \mathbf{M} = \left( (\mathbf{KN}) \mathbf{W}^{-1} (\mathbf{KN})^T + \alpha \mathbf{H} \right)^+$$
3. For  $j = 1 \dots N_V$  do:
 
$$\text{Eq. 59:} \quad W_j = \sqrt{(\mathbf{KN})_j^T \mathbf{M} (\mathbf{KN})_j}$$
4. Go to step 2 until convergence (negligible changes in  $\mathbf{W}$ ).

Finally, the diagonal matrix  $\mathbf{W}$  produced by this algorithm should be plugged into the pseudoinverse matrix  $\mathbf{T}$  (in Eq. 52). This is denoted as the eLORETA inverse solution.

#### 7.4. eLORETA for MEG with fully unknown current density vector field

This case follows the same derivations as given above for the case "EEG with fully unknown current density vector field".

The forward MEG equation has similar form to Eq. 14. For the MEG case,  $\Phi$  would represent the magnetometer or gradiometer measurements,  $\mathbf{K}$  would represent the magnetic lead field, and  $\mathbf{J}$  is exactly the same current density vector field (common to both EEG and MEG).

In the MEG case, there is no reference electrode constant to be accounted for. The consequence is that the EEG regularization term ( $\alpha\mathbf{H}$ ) appearing in some of the equations above (Eq. 37 to Eq. 44) must be changed to the MEG regularization term ( $\alpha\mathbf{I}$ ), where  $\mathbf{I}$  is the identity matrix.

In the case of spherical head models, care must be taken in the MEG case because only the tangential part of the current density vector field is non-silent. The same occurs in realistic head models, in areas that are quasi-spherical. This implies that all calculations at the voxel level have only rank=2 for MEG. Therefore, inverse and symmetric square-root matrix computations should be made via the singular value decomposition (SVD), ignoring the smallest eigenvalue if it is numerically negligible relative to the largest eigenvalue.

In particular, consider the algorithm involving Eq. 44 and Eq. 45. Note that Eq. 44 makes use of the inverse of the weight matrix, which consists of the inverses of all  $3 \times 3$  block-diagonal submatrices. In the quasi-spherical MEG case, these submatrices have rank=2. Referring to Eq. 45, consider the SVD of the matrix of interest:

$$\text{Eq. 60:} \quad [\mathbf{K}_j^T \mathbf{M} \mathbf{K}_j] = \sum_{i=1}^3 \lambda_i \Gamma_i \Gamma_i^T$$

where  $\Gamma_i \in \mathbb{R}^{3 \times 1}$  are the orthonormal eigenvectors, and  $\lambda_1 \geq \lambda_2 \geq \lambda_3$  are the eigenvalues. Then Eq. 45 should be replaced by:

$$\text{Eq. 61:} \quad \mathbf{W}_j = [\mathbf{K}_j^T \mathbf{M} \mathbf{K}_j]^{\text{SymmSqrt}} = \begin{cases} \sum_{i=1}^2 \sqrt{\lambda_i} \Gamma_i \Gamma_i^T, & \text{if } (\lambda_3 / \lambda_1) < \varepsilon \\ \sum_{i=1}^3 \sqrt{\lambda_i} \Gamma_i \Gamma_i^T, & \text{otherwise} \end{cases}$$

where  $\varepsilon$  depends on the numerical precision of the calculations (typically  $\varepsilon \leq 10^{-5}$ ).

Moreover, the inverse of  $\mathbf{W}_j$  (see Eq. 61), which is needed in Eq. 44, and later on after convergence in Eq. 37 for the final inverse solution, should be calculated as the Moore-Penrose pseudoinverse (ignoring the smallest eigenvalue if it is numerically negligible relative to the largest eigenvalue).

Given these provisions and modifications, the discrete 3D distributed linear solution known as eLORETA is given by Eq. 36 and Eq. 37 with the weights defined by the solution the problem in Eq. 42, obtained with the algorithm specified by Eq. 44 and Eq. 45.

### 7.5. **eLORETA for MEG with known current density vector orientation, unknown amplitude**

This case follows the same derivations as given above for the case "EEG with known current density vector orientation, unknown amplitude".

The forward MEG equation in this case has similar form to Eq. 46. For the MEG case,  $\Phi$  would represent the magnetometer or gradiometer measurements,  $\mathbf{K}$  would represent the magnetic lead field,  $\mathbf{N}$  would contain the outward normal vectors to the cortical surface at each voxel (assumed known), and  $\mathbf{L}$  contains exactly the same current density amplitudes (common to both EEG and MEG).

As explained previously, the EEG regularization term ( $\alpha\mathbf{H}$ ) appearing in some of the equations above (Eq. 52 to Eq. 58) must be changed to the MEG regularization term ( $\alpha\mathbf{I}$ ), where  $\mathbf{I}$  is the identity matrix.

The existence of silent MEG sources might occur in practice, especially for quasi-radial sources in quasi-spherical head geometry. Care should be taken to exclude these possible silent sources from the solution space, even if this implies that there are missing cortical patches for the MEG solution space.

Given these provisions and modifications, the discrete 3D distributed linear solution known as eLORETA is given by Eq. 51 and Eq. 52 with the weights defined by the solution to the problem in Eq. 56, obtained with the algorithm specified by Eq. 58 and Eq. 59.

The general family of linear imaging methods is further extended to include data-dependent (adaptive) quasi-linear imaging methods, also with the exact, zero error localization property.

## 8. ***A family of discrete, 3D distributed quasi-linear imaging methods with exact, zero error localization: data-dependent (adaptive) methods***

Formally, this family of methods is identical to the one defined by Eq. 15, except that now, instead of defining the parameter matrix  $\mathbf{C}$  as a given, fixed matrix, it can, for example, be taken as the inverse covariance matrix for the measurements, i.e:

$$\text{Eq. 62:} \quad \mathbf{C} = \left[ \frac{1}{N_K} \sum_{k=1}^{N_K} (\Phi_k - \bar{\Phi})(\Phi_k - \bar{\Phi})^T \right]^+$$

where the subscript "k" may index time or any other form of repeated measurements for the data. Note that in the case of MEG,  $\mathbf{C}$  must be non-singular. In the case of EEG,  $\mathbf{C}$  must be of rank  $(N_E - 1)$ , with its null eigenvector equal to a vector of ones (accounting for the

reference constant). If the data so happens to be insufficient, i.e.  $N_K \leq N_E$ , or the data happens to be almost deterministic, resulting in a low rank matrix  $\mathbf{C}$ , then the method will not have exact, zero error localization.

Eq. 62 corresponds to a single example illustrating the "adaptive" character of this family of methods. Any data dependent matrix  $\mathbf{C}$  can be used, such as, for example, the squared inverse covariance matrix for the measurements.

Rigorously speaking, this method is not linear because the transformation depends on the data on which imaging is being carried out.

## 9. Conclusions

In Pascual-Marqui (1995, 1999, and 2002), the following arguments were used for selecting the best discrete, 3D distributed, linear tomography:

1. The aim of functional imaging is localization. Therefore, the best tomography is the one with minimum localization error.
2. In a linear tomography, the localization properties can be determined by using point-test sources, based on the principles of linearity and superposition.
3. If a linear tomography is incapable of zero error localization to point-test sources that are active one at a time, then the tomography will certainly be incapable of zero error localization to two or more simultaneously active sources.

Here we present a general family of linear imaging methods with exact, zero error localization to point-test sources.

We show that one particular member of this family, sLORETA (Pascual-Marqui 2002) has no localization bias in the presence of measurement noise and biological noise.

We introduce a new particular member of this family, denoted eLORETA. This is a genuine inverse solution and not merely a linear imaging method. We show that it has exact, zero error localization in the presence of measurement and structured biological noise. We derive and construct the method using two different approaches, and give practical algorithms for its estimation.

We present a general family of quasi-linear imaging methods that are data-dependent (adaptive). We also show that they are endowed with the exact, zero error localization property.

These results are expected to be of value to the EEG/MEG neuroimaging community.

## 10. References

- Baillet S, Mosher JC, Leahy RM: Electromagnetic Brain Mapping. IEEE Signal Processing Magazine 18:14-30, 2001.
- Dale AM, Liu AK, Fischl BR, Buckner RL, Belliveau JW, Lewine JD, Halgren E: Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. Neuron 26: 55-67, 2000.
- Greenblatt, R.E. Ossadtchi, A. Pflieger, M.E. Local Linear Estimators for the Bioelectromagnetic Inverse Problem. IEEE transactions on signal processing 2005, 53: 3403-3412.
- Haalman I, Vaadia E: Dynamics of neuronal interactions: relation to behavior, firing rates, and distance between neurons. Human Brain Mapping 5: 249-253, 1997.
- Hämäläinen, M.S., and Ilmoniemi, R.J. Interpreting measured magnetic fields of the brain: estimates of current distributions. Tech. Rep. TKK-F-A559, Helsinki University of Technology, Espoo, 1984.
- Hämäläinen MS, Hari R, Ilmoniemi RJ, Knuutila J, Lounasmaa OV: Magnetoencephalography - theory, instrumentation, and applications to noninvasive studies of the working human brain. Rev. Mod. Phys. 65: 413-497, 1993.
- Llinas RR: The intrinsic electrophysiological properties of mammalian neurons: insights into central nervous system function. Science 242: 1654-1664, 1988.
- Martin JH: The collective electrical behavior of cortical neurons: The electroencephalogram and the mechanisms of epilepsy. In Kandel ER, Schwartz JH, Jessell TM (Eds.) Principles of Neural Science, Prentice Hall International, London, pp 777-791, 1991.
- Mitzdorf U. Current source-density method and application in cat cerebral cortex: investigation of evoked potentials and EEG phenomena. Physiol Rev 1985;65:37-100.
- Pascual-Marqui RD, Mishel CM, Lehmann D: Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. International Journal of Psychophysiology. 1994, 18: 49-65.
- Pascual-Marqui RD. Review of methods for solving the EEG inverse problem. International Journal of Bioelectromagnetism 1999; 1:75-86.  
(<http://www.unizh.ch/keyinst/NewLORETA/TechnicalDetails/TechnicalDetails.pdf>) and (<http://www.ijbem.org/>).
- Pascual-Marqui RD: Reply to comments by Hämäläinen, Ilmoniemi and Nunez. In Source Localization: Continuing Discussion of the Inverse Problem (W. Skrandies, Ed.), pp. 16-28, ISBET Newsletter No.6 (ISSN 0947-5133), 1995.  
(<http://www.unizh.ch/keyinst/NewLORETA/BriefHistory/LORETA-NewsLett2b.pdf>).

Pascual-Marqui, R.D., 2002. Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. *Methods Find. Exp. Clin. Pharmacol.* 24 (Suppl D.), 5 –12. (<http://www.unizh.ch/keyinst/NewLORETA/sLORETA/sLORETA-Mathoi.pdf>).

Roberto D. Pascual-Marqui, Alberto D. Pascual-Montano, Dietrich Lehmann, Kieko Kochi, Michaela Esslen, Lutz Jancke, Peter Anderer, Bernd Saletu, Hideaki Tanaka, Koichi Hirata, E. Roy John, Leslie Prichep. Exact low resolution brain electromagnetic tomography (eLORETA). *Neuroimage* 2006, Vol 31, Suppl. 1, page:S86

J. Sarvas, "Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem," *Phys. Med. Biol.*, vol. 32, pp. 11-22, 1987.

K. Sekihara, M. Sahani, and S. S. Nagarajan, "Localization bias and spatial resolution of adaptive and nonadaptive spatial filters for MEG source reconstruction," *Neuroimag.* 2005, 25:1056-1067.

Sukov W, Barth DS: Three-dimensional analysis of spontaneous and thalamically evoked gamma oscillations in auditory cortex. *J. Neurophysiol.* 79: 2875–2884, 1998.

# Instantaneous and lagged measurements of linear and nonlinear dependence between groups of multivariate time series: frequency decomposition

Roberto D. Pascual-Marqui

The KEY Institute for Brain-Mind Research  
University Hospital of Psychiatry  
Lenggstr. 31, CH-8032 Zurich, Switzerland  
pascualm at key.uzh.ch  
[www.keyinst.uzh.ch/loreta](http://www.keyinst.uzh.ch/loreta)

## 1. Abstract

Measures of linear dependence (coherence) and nonlinear dependence (phase synchronization) between any number of multivariate time series are defined. The measures are expressed as the sum of lagged dependence and instantaneous dependence. The measures are non-negative, and take the value zero only when there is independence of the pertinent type. These measures are defined in the frequency domain and are applicable to stationary and non-stationary time series. These new results extend and refine significantly those presented in a previous technical report (Pascual-Marqui 2007, arXiv:0706.1776 [stat.ME], <http://arxiv.org/abs/0706.1776> ), and have been largely motivated by the seminal paper on linear feedback by Geweke (1982 JASA 77:304-313). One important field of application is neurophysiology, where the time series consist of electric neuronal activity at several brain locations. Coherence and phase synchronization are interpreted as "connectivity" between locations. However, any measure of dependence is highly contaminated with an instantaneous, non-physiological contribution due to volume conduction and low spatial resolution. The new techniques remove this confounding factor considerably. Moreover, the measures of dependence can be applied to any number of brain areas jointly, i.e. distributed cortical networks, whose activity can be estimated with eLORETA (Pascual-Marqui 2007, arXiv:0710.3341 [math-ph], <http://arxiv.org/abs/0710.3341> ).

## 2. Introduction

This study extends and refines significantly the results presented in a previous technical report (Pascual-Marqui 2007a). Some results from that previous paper will be repeated here for the sake of completeness.

### 2.1. The discrete Fourier transform for multivariate time series

The terms "multivariate time series", "multiple time series", and "vector time series" have identical meaning in this paper.

For general notation and definitions, see e.g. Brillinger (1981) for stationary multivariate time series analysis, and see e.g. Mardia et al (1979) for general multivariate statistics.

Let  $\mathbf{X}_{jt} \in \mathbb{R}^{p \times 1}$  and  $\mathbf{Y}_{jt} \in \mathbb{R}^{q \times 1}$  denote two stationary multivariate time series, for discrete time  $t = 0 \dots N_T - 1$ , with  $j = 1 \dots N_R$  denoting the  $j$ -th time segment. The discrete Fourier transforms are denoted as  $\mathbf{X}_{j\omega} \in \mathbb{C}^{p \times 1}$  and  $\mathbf{Y}_{j\omega} \in \mathbb{C}^{q \times 1}$ , and defined as:

$$\text{Eq. 1} \quad \mathbf{X}_{j\omega} = \sum_{t=0}^{N_T-1} \mathbf{X}_{jt} e^{-2\pi i \omega t / N_T}$$

$$\text{Eq. 2} \quad \mathbf{Y}_{j\omega} = \sum_{t=0}^{N_T-1} \mathbf{Y}_{jt} e^{-2\pi i \omega t / N_T}$$

for discrete frequencies  $\omega = 0 \dots N_T - 1$ , and where  $i = \sqrt{-1}$ .

It will be assumed throughout that  $\mathbf{X}_\omega$  and  $\mathbf{Y}_\omega$  each have zero mean.

## 2.2. Classical cross-spectra

Let:

$$\text{Eq. 3} \quad \mathbf{S}_{\mathbf{XX}\omega} = \frac{1}{N_R} \sum_{j=1}^{N_R} \mathbf{X}_{j\omega} \mathbf{X}_{j\omega}^*$$

$$\text{Eq. 4} \quad \mathbf{S}_{\mathbf{YY}\omega} = \frac{1}{N_R} \sum_{j=1}^{N_R} \mathbf{Y}_{j\omega} \mathbf{Y}_{j\omega}^*$$

$$\text{Eq. 5} \quad \mathbf{S}_{\mathbf{XY}\omega} = \frac{1}{N_R} \sum_{j=1}^{N_R} \mathbf{X}_{j\omega} \mathbf{Y}_{j\omega}^*$$

$$\text{Eq. 6} \quad \mathbf{S}_{\mathbf{YX}\omega} = \mathbf{S}_{\mathbf{XY}\omega}^* = \frac{1}{N_R} \sum_{j=1}^{N_R} \mathbf{Y}_{j\omega} \mathbf{X}_{j\omega}^*$$

denote complex valued covariance matrices, where the superscript "\*" denotes vector/matrix transposition and complex conjugation. Note that  $\mathbf{S}_{\mathbf{XX}\omega}$  and  $\mathbf{S}_{\mathbf{YY}\omega}$  are Hermitian matrices, satisfying  $\mathbf{S} = \mathbf{S}^*$ . When multiplied by the factor  $(2\pi N_T)^{-1}$ , these matrices correspond to the classical cross-spectral density matrices.

## 2.3. Phase-information cross-spectra

The discrete Fourier transforms in Eq. 1 and Eq. 2 contain both phase and amplitude information, which carries over to the covariance matrices in Eq. 3, Eq. 4, Eq. 5, and Eq. 6. This means that for the analysis of phase information, the amplitudes must be factored out by an appropriate normalization method. This is achieved by using the following definition for the normalized complex-valued discrete Fourier transform vector:

$$\text{Eq. 7} \quad \tilde{\mathbf{X}}_{j\omega} = \left( \mathbf{X}_{j\omega}^* \mathbf{X}_{j\omega} \right)^{-1/2} \mathbf{X}_{j\omega}$$

and:

$$\text{Eq. 8} \quad \tilde{\mathbf{Y}}_{j\omega} = \left( \mathbf{Y}_{j\omega}^* \mathbf{Y}_{j\omega} \right)^{-1/2} \mathbf{Y}_{j\omega}$$

Note that this normalization operation, although deceptively simple, is a highly nonlinear transformation.

The corresponding covariance matrices containing phase information (without amplitude information) are:

$$\text{Eq. 9} \quad \mathbf{S}_{\tilde{\mathbf{X}}\tilde{\mathbf{X}}\omega} = \frac{1}{N_R} \sum_{j=1}^{N_R} \tilde{\mathbf{X}}_{j\omega} \tilde{\mathbf{X}}_{j\omega}^*$$

$$\text{Eq. 10} \quad \mathbf{S}_{\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}\omega} = \frac{1}{N_R} \sum_{j=1}^{N_R} \tilde{\mathbf{Y}}_{j\omega} \tilde{\mathbf{Y}}_{j\omega}^*$$

$$\text{Eq. 11} \quad \mathbf{S}_{\tilde{\mathbf{X}}\tilde{\mathbf{Y}}\omega} = \frac{1}{N_R} \sum_{j=1}^{N_R} \tilde{\mathbf{X}}_{j\omega} \tilde{\mathbf{Y}}_{j\omega}^*$$

$$\text{Eq. 12} \quad \mathbf{S}_{\tilde{\mathbf{Y}}\tilde{\mathbf{X}}\omega} = \mathbf{S}_{\tilde{\mathbf{X}}\tilde{\mathbf{Y}}\omega}^* = \frac{1}{N_R} \sum_{j=1}^{N_R} \tilde{\mathbf{Y}}_{j\omega} \tilde{\mathbf{X}}_{j\omega}^*$$

Note that the normalization used in Eq. 7 and Eq. 8 will be the basis for the analysis of phase synchronization between the multivariate time series  $\mathbf{X}$  and  $\mathbf{Y}$ .

Note that  $\mathbf{S}_{\tilde{\mathbf{X}}\tilde{\mathbf{X}}\omega}$  and  $\mathbf{S}_{\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}\omega}$  are Hermitian matrices. When multiplied by the factor  $(2\pi N_T)^{-1}$ , these matrices correspond to what is defined here as the phase-information cross-spectra.

## 2.4. Instantaneous, zero-phase, zero-lag covariance

The instantaneous, zero-phase, zero-lag covariance matrix corresponding to a multivariate time series at frequency  $\omega$ , is simply the real part of the Hermitian covariance matrix at frequency  $\omega$ , i.e.  $\text{Re}(\mathbf{S}_\omega)$ .

To justify this, consider the multivariate time series  $\mathbf{X}_{jt} \in \mathbb{R}^{p \times 1}$ , for discrete time  $t = 0 \dots N_T - 1$ , with  $j = 1 \dots N_R$  denoting the  $j$ -th time segment.

In a first step, filter the time series to leave exclusively the frequency  $\omega$  component. Denote the filtered time series as  $(\mathbf{X}_{jt}^{\omega \text{Filtered}})$ . Note that, by construction, the spectral density of  $(\mathbf{X}_{jt}^{\omega \text{Filtered}})$  is zero everywhere except at frequency  $\omega$ .

In a second step, compute the instantaneous, zero-lag, zero phase shifted, time domain, symmetric covariance matrix for the filtered time series  $(\mathbf{X}_{jt}^{\omega \text{Filtered}})$  at frequency  $\omega$ :

$$\text{Eq. 13} \quad \mathbf{A}_\omega = \frac{1}{N_T N_R} \sum_{j=1}^{N_R} \sum_{t=1}^{N_t} (\mathbf{X}_{jt}^{\omega \text{Filtered}}) (\mathbf{X}_{jt}^{\omega \text{Filtered}})^T \in \mathbb{R}^{r \times r}$$

Finally, by making use of Parseval's theorem for the filtered time series, the following relation holds:

$$\text{Eq. 14} \quad \text{Re}(\mathbf{S}_{\mathbf{XX}\omega}) = \frac{N_T^2}{2} \mathbf{A}_\omega$$

where  $\text{Re}(\mathbf{S}_{\mathbf{XX}\omega})$  denotes the real part of  $\mathbf{S}_{\mathbf{XX}\omega}$  given by Eq. 3 above.

These arguments apply identically to the normalized time series, as in Eq. 7 to Eq. 12 above, when considering the phase-information cross-spectra. This means that the instantaneous, zero-phase, zero-lag covariance matrix corresponding to a normalized multivariate time series  $\mathbf{X}$  at frequency  $\omega$ , is simply the real part of the phase-information Hermitian covariance matrix at frequency  $\omega$ , i.e.  $\text{Re}(\mathbf{S}_{\mathbf{XX}\omega})$ .

The section entitled "Appendix 1" gives a brief description of the problems that arise in neurophysiology, where any measure of dependence is highly contaminated with an instantaneous, non-physiological contribution due to volume conduction and low spatial resolution.

### 3. Measures of linear dependence (coherence-type) between two multivariate time series

The definitions presented here are largely motivated by the seminal paper on linear feedback by Geweke (1982).

The measure of linear dependence between time series  $\mathbf{X}$  and  $\mathbf{Y}$  at frequency  $\omega$  is defined as:

$$\text{Eq. 15} \quad F_{\mathbf{X},\mathbf{Y}}(\omega) = \ln \frac{\left| \begin{pmatrix} \mathbf{S}_{\mathbf{YY}\omega} & \mathbf{0} \\ \mathbf{0}^T & \mathbf{S}_{\mathbf{XX}\omega} \end{pmatrix} \right|}{\left| \begin{pmatrix} \mathbf{S}_{\mathbf{YY}\omega} & \mathbf{S}_{\mathbf{YX}\omega} \\ \mathbf{S}_{\mathbf{XY}\omega} & \mathbf{S}_{\mathbf{XX}\omega} \end{pmatrix} \right|}$$

where  $|\mathbf{M}|$  denotes the determinant of  $\mathbf{M}$ . The matrix in the numerator of Eq. 15 is a block-diagonal matrix, with  $\mathbf{0}$  denoting a matrix of zeros, which in this case is of dimension  $q \times p$ .

This measure of linear dependence is expressed as the sum of the lagged linear dependence  $F_{\mathbf{X} \rightleftharpoons \mathbf{Y}}(\omega)$  and instantaneous linear dependence  $F_{\mathbf{X},\mathbf{Y}}(\omega)$ :

$$\text{Eq. 16} \quad F_{\mathbf{X},\mathbf{Y}}(\omega) = F_{\mathbf{X} \rightleftharpoons \mathbf{Y}}(\omega) + F_{\mathbf{X},\mathbf{Y}}(\omega)$$

The measure of instantaneous linear dependence is defined as:

$$\text{Eq. 17} \quad F_{\mathbf{X},\mathbf{Y}}(\omega) = \ln \frac{\left| \text{Re} \begin{pmatrix} \mathbf{S}_{\mathbf{YY}\omega} & \mathbf{0} \\ \mathbf{0}^T & \mathbf{S}_{\mathbf{XX}\omega} \end{pmatrix} \right|}{\left| \text{Re} \begin{pmatrix} \mathbf{S}_{\mathbf{YY}\omega} & \mathbf{S}_{\mathbf{YX}\omega} \\ \mathbf{S}_{\mathbf{XY}\omega} & \mathbf{S}_{\mathbf{XX}\omega} \end{pmatrix} \right|}$$

where  $\text{Re}(\mathbf{M})$  denotes the real part of  $\mathbf{M}$ .

