Journal of Optimization Theory and Applications (2026) 209:32
https://doi.org/10.1007/s10957-026-02967-7

Sparse Optimization of Cross-Power Spectra in Linear Inverse Models from Brain Connectivity

Laura Carini[1] · **Isabella Furci**[1] · **Sara Sommariva**[1]

Received: 28 November 2024 / Accepted: 1 March 2026

Abstract
In this work, we present a computationally efficient linear optimization approach for estimating the cross–power spectrum of a hidden multivariate stochastic process from that of another observed process. Sparsity in the resulting estimator of the cross–power is induced through ℓ_1 regularization and the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) is used for computing such an estimator. With respect to a standard implementation, we prove that a proper initialization step is sufficient to guarantee the required symmetric and antisymmetric properties of the involved quantities. Further, we show how structural properties of the forward operator can be exploited within the FISTA update in order to make our approach adequate also for large–scale problems such as those arising in the context of brain functional connectivity. The effectiveness of the proposed approach is shown in a practical scenario where we aim at quantifying the statistical relationships between brain regions in the context of non-invasive electromagnetic field recordings. Our results show that our method provides results with a higher specificity than classical approaches based on a two–step procedure where first the hidden process describing the brain activity is estimated through a linear optimization step and then the cortical cross–power spectrum is computed from the estimated time–series.

Keywords Sparse Linear Optimization · Large Scale Linear Systems · Statistical multivariate analysis · Stochastic Processes · Functional Brain Connectivity

Mathematics Subject Classification 62H12 · 65F50 · 15A69 · 65K10

Communicated by Emanuele Galligani.

✉ Sara Sommariva
sara.sommariva@unige.it

Laura Carini
laura.carini@edu.unige.it

Isabella Furci
isabella.furci@edu.unige.it

[1] Dipartimento di Matematica, Universitá degli Studi di Genova, Genova, Italy

Published online: 11 April 2026

1 Introduction

We consider the problem of computing sparse estimation of the complex-valued cross-power spectrum of a multivariate stochastic process $\{\mathbf{X}(t)\}_{t\in\mathbb{R}}$, whose realizations, $\{\mathbf{x}(t)\}_{t\in\mathbb{R}}$, can only be observed indirectly through the realization, $\{\mathbf{y}(t)\}_{t\in\mathbb{R}}$, of another observable stochastic process $\{\mathbf{Y}(t)\}_{t\in\mathbb{R}}$. In particular, we show that if $\mathbf{Y}(t)$ is a linear combination of $\mathbf{X}(t)$ corrupted by additive noise, then the problem can be formulated as a large-scale linear optimization problem [33], whose dimension depends on the number of both the observations and the unknown variables.

Our work is motivated by the study of cortical functional connectivity from magneto-/electro-encephalographic (M/EEG) data, which consists in quantifying the statistical relationships between the activity of multiple brain regions from the electromagnetic field non-invasively recorded outside the scalp [22, 23]. From a mathematical point of view, this is typically achieved through a two-step procedure [25]. First a time–series describing the activity of different brain regions is estimated from the recorded electromagnetic field by solving an ill-posed linear optimization problem [2, 18]. Then a proper connectivity metric is computed between the estimated time–series. Due to the oscillatory nature of the neural sources [8], many connectivity metrics are defined in the frequency domain, starting from the cross-power spectrum of the time–series describing the estimated brain activity [20, 27].

A number of recent studies have demonstrated that this two–step procedure is inherently sub-optimal. When the initial optimization step is achieved through Tikhonov regularization [12] a regularization parameter has to be set, and this is typically done in order to obtain the best possible estimate of the brain activity. However, the cross-power spectrum computed from the time–series estimated with this value of the regularization parameter is usually sub-optimal, in the sense that better estimate can be obtained from time–series computed using a different value of the regularization parameter [32, 33] depending on features of the neural time-courses themselves, such as their spectral complexity [34]. Similar results were obtained in a simulation setup where Tikhonov regularization was used to estimate the power spectrum of the neural sources, and connectivity was quantified by using coherence (a normalized version of the cross-power spectrum) [15].