Finally, the measure of lagged linear dependence is:

$$\text{Eq. 18} \quad F_{X \rightleftharpoons Y}(\omega) = F_{X,Y}(\omega) - F_{X \cdot Y}(\omega) = \ln \frac{\left\{ \text{Re} \begin{pmatrix} \mathbf{S}_{YY\omega} & \mathbf{S}_{YX\omega} \\ \mathbf{S}_{XY\omega} & \mathbf{S}_{XX\omega} \end{pmatrix} \right\} / \left\{ \text{Re} \begin{pmatrix} \mathbf{S}_{YY\omega} & \mathbf{o} \\ \mathbf{o}^T & \mathbf{S}_{XX\omega} \end{pmatrix} \right\}}{\left\{ \begin{pmatrix} \mathbf{S}_{YY\omega} & \mathbf{S}_{YX\omega} \\ \mathbf{S}_{XY\omega} & \mathbf{S}_{XX\omega} \end{pmatrix} \right\} / \left\{ \begin{pmatrix} \mathbf{S}_{YY\omega} & \mathbf{o} \\ \mathbf{o}^T & \mathbf{S}_{XX\omega} \end{pmatrix} \right\}}$$

All three measures are non-negative. They take the value zero only when there is independence of the pertinent type (lagged, instantaneous, or both).

Not that the measure of linear dependence  $F_{X,Y}(\omega)$  in Eq. 15 can be interpreted as follows:

$$\text{Eq. 19} \quad \rho_{X,Y}^2(\omega) = 1 - \exp(-F_{X,Y}(\omega))$$

where  $\rho_{X,Y}(\omega)$  was defined as the general coherence in Pascual-Marqui (2007a; see Eq. 7 therein):

$$\text{Eq. 20} \quad \rho_{X,Y}^2(\omega) = \rho_G^2 = 1 - \frac{|\mathbf{S}_{YY\omega} - \mathbf{S}_{YX\omega} \mathbf{S}_{XX\omega}^{-1} \mathbf{S}_{XY\omega}|}{|\mathbf{S}_{YY\omega}|}$$

Some relevant literature that motivated the definition of the general coherence  $\rho_{X,Y}^2(\omega)$  in the previous study (Pascual-Marqui 2007a) follows. In the case of real-valued stochastic variables, Mardia et al (1979) review several "measures of correlation between vectors". In particular, Kent (1983) proposed a general measure of correlation that is closely related to the vector alienation coefficient (Hotelling 1936, Mardia et al 1979). This measure of general coherence is also equivalent to the coefficient of determination as defined by Pierce (1982). All these definitions can be straightforwardly generalized to the complex valued domain.

In order to illustrate and further motivate these measures of linear dependence, a detailed analysis for the simple case of two univariate time series is presented.

In the case that the two time series are univariate, the measure of linear dependence  $F_{X,Y}(\omega)$  in Eq. 15 is:

$$\text{Eq. 21} \quad F_{X,Y}(\omega) = \ln \frac{s_{yy\omega} s_{xx\omega}}{s_{yy\omega} s_{xx\omega} - [\text{Re}(s_{yx\omega})]^2 - [\text{Im}(s_{yx\omega})]^2} = -\ln(1 - \rho^2)$$

where:

$$\text{Eq. 22} \quad \rho^2 = \frac{([\text{Re}(s_{yx\omega})]^2 + [\text{Im}(s_{yx\omega})]^2)}{s_{yy\omega} s_{xx\omega}}$$

In Eq. 22,  $\rho$  is the ordinary squared coherence (see e.g. Equation 3 in Nolte et al 2004).

The measure of instantaneous linear dependence is:

$$\text{Eq. 23} \quad F_{X \cdot Y}(\omega) = \ln \frac{s_{yy\omega} s_{xx\omega}}{s_{yy\omega} s_{xx\omega} - [\text{Re}(s_{yx\omega})]^2}$$

Note that we can define the instantaneous coherence  $\rho_{X,Y}(\omega)$  as:

$$\text{Eq. 24} \quad F_{X,Y}(\omega) = -\ln(1 - \rho_{X,Y}^2(\omega))$$

In general, this gives:

$$\text{Eq. 25} \quad \rho_{X,Y}^2(\omega) = 1 - \exp[-F_{X,Y}(\omega)] = 1 - \frac{\left| \text{Re} \begin{pmatrix} \mathbf{S}_{YY\omega} & \mathbf{S}_{YX\omega} \\ \mathbf{S}_{XY\omega} & \mathbf{S}_{XX\omega} \end{pmatrix} \right|}{\left| \text{Re} \begin{pmatrix} \mathbf{S}_{YY\omega} & \mathbf{0} \\ \mathbf{0}^T & \mathbf{S}_{XX\omega} \end{pmatrix} \right|}$$

and in the case of univariate time series it simplifies to:

$$\text{Eq. 26} \quad \rho_{X,Y}^2(\omega) = \frac{[\text{Re}(s_{yx\omega})]^2}{s_{yy\omega}s_{xx\omega}}$$

which, not surprisingly, is directly related to the real part of the complex valued coherency.

Finally, in the particular case of univariate time series, the measure of lagged linear dependence is:

$$\text{Eq. 27} \quad \left\{ \begin{aligned} F_{X \rightleftharpoons Y}(\omega) &= F_{X,Y}(\omega) - F_{Y,X}(\omega) \\ &= \ln \frac{s_{yy\omega}s_{xx\omega}}{s_{yy\omega}s_{xx\omega} - [\text{Re}(s_{yx\omega})]^2 - [\text{Im}(s_{yx\omega})]^2} - \ln \frac{s_{yy\omega}s_{xx\omega}}{s_{yy\omega}s_{xx\omega} - [\text{Re}(s_{yx\omega})]^2} \\ &= \ln \left[ \frac{s_{yy\omega}s_{xx\omega} - [\text{Re}(s_{yx\omega})]^2}{s_{yy\omega}s_{xx\omega} - [\text{Re}(s_{yx\omega})]^2 - [\text{Im}(s_{yx\omega})]^2} \right] \\ &= -\ln(1 - \rho_{X \rightleftharpoons Y}^2(\omega)) \end{aligned} \right.$$

with:

$$\text{Eq. 28} \quad \rho_{X \rightleftharpoons Y}^2(\omega) = \frac{[\text{Im}(s_{yx\omega})]^2}{s_{yy\omega}s_{xx\omega} - [\text{Re}(s_{yx\omega})]^2}$$

In Eq. 28, for the particular case of univariate time series,  $\rho_{X \rightleftharpoons Y}^2(\omega)$  is equal to the "zero-lag removed general coherence"  $\rho_{GL}$  defined in Pascual-Marqui (2007a).

In our previous related study (Pascual-Marqui 2007a), the general definition given there for the "zero-lag removed coherence" (see Eq. 22 therein) was:

$$\text{Eq. 29} \quad \rho_{GL}^2 = 1 - \frac{\left| \begin{pmatrix} \mathbf{S}_{YY\omega} & \mathbf{S}_{YX\omega} \\ \mathbf{S}_{XY\omega} & \mathbf{S}_{XX\omega} \end{pmatrix} \right|}{\left| \text{Re} \begin{pmatrix} \mathbf{S}_{YY\omega} & \mathbf{S}_{YX\omega} \\ \mathbf{S}_{XY\omega} & \mathbf{S}_{XX\omega} \end{pmatrix} \right|}$$

The new definition given here for the lagged coherence follows from the relation:

$$\text{Eq. 30} \quad F_{X \rightleftharpoons Y}(\omega) = -\ln(1 - \rho_{X \rightleftharpoons Y}^2(\omega))$$

which gives:

$$\text{Eq. 31} \quad \rho_{X \Leftrightarrow Y}^2(\omega) = 1 - \exp[-F_{X \Leftrightarrow Y}(\omega)] = 1 - \frac{\left\{ \left| \begin{pmatrix} \mathbf{S}_{YY\omega} & \mathbf{S}_{YX\omega} \\ \mathbf{S}_{XY\omega} & \mathbf{S}_{XX\omega} \end{pmatrix} \right| \right\}}{\left\{ \left| \begin{pmatrix} \mathbf{S}_{YY\omega} & \mathbf{0} \\ \mathbf{0}^T & \mathbf{S}_{XX\omega} \end{pmatrix} \right| \right\}} \Bigg/ \frac{\left\{ \text{Re} \left( \begin{pmatrix} \mathbf{S}_{YY\omega} & \mathbf{S}_{YX\omega} \\ \mathbf{S}_{XY\omega} & \mathbf{S}_{XX\omega} \end{pmatrix} \right) \right\}}{\left\{ \text{Re} \left( \begin{pmatrix} \mathbf{S}_{YY\omega} & \mathbf{0} \\ \mathbf{0}^T & \mathbf{S}_{XX\omega} \end{pmatrix} \right) \right\}}$$

Both definitions (Eq. 29 and Eq. 31) are identical for the case of two univariate time series. However, they are different for the multivariate case. Whereas the old definition in Eq. 29 lumps together all variables from  $\mathbf{X}$  and  $\mathbf{Y}$ , the new definition given here in Eq. 31 conserves the multivariate structure of the two multivariate time series. The improvement of the new lagged coherence in Eq. 31 is that it measures the lagged linear dependence between the two multivariate time series without being affected by the covariance structure within each multivariate time series. The shortcoming of the old definition from our previous study (Pascual-Marqui 2007a), shown in Eq. 29, is that it is contaminated by the dependence structures of the univariate time series within  $\mathbf{X}$  and within  $\mathbf{Y}$ .

Another point worth stressing is the asymmetry in the results for the instantaneous coherence  $\rho_{X,Y}^2(\omega)$  (Eq. 26) and the lagged coherence  $\rho_{X \Leftrightarrow Y}^2(\omega)$  (Eq. 28). While the instantaneous coherence is the real part of the complex valued coherency, the lagged coherence *is not* the imaginary part of the complex valued coherency. Ideally, the lagged coherence is a measure that is not affected by instantaneous dependence, whereas the imaginary part of the complex valued coherency (Nolte et al 2004) is more affected by instantaneous dependence (Pascual-Marqui 2007a). This makes the lagged coherence (Eq. 31) a much more adequate measure of electrophysiological connectivity, because it removes the confounding effect of instantaneous dependence due to volume conduction and low spatial resolution (Pascual-Marqui 2007a).

Note that the measures of linear dependence defined by Eq. 15, Eq. 17, and Eq. 18 each have the form of a ratio of variances, which compares the residuals of different models (i.e. different dependent and independent variables). Under the assumption that the time series are wide-sense stationary, large sample distribution theory can be used to test the null hypothesis that a given measure of linear dependence is zero. Following the same methodology as in Geweke (1982), the asymptotic distributions are:

$$\text{Eq. 32} \quad \left\{ \begin{array}{l} \text{Under } H_0 : F_{X,Y}(\omega) = 0, \quad N_T \hat{F}_{X,Y}(\omega) \overset{a}{\sim} \chi^2(2pq) \\ \text{Under } H_0 : F_{X \Leftrightarrow Y}(\omega) = 0, \quad N_T \hat{F}_{X \Leftrightarrow Y}(\omega) \overset{a}{\sim} \chi^2(pq) \\ \text{Under } H_0 : F_{X \cdot Y}(\omega) = 0, \quad N_T \hat{F}_{X \cdot Y}(\omega) \overset{a}{\sim} \chi^2(pq) \end{array} \right\}$$

#### 4. Measures of linear dependence (coherence-type) between groups of multivariate time series

Consider the case of three multivariate time series  $\mathbf{X}_{jt} \in \mathbb{R}^{p \times 1}$ ,  $\mathbf{Y}_{jt} \in \mathbb{R}^{q \times 1}$ , and  $\mathbf{Z}_{jt} \in \mathbb{R}^{r \times 1}$ , for discrete time  $t = 0 \dots N_T - 1$ , with  $j = 1 \dots N_R$  denoting the  $j$ -th time segment.

The measures of linear dependence between the three multivariate time series are related in the usual way:

$$\text{Eq. 33} \quad F_{X,Y,Z}(\omega) = F_{X \rightleftharpoons Y \rightleftharpoons Z}(\omega) + F_{X \cdot Y \cdot Z}(\omega)$$

and are given by:

$$\text{Eq. 34} \quad F_{X,Y,Z}(\omega) = \ln \frac{\begin{vmatrix} \mathbf{S}_{YY\omega} & \mathbf{o} & \mathbf{o} \\ \mathbf{o} & \mathbf{S}_{XX\omega} & \mathbf{o} \\ \mathbf{o} & \mathbf{o} & \mathbf{S}_{ZZ\omega} \end{vmatrix}}{\begin{vmatrix} \mathbf{S}_{YY\omega} & \mathbf{S}_{YX\omega} & \mathbf{S}_{YZ\omega} \\ \mathbf{S}_{XY\omega} & \mathbf{S}_{XX\omega} & \mathbf{S}_{XZ\omega} \\ \mathbf{S}_{ZY\omega} & \mathbf{S}_{ZX\omega} & \mathbf{S}_{ZZ\omega} \end{vmatrix}}$$

$$\text{Eq. 35} \quad F_{X \cdot Y \cdot Z}(\omega) = \ln \frac{\text{Re} \begin{vmatrix} \mathbf{S}_{YY\omega} & \mathbf{o} & \mathbf{o} \\ \mathbf{o} & \mathbf{S}_{XX\omega} & \mathbf{o} \\ \mathbf{o} & \mathbf{o} & \mathbf{S}_{ZZ\omega} \end{vmatrix}}{\text{Re} \begin{vmatrix} \mathbf{S}_{YY\omega} & \mathbf{S}_{YX\omega} & \mathbf{S}_{YZ\omega} \\ \mathbf{S}_{XY\omega} & \mathbf{S}_{XX\omega} & \mathbf{S}_{XZ\omega} \\ \mathbf{S}_{ZY\omega} & \mathbf{S}_{ZX\omega} & \mathbf{S}_{ZZ\omega} \end{vmatrix}}$$

and:

$$\text{Eq. 36} \quad F_{X \rightleftharpoons Y \rightleftharpoons Z}(\omega) = \ln \frac{\text{Re} \begin{vmatrix} \mathbf{S}_{YY\omega} & \mathbf{S}_{YX\omega} & \mathbf{S}_{YZ\omega} \\ \mathbf{S}_{XY\omega} & \mathbf{S}_{XX\omega} & \mathbf{S}_{XZ\omega} \\ \mathbf{S}_{ZY\omega} & \mathbf{S}_{ZX\omega} & \mathbf{S}_{ZZ\omega} \end{vmatrix}}{\left\{ \begin{vmatrix} \mathbf{S}_{YY\omega} & \mathbf{S}_{YX\omega} & \mathbf{S}_{YZ\omega} \\ \mathbf{S}_{XY\omega} & \mathbf{S}_{XX\omega} & \mathbf{S}_{XZ\omega} \\ \mathbf{S}_{ZY\omega} & \mathbf{S}_{ZX\omega} & \mathbf{S}_{ZZ\omega} \end{vmatrix} \middle/ \begin{vmatrix} \mathbf{S}_{YY\omega} & \mathbf{o} & \mathbf{o} \\ \mathbf{o} & \mathbf{S}_{XX\omega} & \mathbf{o} \\ \mathbf{o} & \mathbf{o} & \mathbf{S}_{ZZ\omega} \end{vmatrix} \right\}}$$

Coherences for each type of measure of linear dependence in Eq. 33 are defined by the general relation (see e.g. Pierce 1982):

$$\text{Eq. 37} \quad \rho^2(\omega) = 1 - \exp[-F(\omega)]$$

As previously argued, under the assumption that the time series are wide-sense stationary, large sample distribution theory can be used to test the null hypothesis that a given measure of linear dependence is zero. In this case, the asymptotic distributions are:

$$\text{Eq. 38} \quad \left\{ \begin{array}{l} \text{Under } H_0 : F_{X,Y,Z}(\omega) = 0, \quad N_T \hat{F}_{X,Y,Z}(\omega) \stackrel{a}{\sim} \chi^2(2pq + 2pr + 2qr) \\ \text{Under } H_0 : F_{X \rightleftharpoons Y \rightleftharpoons Z}(\omega) = 0, \quad N_T \hat{F}_{X \rightleftharpoons Y \rightleftharpoons Z}(\omega) \stackrel{a}{\sim} \chi^2(pq + pr + qr) \\ \text{Under } H_0 : F_{X \cdot Y \cdot Z}(\omega) = 0, \quad N_T \hat{F}_{X \cdot Y \cdot Z}(\omega) \stackrel{a}{\sim} \chi^2(pq + pr + qr) \end{array} \right\}$$

The generalization of these definitions to any number of multivariate time series is straightforward.

It is important to emphasize here that these measures of linear dependence for groups of multivariate time series can be applied in the field of neurophysiology. In this

case, the time series consist of electric neuronal activity at several brain locations, and the measures of dependence are interpreted as "connectivity" between locations. When considering several brain locations, these new measures can be used to test for the existence of distributed cortical networks, whose activity can be estimated with exact low resolution brain electromagnetic tomography (Pascual-Marqui 2007b).

## 5. Measures of linear dependence (coherence-type) between all univariate time series

A particular case of interest consists of measuring the linear dependence between all the univariate time series that form part of the vector time series. For instance, consider the vector time series  $\mathbf{X}_{jt} \in \mathbb{R}^{p \times 1}$ . Then the measures of linear dependence between all "p" univariate time series of  $\mathbf{X}$  are:

$$\text{Eq. 39} \quad F_{\mathbf{X},\mathbf{X}}(\omega) = F_{\mathbf{X} \rightleftharpoons \mathbf{X}}(\omega) + F_{\mathbf{X} \cdot \mathbf{X}}(\omega)$$

$$\text{Eq. 40} \quad F_{\mathbf{X},\mathbf{X}}(\omega) = \ln \frac{|Diag(\mathbf{S}_{\mathbf{X}\mathbf{X}\omega})|}{|\mathbf{S}_{\mathbf{X}\mathbf{X}\omega}|}$$

$$\text{Eq. 41} \quad F_{\mathbf{X} \cdot \mathbf{X}}(\omega) = \ln \frac{|Diag(\mathbf{S}_{\mathbf{X}\mathbf{X}\omega})|}{|\text{Re}(\mathbf{S}_{\mathbf{X}\mathbf{X}\omega})|}$$

$$\text{Eq. 42} \quad F_{\mathbf{X} \rightleftharpoons \mathbf{X}}(\omega) = F_{\mathbf{X},\mathbf{X}}(\omega) - F_{\mathbf{X} \cdot \mathbf{X}}(\omega) = \ln \frac{|\text{Re}(\mathbf{S}_{\mathbf{X}\mathbf{X}\omega})|}{|\mathbf{S}_{\mathbf{X}\mathbf{X}\omega}|}$$

Coherences for each type of measure of linear dependence in Eq. 39 are defined by the general relation (see e.g. Pierce 1982):

$$\text{Eq. 43} \quad \rho^2(\omega) = 1 - \exp[-F(\omega)]$$

In Eq. 40 and Eq. 41, the notation  $Diag(\mathbf{M})$  denotes a diagonal matrix formed by the diagonal elements of  $\mathbf{M}$ . Note that for Hermitian matrices, such as  $\mathbf{S}_{\mathbf{X}\mathbf{X}\omega}$ , the diagonal elements are pure real, which implies that:

$$\text{Eq. 44} \quad Diag(\mathbf{S}_{\mathbf{X}\mathbf{X}\omega}) = Diag(\text{Re}(\mathbf{S}_{\mathbf{X}\mathbf{X}\omega})) = \text{Re}(Diag(\mathbf{S}_{\mathbf{X}\mathbf{X}\omega}))$$

As a consistency check, it can easily be verified that when these definitions are applied to a vector time series with 2 components, the same results are obtained as in the case of two univariate time series (Eq. 21, Eq. 23, and Eq. 27).

Under the assumption that the time series are wide-sense stationary, large sample distribution theory can be used to test the null hypothesis that a given measure of linear dependence is zero. In this case, the asymptotic distributions are:

$$\text{Eq. 45} \quad \left\{ \begin{array}{l} \text{Under } H_0 : F_{\mathbf{X},\mathbf{X}}(\omega) = 0, \quad N_T \hat{F}_{\mathbf{X},\mathbf{X}}(\omega) \overset{a}{\sim} \chi^2(p(p-1)) \\ \text{Under } H_0 : F_{\mathbf{X} \rightleftharpoons \mathbf{X}}(\omega) = 0, \quad N_T \hat{F}_{\mathbf{X} \rightleftharpoons \mathbf{X}}(\omega) \overset{a}{\sim} \chi^2(p(p-1)/2) \\ \text{Under } H_0 : F_{\mathbf{X} \cdot \mathbf{X}}(\omega) = 0, \quad N_T \hat{F}_{\mathbf{X} \cdot \mathbf{X}}(\omega) \overset{a}{\sim} \chi^2(p(p-1)/2) \end{array} \right\}$$

As a further consistency check, note that the test  $H_0 : F_{X,X}(\omega) = 0$  corresponds to the classical case of testing if a real-valued correlation matrix is the identity matrix. The statistic given above is precisely the log-likelihood ratio statistic, which is asymptotically chi-square with the specified degrees of freedom (Kullback 1967).

## **6. Measures of nonlinear dependence (phase synchronization type) between two multivariate time series**

The term "phase synchronization" has a very rigorous physics definition (see e.g. Rosenblum et al 1996). The basic idea behind this definition has been adapted and used to great advantage in the neurosciences (Tass et al 1998, Quian-Quiroga et al 2002, Pereda et al 2005, Stam et al 2007), as in for example, the analysis of pairs of time series of measured scalp electric potentials differences (i.e. EEG: electroencephalogram). Other equivalent descriptive names for "phase synchronization" that appear in the neurosciences are phase locking, phase locking value, phase locking index, phase coherence, and so on.

An informal definition for the statistical "phase synchronization" model will now be given. In order to simplify this informal definition even further, it will be assumed that there are two univariate stationary time series (i.e.  $p = q = 1$ ) of interest. At a given discrete frequency  $\omega$ , the sample data in the frequency domain (using the discrete Fourier transform) is denoted as  $x_{j\omega}, y_{j\omega} \in \mathbb{C}$ , with  $j = 1 \dots N_R$  denoting the  $j$ -th time segment. If the phase difference  $\Delta\phi_j = \phi_j^x - \phi_j^y$  is "stable" over time segments  $j$ , regardless of the amplitudes, then there is a "connection" between the locations at which the measurements were made. A measure of stability of phase difference is precisely "phase synchronization". It can as well be defined for the non-stationary case, using concepts of time-varying instantaneous phase, and defining stability over time (instead of stability over time segments).

In the case of univariate time series, i.e.  $p = q = 1$ , phase synchronization can be viewed as the modulus (absolute value) of the complex valued (Hermitian) coherency between the *normalized* Fourier transforms. These variables are normalized prior to the coherency calculation in order to remove from the outset any amplitude effect, leaving only phase information. This normalization operation is highly nonlinear.

The modulus of the coherency is used as a measure for phase synchronization because it is conveniently bounded in the range zero (no synchronization) to one (perfect synchronization).

Based on the foregoing arguments, a natural definition for the measures of nonlinear dependence (phase synchronization type) between two multivariate time series is exactly the same definitions as developed in the previous sections of this study, but applied to the phase-information cross-spectra (Eq. 7 to Eq. 12). The phase-information cross-spectra are based on normalized Fourier transform vectors, which is the particular requirement in this case (without amplitude information).

For two multivariate time series, the measure of nonlinear dependence  $G_{X,Y}(\omega)$  is expressed as the sum of lagged nonlinear dependence  $G_{X \rightleftharpoons Y}(\omega)$  and instantaneous nonlinear dependence  $G_{X \cdot Y}(\omega)$ :

$$\text{Eq. 46} \quad G_{X,Y}(\omega) = G_{X \rightleftharpoons Y}(\omega) + G_{X \cdot Y}(\omega)$$

with:

$$\text{Eq. 47} \quad G_{X,Y}(\omega) = \ln \frac{\left| \begin{pmatrix} \mathbf{S}_{\bar{Y}\bar{Y}\omega} & \mathbf{o} \\ \mathbf{o}^T & \mathbf{S}_{\bar{X}\bar{X}\omega} \end{pmatrix} \right|}{\left| \begin{pmatrix} \mathbf{S}_{\bar{Y}\bar{Y}\omega} & \mathbf{S}_{\bar{Y}\bar{X}\omega} \\ \mathbf{S}_{\bar{X}\bar{Y}\omega} & \mathbf{S}_{\bar{X}\bar{X}\omega} \end{pmatrix} \right|}$$

$$\text{Eq. 48} \quad G_{X \cdot Y}(\omega) = \ln \frac{\left| \text{Re} \begin{pmatrix} \mathbf{S}_{\bar{Y}\bar{Y}\omega} & \mathbf{o} \\ \mathbf{o}^T & \mathbf{S}_{\bar{X}\bar{X}\omega} \end{pmatrix} \right|}{\left| \text{Re} \begin{pmatrix} \mathbf{S}_{\bar{Y}\bar{Y}\omega} & \mathbf{S}_{\bar{Y}\bar{X}\omega} \\ \mathbf{S}_{\bar{X}\bar{Y}\omega} & \mathbf{S}_{\bar{X}\bar{X}\omega} \end{pmatrix} \right|}$$

and:

$$\text{Eq. 49} \quad G_{X \rightleftharpoons Y}(\omega) = G_{X,Y}(\omega) - G_{X \cdot Y}(\omega) = \ln \frac{\left\{ \left| \text{Re} \begin{pmatrix} \mathbf{S}_{\bar{Y}\bar{Y}\omega} & \mathbf{S}_{\bar{Y}\bar{X}\omega} \\ \mathbf{S}_{\bar{X}\bar{Y}\omega} & \mathbf{S}_{\bar{X}\bar{X}\omega} \end{pmatrix} \right| \right\} / \left\{ \left| \text{Re} \begin{pmatrix} \mathbf{S}_{\bar{Y}\bar{Y}\omega} & \mathbf{o} \\ \mathbf{o}^T & \mathbf{S}_{\bar{X}\bar{X}\omega} \end{pmatrix} \right| \right\}}{\left\{ \left| \begin{pmatrix} \mathbf{S}_{\bar{Y}\bar{Y}\omega} & \mathbf{S}_{\bar{Y}\bar{X}\omega} \\ \mathbf{S}_{\bar{X}\bar{Y}\omega} & \mathbf{S}_{\bar{X}\bar{X}\omega} \end{pmatrix} \right| \right\} / \left\{ \left| \begin{pmatrix} \mathbf{S}_{\bar{Y}\bar{Y}\omega} & \mathbf{o} \\ \mathbf{o}^T & \mathbf{S}_{\bar{X}\bar{X}\omega} \end{pmatrix} \right| \right\}}$$

In Eq. 47, Eq. 48, and Eq. 49, the Hermitian covariance matrices are defined for the normalized discrete Fourier transform vectors (Eq. 7 to Eq. 12).

All three measures are non-negative. They take the value zero only when there is independence of the pertinent type (lagged, instantaneous, or both).

These measures of nonlinear dependence can be associated with measures phase synchronization  $\varphi$  as follows.

The phase synchronization between two multivariate time series is:

$$\text{Eq. 50} \quad \varphi_{X,Y}^2(\omega) = 1 - \exp(-G_{X,Y}(\omega)) = 1 - \frac{\left| \begin{pmatrix} \mathbf{S}_{\bar{Y}\bar{Y}\omega} & \mathbf{S}_{\bar{Y}\bar{X}\omega} \\ \mathbf{S}_{\bar{X}\bar{Y}\omega} & \mathbf{S}_{\bar{X}\bar{X}\omega} \end{pmatrix} \right|}{\left| \begin{pmatrix} \mathbf{S}_{\bar{Y}\bar{Y}\omega} & \mathbf{o} \\ \mathbf{o}^T & \mathbf{S}_{\bar{X}\bar{X}\omega} \end{pmatrix} \right|}$$

The instantaneous phase synchronization between two multivariate time series is:

$$\text{Eq. 51} \quad \varphi_{X \cdot Y}^2(\omega) = 1 - \exp(-G_{X \cdot Y}(\omega)) = 1 - \frac{\left| \text{Re} \begin{pmatrix} \mathbf{S}_{\bar{Y}\bar{Y}\omega} & \mathbf{S}_{\bar{Y}\bar{X}\omega} \\ \mathbf{S}_{\bar{X}\bar{Y}\omega} & \mathbf{S}_{\bar{X}\bar{X}\omega} \end{pmatrix} \right|}{\left| \text{Re} \begin{pmatrix} \mathbf{S}_{\bar{Y}\bar{Y}\omega} & \mathbf{o} \\ \mathbf{o}^T & \mathbf{S}_{\bar{X}\bar{X}\omega} \end{pmatrix} \right|}$$

The lagged phase synchronization between two multivariate time series is:

$$\text{Eq. 52} \quad \varphi_{X \leftrightarrow Y}^2(\omega) = 1 - \exp(-G_{X \leftrightarrow Y}(\omega)) = 1 - \frac{\left\| \begin{pmatrix} \mathbf{S}_{\bar{Y}\bar{Y}\omega} & \mathbf{S}_{\bar{Y}\bar{X}\omega} \\ \mathbf{S}_{\bar{X}\bar{Y}\omega} & \mathbf{S}_{\bar{X}\bar{X}\omega} \end{pmatrix} \right\|}{\left\| \begin{pmatrix} \mathbf{S}_{\bar{Y}\bar{Y}\omega} & \mathbf{0} \\ \mathbf{0}^T & \mathbf{S}_{\bar{X}\bar{X}\omega} \end{pmatrix} \right\|}}{\left\| \text{Re} \begin{pmatrix} \mathbf{S}_{\bar{Y}\bar{Y}\omega} & \mathbf{S}_{\bar{Y}\bar{X}\omega} \\ \mathbf{S}_{\bar{X}\bar{Y}\omega} & \mathbf{S}_{\bar{X}\bar{X}\omega} \end{pmatrix} \right\|}{\left\| \text{Re} \begin{pmatrix} \mathbf{S}_{\bar{Y}\bar{Y}\omega} & \mathbf{0} \\ \mathbf{0}^T & \mathbf{S}_{\bar{X}\bar{X}\omega} \end{pmatrix} \right\|}}$$

The phase synchronization between two multivariate time series  $\varphi_{X,Y}^2(\omega)$  given by Eq. 50 corresponds to the square of the “general phase synchronization” previously defined in Pascual-Marqui (2007a; see Eq. 15 therein).

In order to illustrate and further motivate these measures of nonlinear dependence, a detailed analysis for the simple case of two univariate time series is presented.

In the case that the two time series are univariate, the measure of nonlinear dependence  $G_{X,Y}(\omega)$  in Eq. 47 is:

$$\text{Eq. 53} \quad G_{X,Y}(\omega) = \ln \frac{1}{1 - [\text{Re}(s_{\bar{x}\bar{y}\omega})]^2 - [\text{Im}(s_{\bar{x}\bar{y}\omega})]^2} = -\ln(1 - \varphi_{X,Y}^2(\omega))$$

with phase synchronization:

$$\text{Eq. 54} \quad \varphi_{X,Y}^2(\omega) = [\text{Re}(s_{\bar{x}\bar{y}\omega})]^2 + [\text{Im}(s_{\bar{x}\bar{y}\omega})]^2 = \left| \frac{1}{N_R} \sum_{j=1}^{N_R} \tilde{x}_{j\omega} \tilde{y}_{j\omega}^* \right|^2$$

Note that by definition, due to the normalization,  $s_{\bar{x}\bar{x}\omega} = s_{\bar{y}\bar{y}\omega} = 1$ . In Eq. 54,  $\varphi_{X,Y}$  is the classical measure of phase synchronization.

The measure of instantaneous nonlinear dependence is:

$$\text{Eq. 55} \quad G_{X,Y}(\omega) = \ln \frac{1}{1 - [\text{Re}(s_{\bar{x}\bar{y}\omega})]^2} = -\ln(1 - \varphi_{X,Y}^2(\omega))$$

with instantaneous phase synchronization:

$$\text{Eq. 56} \quad \varphi_{X,Y}^2(\omega) = [\text{Re}(s_{\bar{x}\bar{y}\omega})]^2$$

which, not surprisingly, is directly related to the real part of the complex valued coherency of the normalized time series.

Finally, in the particular case of univariate time series, the measure of lagged nonlinear dependence is:

$$\text{Eq. 57} \quad G_{X \leftrightarrow Y}(\omega) = \ln \frac{1 - [\text{Re}(s_{\bar{x}\bar{y}\omega})]^2}{1 - [\text{Re}(s_{\bar{x}\bar{y}\omega})]^2 - [\text{Im}(s_{\bar{x}\bar{y}\omega})]^2} = -\ln(1 - \varphi_{X \leftrightarrow Y}^2(\omega))$$

with lagged phase synchronization:

$$\text{Eq. 58} \quad \varphi_{X \leftrightarrow Y}^2(\omega) = \frac{[\text{Im}(s_{\bar{x}\bar{y}\omega})]^2}{1 - [\text{Re}(s_{\bar{x}\bar{y}\omega})]^2}$$

The lagged phase synchronization between two univariate time series  $\varphi_{X \leftrightarrow Y}(\omega)$  given by Eq. 58 corresponds to the “general lagged phase synchronization” (i.e. the “zero-lag

removed" general phase synchronization)" previously defined in Pascual-Marqui (2007a), see Eq. 33 therein.

It is worth stressing the asymmetry in the results for the instantaneous phase synchronization  $\varphi_{X,Y}^2(\omega)$  (Eq. 56) and the lagged phase synchronization  $\varphi_{X \rightleftharpoons Y}^2(\omega)$  (Eq. 58). While the instantaneous phase synchronization is the real part of the complex valued coherency for the normalized time series, the lagged phase synchronization *is not* the imaginary part. Ideally, the lagged phase synchronization is a measure that is less affected by instantaneous nonlinear dependence.

In our previous related study (Pascual-Marqui 2007a), the definition given there for the "zero-lag removed general phase synchronization" (see Eq. 28 therein) was:

$$\text{Eq. 59} \quad PS_{GL} = \rho_{GL} = \sqrt{1 - \frac{\left| \begin{pmatrix} \mathbf{S}_{YY\omega} & \mathbf{S}_{YX\omega} \\ \mathbf{S}_{XY\omega} & \mathbf{S}_{XX\omega} \end{pmatrix} \right|}{\left| \text{Re} \begin{pmatrix} \mathbf{S}_{YY\omega} & \mathbf{S}_{YX\omega} \\ \mathbf{S}_{XY\omega} & \mathbf{S}_{XX\omega} \end{pmatrix} \right|}}$$

The new definition given here for the lagged phase synchronization  $\varphi_{X \rightleftharpoons Y}^2(\omega)$  is given by Eq. 52. Both definitions (Eq. 52 and Eq. 59) are identical for the case of two univariate time series. However, they are different for the multivariate case. Whereas the old definition in Eq. 59 lumps together all variables from X and Y, the new definition given here in Eq. 52 conserves the multivariate structure of the two multivariate time series. The improvement of the new lagged phase synchronization in Eq. 52 is that it measures the lagged nonlinear dependence between the two multivariate time series without being affected by the covariance structure within each multivariate time series. The shortcoming of the old definition from our previous study (Pascual-Marqui 2007a), shown in Eq. 59, is that it is contaminated by the dependence structures of the univariate time series within X and within Y.

## 7. Measures of nonlinear dependence (phase synchronization type) between groups of multivariate time series

Consider the case of three multivariate time series  $\mathbf{X}_{jt} \in \mathbb{R}^{p \times 1}$ ,  $\mathbf{Y}_{jt} \in \mathbb{R}^{q \times 1}$ , and  $\mathbf{Z}_{jt} \in \mathbb{R}^{r \times 1}$ , for discrete time  $t = 0 \dots N_T - 1$ , with  $j = 1 \dots N_R$  denoting the  $j$ -th time segment.

The measures of nonlinear dependence between the three multivariate time series are related in the usual way:

$$\text{Eq. 60} \quad G_{X,Y,Z}(\omega) = G_{X \rightleftharpoons Y \rightleftharpoons Z}(\omega) + G_{X,Y,Z}(\omega)$$

and are given by:

$$\text{Eq. 61} \quad G_{X,Y,Z}(\omega) = \ln \frac{\begin{pmatrix} \mathbf{S}_{\bar{Y}\bar{Y}\omega} & \mathbf{o} & \mathbf{o} \\ \mathbf{o} & \mathbf{S}_{\bar{X}\bar{X}\omega} & \mathbf{o} \\ \mathbf{o} & \mathbf{o} & \mathbf{S}_{\bar{Z}\bar{Z}\omega} \end{pmatrix}}{\begin{pmatrix} \mathbf{S}_{\bar{Y}\bar{Y}\omega} & \mathbf{S}_{\bar{Y}\bar{X}\omega} & \mathbf{S}_{\bar{Y}\bar{Z}\omega} \\ \mathbf{S}_{\bar{X}\bar{Y}\omega} & \mathbf{S}_{\bar{X}\bar{X}\omega} & \mathbf{S}_{\bar{X}\bar{Z}\omega} \\ \mathbf{S}_{\bar{Z}\bar{Y}\omega} & \mathbf{S}_{\bar{Z}\bar{X}\omega} & \mathbf{S}_{\bar{Z}\bar{Z}\omega} \end{pmatrix}}$$

$$\text{Eq. 62} \quad G_{X \cdot Y \cdot Z}(\omega) = \ln \frac{\left| \text{Re} \begin{pmatrix} \mathbf{S}_{\bar{Y}\bar{Y}\omega} & \mathbf{o} & \mathbf{o} \\ \mathbf{o} & \mathbf{S}_{\bar{X}\bar{X}\omega} & \mathbf{o} \\ \mathbf{o} & \mathbf{o} & \mathbf{S}_{\bar{Z}\bar{Z}\omega} \end{pmatrix} \right|}{\left| \text{Re} \begin{pmatrix} \mathbf{S}_{\bar{Y}\bar{Y}\omega} & \mathbf{S}_{\bar{Y}\bar{X}\omega} & \mathbf{S}_{\bar{Y}\bar{Z}\omega} \\ \mathbf{S}_{\bar{X}\bar{Y}\omega} & \mathbf{S}_{\bar{X}\bar{X}\omega} & \mathbf{S}_{\bar{X}\bar{Z}\omega} \\ \mathbf{S}_{\bar{Z}\bar{Y}\omega} & \mathbf{S}_{\bar{Z}\bar{X}\omega} & \mathbf{S}_{\bar{Z}\bar{Z}\omega} \end{pmatrix} \right|}$$

$$\text{Eq. 63} \quad G_{X \Leftrightarrow Y \Leftrightarrow Z}(\omega) = \ln \frac{\left| \text{Re} \begin{pmatrix} \mathbf{S}_{\bar{Y}\bar{Y}\omega} & \mathbf{o} & \mathbf{o} \\ \mathbf{o} & \mathbf{S}_{\bar{X}\bar{X}\omega} & \mathbf{o} \\ \mathbf{o} & \mathbf{o} & \mathbf{S}_{\bar{Z}\bar{Z}\omega} \end{pmatrix} \right|}{\left| \text{Re} \begin{pmatrix} \mathbf{S}_{\bar{Y}\bar{Y}\omega} & \mathbf{o} & \mathbf{o} \\ \mathbf{o} & \mathbf{S}_{\bar{X}\bar{X}\omega} & \mathbf{o} \\ \mathbf{o} & \mathbf{o} & \mathbf{S}_{\bar{Z}\bar{Z}\omega} \end{pmatrix} \right|} \left/ \frac{\left| \begin{pmatrix} \mathbf{S}_{\bar{Y}\bar{Y}\omega} & \mathbf{S}_{\bar{Y}\bar{X}\omega} & \mathbf{S}_{\bar{Y}\bar{Z}\omega} \\ \mathbf{S}_{\bar{X}\bar{Y}\omega} & \mathbf{S}_{\bar{X}\bar{X}\omega} & \mathbf{S}_{\bar{X}\bar{Z}\omega} \\ \mathbf{S}_{\bar{Z}\bar{Y}\omega} & \mathbf{S}_{\bar{Z}\bar{X}\omega} & \mathbf{S}_{\bar{Z}\bar{Z}\omega} \end{pmatrix} \right|}{\left| \begin{pmatrix} \mathbf{S}_{\bar{Y}\bar{Y}\omega} & \mathbf{o} & \mathbf{o} \\ \mathbf{o} & \mathbf{S}_{\bar{X}\bar{X}\omega} & \mathbf{o} \\ \mathbf{o} & \mathbf{o} & \mathbf{S}_{\bar{Z}\bar{Z}\omega} \end{pmatrix} \right|} \right\}$$

Phase synchronization for each type of measure of linear dependence in Eq. 60 can be defined by the general relation (see e.g. Pierce 1982):

$$\text{Eq. 64} \quad \varphi^2(\omega) = 1 - \exp[-G(\omega)]$$

The generalization of these definitions to any number of multivariate time series is straightforward.

It is important to emphasize here that these measures of nonlinear dependence for groups of multivariate time series can be applied in the field of neurophysiology. In this case, the time series consist of electric neuronal activity at several brain locations, and the measures of dependence are interpreted as “connectivity” between locations. When considering several brain locations, these new measures can be used to test for the existence of distributed cortical networks, whose activity can be estimated with exact low resolution brain electromagnetic tomography (Pascual-Marqui 2007b).

## 8. Measures of nonlinear dependence (phase synchronization type) between all univariate time series

A particular case of interest consists of measuring the nonlinear dependence between all the univariate time series that form part of the vector time series. For instance, consider the vector time series  $\mathbf{X}_{jt} \in \mathbb{R}^{p \times 1}$ . In this case, since each univariate time series on its own is

of interest, each one must be normalized. For this particular purpose we adopt the definition:

$$\text{Eq. 65} \quad \hat{\mathbf{X}}_{j\omega} = \left[ \text{Diag}(\mathbf{X}_{j\omega} \mathbf{X}_{j\omega}^*) \right]^{-1/2} \mathbf{X}_{j\omega}$$

which normalizes each variable. The corresponding covariance matrix is:

$$\text{Eq. 66} \quad \mathbf{S}_{\hat{\mathbf{X}}\hat{\mathbf{X}}\omega} = \frac{1}{N_R} \sum_{j=1}^{N_R} \hat{\mathbf{X}}_{j\omega} \hat{\mathbf{X}}_{j\omega}^*$$

Then the measures of nonlinear dependence between all “ $p$ ” univariate time series of  $\mathbf{X}$  are:

$$\text{Eq. 67} \quad G_{\mathbf{X},\mathbf{X}}(\omega) = G_{\mathbf{X} \leftrightarrow \mathbf{X}}(\omega) + G_{\mathbf{X} \cdot \mathbf{X}}(\omega)$$

$$\text{Eq. 68} \quad G_{\mathbf{X},\mathbf{X}}(\omega) = -\ln |\mathbf{S}_{\hat{\mathbf{X}}\hat{\mathbf{X}}\omega}|$$

$$\text{Eq. 69} \quad G_{\mathbf{X} \cdot \mathbf{X}}(\omega) = -\ln |\text{Re}(\mathbf{S}_{\hat{\mathbf{X}}\hat{\mathbf{X}}\omega})|$$

$$\text{Eq. 70} \quad G_{\mathbf{X} \leftrightarrow \mathbf{X}}(\omega) = G_{\mathbf{X},\mathbf{X}}(\omega) - G_{\mathbf{X} \cdot \mathbf{X}}(\omega) = \ln \frac{|\text{Re}(\mathbf{S}_{\hat{\mathbf{X}}\hat{\mathbf{X}}\omega})|}{|\mathbf{S}_{\hat{\mathbf{X}}\hat{\mathbf{X}}\omega}|}$$

Phase synchronization for each type of measure of linear dependence in Eq. 67 can be defined by the general relation (see e.g. Pierce 1982):

$$\text{Eq. 71} \quad \varphi^2(\omega) = 1 - \exp[-G(\omega)]$$

As a consistency check, it can easily be verified that when these definitions are applied to a vector time series with 2 components, the same results are obtained as in the case of two univariate time series (Eq. 53, Eq. 55, and Eq. 57).

## 9. Conclusions

1. Previous related work (Pascual-Marqui 2007a) was limited to measures of dependence between two multivariate time series. This study generalizes the definitions to include measures of dependence between any number of multivariate time series.

2. Previous measures for lagged dependence between two vector time series (Pascual-Marqui 2007a) were inadequately affected by the dependence structure of the univariate time series within each vector time series. This study adequately partials out the dependence structures within each vector time series.

3. A new measure for instantaneous linear and non-linear dependence is introduced.

4. The measures of dependence introduced here have been developed for discrete frequency components. However, they can as well be applied to any frequency band, defined as a set of discrete frequencies (which can even be disjoint). In this case, the Hermitian covariance matrices to be used in the equations for the measures of dependence should now correspond to the pooled matrices (i.e. the average Hermitian covariance over all discrete frequencies in the set defining the frequency band).

5. Inference methods for the measures of linear dependence are described.

6. All the measures of dependence can be based on any form of time-varying Fourier transforms or wavelets, such as, for instance, Gabor or Morlet transforms.

7. The new measures of dependence between any number of multivariate time series can be applied to the study of brain electrical activity, which can be estimated non-invasively from EEG/MEG recordings with methods such as eLORETA (Pascual-Marqui 2007b). When considering several brain locations jointly, these new measures can be used to test for the existence of distributed cortical networks. Previous methodology explores the connections between all possible pairs of locations, while the new "network approach" can test the joint dependence of several locations.

## ***Appendix 1: Zero-lag contribution to coherence and phase synchronization: problem description***

In some fields of application, the coherence or phase synchronization between two time series corresponding to two different spatial locations is interpreted as a measure of the "connectivity" between those two locations.

For example, consider the time series of scalp electric potential differences (EEG: electroencephalogram) at two locations. The coherence or phase synchronization is interpreted by some researchers as a measure of "connectivity" between the underlying cortices (see e.g. Nolte et al 2004 and Stam et al 2007).

However, even if the underlying cortices are not actually connected, significantly high coherence or phase synchronization might still occur due to the volume conduction effect: activity at any cortical area will be observed instantaneously (zero-lag) by all scalp electrodes.

As a possible solution to this problem, the electric neuronal activity distributed throughout the cortex can be estimated from the EEG by using imaging techniques such as standardized or exact low resolution brain electromagnetic tomography (sLORETA, eLORETA) (Pascual-Marqui et al 2002; Pascual-Marqui 2007b). At each voxel in the cortical grey matter, a 3-component vector time series is computed, corresponding to the current density vector with dipole moments along axes X, Y, and Z. This tomography has the unique properties of being linear, of having zero localization error, but of having low spatial resolution. Due to such spatial "blurring", the time series will again suffer from non-physiological inflated values of zero-lag coherence and phase synchronization.

Formally, consider two different spatial locations where there is no actual activity. However, due to a third truly active location, and because of low spatial resolution (or volume conductor type effect), there is some measured activity at these locations:

$$\text{Eq. 72: } \begin{cases} \mathbf{X}_{jt} = \mathbf{CZ}_{jt} + \boldsymbol{\varepsilon}_{jt}^x \\ \mathbf{Y}_{jt} = \mathbf{DZ}_{jt} + \boldsymbol{\varepsilon}_{jt}^y \end{cases}$$

where  $\mathbf{Z}_{jt}$  is the time series of the truly active location;  $\mathbf{C}$  and  $\mathbf{D}$  are matrices determined by the properties of the low spatial resolution problem; and  $\varepsilon_{jt}^x$  and  $\varepsilon_{jt}^y$  are independent and identically distributed random white noise.