Furthermore, the estimates of the cortical cross-power spectrum obtained through this two–step procedure are usually affected by a large number of false positives, that is they identify statistical interactions between pairs of brain regions that are actually independent. One of the main causes of these spurious interactions is the residual mixing (or source–leakage) of the hidden process describing brain activity [25]. To mitigate this issue, metrics insensitive to zero-lag interactions can be computed from the estimated cross–power spectrum, such as the imaginary part of coherency [20] or the weighted phase lag index [28, 35]. However, such metrics have the obvious drawback that they also ignore instantaneous (or nearly instantaneous) true interactions [21].

In the last years, few optimization techniques have been proposed to avoid such two–step procedure. In [9] the authors propose a new method of MEG source reconstruction that simultaneously estimates the source amplitudes and interactions across the whole brain by a variational Bayesian algorithm; in particular, they use a multi-

variate autoregressive (AR) model to represent directed interaction between sources together with a prior knowledge on structural brain connectivity inferred from diffusion magnetic resonance imaging (dMRI). In [31] a Kalman filter is used to achieve a simultaneous estimation of source activities and their dynamic functional connectivity. However, also in this case AR models are used for representing source interactions. Ossadtchi and colleagues [21] introduce a novel projection matrix that operates on the cross–power spectrum of the observable process in order to mitigate the contribution of source–leakage to its real part. Once applied the projector, an optimization step still needs to be carried for estimating the cortical cross–power spectrum from the modified sensor–level data.

In this work we suggest an alternative approach where the cross–power spectrum of the hidden process describing the cortical activity is directly estimated from that of observed M/EEG time–series by setting up a linear optimization problem. In order to reduce the number of false positive, ℓ_1 regularization [7] is used so as to promote sparsity in the resulting estimator of the cortical cross–power spectrum. The Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) [4] is a classical approach for solving linear optimization problems with ℓ_1 penalty. However, its application to our problem is not trivial due to the high dimension of the matrix describing the forward operator which typically includes few hundreds of rows and few thousands of columns. Here we develop a computational strategy for efficiently carrying on the FISTA update by exploiting structural properties of the forward operator. The proposed method is validated on a large set of simulated data showing that our approach is capable of detecting the truly interacting sources but with a higher specificity than the classical two–step approach. The Python codes implementing the proposed approach are freely available at the GitHub repository https://github.com/theMIDAgroup/fista_cps_conn.

The paper is organised as follows. In Section 2 we derive the forward model relating the cross–power spectrum of the hidden process $\{\mathbf{X}(t)\}_{t\in\mathbb{R}}$ to that of the observable one. In Section 3 we present the novel proposed approach for estimating the cross–power spectrum of $\{\mathbf{X}(t)\}_{t\in\mathbb{R}}$ and we recall a classical two-step approach used as benchmark, while Section 4 focuses on the strategy we develop to make our approach computationally affordable. Finally, numerical validations of the presented approach are shown in Section 5 and our conclusions are discussed in Section 6.

2 Problem Formulation and Forward Modeling

Throughout the paper we assume without loss of generality that the mean values of all the considered stochastic processes are zero.

Definition 2.1 Let $\{\mathbf{X}(t)\}_{t\in\mathbb{R}}$ and $\{\mathbf{E}(t)\}_{t\in\mathbb{R}}$ be two real-valued, multivariate, stationary stochastic processes of dimension n and m, respectively. Denoted with $\left\{\Gamma^{\mathbf{XE}}(\tau)\right\}_{\tau\in\mathbb{R}}$ the corresponding covariance function, that is $\Gamma^{\mathbf{XE}}(\tau) = \mathbb{E}\left[\mathbf{X}(t)\mathbf{E}(t+\tau)^\top\right]$, we assume $\Gamma^{\mathbf{XE}}_{j,k}(\tau)$ to be absolutely integrable for all $j = 1,\dots,n$ and $k = 1,\dots,m$.

Then, the cross-power spectrum between $\mathbf{X}(t)$ and $\mathbf{E}(t)$ is a one–parameter family of complex–valued matrices $\mathbf{S}^{\mathbf{XE}}(f) \in \mathbb{C}^{n\times m}$ defined as [5]

$$S^{\mathbf{XE}}(f) = \int_{-\infty}^{+\infty} \Gamma^{\mathbf{XE}}(\tau) e^{-2\pi i f \tau} \mathrm{d}\tau,$$

where the integral is computed component-wise.