In this model, although  $\mathbf{X}$  and  $\mathbf{Y}$  are not "connected", coherence and phase synchronization will indicate some connection, due to zero-lag spatial blurring.

Things can get even worse due to the zero-lag effect. Suppose that two time series are measured under two different conditions in which the zero-lag blurring effect is constant. The goal is to perform a statistical test to compare if there is a change in connectivity. Since the zero-lag effect is the same in both conditions, then it should seemingly not account for any significant difference in coherence or phase synchronization. However, this might be very misleading. In the model in Eq. 72, a simple increase in the signal to noise ratio (e.g. by increasing the norms of  $\mathbf{C}$  and  $\mathbf{D}$ ) will produce an increase in coherence and phase synchronization, due again to the zero-lag effect. This example shows that the zero-lag effect can render meaningless a comparison of two or more conditions.

## Acknowledgements

I have had extremely useful discussions with G. Nolte, who pointed out a number of embarrassing inconsistencies I wrote into the first draft of the previous related technical report (Pascual-Marqui 2007a). Those discussions partly motivated the new methods developed in this study.

## References

- C Allefeld, J Kurths (2004); Testing for phase synchronization. *International Journal of Bifurcation and Chaos*, 14: 405-416.
- DR Brillinger (1981): *Time series: data analysis and theory*. McGraw-Hill, New York.
- J Geweke (1982): Measurement of Linear Dependence and Feedback Between Multiple Time Series. *Journal of the American Statistical Association*, 77: 304-313.
- H Hotelling (1936): Relations between two sets of variables. *Biometrika* 28: 321-377.
- JT Kent (1983): Information gain and a general measure of correlation. *Biometrika*, 70: 163-173.
- S Kullback (1967): On Testing Correlation Matrices. *Applied Statistics*, 16: 80-85.
- BFJ Manly (1997): *Randomization, bootstrap and MonteCarlo methods in biology*. Chapman & Hall, London.
- KV Mardia, JT Kent, JM Bibby (1979): *Multivariate Analysis*. Academic Press, London.
- TE Nichols, AP Holmes (2001): Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human Brain Mapping*, 15: 1-25.
- G Nolte, O Bai, L Wheaton, Z Mari, S Vorbach, M Hallett (2004): Identifying true brain interaction from EEG data using the imaginary part of coherency. *Clin Neurophysiol.*, 115: 2292-2307.
- RD Pascual-Marqui (2002): Standardized low resolution brain electromagnetic tomography (sLORETA): technical details. *Methods & Findings in Experimental & Clinical Pharmacology*, 24D: 5-12.

RD Pascual-Marqui (2007a): Coherence and phase synchronization: generalization to pairs of multivariate time series, and removal of zero-lag contributions. arXiv:0706.1776v3 [stat.ME] 12 July 2007, <http://arxiv.org/abs/0706.1776> .

RD Pascual-Marqui (2007b): Discrete, 3D distributed, linear imaging methods of electric neuronal activity. Part 1: exact, zero error localization. arXiv:0710.3341 [math-ph], 2007-October-17, <http://arxiv.org/abs/0710.3341> .

E Pereda, R Quian-Quiroga, J Bhattacharya (2005): Nonlinear multivariate analysis of neurophysiological signals. *Prog Neurobiol.*, 77: 1-37.

DA Pierce (1982): Comment on J Geweke's "Measurement of Linear Dependence and Feedback Between Multiple Time Series". *Journal of the American Statistical Association*, 77: 315-316.

R Quian Quiroga, A Kraskov, T Kreuz, P Grassberger (2002): Performance of different synchronization measures in real data: a case study on electroencephalographic signals. *Phys. Rev. E*, 65: 041903.

MG Rosenblum, AS Pikovsky, J Kurths(1996): Phase synchronization of chaotic oscillators. *Phys. Rev. Lett.*, 76: 1804-1807.

CJ Stam, G Nolte, A Daffertshofer (2007): Phase lag index: Assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources. *Human Brain Mapping*, 28: 1178-1193.

P Tass, MG Rosenblum, J Weule, J Kurths, A Pikovsky, J Volkmann, A Schnitzler, HJ Freund (1998): Detection of n:m phase locking from noisy data: application to magnetoencephalography. *Phys. Rev. Lett.*, 81: 3291-3294.

# Dynamic interactions in terms of senders, hubs, and receivers (SHR) using the singular value decomposition of time series: Theory and brain connectivity applications

Roberto D. Pascual-Marqui<sup>1,2</sup> and Rolando J. Biscay-Lirio<sup>3,4</sup>

1: The KEY Institute for Brain-Mind Research, University Hospital of Psychiatry, Zurich, Switzerland

2: Department of Neuropsychiatry, Kansai Medical University, Osaka, Japan

3: Institute for Cybernetics, Mathematics, and Physics, Havana, Cuba

4: DEUV-CIMFAV, Facultad de Ciencias, Universidad de Valparaiso, Chile

Corresponding author:

Roberto D. Pascual-Marqui  
The KEY Institute for Brain-Mind Research  
University Hospital of Psychiatry  
Lenggstrasse 31  
CH-8032, Zurich  
Switzerland  
[pascualm@key.uzh.ch](mailto:pascualm@key.uzh.ch)

**Abstract:** Understanding of normal and pathological brain function requires the identification and localization of functional connections between specialized regions. The availability of high time resolution signals of electric neuronal activity at several regions offers information for quantifying the connections in terms of information flow. When the signals cover the whole cortex, the number of connections is very large, making visualization and interpretation very difficult. We introduce here the singular value decomposition of time-lagged multiple signals, which localizes the senders, hubs, and receivers (SHR) of information transmission. Unlike methods that operate on large connectivity matrices, such as correlation thresholding and graph-theoretic analyses, this method operates on the multiple time series directly, providing 3D brain images that assign a score to each location in terms of its sending, relaying, and receiving capacity. The scope of the method is general and encompasses other applications outside the field of brain connectivity.

## Introduction

Electric neuronal activity can be recorded invasively from the human brain by means of intracranial electrodes (see e.g. Crone et al, 2009). The time series of electric potentials provided by such electrodes can have very high time resolution. In particular, if a pair of such electrodes is placed with a small separation distance, then the local electric potential difference is a gradient, and is proportional to the current density vector projected onto the line joining the electrodes, according to Ohm's law (see e.g. Sarvas, 1987):

Eq. 1  $-\sigma \nabla \varphi = \mathbf{J}$

where the electric neuronal activity is given by the current density  $\mathbf{J}$ ,  $\sigma$  is conductivity,  $\varphi$  is electric potential, and  $\nabla$  is the gradient operator.

Such time series contain local information on brain function.

In practice, it is very desirable to be able to obtain such information non-invasively. This can be achieved by computing the current density in the brain, from non-invasive electric potential differences recorded on the scalp, i.e. from the EEG.

In particular, the method of choice in this paper for solving this inverse problem is exact low resolution electromagnetic tomography (eLORETA), see e.g. Pascual-Marqui (2007, 2009). This method is a multivariate linear solution to the EEG inverse problem, and is endowed with the property of exact localization response when probed with Dirac deltas (which in this case are test dipoles located anywhere in the brain). Due to the principles of linearity and superposition, the method will perform well with any distribution of current density, albeit with low spatial resolution.

Regardless of the technique used for obtaining the current density, either invasive (with intracranial electrodes) or non-invasive (computed from the EEG by means of a validated inverse solution), the method to be described here can be used for revealing connections in the brain. In particular, the new method is capable of localizing and distinguishing senders, hubs, and receivers of information transmission in the brain.

### Stationary case

Let  $U_{i,t}$  denote the current density at the  $i$ -th voxel, at time  $t$ ; with  $i=1\dots N_v$ ,  $N_v$  denoting the number of cortical voxels;  $t=1\dots N_T$ ,  $N_T$  denoting the number of time frames (discrete time samples). At the  $i$ -th voxel, consider the univariate autoregressive model:

$$\text{Eq. 2} \quad U_{i,t} = \sum_{k=1}^Q a_{i,k} U_{i,t-k} + V_{i,t}$$

where  $Q \geq 2$  denotes the global autoregressive order,  $a_{i,k}$  are the auto-regression coefficients, and  $V_{i,t}$  is the innovation time series.

In a first step, it will be required to fit the model in Eq. 2 separately to each voxel, in order to obtain the innovations. The vector of innovation time series containing all voxels is denoted as:

$$\text{Eq. 3} \quad \mathbf{V}_t \in \mathbb{R}^{N_v \times 1}$$

defined for  $t = (Q+1)\dots N_T$ .

Now define the vector:

$$\text{Eq. 4} \quad \mathbf{Z}_t = \begin{pmatrix} \mathbf{V}_t \\ \mathbf{V}_{t-1} \\ \vdots \\ \mathbf{V}_{t-Q} \end{pmatrix} \in \mathbb{R}^{[(Q+1)N_v] \times 1}$$

defined for  $t = (2Q+1)\dots N_T$ , which contains the innovation current density at time  $t$  and its past, back to  $Q$  time units into the past.

Next, form the data matrix:

$$\text{Eq. 5} \quad \mathbf{Z} = \begin{pmatrix} \mathbf{Z}_{2Q+1} & \mathbf{Z}_{2Q+2} & \dots & \mathbf{Z}_{N_T} \end{pmatrix} \in \mathbb{R}^{[(Q+1)N_v] \times (N_T - 2Q)}$$

and normalize each row, i.e. the time series at each row of the data matrix  $\mathbf{Z}$  should have zero mean and unit variance. When  $\mathbf{Z}$  is normalized in this way, then  $\mathbf{Z}\mathbf{Z}^T$  corresponds to the common cross-correlation matrix for multivariate time series.

Consider the singular value decomposition (SVD) of  $\mathbf{Z}$ :

Eq. 6  $\mathbf{Z} = \mathbf{L}\mathbf{\Lambda}\mathbf{R}^T$

with  $\mathbf{L} \in \mathbb{R}^{[(Q+1)N_v] \times K}$  containing the left eigenvectors,  $\mathbf{R} \in \mathbb{R}^{(N_t-Q) \times K}$  containing the right eigenvectors,  $\mathbf{\Lambda} \in \mathbb{R}^{K \times K}$  is diagonal and contains the eigenvalues in descending order, and  $K = \min\left\{\left[(Q+1)N_v\right], (N_t-2Q)\right\}$ . Both  $\mathbf{L}$  and  $\mathbf{R}$  are orthonormal, i.e.  $\mathbf{L}^T\mathbf{L} = \mathbf{I}$  and  $\mathbf{R}^T\mathbf{R} = \mathbf{I}$ .

The main feature of interest is the first column of  $\mathbf{L}$ , i.e. the first left eigenvector denoted as  $\mathbf{\Gamma} \in \mathbb{R}^{[(Q+1)N_v] \times 1}$ , corresponding to the largest eigenvalue. As can be seen from the structure of the data matrix defined by Eq. 4, we note that the first  $N_v$  elements of  $\mathbf{\Gamma}$  correspond to the present, the next  $N_v$  elements correspond to one step in the past, and so on:

Eq. 7  $\mathbf{\Gamma} = \begin{pmatrix} \mathbf{\Gamma}_t \\ \mathbf{\Gamma}_{t-1} \\ \vdots \\ \mathbf{\Gamma}_{t-Q} \end{pmatrix} \in \mathbb{R}^{[(Q+1)N_v] \times 1}$

Definitions:

1. The elements of  $\mathbf{\Gamma}_t \in \mathbb{R}^{N_v \times 1}$  quantify the “receiving” function at each voxel.
2. The elements of  $\mathbf{\Gamma}_{t-Q} \in \mathbb{R}^{N_v \times 1}$  quantify the “sending” function at each voxel.
3. The elements of  $\mathbf{\Gamma}_{t-k} \in \mathbb{R}^{N_v \times 1}$ , for  $k = 1 \dots (Q-1)$ , quantify the “hub” function at each voxel.

The first right eigenvector of  $\mathbf{R}$ , denoted as  $\mathbf{R}_1 \in \mathbb{R}^{(N_t-2Q) \times 1}$ , is the time series that expresses the dynamics of the senders, hubs, and receivers given by  $\mathbf{\Gamma}$ .

### Notes and motivation

At the heart of the concepts of “senders, hubs, and receivers” (SHR), rests Granger causality (1967), which is based on the analysis of the innovation time series, and not the original time series. This is one reason for computing the SVD on the innovation time series.

Furthermore, the explicit dependence of the current density at a given voxel with its own past might be a confounding factor for the hub function: a voxel sending and receiving large amounts of information is by definition a hub, even if it sends and receives from its own self. Therefore, by partialling out the univariate auto-dependence at each voxel, the residuals should be better at characterizing the hub.

At the heart of using the largest left eigenvector of the data matrix constructed in Eq. 5 is the method advocated by Worsley et al (2005). In that paper, it is shown that this eigenvector reveals functional connectivity between voxels. By augmenting the time series at each voxel with their time-shifted pasts, we now additionally have information on “Granger-causal functional connectivity”.

The structure of the data matrix as defined in Eq. 4 carries over to the left eigenvector. This allows the interpretation of its different blocks in terms of senders, hubs, and receivers. For instance, note that:

1.  $\Gamma_t$  is the ultimate receiver, (Granger-) causally being influenced by the pasts  $\Gamma_{t-1} \dots \Gamma_{t-Q}$ .
2.  $\Gamma_{t-Q}$  is the ultimate sender, (Granger-) causally influencing all the immediate future  $\Gamma_t \dots \Gamma_{t-Q+1}$ .
3. The intermediate time blocks  $\Gamma_{t-1} \dots \Gamma_{t-Q+1}$  can send and receive, and are therefore related to hubs, i.e. to relay stations.

If the global auto-regressive order  $Q$  is set to 1, then the hub function is not defined, although senders and receivers remain well defined.

In practice, when the number of voxels is very large compared to the number of time samples, it might be efficient to compute only the largest left and right eigenvectors by applying the power method to the matrix  $\mathbf{Z}$ .

In the individual univariate time series for each voxel (Eq. 2), it is possible to consider the use of different autoregressive order values, as long as they are at least as large as the global  $Q$ .

### Locally stationary case

It will be assumed that the current density time series can be repeatedly sampled, and that each sample starts at a repeated event. For instance, each sample might correspond to the presentation of a visual stimulus, and all samples are time locked to the moment of stimulus onset.

Now let  $U_{i,t,j}$  denote the current density at the  $i$ -th voxel, at time  $t$ , for the  $j$ -th sample (e.g. for the  $j$ -th stimulus presentation), with  $j=1 \dots N_s$ ,  $N_s$  denoting the number of samples (e.g. the number of stimuli).

We will consider local univariate autoregressive models of order  $Q$ , specified at the target time  $\tau$ , defined at values  $\tau=(2Q+1) \dots N_s$ . The model will be locally valid for time instants in the immediate past of the target time, for  $(\tau-2Q) \leq t \leq \tau$ .

In what follows, the target time  $\tau$  is considered fixed.

For the  $i$ -th voxel, the model is:

$$\text{Eq. 8} \quad U_{i,t,j} = \sum_{k=1}^Q a_{\tau,i,k} U_{i,t-k,j} + V_{\tau,i,t,j}, \text{ for } (\tau-2Q) \leq t \leq \tau$$

where  $a_{\tau,i,k}$  are the auto-regression coefficients, and  $V_{\tau,i,t,j}$  is the innovation time series.

Note that the available data for estimating this model consists of all the samples  $j=1 \dots N_s$ , but only the local time instants  $(\tau-2Q) \leq t \leq \tau$ . From here we estimate the coefficients  $a_{\tau,i,k}$ , and the innovations for all samples ( $j=1 \dots N_s$ ), at  $t=\tau$ ,  $t=\tau-1$ , ...,  $t=\tau-Q$ . The vector of innovation time series containing all voxels is denoted as:

**Eq. 9**  $\mathbf{V}_{\tau,t,j} \in \mathbb{R}^{N_V \times 1}$

defined for local times  $\tau - Q \leq t \leq \tau$ , and for all samples  $j = 1 \dots N_S$ .

Now define the vector:

**Eq. 10**  $\mathbf{Z}_{\tau,j} = \begin{pmatrix} \mathbf{V}_{\tau,t,j} \\ \mathbf{V}_{\tau,t-1,j} \\ \vdots \\ \mathbf{V}_{\tau,t-Q,j} \end{pmatrix} \in \mathbb{R}^{[(Q+1)N_V] \times 1}$

which contains the innovation current density for the  $j$ -th sample, at time  $t$  and its past, back to  $Q$  time units into the past.

Next, form the data matrix:

**Eq. 11**  $\mathbf{Z}_{\tau} = \begin{pmatrix} \mathbf{Z}_{\tau,1} & \mathbf{Z}_{\tau,2} & \dots & \mathbf{Z}_{\tau,N_S} \end{pmatrix} \in \mathbb{R}^{[(Q+1)N_V] \times N_S}$

and normalize each row, i.e. the sample values at each row of the data matrix  $\mathbf{Z}_{\tau}$  should have zero mean and unit variance. When  $\mathbf{Z}_{\tau}$  is normalized in this way, then  $\mathbf{Z}_{\tau} \mathbf{Z}_{\tau}^T$  corresponds to the cross-correlation matrix at target time  $\tau$  for local multivariate time series.

As in the stationary case above, the first left eigenvector of the SVD of  $\mathbf{Z}_{\tau}$ , now denoted as  $\mathbf{\Gamma}_{\tau}$ , contains all the relevant information on senders, hubs, and receivers, at each target time  $\tau$ .

## References

Crone NE, Korzeniewska A, Ray S, and Franaszczuk PJ. Cortical function mapping with intracranial EEG. Chapter 14, pp: 369-399. In "Quantitative EEG Analysis: Methods and Applications", Eds: S. Tong and N. Thakor; 2009 Artech House, Boston

Granger CWJ. Investigating causal relations by econometric models and crossspectral methods. *Econometrica*, 1969; 37: 424-438.

Pascual-Marqui RD. Discrete, 3D Distributed, Linear Imaging Methods of Electric Neuronal Activity. Part 1: Exact, Zero Error Localization. 2007. arXiv:0710.3341 [math-ph], 2007-October-17, <http://arxiv.org/pdf/0710.3341>.

Pascual-Marqui RD: Theory of the EEG Inverse Problem, Chapter 5, pp: 121-140. In "Quantitative EEG Analysis: Methods and Applications", Eds: S. Tong and N. Thakor; 2009 Artech House, Boston

Sarvas J. Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. *Physics in Medicine and Biology* 1987; 32: 11-22.

Worsley KJ, Chen JI, Lerch J, Evans AC. Comparing functional connectivity via thresholding correlations and singular value decomposition. *Phil. Trans. R. Soc. B*, 2005; 360: 913-920.

## Interaction patterns of brain activity across space, time and frequency. Part I: methods

Roberto D. Pascual-Marqui<sup>1,2</sup> and Rolando J. Biscay-Lirio<sup>3,4</sup>

1: The KEY Institute for Brain-Mind Research, University Hospital of Psychiatry, Zurich, Switzerland

2: Department of Neuropsychiatry, Kansai Medical University Hospital, Osaka, Japan

3: Institute for Cybernetics, Mathematics, and Physics, Havana, Cuba

4: DEUV-CIMFAV, Facultad de Ciencias, Universidad de Valparaíso, Chile

Corresponding author:

Roberto D. Pascual-Marqui

The KEY Institute for Brain-Mind Research  
University Hospital of Psychiatry  
Lenggstrasse 31  
CH-8032 Zurich  
Switzerland  
pascualm {at} key.uzh.ch  
[www.keyinst.uzh.ch/loreta](http://www.keyinst.uzh.ch/loreta)  
and:

Department of Neuropsychiatry  
Kansai Medical University Hospital  
10-15, Fumizono-cho, Moriguchi  
Osaka, 570-8507  
Japan  
pascualr {at} takii.kmu.ac.jp

### Abstract

We consider exploratory methods for the discovery of cortical functional connectivity. Typically, data for the  $i$ -th subject ( $i = 1 \dots N_s$ ) is represented as  $\mathbf{X}_i \in \mathbb{R}^{N_v \times N_t}$ , corresponding to brain activity sampled at  $N_t$  moments in time from  $N_v$  cortical voxels. A widely used method of analysis first concatenates all subjects along the temporal dimension, and then performs an independent component analysis (ICA) for estimating the common cortical patterns of functional connectivity. There exist many other interesting variations of this technique, as reviewed in [Calhoun et al. 2009 Neuroimage 45: S163-172].

We present methods for the more general problem of discovering functional connectivity occurring at all possible time lags. For this purpose, brain activity is viewed as a function of space and time, which allows the use of the relatively new techniques of functional data analysis [Ramsay & Silverman 2005: Functional data analysis. New York: Springer]. In essence, our method first vectorizes the data from each subject  $\mathbf{X}_i^{vec} \in \mathbb{R}^{(N_t N_v) \times 1}$ , which constitutes the natural discrete representation of a function of several variables, followed by concatenation of all subjects. The singular value decomposition (SVD), as well as the ICA of this new matrix will reveal spatio-temporal patterns of connectivity. As a further example, in the case of EEG neuroimaging,  $\mathbf{X}_i \in \mathbb{R}^{N_v \times N_\Omega}$  may represent spectral density for electric neuronal activity at  $N_\Omega$  discrete frequencies from  $N_v$  cortical voxels, from the  $i$ -th EEG epoch. In this case our functional data analysis approach would reveal coupling of brain regions at possibly different frequencies.

## 1. Introduction

For the sake of simplicity, a particular example will be used for explaining the methods. Straightforward generalizations will be considered in a later Section.

Let  $\mathbf{X}_i \in \mathbb{R}^{N_v \times N_T}$  denote brain activity for the  $i$ -th subject ( $i=1\dots N_s$ ), sampled at  $N_T$  moments in time from  $N_v$  cortical voxels. Based on such data, it is of interest to find the interactions between different brain regions. This is the topic of “functional connectivity”.

Many methods of analysis exist for the study of functional connectivity. Recent reviews are presented in [1] and [2]. Two methods are of particular interest here: one based on the SVD [3], and the other based on group ICA [1].

Let:

$$\text{Eq. 1} \quad \mathbf{Y} = (\mathbf{X}_1 \quad \mathbf{X}_2 \quad \dots \quad \mathbf{X}_{N_s}) \in \mathbb{R}^{N_v \times (N_s N_T)}$$

denote the matrix obtained by the temporal concatenation of the subjects. It will be required that the elements of each row have zero mean, i.e. that the concatenated time series at each voxel have zero mean:

$$\text{Eq. 2} \quad \mathbf{Y}\mathbf{1} = \mathbf{0}$$

where  $\mathbf{1}$  and  $\mathbf{0}$  are vectors of ones and zeros, respectively.

Consider the spatial covariance matrix:

$$\text{Eq. 3} \quad \mathbf{C}_{\mathbf{Y}\mathbf{Y}} = \frac{1}{N_s N_T} \mathbf{Y}\mathbf{Y}^T \in \mathbb{R}^{N_v \times N_v}$$

and its corresponding correlation matrix:

$$\text{Eq. 4} \quad \mathbf{R}_{\mathbf{Y}\mathbf{Y}} = [\text{diag}(\mathbf{C}_{\mathbf{Y}\mathbf{Y}})]^{-1/2} \mathbf{C}_{\mathbf{Y}\mathbf{Y}} [\text{diag}(\mathbf{C}_{\mathbf{Y}\mathbf{Y}})]^{-1/2}$$

where the “diag” operator returns a diagonal matrix by setting all off-diagonal elements to zero.

Then, as demonstrated by Worsley et al [3], the largest normalized eigenvector of  $\mathbf{R}_{\mathbf{Y}\mathbf{Y}}$ , denoted as  $\boldsymbol{\Gamma}_y \in \mathbb{R}^{N_v \times 1}$ , will detect regions of correlated voxels. In practice, this is achieved by thresholding the brain image corresponding to the eigenvector. Those elements with large absolute value will convey information on the correlated brain regions. The method of Worsley et al [3] was recently extended for the detection of senders, hubs, and receivers of cortical information transactions [4].

A commonly used related approach is known as group ICA with temporal concatenation [1], where the matrix  $\mathbf{Y}$  in Eq. 1, which must satisfy the condition in Eq. 2, can be factorized as:

$$\text{Eq. 5} \quad \mathbf{Y} = \mathbf{A}_y \mathbf{S}_y$$

with  $\mathbf{A}_y \in \mathbb{R}^{N_v \times K}$ ,  $\mathbf{S}_y \in \mathbb{R}^{K \times (N_s N_T)}$ , and  $K \leq \text{rank}(\mathbf{Y})$  denoting the number of components. This form of factorization is typical with EEG related data, while the factorization for the transposed of  $\mathbf{Y}$  is typical of fMRI data (see e.g. [5]). Ideally, the  $K$  time series in the matrix  $\mathbf{S}_y$  should be statistically independent in a strict sense, which can be approximately achieved in many different ways (see e.g. [6]). The columns of the matrix  $\mathbf{A}_y$  contain the spatial components, where each one provides information on the correlated brain regions.

Note: The temporal concatenation of data in Eq. 1 is just one possibility. Spatial concatenation is another example. Other data organization schemes are also possible, such as the three dimensional array in tensorial or PARAFAC analyses (see review in e.g. [1]).

These methodologies are of proven value in the discovery of functional connectivity. When they are interpreted from the point of view of functional data analysis [7], new generalizations can be derived, giving detailed temporal information about the nature of the connectivity patterns. The aim of this study is to present a functional data analysis approach to functional connectivity that allows the discovery of brain interactions across space (cortical locations), time, and frequency.

## 2. Functional data analysis perspective

Typically, measures of connectivity are based on the “similarity” between the time series recorded at two different locations. A simple similarity index is, for instance, the cross-correlation coefficient. However, it is nearly impossible to analyze the massive number of similarities when one considers all possible pairs of voxels at all possible time lags.

A solution to this problem can be obtained by considering the basic data as a function of several variables: space (cortical voxels) and time. This is the approach used in functional data analysis [7]. The data from each subject, consisting of brain activity, is now represented as a vector:

$$\text{Eq. 6} \quad \mathbf{X}_i \in \mathbb{R}^{N_V \times N_T} \rightarrow \mathbf{X}_i^{vec} \in \mathbb{R}^{(N_T N_V) \times 1}$$

where the “vec” operator transforms a matrix into a vector by stacking the columns of the matrix one underneath the other [8]. Thus, the elements of the vector correspond to brain activity values sampled at points in the (space, time) hyperplane.

The new group data matrix is now defined as follows:

$$\text{Eq. 7} \quad \mathbf{Z} = \begin{pmatrix} \mathbf{X}_1^{vec} & \mathbf{X}_2^{vec} & \dots & \mathbf{X}_{N_S}^{vec} \end{pmatrix} \in \mathbb{R}^{(N_T N_V) \times N_S}$$

This is the basic idea behind functional data analysis, and it may seem deceptively simple, but in fact it is radically different from any other published form of group analysis [1], [9], [10].

## 3. The functional singular value decomposition (fSVD) approach

Here we apply the SVD method described in [3] to the functional data defined in Eq. 7. The first requirement is to center the data to have zero mean value for the elements of each row, such that:

$$\text{Eq. 8} \quad \mathbf{Z}\mathbf{1} = \mathbf{0}$$

as in Eq. 2. Next, consider the very high dimensional covariance matrix:

$$\mathbf{C}_{\mathbf{ZZ}} = \frac{1}{N_S} \mathbf{Z}\mathbf{Z}^T \in \mathbb{R}^{(N_T N_V) \times (N_T N_V)}$$

and its corresponding correlation matrix:

$$\text{Eq. 9} \quad \mathbf{R}_{\mathbf{ZZ}} = [\text{diag}(\mathbf{C}_{\mathbf{ZZ}})]^{-1/2} \mathbf{C}_{\mathbf{ZZ}} [\text{diag}(\mathbf{C}_{\mathbf{ZZ}})]^{-1/2}$$

Then, based on the method of Worsley et al [3], the largest normalized eigenvector of  $\mathbf{R}_{zz}$ , denoted as  $\Gamma_z \in \mathbb{R}^{(N_T N_V) \times 1}$ , will detect the time course of the regions of correlated voxels. In practice, this is achieved by thresholding the time varying brain images corresponding to the eigenvector. Those elements with large absolute value will convey information on the time course of the correlated brain regions.

### 3.1. Interpretation example

For instance, after appropriately thresholding  $\Gamma_z$ , if brain region  $A$  at an early latency  $\tau_A$  has high values, and is followed by high values in a different brain region  $B$  at a later latency  $\tau_B$ , then the interpretation is that brain regions  $A$  and  $B$  are cross-correlated with the time lag  $|\tau_A - \tau_B|$ .

Such cross-spatial and cross-temporal connections can be exposed without having to explore nor calculate and analyze explicitly all pairwise cross-correlations.

### 3.2. A practical algorithm

With respect to the practical aspect of computations, note that it is not necessary to perform the SVD on the very high dimensional correlation matrix  $\mathbf{R}_{zz}$ . All that is needed is the largest left eigenvector of the matrix:

$$\text{Eq. 10 } \mathbf{U} = \left\{ \left[ \text{diag}(\mathbf{C}_{zz}) \right]^{-1/2} \mathbf{Z} \right\} \in \mathbb{R}^{(N_T N_V) \times N_s}$$

where  $\mathbf{Z}$  must satisfy Eq. 8. Typically,  $N_s \ll (N_T N_V)$ , which allows for a very efficient calculation, using for instance, the iterative power method.

## 4. The functional independent component analysis (fICA) approach

Consider the functional data matrix defined in Eq. 7, satisfying Eq. 8. The functional ICA model is:

$$\text{Eq. 11 } \mathbf{Z} = \mathbf{A}_z \mathbf{S}_z$$

with  $\mathbf{A}_z \in \mathbb{R}^{(N_T N_V) \times K}$ ,  $\mathbf{S}_z \in \mathbb{R}^{K \times N_s}$ , and  $K \leq \text{rank}(\mathbf{Z})$  denoting the number of components. As above (see Eq. 5), it is required that the  $K$  rows in the matrix  $\mathbf{S}_z$  should be statistically independent in a strict sense, which can be approximately achieved in many different ways (see e.g. [6]). Thus, each component, corresponding to a column of the matrix  $\mathbf{A}_z$ , conveys information on the time course of the correlated brain regions related to that component. This means that the interpretation explained above in Subsection 3.1 applies for each independent component here.

## 5. Generalizations

The methods presented here can be applied to other brain activity data, especially when using an EEG tomography such as LORETA [11-13].

In one example the basic data  $\mathbf{X}_i \in \mathbb{R}^{N_V \times N_\Omega}$  may represent spectral density for electric neuronal activity at  $N_\Omega$  discrete frequencies from  $N_V$  cortical voxels, from the  $i$ -th EEG epoch. In this case our functional data analysis approach would reveal coupling of brain regions at

possibly different frequencies. For instance, the method may reveal coupling of frontal gamma activity with occipital theta activity.

In an event related potential (ERP) experiment, when analyzing the collection of single trial epochs, a time-varying spectral analysis will produce extremely high dimensional functional data, consisting of spectral density for electric neuronal activity at  $N_\omega$  discrete frequencies from  $N_v$  cortical voxels, as a function of time (relative to stimulus onset), for each stimulus (i.e. each epoch). In this case the functional components, either with fSVD or fICA may reveal coupling of different frequencies at different moments in time between different brain regions.

## References

- 1 Calhoun VD, Liu J, Adali T. A review of group ica for fmri data and ica for joint inference of imaging, genetic, and erp data. *Neuroimage*. 2009; **45**: S163-172.
- 2 Li K, Guo L, Nie J, Li G, Liu T. Review of methods for functional brain connectivity detection using fmri. *Comput Med Imaging Graph*. 2009; **33**: 131-139.
- 3 Worsley KJ, Chen JI, Lerch J, Evans AC. Comparing functional connectivity via thresholding correlations and singular value decomposition. *Philos Trans R Soc Lond B Biol Sci*. 2005; **360**: 913-920.
- 4 Pascual-Marqui RD, Biscay-Lirio RJ. Dynamic interactions in terms of senders, hubs, and receivers (shr) using the singular value decomposition of time series: Theory and brain connectivity applications. *arXiv:10090796v2 [statME]*. 2010; <http://arxiv.org/abs/1009.0796>.
- 5 Eichele T, Calhoun VD, Debener S. Mining eeg-fmri using independent component analysis. *Int J Psychophysiol*. 2009; **73**: 53-61.
- 6 Cichocki A, Amari S. *Adaptive blind signal and image processing: Learning algorithms and applications*. New York: John Wiley & Sons, 2002.
- 7 Ramsay JO, Silverman BW. *Functional data analysis*. 2nd edn. New York: Springer, 2005.
- 8 Magnus JR, Neudecker H. *Matrix differential calculus with applications in statistics and econometrics*. Rev. edn. New York: John Wiley, 1999.
- 9 Guo Y, Pagnoni G. A unified framework for group independent component analysis for multi-subject fmri data. *Neuroimage*. 2008; **42**: 1078-1093.
- 10 Varoquaux G, Sadaghiani S, Pinel P, Kleinschmidt A, Poline JB, Thirion B. A group model for stable multi-subject ica on fmri datasets. *Neuroimage*. 2010; **51**: 288-299.
- 11 Pascual-Marqui RD. Standardized low-resolution brain electromagnetic tomography (sloreta): Technical details. *Methods Find Exp Clin Pharmacol*. 2002; **24 Suppl D**: 5-12.
- 12 Pascual-Marqui RD. Theory of the eeg inverse problem. In: Tong S, Thakor N, eds. *Quantitative eeg analysis: Methods and clinical applications*. Boston: Artech House 2009; 121-140.
- 13 Pascual-Marqui RD, Michel CM, Lehmann D. Low resolution electromagnetic tomography: A new method for localizing electrical activity in the brain. *Int J Psychophysiol*. 1994; **18**: 49-65.

# Cortical current source connectivity by means of partial coherence fields

Roberto D. Pascual-Marqui<sup>1,2</sup>, Rolando J. Biscay<sup>3</sup>, Pedro A. Valdes-Sosa<sup>4</sup>, Jorge Bosch-Bayard<sup>5</sup>, Jorge J. Riera-Diaz<sup>6</sup>

1: The KEY Institute for Brain-Mind Research, University Hospital of Psychiatry, Zurich, Switzerland (pascualm@key.uzh.ch)

2: Department of Neuropsychiatry, Kansai Medical University Hospital, Osaka, Japan (pascualr@takii.kmu.ac.jp)

3: DEUV-CIMFAV, Facultad de Ciencias, Universidad de Valparaiso, Chile (rolando.biscay@uv.cl)

4: Cuban Neuroscience Center, Havana, Cuba (peter@cneuro.edu.cu)

5: Cuban Neuroscience Center, Havana, Cuba (bosch@cneuro.edu.cu)

6: Institute of Development, Aging and Cancer; Tohoku University, Sendai, Japan(riera@idac.tohoku.ac.jp)

## 1. Abstract

An important field of research in functional neuroimaging is the discovery of integrated, distributed brain systems and networks, whose different regions need to work in unison for normal functioning.

The EEG is a non-invasive technique that can provide information for massive connectivity analyses. Cortical signals of time varying electric neuronal activity can be estimated from the EEG. Although such techniques have very high time resolution, two cortical signals even at distant locations will appear to be highly similar due to the low spatial resolution nature of the EEG.

In this study a method for eliminating the effect of common sources due to low spatial resolution is presented. It is based on an efficient estimation of the whole-cortex partial coherence matrix. Using as a starting point any linear EEG tomography that satisfies the EEG forward equation, it is shown that the generalized partial coherences for the cortical grey matter current density time series are invariant to the selected tomography. It is empirically shown with simulation experiments that the generalized partial coherences have higher spatial resolution than the classical coherences. The results demonstrate that with as little as 19 electrodes, lag-connected brain regions can often be missed and misplaced even with lagged coherence measures, while the new method detects and localizes correctly the connected regions using the lagged partial coherences.

## 2. Introduction

In its early development, methods of analysis in functional neuroimaging were aimed at the “localization” of “effects”. For instance, by comparing normal control subjects and patients suffering schizophrenia (“effect”), a small number of brain regions (“localization”) were claimed to be responsible for the disorder. The field has evolved and moved on, and it is now more frequent to find methods aimed at the discovery of integrated distributed systems and

networks, whose different brain regions need to work in unison for normal functioning. An example of such methods used for this purpose is based on the analysis of massive amounts of pairwise similarity measures between cortical signals, i.e. on the analysis of the extremely high dimensional similarity matrix given by correlations or coherences.

Extensive general reviews on these types of methods can be found in Valdes-Sosa et al (2011) and Sporns (2011).

In the case of metabolic functional neuroimaging methods such as fMRI and PET, there is high spatial resolution, but low time resolution, with signals having most of their spectral power concentrated at frequencies lower than 0.1 Hz. In the case of EEG-based neuroimaging, the spatial resolution is lower, but with very high time resolution, with effective sampling rates typically higher than 100 Hz.

The problem of interest in this study consists of estimating intracortical connectivities from computed signals of electric neuronal activity, obtained from non-invasive scalp EEG recordings. The correlation or coherence matrices have very high dimension, equal to the number of cortical grey matter voxels at which the current density is computed (typically more than 6000), but these matrices are of low rank, due to the small number of scalp electrodes (typically ranging from 19 to 128). Due to the low spatial resolution nature of linear, discrete, distributed EEG inverse solutions, the similarity between pairs of computed cortical signals will be much higher than the true value.

The solution presented here consists of computing partial coherences, which by themselves are of great interest because they provide information on non-mediated direct connections. In addition, partial coherences would interpret the low spatial resolution effect as “common sources”, thus decreasing this effect in the estimated connectivities. The whole cortex partial coherence can be obtained from the inverse of the coherence matrix. However, since this is a very low rank matrix, the inverse does not exist. In its place a generalized inverse coherence matrix is computed, thus producing the generalized whole cortex partial coherence, endowed with higher spatial resolution. Very simple and efficient equations for the whole-cortex partial coherence are derived.

### 3. Method

#### 3.1. Basic equations and definitions

Let  $\Phi_{t,i} \in \mathbb{R}^{N_E \times 1}$  denote the time domain EEG, for  $N_E$  electrodes,  $N_T$  discrete time samples (with  $t = 1 \dots N_T$ ), for  $N_S$  epochs (with  $i = 1 \dots N_S$ ).

The EEG forward equation is:

**Eq. 1**  $\Phi_{t,i} = \mathbf{K} \mathbf{J}_{t,i}$

where  $\mathbf{J}_{t,i} \in \mathbb{R}^{N_V \times 1}$  is the current density,  $\mathbf{K} \in \mathbb{R}^{N_E \times N_V}$  denotes the lead field matrix, and  $N_V$  is the number of cortical grey matter voxels. Typically, the number of voxels is much larger than the number of scalp measurements, i.e.  $N_V \gg N_E$ . And without loss of generality, it will be assumed that the lead field matrix is of full row rank, i.e. for the typical situation when  $N_V \gg N_E$ ,  $\text{rank}(\mathbf{K}) = N_E$ .

Without loss of generality, the derivations presented here refer to EEG, but they equivalently are valid for MEG.

Any time domain linear transform, such as the discrete Fourier transform, can be applied to Eq. 1, giving:

$$\text{Eq. 2} \quad \Phi_{\omega,i} = \mathbf{K} \mathbf{J}_{\omega,i}$$

where  $\Phi_{\omega,i} \in \mathbb{C}^{N_E \times 1}$  and  $\mathbf{J}_{\omega,i} \in \mathbb{C}^{N_V \times 1}$  denote the corresponding discrete Fourier transforms at the discrete frequency  $\omega$ . Other transforms can be accommodated, such as wavelets or time varying Fourier transforms.

The generalized linear inverse solution is:

$$\text{Eq. 3} \quad \hat{\mathbf{J}}_{\omega,i} = \mathbf{T} \Phi_{\omega,i}$$

for any matrix  $\mathbf{T} \in \mathbb{R}^{N_V \times N_E}$  that satisfies:

$$\text{Eq. 4} \quad \mathbf{K} \mathbf{T} = \mathbf{I}$$

where  $\mathbf{I}$  is the identity matrix. This is equivalent to:

$$\text{Eq. 5} \quad \mathbf{K} \mathbf{T} \mathbf{K} = \mathbf{K}$$

due to the fact that  $\mathbf{K}$  is of full row rank (see e.g. Lutkepohl 1996, pp. 32, property 3g).

It is of the essence to note that in general,  $\hat{\mathbf{J}}_{\omega,i}$  is a solution to Eq. 2, as can be confirmed by plugging Eq. 3 into Eq. 2 and making use of Eq. 4:

$$\text{Eq. 6} \quad \Phi_{\omega,i} = \mathbf{K} \hat{\mathbf{J}}_{\omega,i} = \mathbf{K} (\mathbf{T} \Phi_{\omega,i}) \equiv \Phi_{\omega,i}$$

From Eq. 3, the Hermitian covariance matrix, which is proportional to the cross-spectral density matrix, satisfies:

$$\text{Eq. 7} \quad \mathbf{S}_{\mathbf{J}\omega} = \mathbf{T} \mathbf{S}_{\Phi\omega} \mathbf{T}^T$$

with:

$$\text{Eq. 8} \quad \mathbf{S}_{\Phi\omega} = \frac{1}{N_S} \sum_{i=1}^{N_S} \Phi_{\omega,i} \Phi_{\omega,i}^*$$

where  $\mathbf{S}_{\Phi\omega} \in \mathbb{C}^{N_E \times N_E}$  is the measurement-based scalp EEG Hermitian covariance, and  $\mathbf{S}_{\mathbf{J}\omega} \in \mathbb{C}^{N_V \times N_V}$  is the estimated current density Hermitian covariance.

In general, the superscript “ $T$ ” denotes transpose, and the superscript “ $*$ ” denotes complex conjugate and transpose (as used e.g. in Eq. 7 and Eq. 8, respectively).

It is important to emphasize that the source covariance estimator  $\mathbf{S}_{\mathbf{J}\omega}$  in Eq. 7 depends on the choice of inverse matrix  $\mathbf{T}$  and on the measurements via Eq. 8. This is to be contrasted with the particular estimator for the inverse source covariance below, which is independent of any linear inverse used (i.e. independent of the choice of  $\mathbf{T}$ ).

The estimated intracranial coherences are obtained from the current density covariance in Eq. 7 as:

$$\text{Eq. 9} \quad \mathbf{R}_{\mathbf{J}\omega} = \mathbf{D}_{\mathbf{J}\omega} \mathbf{S}_{\mathbf{J}\omega} \mathbf{D}_{\mathbf{J}\omega}$$

with:

$$\text{Eq. 10} \quad \mathbf{D}_{\mathbf{J}\omega} = \left[ \text{diag}(\mathbf{S}_{\mathbf{J}\omega}) \right]^{-1/2}$$

where the “diag” operator returns a diagonal matrix (off-diagonal elements set to zero). Note that Eq. 10 is valid only if the diagonal elements:

$$\text{Eq. 11} \quad \left[ \mathbf{S}_{\mathbf{j}_\omega} \right]_{ii} > 0 \quad \text{for } i = 1 \dots N_v$$

It is of interest to minimize the effect of common sources when analyzing connectivity. This can be achieved by computing the partial coherence matrix. In practice, the computations use the well-known fact that the inverse covariance matrix corresponds to the matrix of partial covariances. From the matrix of partial covariances, the partial coherence matrix is obtained by scaling to partial covariances with the corresponding partial standard deviations.

In simple matrix algebra terms, if  $\mathbf{S}$  denotes a positive definite covariance matrix, and:

$$\text{Eq. 12} \quad \mathbf{Q} = \mathbf{S}^{-1}$$

denotes its inverse, then the partial coherence is:

$$\text{Eq. 13} \quad \mathbf{P} = \mathbf{E} \mathbf{Q} \mathbf{E}$$

with:

$$\text{Eq. 14} \quad \mathbf{E} = \left[ \text{diag}(\mathbf{Q}) \right]^{-1/2}$$

This definition can be generalized to the case of a non-negative definite covariance matrix, such as the estimated current density covariance in Eq. 7, which is singular, with  $\text{rank}(\mathbf{S}_{\mathbf{j}_\omega}) = N_E \ll N_v$ .

### 3.2. The inverse source covariance

We propose as an estimator for the inverse source covariance the particular form:

$$\text{Eq. 15} \quad \mathbf{S}_{\mathbf{j}_\omega}^{-r} = \left( \mathbf{T} \mathbf{S}_{\Phi\omega} \mathbf{T}^T \right)^{-r} = \mathbf{K}^T \mathbf{S}_{\Phi\omega}^{-1} \mathbf{K}$$

where the superscript “-r” denotes a reflexive generalized inverse. Note that this generalized inverse does indeed satisfy the two properties that characterize a reflexive inverse:

$$\text{Eq. 16} \quad \mathbf{S}_{\mathbf{j}_\omega} \mathbf{S}_{\mathbf{j}_\omega}^{-r} \mathbf{S}_{\mathbf{j}_\omega} = \mathbf{S}_{\mathbf{j}_\omega} \quad \left[ \text{based on } \mathbf{T} \mathbf{S}_{\Phi\omega} \mathbf{T}^T \mathbf{K}^T \mathbf{S}_{\Phi\omega}^+ \mathbf{K} \mathbf{T} \mathbf{S}_{\Phi\omega} \mathbf{T}^T = \mathbf{T} \mathbf{S}_{\Phi\omega} \mathbf{T}^T \right]$$

$$\text{Eq. 17} \quad \mathbf{S}_{\mathbf{j}_\omega}^{-r} \mathbf{S}_{\mathbf{j}_\omega} \mathbf{S}_{\mathbf{j}_\omega}^{-r} = \mathbf{S}_{\mathbf{j}_\omega}^{-r} \quad \left[ \text{based on } \mathbf{K}^T \mathbf{S}_{\Phi\omega}^{-1} \mathbf{K} \mathbf{T} \mathbf{S}_{\Phi\omega} \mathbf{T}^T \mathbf{K}^T \mathbf{S}_{\Phi\omega}^{-1} \mathbf{K} = \mathbf{K}^T \mathbf{S}_{\Phi\omega}^{-1} \mathbf{K} \right]$$

Eq. 15 defines the *partial covariance field*. And the *partial coherence field* is obtained by standardization of Eq. 15 as described in Eq. 13 and Eq. 14.

Note that Eq. 15 requires the existence of the inverse of the EEG covariance. If such an inverse does not exist because, e.g., the number of EEG epochs is smaller than the number of electrodes (i.e.  $N_s < N_E$ ), we propose the use of a generalized inverse for the EEG covariance, thus giving the estimator for the inverse source covariance as:

$$\text{Eq. 18} \quad \mathbf{S}_{\mathbf{j}_\omega}^{-r} = \left( \mathbf{T} \mathbf{S}_{\Phi\omega} \mathbf{T}^T \right)^{-r} = \mathbf{K}^T \mathbf{S}_{\Phi\omega}^+ \mathbf{K}$$

where the superscript “+” denotes the particular generalized inverse given by the Moore-Penrose inverse.

It should be noted that this estimator (Eq. 15) is based on the reflexive inverse, and not on the Moore-Penrose inverse, which would need to satisfy additional conditions, one of which is:

$$\text{Eq. 19} \quad \left( \mathbf{S}_{j\omega} \mathbf{S}_{j\omega}^{-r} \right)^* = \mathbf{S}_{j\omega} \mathbf{S}_{j\omega}^{-r}$$

Using Eq. 7 and Eq. 15, this gives the requirement that:

$$\text{Eq. 20} \quad \mathbf{K}^T \mathbf{T}^T = \mathbf{T} \mathbf{K}$$

However, Eq. 20 is not satisfied in general. One example suffices for the proof, and easily given by any non-identity weighted inverse solution:

$$\text{Eq. 21} \quad \mathbf{T} = \mathbf{W} \mathbf{K}^T (\mathbf{K} \mathbf{W} \mathbf{K}^T)^{-1}$$

for  $\mathbf{W} \neq \mathbf{I}$ .