Remark 2.1 Following Definition 2.1, we define the cross-power spectrum of the process $\{\mathbf{X}(t)\}_{t\in\mathbb{R}}$ as a one–parameter family of Hermitian matrices of size $n \times n$ denoted as $\mathbf{S}^{\mathbf{X}}(f) := \mathbf{S}^{\mathbf{XX}}(f)$. Analogously, we denote $\Gamma^{\mathbf{X}}(\tau) := \Gamma^{\mathbf{XX}}(\tau)$.

Theorem 2.1 *Let $\{\mathbf{X}(t)\}_{t\in\mathbb{R}}$ and $\{\mathbf{Y}(t)\}_{t\in\mathbb{R}}$ be two real-valued, multivariate, stationary stochastic processes of dimension n and m, respectively. We further assume that $\mathbf{Y}(t)$ is a linear mixture of the components of $\mathbf{X}(t)$ corrupted by independent additive noise, that is*

$$\mathbf{Y}(t) = \mathbf{G}\mathbf{X}(t) + \mathbf{E}(t)\,, \tag{1}$$

where $\mathbf{G} \in \mathbb{R}^{m\times n}$ and $\{\mathbf{E}(t)\}_{t\in\mathbb{R}}$ is a multivariate, stationary stochastic process independent from $\{\mathbf{X}(t)\}_{t\in\mathbb{R}}$. Then,

$$\mathbf{S}^{\mathbf{Y}}(f) = \mathbf{G}\mathbf{S}^{\mathbf{X}}(f)\mathbf{G}^{\top} + \mathbf{S}^{\mathbf{E}}(f)\,. \tag{2}$$

Proof We observe that from the linearity of the expectation and the hypothesis on $\mathbf{Y}(t)$ it follows

$$\begin{aligned}
\Gamma^{\mathbf{Y}}(\tau) &= \mathbb{E}\Big[(\mathbf{G}\mathbf{X}(t) + \mathbf{E}(t))(\mathbf{G}\mathbf{X}(t+\tau) + \mathbf{E}(t+\tau))^{\top}\Big] \\
&= \mathbf{G}\Gamma^{\mathbf{X}}(\tau)\mathbf{G}^{\top} + \mathbf{G}\Gamma^{\mathbf{XE}}(\tau) + \Big(\mathbf{G}\Gamma^{\mathbf{XE}}(\tau)\Big)^{\top} + \Gamma^{\mathbf{E}}(\tau) \\
&= \mathbf{G}\Gamma^{\mathbf{X}}(\tau)\mathbf{G}^{\top} + \Gamma^{\mathbf{E}}(\tau),
\end{aligned}$$

where the last equality holds due to the fact that $\{\mathbf{X}(t)\}_{t\in\mathbb{R}}$ and $\{\mathbf{E}(t)\}_{t\in\mathbb{R}}$ are independent and thus $\Gamma^{\mathbf{XE}}(\tau) = 0$ for all τ. The thesis follows from Definition 2.1 by exploiting the linearity of the Fourier transform. □

Theorem 2.1 defines a linear relationship between the cross-power spectrum of the observable process $\{\mathbf{Y}(t)\}_{t\in\mathbb{R}}$ and that of the unknown process $\{\mathbf{X}(t)\}_{t\in\mathbb{R}}$. More explicitly, we denote with $\mathcal{S}^{\mathbf{X}}(f) \in \mathbb{C}^{n^2}$, and $\mathcal{S}^{\mathbf{Y}}(f), \mathcal{S}^{\mathbf{E}}(f) \in \mathbb{C}^{m^2}$ the vector obtained by stacking the columns (vec$(\cdot)$ operator) of matrices $\mathbf{S}^{\mathbf{X}}(f)$, $\mathbf{S}^{\mathbf{Y}}(f)$, and $\mathbf{S}^{\mathbf{E}}(f)$, respectively. From Eq. (2) it follows

$$\mathcal{S}^{\mathbf{Y}}(f) = (\mathbf{G}\otimes\mathbf{G})\,\mathcal{S}^{\mathbf{X}}(f) + \mathcal{S}^{\mathbf{E}}(f), \tag{3}$$

where $\otimes$ is the Kronecker product.