### 3.3. Essential properties of the inverse source covariance estimator

#### 3.3.1. The partial coherence field estimator is independent of the choice of linear inverse solution

As seen from Eq. 15, a first property is that the inverse source covariance is independent of any particular inverse solution, i.e., it is independent of any choice of  $\mathbf{T}$  in Eq. 3 that satisfies Eq. 4.

#### 3.3.2. The partial coherence estimator for a particular pair of cortical voxels is explicitly independent of all other voxels

As seen from Eq. 15, a second property is that for any given pair of voxels, their inverse covariance, which is equivalent to their partial covariance after having accounted for the effect of all other voxels, is explicitly independent of all other voxels.

In formal terms, consider the voxel pair indexed by  $(k, l)$ . Let  $\mathbf{k}_i \in \mathbb{R}^{N_E \times 1}$  denote the  $i$ -th column of the lead field matrix in Eq. 1. Then, from Eq. 15, the inverse covariance (i.e. partial covariance) of voxels  $(k, l)$  is:

$$\text{Eq. 22} \quad \left[ \mathbf{S}_{j\omega}^{-r} \right]_{kl} = \mathbf{k}_k^T \mathbf{S}_{\Phi\omega}^{-1} \mathbf{k}_l$$

and the partial coherence is:

$$\text{Eq. 23} \quad \left[ \mathbf{P}_{j\omega}^{-r} \right]_{kl} = \frac{\mathbf{k}_k^T \mathbf{S}_{\Phi\omega}^{-1} \mathbf{k}_l}{\sqrt{(\mathbf{k}_k^T \mathbf{S}_{\Phi\omega}^{-1} \mathbf{k}_k)(\mathbf{k}_l^T \mathbf{S}_{\Phi\omega}^{-1} \mathbf{k}_l)}}$$

which are independent of the lead fields at other voxels.

This remarkable second property allows for very efficient computations of the inverse source covariance, without the need of an actual inversion of the enormous source covariance matrix in Eq. 7.

It is important to note that the inverse source covariance implicitly depends on the whole cortex covariance via the inverse of the EEG covariance.

#### 3.3.3. The resolution of the partial covariance field estimator

Let  $\mathbf{S}_{j\omega} \in \mathbb{C}^{N_V \times N_V}$  denote the true source covariance, which will be assumed to be positive definite. Then, from Eq. 2, the true EEG covariance is:

$$\text{Eq. 24 } \mathbf{S}_{\Phi\omega} = \mathbf{K} \mathbf{S}_{\mathbf{J}\omega} \mathbf{K}^T$$

and the estimated source covariance (in general, for any  $\mathbf{T}$ ) from Eq. 7 is:

$$\text{Eq. 25 } \mathbf{S}_{\mathbf{J}\omega} = \mathbf{T} \mathbf{S}_{\Phi\omega} \mathbf{T}^T$$

Using Eq. 24, the estimated reflexive generalized inverse of Eq. 25 is:

$$\text{Eq. 26 } \mathbf{S}_{\mathbf{J}\omega}^{-r} = (\mathbf{T} \mathbf{K} \mathbf{S}_{\mathbf{J}\omega} \mathbf{K}^T \mathbf{T}^T)^{-r}$$

which is independent of the choice of the inverse solution  $\mathbf{T}$ .

Since any choice of  $\mathbf{T}$  must give the same solution, we may choose the simplest Moore-Penrose (minimum norm) inverse solution:

$$\text{Eq. 27 } \mathbf{T}_{MinNorm} = \mathbf{K}^T (\mathbf{K} \mathbf{K}^T)^{-1}$$

which gives:

$$\text{Eq. 28 } \mathbf{S}_{\mathbf{J}\omega}^{-r} = (\mathbf{H} \mathbf{S}_{\mathbf{J}\omega} \mathbf{H})^{-r}$$

which is independent of the choice of the inverse solution  $\mathbf{T}$ , with:

$$\text{Eq. 29 } \mathbf{H} = \mathbf{K}^T (\mathbf{K} \mathbf{K}^T)^{-1} \mathbf{K}$$

Note that from Eq. 29,  $\mathbf{H}$  is an idempotent projection matrix, and corresponds to the well-known resolution matrix of the minimum norm inverse solution.

Eq. 28 demonstrates the relation between the estimated inverse source covariance  $\mathbf{S}_{\mathbf{J}\omega}^{-r}$  and the actual source covariance  $\mathbf{S}_{\mathbf{J}\omega}$ . The estimator  $\mathbf{S}_{\mathbf{J}\omega}^{-r}$  is a function of the spatially filtered covariance, as “seen” through the filter given by the resolution matrix  $\mathbf{H}$  (Eq. 29), which is known to have the effect of low spatial resolution.

#### 4. An algorithm for the whole cortex electromagnetic connectivity

Step 0: Given an EEG Hermitian covariance matrix  $\mathbf{S}_{\Phi\omega}$  (e.g. as in Eq. 8). Given the lead field matrix  $\mathbf{K}$  (e.g. as used in Eq. 1).

Step 1: Compute the singular value decomposition of the EEG covariance matrix ( $N_E \times N_E$ ):

$$\text{Eq. 30 } \mathbf{S}_{\Phi\omega} = \mathbf{\Gamma} \mathbf{\Lambda} \mathbf{\Gamma}^*$$

with  $\mathbf{\Gamma}$  denoting the eigenvectors, and the diagonal  $\mathbf{\Lambda}$  the non-zero eigenvalues.

Step 3: Compute the matrix  $\mathbf{U}$ , corresponding to the Hermitian square root inverse of the EEG covariance:

$$\text{Eq. 31 } \mathbf{U} = \mathbf{\Gamma} \mathbf{\Lambda}^{-1/2} \mathbf{\Gamma}^*$$

NOTE: other inverse square root choices are possible.

Step 4: Compute:

$$\text{Eq. 32 } \mathbf{V} = \mathbf{K}^T \mathbf{U}$$

and normalize each row. This can be expressed as:

$$\text{Eq. 33 } \mathbf{W} = \mathbf{E} \mathbf{V}$$

with:

$$\text{Eq. 34} \quad \mathbf{E} = \left[ \text{diag}(\mathbf{V}\mathbf{V}^T) \right]^{-1/2}$$

NOTE: The matrices  $\mathbf{V}$  and  $\mathbf{W}$  are of size  $(N_v \times N_e)$ ; and in Eq. 34, only the diagonal elements of  $(\mathbf{V}\mathbf{V}^T)$  are needed, not the full matrix.

Step 5: The whole-cortex partial coherence matrix is:

$$\text{Eq. 35} \quad \mathbf{P} = \mathbf{W}\mathbf{W}^T$$

## 5. Some computational notes

These equations show that all the relevant connectivity information is contained in the low-rank square root matrix  $\mathbf{W}$  of dimension  $N_v \times N_e$  defined in Eq. 33. For instance, the largest left eigenvector can be used to compute distributed connections using the methodology and interpretation of Worsley et al (2005).

These equations are especially efficient when the aim is to view the whole cortex connections to a single point. For instance: how does the right auditory cortex connect to the rest of the cortex? In this case, the whole brain connectivities correspond to a single row (or column) of the full connectivity matrix Eq. 35, which is very easy to compute.

## 6. Results

A very simple simulation test was performed. Two cortical point sources were considered:  
 $x_t$  : on cortex under scalp electrode Fp1, left frontal  
 $y_t$  : on cortex under scalp electrode O2, right occipital

The time series for these true cortical current densities were generated as:

$$\text{Eq. 36} \quad \begin{cases} x_t = \varepsilon_{x,t} \\ y_t = 0.5x_{t-1} + \varepsilon_{y,t} \end{cases}$$

where both noise series  $\varepsilon$  are independent and identically distributed uniformly in the interval  $(-0.15 \dots +0.15)$ , i.e.  $IID \sim U(-0.15 \dots +0.15)$ .

Additive biological noise, with 57 cortical neurons firing as  $IID \sim U(-0.05 \dots +0.05)$ , was superimposed on the signals defined Eq. 36. Given such cortical current densities, the scalp potentials (EEG) were computed from Eq. 1 on 19 electrodes ( $N_e = 19$ ) corresponding to the 10/20 system locations. Another layer of noise was added, consisting of measurement noise for the scalp potentials, as  $IID \sim U(-0.05 \dots +0.05)$  at each moment in time and at each electrode.

Using this procedure, 100 EEG epochs were generated ( $N_s = 100$ ), each one consisting of 64 discrete time samples ( $N_t = 64$ ). Assuming a sampling rate of 64 Hz, the EEG cross-spectrum was computed (see Eq. 8). Finally, using the EEG cross-spectral matrix for the alpha band (8-12 Hz), the classical (Eq. 9) and the new (Eq. 30 to Eq. 35) connectivities were computed, on 6239 cortical grey matter voxels ( $N_v = 6239$ ).

In the figures, only “lagged classical” and “lagged partial” connectivities (as defined in Pascual-Marqui 2007; and Pascual-Marqui et al 2011) are displayed.

Define a “3D seeded connectivity map” as the connectivity values between a given cortical seed point with all other cortical voxels.

For illustration purposes, 19 seed points were used, consisting of the cortex underlying each electrode of the 10/20 system. The figures show the maximum connectivity at each voxel, over the set of 19 seeded connectivity maps. Figures 1 and 2 show the results for classical and partial connectivities, respectively.

The color scales of the images are proportionally identical, in such a way that the scale’s maximum value was adjusted to 95% of the extreme connectivity value, thus allowing for a fair, unbiased comparison of the two competing methods.

For both types of connectivity measures (lagged classical and lagged partial), the maximum connectivities are in frontal and posterior regions. However, classical connectivity is extremely low resolution, with actual maxima not exactly located under Fp1 and O2. In contrast, and despite such noisy data and so few electrodes, the new partial connectivity measures have exact localization with very high resolution.

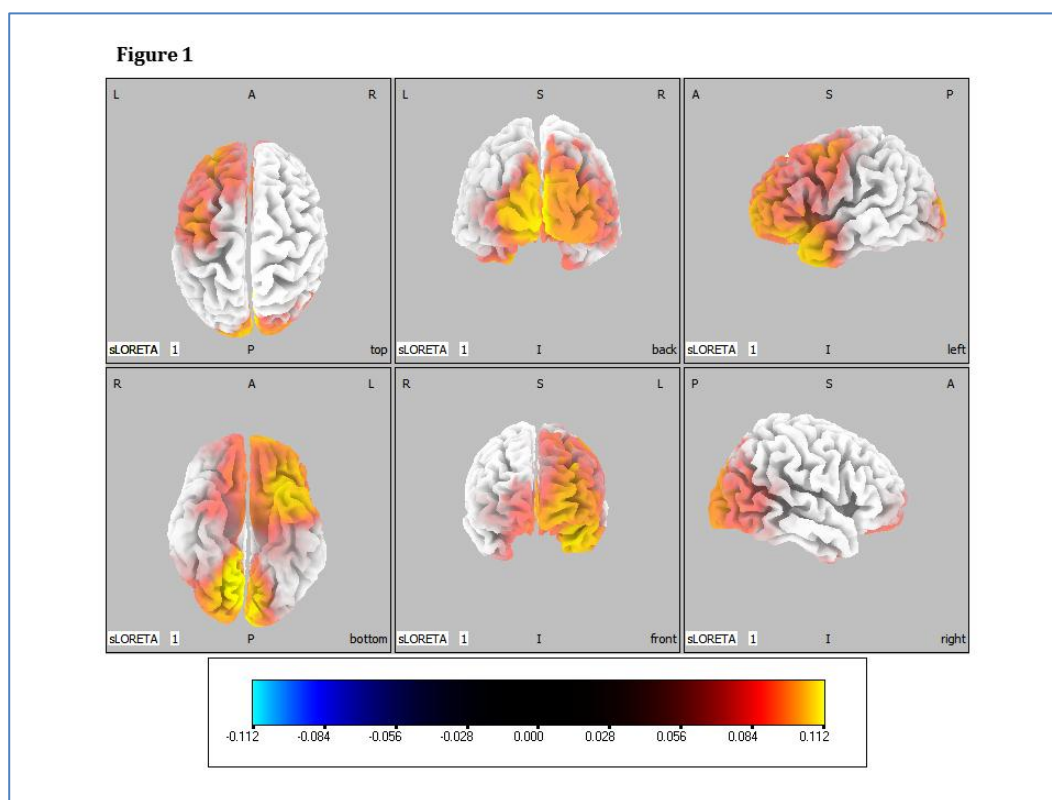


Figure 1: Classical connectivity map. L=left; R=right; A=anterior; P=posterior; I=inferior; S=superior. Actual connections correspond only to left frontal with right occipital.

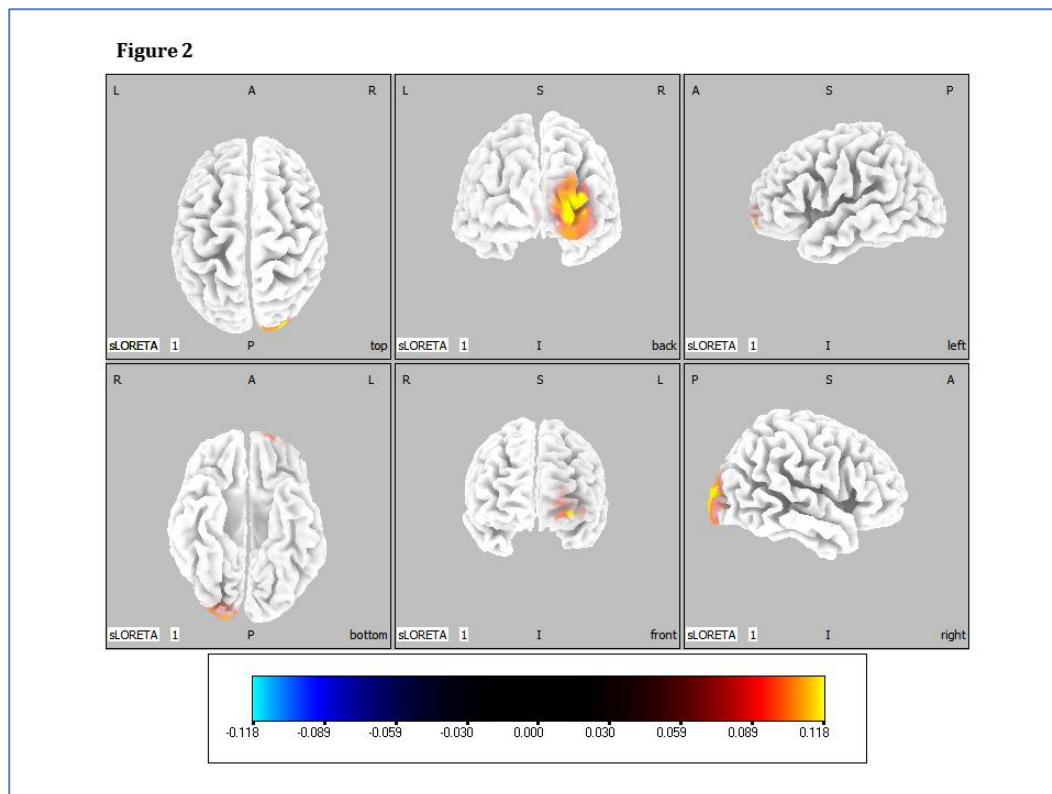


Figure 2: New partial connectivity map. L=left; R=right; A=anterior; P=posterior; I=inferior; S=superior. Actual connections correspond only to left frontal with right occipital.

## 7. Concluding remarks

Whole cortex connectivities can be estimated non-invasively with EEG and MEG. The new method uses a generalized reflexive inverse for the source covariance, which is endowed with the property of being invariant to any type of linear tomography. The method is very simple and efficient from a computational point of view.

The simulation results, based on only 19 EEG electrodes, with significant corruption by both biological and measurement noise, empirically show that the new method has much higher resolution than the classical connectivity measures.

The new method can be extended to time domain data, including the case of lagged partial correlation fields, thus providing information akin to Granger-type causality. Moreover, the general application field of this new method reaches beyond neuroimaging of connectivity fields.

## 8. References

1. Pedro A. Valdes-Sosa, AlardRoebroek, JeasDunizeua, Karl Friston. Effective connectivity: Influence, causality and biophysical modeling. NeuroImage, In Press, Corrected Proof, Available online 6 April 2011.
2. Olaf Sporns. Networks of the Brain. The MIT Press, Cambridge, Massachusetts, 2011.

3. Roberto D. Pascual-Marqui, Dietrich Lehmann, Martha Koukkou, Kieko Kochi, Peter Anderer, Bernd Saletu, Hideaki Tanaka, Koichi Hirata, E. Roy John, Leslie Prichep, Rolando Biscay-Lirio, Toshihiko Kinoshita. Assessing interactions in the brain with exact low resolution electromagnetic tomography (eLORETA). Philosophical Transactions of the Royal Society A (accepted and in press, 2011).
4. Worsley KJ, Chen JI, Lerch J, Evans AC. Comparing functional connectivity via thresholding correlations and singular value decomposition. Phil. Trans. R. Soc. B, 2005; 360: 913-920.
5. Pascual-Marqui RD. Instantaneous and lagged measurements of linear and nonlinear dependence between groups of multivariate time series: Frequency decomposition. arXiv:07111455 [statME]. 2007; <http://arxiv.org/abs/0711.1455>.
6. Roberto D. Pascual-Marqui, Dietrich Lehmann, Martha Koukkou, Kieko Kochi, Peter Anderer, Bernd Saletu, Hideaki Tanaka, Koichi Hirata, E. Roy John, Leslie Prichep, Rolando Biscay-Lirio, Toshihiko Kinoshita. Assessing interactions in the brain with exact low resolution electromagnetic tomography (eLORETA). Philosophical Transactions of the Royal Society A (accepted and in press, 2011).
9. R.D. Pascual-Marqui: Discrete, 3D distributed, linear imaging methods of electric neuronal activity. Part 1: exact, zero error localization. arXiv:0710.3341 [math-ph], 2007-October-17, <http://arxiv.org/pdf/0710.3341>
10. H. Lutkepohl. Handbook of Matrices. 1996 Wiley, New York.

# Isolated effective coherence (iCoh): causal information flow excluding indirect paths

RD Pascual-Marqui<sup>1,2</sup>, RJ Biscay<sup>3</sup>, J Bosch-Bayard<sup>4</sup>, D Lehmann<sup>1</sup>, K Kochi<sup>1</sup>, N Yamada<sup>2</sup>, T Kinoshita<sup>5</sup>, N Sadato<sup>6</sup>

<sup>1</sup>The KEY Institute for Brain-Mind Research, University of Zurich, Switzerland; <sup>2</sup>Department of Psychiatry, Shiga University of Medical Science, Japan; <sup>3</sup>CIMFAV, Universidad de Valparaiso, Chile; <sup>4</sup>Cuban Neuroscience Center, Havana, Cuba; <sup>5</sup>Department of Neuropsychiatry, Kansai Medical University, Japan; <sup>6</sup>Division of Cerebral Integration, National Institute for Physiological Sciences, Okazaki, Japan

Corresponding author:

RD Pascual-Marqui; "pascualm at key.uzh.ch"; "pascualm at belle.shiga-med.ac.jp"; [www.uzh.ch/keyinst/loreta.htm](http://www.uzh.ch/keyinst/loreta.htm)

The KEY Institute for Brain-Mind Research, University Hospital of Psychiatry, Zurich, Switzerland

Department of Psychiatry, Shiga University of Medical Sciences, Shiga, Japan

## 1. Abstract

A problem of great interest in real world systems, where multiple time series measurements are available, is the estimation of the intra-system causal relations. For instance, electric cortical signals are used for studying functional connectivity between brain areas, their directionality, the direct or indirect nature of the connections, and the spectral characteristics (e.g. which oscillations are preferentially transmitted). The earliest spectral measure of causality was Akaike's (1968) seminal work on the noise contribution ratio, reflecting direct and indirect connections. Later, a major breakthrough was the partial directed coherence of Baccala and Sameshima (2001) for direct connections. The simple aim of this study consists of two parts: (1) To expose a major problem with the partial directed coherence, where it is shown that it is affected by irrelevant connections to such an extent that it can misrepresent the frequency response, thus defeating the main purpose for which the measure was developed, and (2) To provide a solution to this problem, namely the "isolated effective coherence", which consists of estimating the partial coherence under a multivariate auto-regressive model, followed by setting all irrelevant associations to zero, other than the particular directional association of interest. Simple, realistic, toy examples illustrate the severity of the problem with the partial directed coherence, and the solution achieved by the isolated effective coherence.

## 2. Introduction

Consider a realistic, non-artificial example where time series of local electric potential differences are measured at five sites on the cortex (electrocorticogram, ECoG). An informal description (later to be made more precise using multivariate autoregression) follows. Site 1 has intrinsic activity at 28 Hz, and sends information to Site 2 with a measurable physiological time lag. Site 2 has intrinsic activity at 16 Hz, and sends information to Sites 1, 3, 4, and 5 with a measurable physiological time lag. Sites 3, 4, and 5 have intrinsic independent activities at 23 Hz. Instantaneous information transmission would require ephaptic conduction (see e.g. Weiss et al 2013), which is not considered to be present in this realistic example.

In this realistic, non-artificial example, we wish to recover from the time series measurements, all the detailed information about the system: the direct connections, their directionality, and the spectral nature of the information being transmitted.

This example illustrates a very general problem, of realistic nature, that has great interest in the field of human brain function. Other fields of research can as well benefit from the solution to this type of problem.

This type of problem has a higher degree of complexity than what is usually being considered in time series of metabolism (e.g. as measured by fMRI), which have oscillations much lower than 0.1 Hz. It is typical in fMRI to focus only on the directness and directionality of the connections, and not on the fundamental problem encountered in electrophysiology regarding the spectral nature of the information being transmitted. For an example on fMRI connectivity research that lacks spectral information, see Marinazzo et al (2011).

A very narrow and focused review of the literature on methods for estimating direct or indirect connections, the directionality, and their spectral nature, reveals two major contributions:

1. The noise contribution ratio (NCR) of Akaike (1967), which has been extensively used under other names by Saito and Harashima (1981), Kaminski and Blinowska (1991), and Baccala and Sameshima (1998). Details are given below. This method discovers indirect and direct connections without differentiating them, their directionality and spectral characteristics.
2. The partial directed coherence (PDC) of Baccala and Sameshima (2001), which is a measure designed to quantify direct connections that are not confounded by indirect paths, their directionality and spectral characteristics. This is a very widely used measure (cited 650 times at the time of this writing according to “Google-Scholar”).

Recently, the PDC has been critically studied by Schelter et al (2009). They pointed out that the normalization used in PDC, i.e. the denominator in the PDC formula (see below) contains all influences from a source node to all other (receiving) nodes, and as a consequence, the PDC decreases in the presence of many nodes, even if the relationship between source and target nodes remains unchanged. The solution to this problem was given in the form of a renormalization of the PDC, using the statistical variance of the strength of the connection.

In this present study, rather the aiming at a re-normalization of the PDC, such as that successfully achieved by Schelter et al (2009), we reformulate the problem from scratch, estimating the partial coherence under a multivariate auto-regressive model, followed by setting all irrelevant associations to zero, other than the particular directional association of interest. This procedure is akin to Pearl’s (2000) “surgical intervention” for studying causality. This approach gives the isolated effective coherence (iCoh).

We give a compelling realistic example that shows how the PDC can give incorrect information about the strength of a connection, and incorrect information on its spectral characteristics. And we show how the iCoh solves this problem.

It is also shown how the iCoh can be obtained from Akaike’s NCR under the disconnection constraints as used in iCoh. This demonstration does not in any way mean that the iCoh is identical to NCR. Rather it shows that there is a common foundation to both iCoh and NCR, which may aid in interpreting the new and distinct measure: “iCoh”.

Finally, it should be mentioned that instantaneous connections can be accommodated in the iCoh measure. As a realistic example of instantaneous connectivity in electrophysiology, consider the use of scalp EEG recordings. These scalp signals should never be used by themselves for studying cortical connectivity (as an example where it is used anyway, see Marinazzo et al 2011), the reason being that the cortical generators do not in general project radially onto the scalp. A

“strong connection” between, say, F3 and F4 signals does not in any way imply that there is a strong connection between the underlying left and right frontal cortices. This is explained and illustrated, for instance, in Lehmann et al (2012). Instead, the EEG signals can be used for estimating the cortical activity signals, using a method such as eLORETA (Pascual-Marqui et al, 2011). However, these estimated cortical signals are instantaneously mixed by volume conduction. Formulations for estimating multivariate autoregressive models that take into account this instantaneous mixing which introduces “apparent” instantaneous connectivity can be found, e.g. in Gomez-Herrero et al (2008). Another approach for modeling instantaneous connections consists of formulating a multivariate autoregressive model that includes zero-lag coefficients, as in Faes et al (2013). In this case, no new measures were derived, since all that is needed are the newly estimated lagged coefficients to be plugged into any of the measures, such as PDF or iCoh.

### 3. Multivariate auto-regression, Granger causality, and the cross-spectral density

General background and notation on multivariate autoregressive models, and the corresponding frequency domain spectral density matrix, can be found, for instance, in Akaike (1968) and Yamashita et al (2005).

Consider the stable, stationary multivariate autoregressive model of order  $p$  written as:

$$\text{Eq. 1} \quad \mathbf{X}(t) = \sum_{k=1}^p \mathbf{A}(k) \mathbf{X}(t-k) + \boldsymbol{\varepsilon}(t)$$

with  $\mathbf{X}(t) \in \mathbb{R}^{q \times 1}$ ,  $q \geq 2$ ,  $\mathbf{A}(k) \in \mathbb{R}^{q \times q}$ ,  $\boldsymbol{\varepsilon}(t) \in \mathbb{R}^{q \times 1}$ , and with discrete time  $t = 0 \dots N_T - 1$ .

Given data sampled in discrete time, the auto-regressive parameters can be estimated by any number of methods, one of which is the simple least squares approach (see e.g. Akaike 1968).

The frequency domain representation is:

$$\text{Eq. 2} \quad \mathbf{X}(\omega) = \mathbf{A}(\omega) \mathbf{X}(\omega) + \boldsymbol{\varepsilon}(\omega)$$

where  $\mathbf{X}(\omega) \in \mathbb{C}^{q \times 1}$ ,  $\mathbf{A}(\omega) \in \mathbb{C}^{q \times q}$ ,  $\boldsymbol{\varepsilon}(\omega) \in \mathbb{C}^{q \times 1}$  are the discrete Fourier transforms, with discrete frequency  $\omega = 0 \dots N_T - 1$ . The discrete Fourier transform for obtaining Eq. 2 in practice is described in detail in the appendix.

In this setting, direct Granger causality is defined as follows:

Time series “ $j$ ” directly causes time series “ $i$ ” if the  $(i, j)$  element of  $\mathbf{A}(\omega)$  is non-zero, i.e.:

$$\text{Eq. 3} \quad [\mathbf{A}(\omega)]_{ij} \neq 0 \Rightarrow j \text{ directly Granger causes } i$$

see e.g. Granger (1969) and Lutkepohl (2005).

Notation: In what follows,  $[\mathbf{M}]_{ij}$  denotes the element  $(i, j)$  of the matrix  $\mathbf{M}$ .

From Eq. 2, the Hermitian covariance, i.e. the spectral density matrix, is:

$$\text{Eq. 4} \quad \mathbf{S}_x(\omega) = (\mathbf{I} - \mathbf{A}(\omega))^{-1} \mathbf{S}_\varepsilon (\mathbf{I} - \mathbf{A}^*(\omega))^{-1} = (\tilde{\mathbf{A}}(\omega))^{-1} \mathbf{S}_\varepsilon (\tilde{\mathbf{A}}^*(\omega))^{-1} = \mathbf{B}(\omega) \mathbf{S}_\varepsilon \mathbf{B}^*(\omega)$$

where the superscript “\*” denotes transpose and complex conjugate,  $\mathbf{I}$  is the identity matrix,  $\mathbf{S}_\varepsilon \in \mathbb{R}^{q \times q}$  is the noise covariance, and:

$$\text{Eq. 5} \quad \tilde{\mathbf{A}}(\omega) = \mathbf{I} - \mathbf{A}(\omega)$$

and:

$$\text{Eq. 6} \quad \mathbf{B}(\omega) = (\mathbf{I} - \mathbf{A}(\omega))^{-1} = (\tilde{\mathbf{A}}(\omega))^{-1}$$

#### 4. The inverse of the spectral density matrix and the partial coherence

The inverse of the Hermitian covariance matrix, i.e. the inverse of the spectral density matrix, is:

$$\text{Eq. 7} \quad \mathbf{S}_x^{-1}(\omega) = \tilde{\mathbf{A}}^*(\omega) \mathbf{S}_\varepsilon^{-1} \tilde{\mathbf{A}}(\omega)$$

Its  $(i, j)$  element is:

$$\text{Eq. 8} \quad [\mathbf{S}_x^{-1}(\omega)]_{ij} = \sum_{k=1}^q \sum_{l=1}^q [\tilde{\mathbf{A}}^*(\omega)]_{ik} [\mathbf{S}_\varepsilon^{-1}]_{kl} [\tilde{\mathbf{A}}(\omega)]_{lj} = \sum_{k=1}^q \sum_{l=1}^q \overline{[\tilde{\mathbf{A}}(\omega)]_{ki}} [\mathbf{S}_\varepsilon^{-1}]_{kl} [\tilde{\mathbf{A}}(\omega)]_{lj}$$

where the complex conjugate of a scalar  $c$  is denoted as  $\bar{c}$ . Note that Eq. 8 can be evaluated for subscripts  $(i, j)$ ,  $(i, i)$ , and  $(j, j)$ .

The coefficient  $[\tilde{\mathbf{A}}(\omega)]_{ij}$  quantifies the effective causal influence for  $j \rightarrow i$ , based on the definition in Eq. 3.

The partial coherence (see e.g. Brillinger 1981) between  $(i, j)$  is:

$$\text{Eq. 9} \quad p_{ij}(\omega) = \frac{[\mathbf{S}_x^{-1}(\omega)]_{ij}}{\sqrt{[\mathbf{S}_x^{-1}(\omega)]_{ii} [\mathbf{S}_x^{-1}(\omega)]_{jj}}}$$

The significance of the partial coherence in a very general setting can be found in Rao (1981). In simple terms, the partial coherence is a measure of association between two complex valued random variables after removing the effect of other measured variables.

#### 5. The isolated effective coherence (iCoh) for $j \rightarrow i$

The term “effective” as is used here in this context, refers to the direct causal influence of one time series on another, conditional on all other time series, i.e. after removing the effect of all other time series.

The isolated effective coherence (iCoh) for  $j \rightarrow i$  is defined under the condition that the only non-zero association between the time series is due to  $[\tilde{\mathbf{A}}(\omega)]_{ij} \neq 0$ . This requires that all other possible associations be set to zero, i.e.:

$$\text{Eq. 10} \quad [\mathbf{A}(\omega)]_{kl} \equiv 0, \text{ for all } (k, l) \text{ such that } (k, l) \neq (i, j) \text{ and } k \neq l$$

and:

$$\text{Eq. 11} \quad [\mathbf{S}_\varepsilon]_{kl} \equiv 0, \text{ for all } (k, l) \text{ such that } k \neq l$$

Note that the diagonal elements of  $\mathbf{S}_\varepsilon$  and  $\mathbf{A}(\omega)$  remain unmodified, since they do not “associate” different nodes.

Emphasis must be placed on the fact that this procedure is meaningful only if the new system with a single association remains stable and stationary.

Plugging Eq. 10 and Eq. 11 into Eq. 8 gives a covariance matrix  $[\mathbf{S}_x^{-1}(\omega)]^{(i \leftarrow j)}$  with elements:

$$\text{Eq. 12} \quad [\mathbf{S}_x^{-1}(\omega)]_{ij}^{(i \leftarrow j)} = [\tilde{\mathbf{A}}(\omega)]_{ij} [\overline{\tilde{\mathbf{A}}(\omega)}]_{ii} [\mathbf{S}_\varepsilon]_{ii}^{-1}$$

$$\text{Eq. 13} \quad [\mathbf{S}_x^{-1}(\omega)]_{ii}^{(i \leftarrow j)} = [\mathbf{S}_\varepsilon]_{ii}^{-1} \left| [\tilde{\mathbf{A}}(\omega)]_{ii} \right|^2$$

$$\text{Eq. 14} \quad [\mathbf{S}_x^{-1}(\omega)]_{jj}^{(i \leftarrow j)} = [\mathbf{S}_\varepsilon]_{ii}^{-1} \left| [\tilde{\mathbf{A}}(\omega)]_{ij} \right|^2 + [\mathbf{S}_\varepsilon]_{jj}^{-1} \left| [\tilde{\mathbf{A}}(\omega)]_{jj} \right|^2$$

where it is assumed that the self-auto-regressions for both nodes are stable, i.e.  $[\tilde{\mathbf{A}}(\omega)]_{ii}$  and  $[\tilde{\mathbf{A}}(\omega)]_{jj}$  must correspond to stable self-auto-regressions.

Plugging Eq. 12, Eq. 13, and Eq. 14 into Eq. 9 defines the isolated effective coherence (iCoh) for  $j \rightarrow i$ , as the squared modulus of the partial coherence:

$$\text{Eq. 15} \quad \kappa_{i \leftarrow j}(\omega) = \frac{\left| [\tilde{\mathbf{A}}(\omega)]_{ij} \right|^2 \left| [\tilde{\mathbf{A}}(\omega)]_{ii} \right|^2 [\mathbf{S}_\varepsilon]_{ii}^{-2}}{[\mathbf{S}_\varepsilon]_{ii}^{-1} \left| [\tilde{\mathbf{A}}(\omega)]_{ii} \right|^2 \left( [\mathbf{S}_\varepsilon]_{ii}^{-1} \left| [\tilde{\mathbf{A}}(\omega)]_{ij} \right|^2 + [\mathbf{S}_\varepsilon]_{jj}^{-1} \left| [\tilde{\mathbf{A}}(\omega)]_{jj} \right|^2 \right)}$$

which clearly satisfies:

$$\text{Eq. 16} \quad 0 \leq \kappa_{i \leftarrow j}(\omega) \leq 1$$

Note that Eq. 15 is a genuine partial coherence, obtained under certain constraints. The numerator contains an off-diagonal element of the inverse covariance matrix, while the denominator contains the corresponding product of diagonal elements.

For practical computations, Eq. 15 simplifies to:

$$\text{Eq. 17} \quad \kappa_{i \leftarrow j}(\omega) = \frac{[\mathbf{S}_\varepsilon]_{ii}^{-1} \left| [\tilde{\mathbf{A}}(\omega)]_{ij} \right|^2}{[\mathbf{S}_\varepsilon]_{ii}^{-1} \left| [\tilde{\mathbf{A}}(\omega)]_{ij} \right|^2 + [\mathbf{S}_\varepsilon]_{jj}^{-1} \left| [\tilde{\mathbf{A}}(\omega)]_{jj} \right|^2}$$

The iCoh can be described as the answer to the following question:

“Given a dynamic linear system characterized by its auto-regressive parameters, what would be the equation for partial coherence if all connections are severed, except for the single one of interest?”

Note that the algorithm for computing the iCoh requires:

- (1) The estimation of the full, joint, multivariate auto-regressive model only once. No re-estimation is required in the following steps.
- (2) Given a pair of nodes and a direction such as  $j \rightarrow i$ , compute Eq. 15 (or equivalently, compute Eq. 17) using the model parameters obtained from step (1).

## 6. Akaike's noise contribution ratio (NCR)

Akaike's (1968) noise contribution ratio (NCR) is based on the spectral representation in Eq. 4, Eq. 5, and Eq. 6. Consider the  $i$ -th node and consider its univariate spectral density:

$$\text{Eq. 18} \quad [\mathbf{S}_x(\omega)]_{ii} = \sum_{k=1}^q \sum_{l=1}^q [\mathbf{B}(\omega)]_{ik} [\mathbf{S}_\varepsilon]_{kl} [\mathbf{B}^*(\omega)]_{li}$$

Consider the case when the innovations are uncorrelated. Then Eq. 18 simplifies to:

$$\text{Eq. 19} \quad [\mathbf{S}_x(\omega)]_{ii} = \sum_{k=1}^q \left| [\mathbf{B}(\omega)]_{ik} \right|^2 [\mathbf{S}_\varepsilon]_{kk}$$

Eq. 19 shows that the spectral power at the node " $i$ " receives an additive contribution from the innovation at the  $j$ -th node  $[\mathbf{S}_\varepsilon]_{jj}$ , weighted by the transfer function  $\left| [\mathbf{B}(\omega)]_{ij} \right|^2$ . The relative contribution (see e.g. Yamashita et al 2005) the  $j$ -th node sends to the  $i$ -th receiving node defines the noise contribution ratio as:

$$\text{Eq. 20} \quad \gamma_{i \leftarrow j}(\omega) = \frac{\left| [\mathbf{B}(\omega)]_{ij} \right|^2 [\mathbf{S}_\varepsilon]_{jj}}{\sum_{k=1}^q \left| [\mathbf{B}(\omega)]_{ik} \right|^2 [\mathbf{S}_\varepsilon]_{kk}}$$

which clearly satisfies:

$$\text{Eq. 21} \quad 0 \leq \gamma_{i \leftarrow j}(\omega) \leq 1$$

Note that the NCR is non-zero if  $[\mathbf{B}(\omega)]_{ij}$  is non-zero, and that this can happen even if the direct causality is zero, i.e. even if  $[\mathbf{A}(\omega)]_{ij}$  is zero. Thus, the NCR is a frequency domain measure of total causality, which includes direct and indirect connections.

## 7. Other measures published after 1968 that are equivalent to Akaike's NCR

Note that a number of frequency domain measures of causality published very much after Akaike's 1968 NCR are actually equivalent to the NCR as is, or are equivalent to the NCR under some simple constraint such as an identity matrix for the innovation covariance.

This assertion can be checked and verified by the interested reader.

In this sense, it is most unfortunate that Akaike's seminal contribution hasn't been properly acknowledged as predating by many years the methods of Saito and Harashima (1981), Kaminski and Blinowska (1991), and Baccala and Sameshima (1998).

## 8. Constraining the NCR and equivalence to the iCoh

As mentioned in Yamashita et al (2005), the NCR in Eq. 20 can be applied in an ad hoc manner for the case of correlated innovations, by simply setting the off-diagonal elements of the innovation covariance matrix  $\mathbf{S}_\varepsilon$  to zero, which is equivalent to enforcing Eq. 11.

Now consider the case when both constraints given by Eq. 10 and Eq. 11 are forced (or used in an ad hoc manner) on the definition for the NCR. As with the iCoh, this corresponds to the condition that the only non-zero association between time series is due to  $[\check{\mathbf{A}}(\omega)]_{ij} \neq 0$ . The matrix  $\mathbf{B}(\omega)$  in Eq. 6 can be explicitly computed due to its simple structure, which has as non-zero elements all diagonal elements  $[\check{\mathbf{A}}(\omega)]_{kk} \neq 0$ , and the element  $[\check{\mathbf{A}}(\omega)]_{ij} \neq 0$ .

It can then be easily shown that the constrained NCR and the iCoh are identical.

Informally, the constrained NCR can be described as the answer to the following question: “Given a dynamic linear system characterized by its auto-regressive parameters, what would be the equation for NCR if all connections are severed, except for the single one of interest?”

## 9. The partial directed coherence (PDC) and the generalized partial directed coherence (gPDC)

These definitions are replicated here for the sake of completeness.

The PDC is:

$$\text{Eq. 22} \quad |\bar{\pi}_{ij}(\omega)|^2 = \frac{[\check{\mathbf{A}}(\omega)]_{ij}^2}{[\check{\mathbf{A}}^*(\omega)\check{\mathbf{A}}(\omega)]_{jj}} = \frac{[\check{\mathbf{A}}(\omega)]_{ij}^2}{\sum_{k=1}^q [\check{\mathbf{A}}(\omega)]_{kj}^2}$$

which corresponds to Baccala and Sameshima (2001), equation #18 therein.

The gPDC is:

$$\text{Eq. 23} \quad |\pi_{ij}^w(\omega)|^2 = \frac{[\mathbf{S}_\varepsilon]_{ii}^{-1} [\check{\mathbf{A}}(\omega)]_{ij}^2}{[\check{\mathbf{A}}^*(\omega)(\text{diag}\mathbf{S}_\varepsilon)^{-1}\check{\mathbf{A}}(\omega)]_{jj}} = \frac{[\mathbf{S}_\varepsilon]_{ii}^{-1} [\check{\mathbf{A}}(\omega)]_{ij}^2}{\sum_{k=1}^q [\mathbf{S}_\varepsilon]_{kk}^{-1} [\check{\mathbf{A}}(\omega)]_{kj}^2}$$

which corresponds to Baccala et al (2007), equation #11 therein. In Eq. 23, the notation “ $\text{diag}\mathbf{M}$ ” denotes a diagonal matrix formed by the diagonal elements of the  $\mathbf{M}$ .

## 10. The PDC and gPDC are neither coherences nor partial coherences

It is important to note that neither the partial directed *coherence*, nor the generalized partial directed *coherence* are *coherences* in any sense. Note that both satisfy:

$$\text{Eq. 24} \quad \sum_{j=1}^q \left| \pi_{ij}^w(\omega) \right|^2 = \sum_{j=1}^q \left| \bar{\pi}_{ij}(\omega) \right|^2 = 1$$

which is a property not related to nor satisfied by a coherence matrix.

It is important to clarify this fact, because even though the PDCs certainly do give frequency information on the direct connections, they do so in such a way which is fundamentally different from a proper, genuine coherence.

## 11. Toy Examples

Two toy examples will be considered. In both cases, 25600 time samples were generated (after discarding the first 1000 time samples) and used for all estimation procedures. Reported frequency values in Hz units correspond to the assumption that the sampling rate is 256 Hz. All frequency domain measures are shown from 1 to 127 Hz.

Multivariate auto-regressive models are estimated from the data by common least squares, assuming an auto-regressive order  $p=3$ , although both toy examples are of actual auto-regressive order  $p=2$ .

### Toy Example 9.1.

Toy Example 9.1. is taken from Baccala and Sameshima (2001), corresponding to their example #5, shown in their figure #4. The Baccala and Sameshima (2001) paper already compares their PDC measure with Akaike's NCR, which is referenced therein as the directed transfer function (DTF) of Kaminski and Blinowska (1991). For this reason, the NCR is not shown in this present study.

Figure 1 is a schematic representation of the direct connections among 5 nodes.

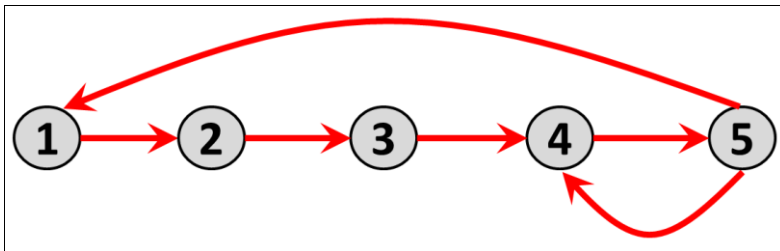


Figure 1: Toy Example 9.1. Schematic representation of the direct wiring among 5 nodes, in example#5 from Baccala and Sameshima (2001).

Table 1 shows the time domain auto-regressive parameters.

|                              |       |     |      |       |      |
|------------------------------|-------|-----|------|-------|------|
| $A(1) =$                     | 1.34  | 0   | 0    | 0     | 0    |
|                              | -0.5  | 0   | 0    | 0     | 0    |
|                              | 0     | 0   | 0    | 0     | 0    |
|                              | 0     | 0   | -0.5 | 0.35  | 0.35 |
|                              | 0     | 0   | 0    | -0.35 | 0.35 |
| $A(2) =$                     | -0.90 | 0   | 0    | 0     | 0.5  |
|                              | 0     | 0   | 0    | 0     | 0    |
|                              | 0     | 0.4 | 0    | 0     | 0    |
|                              | 0     | 0   | 0    | 0     | 0    |
|                              | 0     | 0   | 0    | 0     | 0    |
| $diagonal S_{\varepsilon} =$ |       |     |      |       |      |
|                              | 1     | 1   | 1    | 1     | 1    |

Table 1: Toy Example 9.1. Time domain auto-regressive parameters for 5 nodes, in example#5 from Baccala and Sameshima (2001).

Figure 2 displays 1024 time samples from the time series.

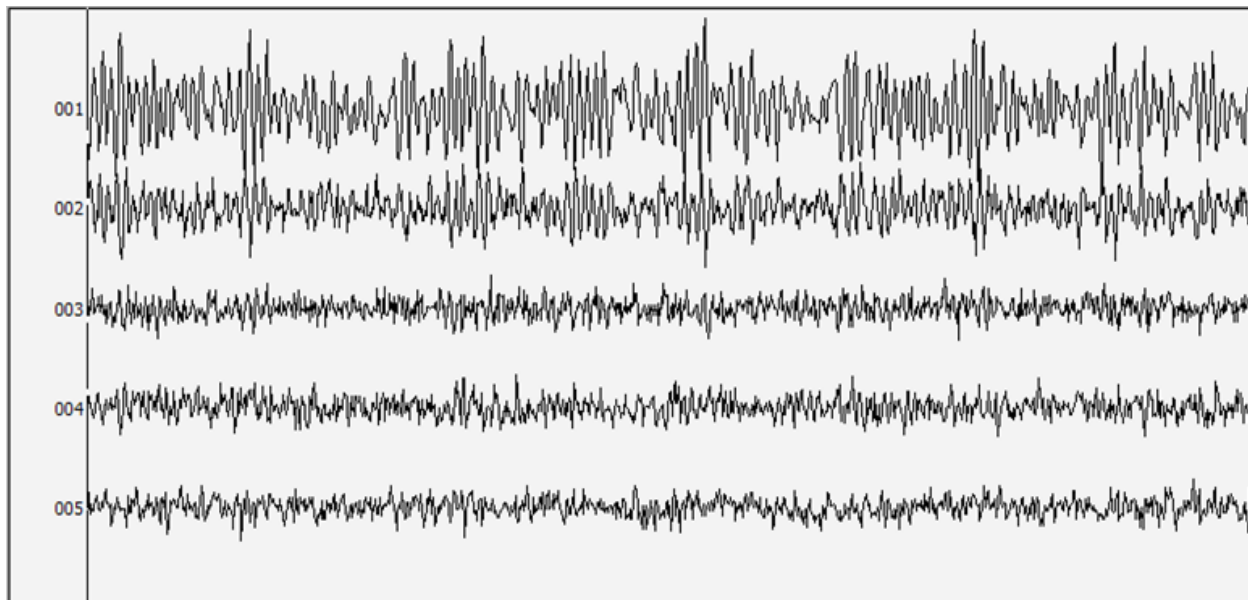


Figure 2: Toy Example 9.1. Time series display of 1024 time samples.

Figure 3 shows the coherence and the spectra.

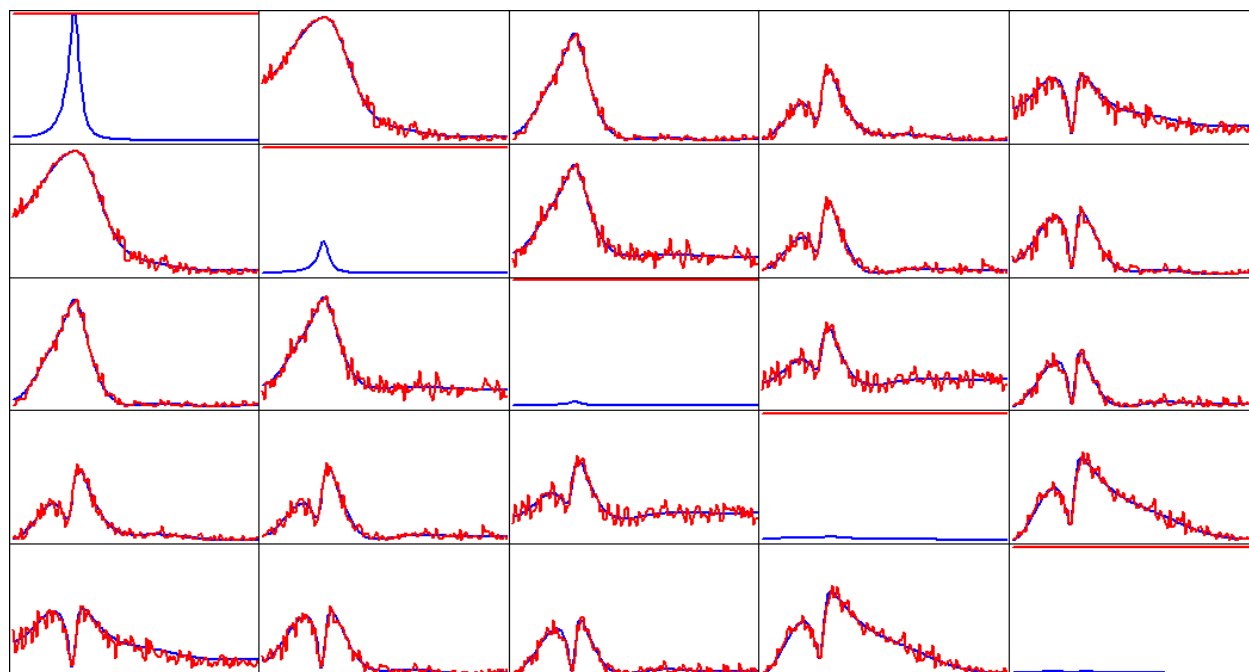


Figure 3: Toy Example 9.1. Auto-regressive spectra (diagonal, scaled to unit maximum power) and squared modulus of the coherence. Vertical axis: 0 to 1. Frequency axis: 1 to 127 Hz. Auto-regression based coherence estimates are shown in blue, while periodogram based estimates are shown in red. Spectral peak at 33 Hz (e.g. row 1, column 1). Coherence peaks at 22 and 35 Hz (e.g. row 4, column 1).

Figure 4 shows the iCoh (Eq. 15) and the gPDC (Eq. 23). Note that in this simple toy example taken from Baccala and Semashima (2001), the two methods give very similar results. These results show that the only non-zero values correctly detect the direction of the directly connected nodes in Figure 1. However, note that iCoh is slightly larger than gPDC for the connections of from node #5 to nodes #1 and #4.

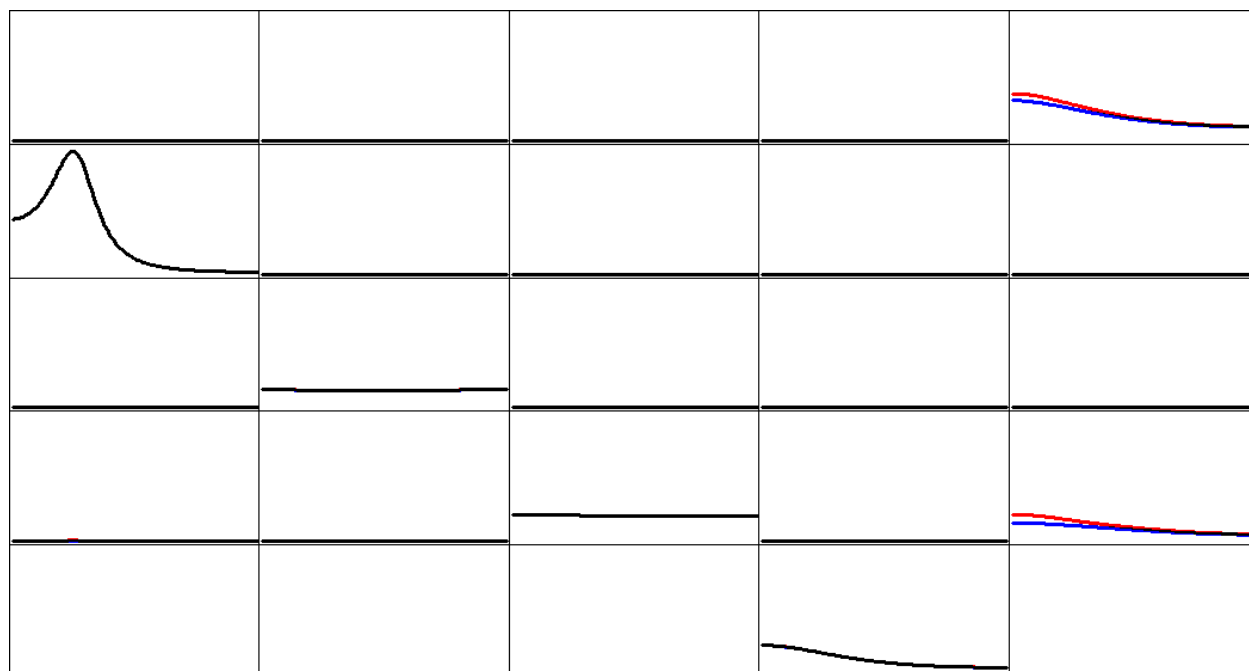


Figure 4: Toy Example 9.1. Isolated effective coherence (iCoh, Eq. 15) shown in RED, and the generalized partial directed coherence (gPDC, Eq. 23) shown in BLUE. Overlap of both curves is shown in BLACK. Vertical axis: 0 to 1. Frequency axis: 1 to 127 Hz. Columns are senders, rows are receivers. Coherence peak at 33 Hz (row 2, column 1).

The fact that the two methods (iCoh and gPDC) give similar results is only due to the simplicity of this example. As will be shown in the next toy example, the two methods can give very different results.

### Toy Example 9.2.

Figure 1 is a schematic representation of the direct connections among 5 nodes for Toy Example 9.2.

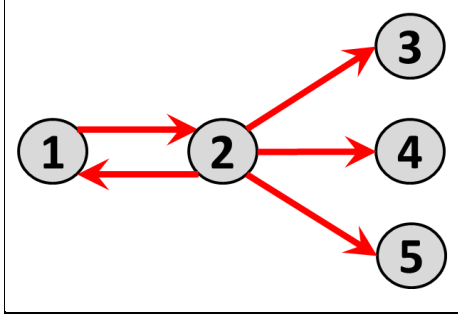


Figure 5: Toy Example 9.2. Schematic representation of the direct wiring among 5 nodes.

Table 2 shows the time domain auto-regressive parameters for Toy Example 9.2.

|                              |       |       |      |      |      |
|------------------------------|-------|-------|------|------|------|
| $A(1) =$                     | 1.5   | -0.25 | 0    | 0    | 0    |
|                              | -0.2  | 1.8   | 0    | 0    | 0    |
|                              | 0     | 0.9   | 1.65 | 0    | 0    |
|                              | 0     | 0.9   | 0    | 1.65 | 0    |
|                              | 0     | 0.9   | 0    | 0    | 1.65 |
| $A(2) =$                     | -0.95 | 0     | 0    | 0    | 0    |
|                              | 0     | -0.96 | 0    | 0    | 0    |
|                              | 0     | 0.8   | 0.95 | 0    | 0    |
|                              | 0     | 0.8   | 0    | 0.95 | 0    |
|                              | 0     | 0.8   | 0    | 0    | 0.95 |
| $diagonal S_{\varepsilon} =$ | 1     | 1     | 1    | 1    | 1    |

Table 2: Toy Example 9.2. Time domain auto-regressive parameters for 5 nodes.

Figure 6 displays 1024 time samples from the time series.

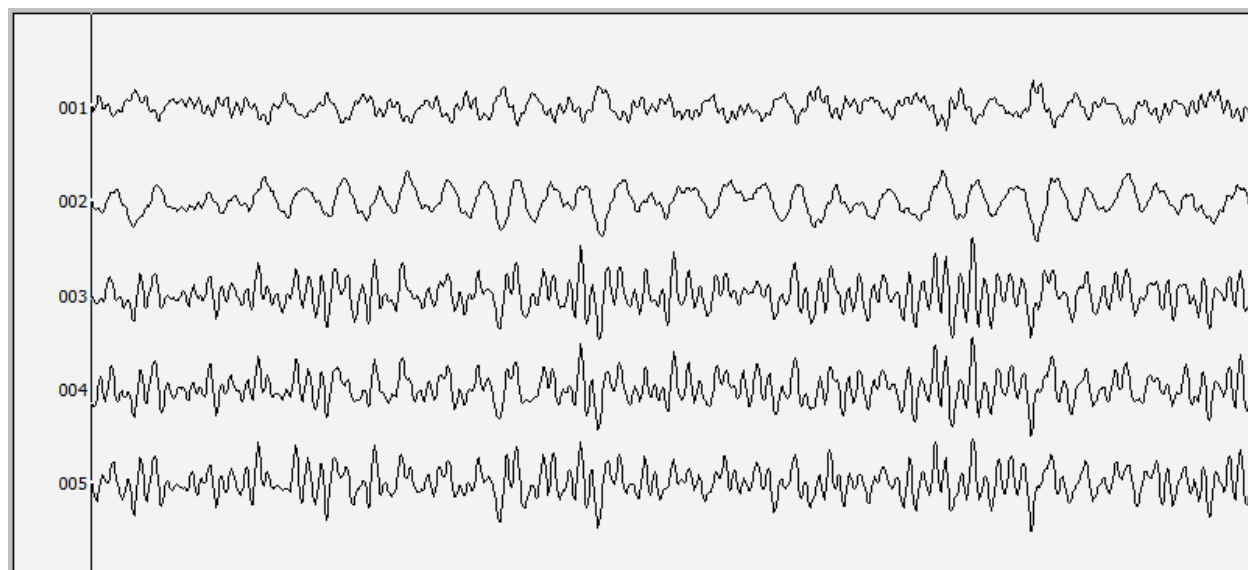


Figure 6: Toy Example 9.2. Time series display of 1024 time samples.

Figure 7 shows the coherence and the spectra. Note that practically all coherences reach very high values (close to 1) at some frequency.

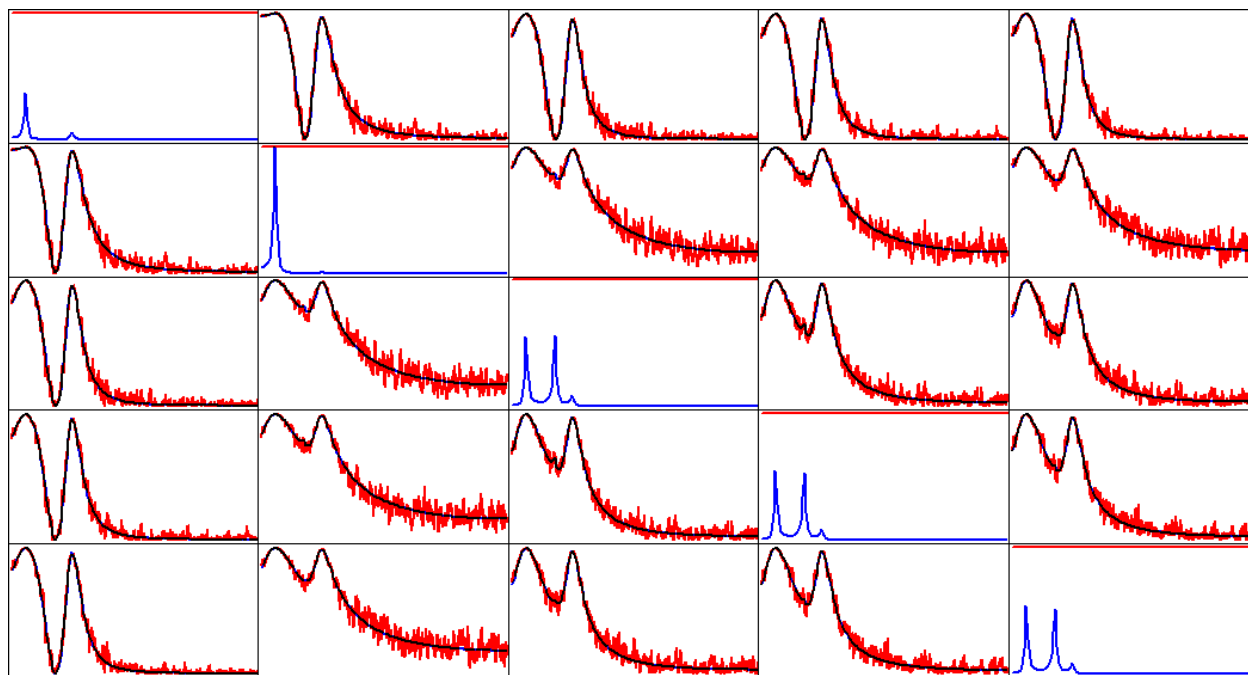


Figure 7: Toy Example 9.2. Auto-regressive spectra (diagonal, scaled to unit maximum power) and squared modulus of the coherence. Vertical axis: 0 to 1. Frequency axis: 1 to 127 Hz. Auto-regression based coherence estimates are shown in blue, while periodogram based estimates are shown in red. Spectral peaks at 8, and 32 Hz present for all diagonals; additional peak at 23 Hz for last three diagonals. Coherence peaks at 8 and 32 Hz.

Figure 8 shows the iCoh (Eq. 15) and the gPDC (Eq. 23). Note that in this toy example, the two methods give very different results with respect to node #2 as sender (column 2).

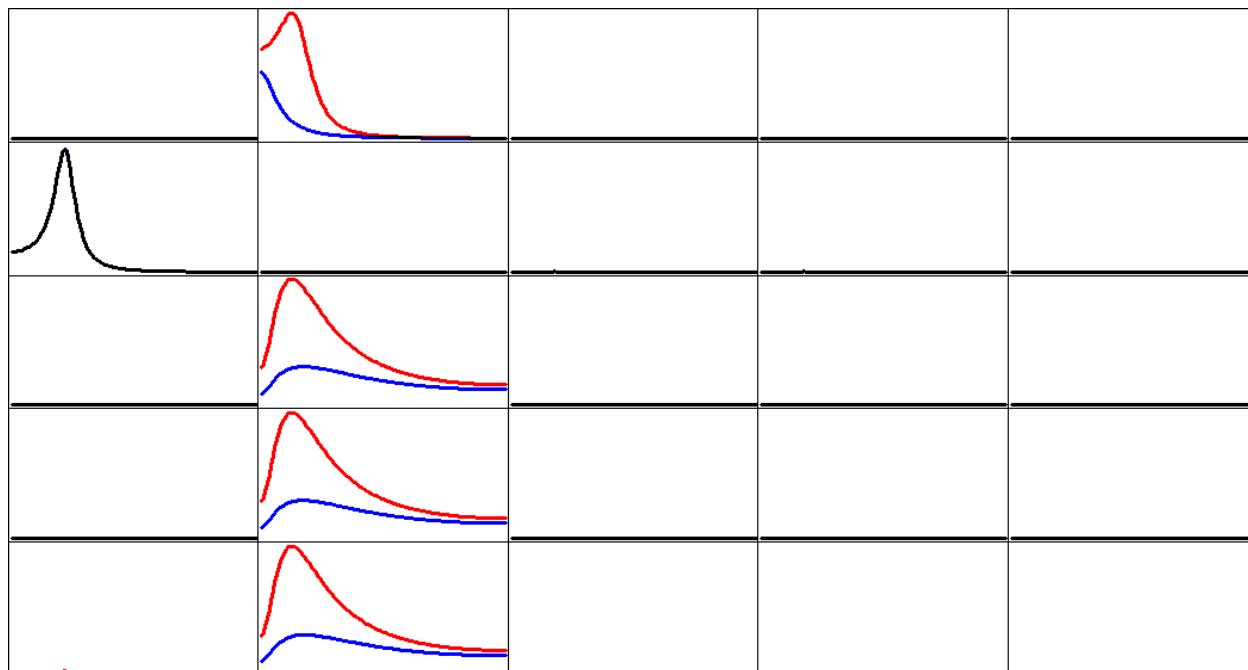


Figure 8: Toy Example 9.2. Isolated effective coherence (iCoh, Eq. 15) shown in **RED**, and the generalized partial directed coherence (gPDC, Eq. 23) shown in **BLUE**. Overlap of both curves is shown in **BLACK**. Vertical axis: 0 to 1. Frequency axis: 1 to 127 Hz. Columns are senders, rows are receivers. Coherence peak in column 2 occurs at 28 Hz. Coherence peak for iCoh in column 2 occurs at 16 Hz. Coherence peak for gPDC in column 2, row 1 occurs at 1 Hz; and Coherence peak for gPDC in column 2, rows 3, 4, and 5 occur at 23 Hz.

## 12. Discussion: the good and the bad of iCoh and gPDC

In general, both iCoh and gPDC give a frequency representation of direct, directional information flow, which is in general theoretical agreement with Granger causality.

However, due to the denominator in the definition of gPDC, this measure can produce very misleading results.

By definition and construction of the auto-regressive coefficients in Table 2, the intrinsic, isolated frequency generated by node #2 is 16 Hz. However, as shown in Toy Example 9.2., the gPDC would lead to the conclusion that node #2 is sending a small amount of low frequency information to node #1 (spectral peak at 1 Hz). This conclusion is incorrect and misleading.

In contrast, iCoh correctly identifies not only the peak frequency of the information being transmitted from node #1 to node #2, but also the strength of the connection, which is, by construction, very high.

The reason for this undesirable behavior of the gPDC is its denominator. When computing the gPDC for  $j \rightarrow i$ , the denominator includes many terms that contain irrelevant information, such as the coefficients corresponding to node “j” as a sender to all other non-related and irrelevant nodes.

The problem with gPDC is actually worse, in that not only misleading frequency characteristics can be produced, but also in that the actual values can be incorrectly low. This is also clear from Toy Example 9.2., where the classical coherences between (2,3), (2,4), and (2,5) are high (reaching values near to 1 at some frequency, see Figure 7), whereas the gPDC reports maximum value lower

than 0.5. In contrast, the iCoh shows correctly higher values, in agreement with the classical coherence.

These results cast doubt on the use of the gPDC in practice (see e.g. Sato et al 2009), since it is possible (but not necessarily always so) that it reports incorrectly both the frequency characteristics of the information being transferred, and the actual strength of the association.

The isolated effective coherence (iCoh) does not share these problems. Furthermore, it is a genuine coherence under certain constraints, unlike the gPDC which is not a proper coherence.

One important feature of iCoh is that it does not show the frequency characteristics that are actually being transmitted, which correspond to 8 and 32 Hz, as shown in Toy Example 9.2., the spectral peaks in Figure 7. By definition and construction of the AR coefficients (Table 2), the intrinsic, isolated frequencies of node #1 and #2 are 28 Hz and 16 Hz, respectively. And it is these isolated, intrinsic frequencies which are correctly detected by iCoh, together with the quantitative strength of the connection.

However, it would be desirable to also have a measure of effective coherence that at the same time provides correct information on: (1) actual frequency characteristics being transmitted, (2) directedness of the connections, and (3) direction of the connections. At the moment of this writing, work is in progress in this direction.

### 13. References

- 1) Akaike, H., 1968. On the use of a linear model for the identification of feedback systems. *Ann. Inst. Stat. Math.* 20, 425–439.
- 2) Baccala LA, Sameshima K, Ballester G, Valle AC, Timo-Iaria C (1998) Studying the interaction between brain structures via directed coherence and Granger causality. *Appl Signal Process* 5: 40-48.
- 3) L. Baccala, K. Sameshima. Partial directed coherence: a new concept in neural structure determination. *Biol. Cybern.*, 84 (2001), pp. 463–474.
- 4) Baccala LA, Takahashi DY, Sameshima K (2007) Generalized partial directed coherence. In: *Cardiff proceedings of the 2007 15th international conference on digital signal processing (DSP 2007)*, pp 162–166.
- 5) D.R. Brillinger, *Time Series: Data Analysis and Theory* (Holden-Day, Inc., San Francisco, 1981).
- 6) Granger, C. W. J. 1969 Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37, 424-438.
- 7) Kaminski MJ, Blinowska KJ (1991) A new method of the description of the information flow in the brain structures. *Biol Cybern* 65: 203-210.
- 8) Lutkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*, Springer, Berlin.
- 9) C. Radhakrishna Rao (1981) A lemma on G-inverse of a matrix and computation of correlation coefficients in the singular case, *Communications in Statistics - Theory and Methods*, 10:1, 1-10.
- 10) Saito Y, Harashima H (1981) Tracking of information within multichannel EEG record - causal analysis in EEG. In: Yamaguchi N, Fujisawa K (eds) *Recent advances in EEG and EMG Data Processing*. Elsevier, Amsterdam, pp 133-146.
- 11) Sato JR, Takahashi DY, Arcuri SM, Sameshima K, Morettin PA, Baccalá LA. 2009 Frequency domain connectivity identification: an application of partial directed coherence in fMRI. *Hum. Brain Mapp.* 30, 452–461.

- 12) Yamashita, N. Sadato, T. Okada, T. Ozaki, Evaluating frequency-wise directed connectivity of BOLD signals applying relative power contribution with the linear multivariate time-series models, *NeuroImage* 25 (2) (2005) 478–490.
- 13) SA Weiss, G McKhann, R Goodman, RG Emerson, A Trevelyan, M Bikson, CA Schevon. Field effects and ictal synchronization: insights from in homine observations. *Front Hum Neurosci.* 2013; 7: 828.
- 14) Marinazzo D, Liao W, Chen H, Stramaglia S. Nonlinear connectivity by Granger causality. *Neuroimage* 2011 58:330-338.
- 15) Schelter B, Timmer J, Eichler M. Assessing the strength of directed influences among neural signals using renormalized partial directed coherence. *J Neurosci Methods* 2009, 179:121-130.
- 16) Pearl, J. (2000): *Causality*. New York: Cambridge University Press.
- 17) Lehmann D, Faber PL, Tei S, Pascual-Marqui RD, Milz P, Kochi K. Reduced functional connectivity between cortical sources in five meditation traditions detected with lagged coherence using EEG tomography. *Neuroimage* 2012, 60:1574-1586.
- 18) Gomez-Herrero G, Atienza M, Egiazarian K, Cantero JL. Measuring directional coupling between EEG sources. *Neuroimage.* 2008, 43:497-508.
- 19) Roberto D. Pascual-Marqui, Dietrich Lehmann, Martha Koukkou, Kieko Kochi, Peter Anderer, Bernd Saletu, Hideaki Tanaka, Koichi Hirata, E. Roy John, Leslie Prichep, Rolando Biscay-Lirio, Toshihiko Kinoshita. Assessing interactions in the brain with exact low resolution electromagnetic tomography (eLORETA). *Philosophical Transactions of the Royal Society A;* (2011) 369, 3768–3784.
- 20) Faes L, Erla S, Porta A, Nollo G. 2013 A framework for assessing frequency domain causality in physiological time series with instantaneous effects. *Phil Trans R Soc A* 371: 20110618.

## 14. Appendix: The discrete Fourier transform

Let  $z(t)$ , with discrete time  $t = 0 \dots N_T - 1$ , denote a time series. Its discrete Fourier transform is defined as:

$$\text{Eq. 25} \quad z(\omega) = \sum_{t=0}^{N_T-1} z(t) \exp\left(-\imath \frac{2\pi t \omega}{N_T}\right)$$

with  $\imath = \sqrt{-1}$ , and with discrete frequency  $\omega = 0 \dots N_T - 1$ .

In the particular case of the discrete Fourier transform of autoregressive coefficients  $b(k)$  defined for  $k = 1 \dots p$ , the coefficients are redefined for  $k = 0 \dots N_T - 1$ , with:

$$b(k) = 0 \text{ for } k = 0 \text{ and } k > p$$

The discrete Fourier transform is calculated for this new series, using Eq. 25, giving  $b(\omega)$  for all discrete frequencies  $\omega = 0 \dots N_T - 1$ .

## 15. Acknowledgment

RJ Biscay was partly support by the CONICYT ACT 1112 and CONACYT 131771 projects.