By splitting real and imaginary parts, Eq. (3) can be rewritten so as to include only real-valued quantities:

$$\begin{pmatrix} \mathrm{Re}(\mathcal{S}^{\mathbf{Y}}(f)) \\ \mathrm{Im}(\mathcal{S}^{\mathbf{Y}}(f)) \end{pmatrix} = \begin{pmatrix} \mathbf{G}\otimes\mathbf{G} & 0 \\ 0 & \mathbf{G}\otimes\mathbf{G} \end{pmatrix} \begin{pmatrix} \mathrm{Re}(\mathcal{S}^{\mathbf{X}}(f)) \\ \mathrm{Im}(\mathcal{S}^{\mathbf{X}}(f)) \end{pmatrix} + \begin{pmatrix} \mathrm{Re}(\mathcal{S}^{\mathbf{E}}(f)) \\ \mathrm{Im}(\mathcal{S}^{\mathbf{E}}(f)) \end{pmatrix} . \tag{4}$$

We finally remark that, since $\mathbf{S}^{\mathbf{X}}(f)$ is a Hermitian matrix, $\mathrm{Re}(\mathcal{S}^{\mathbf{X}}(f))$ and $\mathrm{Im}(\mathcal{S}^{\mathbf{X}}(f))$ are the vectorization of a symmetric and antisymmetric matrix, respectively.

3 Inverse Modeling

Given a realization $\{\mathbf{y}(t)\}_{t\in\mathbb{R}}$ of the observable process $\{\mathbf{Y}(t)\}_{t\in\mathbb{R}}$ our approach aims at providing a sparse estimation of the cross-power spectrum $\mathbf{S}^{\mathbf{X}}(f)$ by exploiting the model described by Eq. (4). When studying cortical functional connectivity from M/EEG data, this gives rise to a high-dimensional optimization problem. In fact, the forward matrix

$$\mathcal{G} := \begin{pmatrix} \mathbf{G}\otimes\mathbf{G} & 0 \\ 0 & \mathbf{G}\otimes\mathbf{G} \end{pmatrix} \tag{5}$$

has size $2m^2 \times 2n^2$, where, in connectivity studies, $m \propto 10^2$ is the number of M/EEG sensors and $n \propto 10^3$ is the number of cortical locations where the brain activity and connectivity are estimated. Hence, the matrix $\mathcal{G}$ has size proportional to $(2\cdot 10^4)\times(2\cdot 10^6)$, making it necessary to develop proper computational strategies, described in Section 4, to carry out any algorithmic step involving such a matrix.

Throughout the manuscript, our approach is compared to a classical two-step approach summarized in the next subsection.

3.1 Benchmark: Classical Two–Step Approach

A widely used approach for estimating the cross-power spectrum $\mathbf{S}^{\mathbf{X}}(f)$ consists of the following two steps [25, 32, 34].

Step 1. A regularised estimate, $\{\mathbf{x}_\lambda(t)\}_{t\in\mathbb{R}}$, of the unknown process $\{\mathbf{X}(t)\}_{t\in\mathbb{R}}$, is obtained by solving the inverse problem associated to Eq. (1). Here we consider the Tikhonov estimator [30], which is defined as

$$\mathbf{x}_\lambda(t) = \arg\min_{\mathbf{x}(t)} \left\{ \|\mathbf{G}\mathbf{x}(t) - \mathbf{y}(t)\|_2^2 + \lambda\|\mathbf{x}(t)\|_2^2 \right\} \quad \forall t \in T\,, \tag{6}$$

where λ is a proper regularisation parameter, $\|\cdot\|_2$ is the ℓ_2-norm, and $T \subset \mathbb{N}$ is the discrete set of time points where the data were collected.

Step 2. The Welch's estimator, $\mathbf{S}^{\mathbf{x}_\lambda}(f)$, of the time–series $\{\mathbf{x}_\lambda(t)\}_{t\in T}$ is computed as follows [36]. First the time–series $\{\mathbf{x}_\lambda(t)\}_{t\in T}$ is partitioned in P overlapping segments of length L, denoted as $\left\{\mathbf{x}_\lambda^{(p)}(\tau)\right\}_{\tau=0,\dots,L-1}$ with $p \in \{1,\dots,P\}$, and the discrete

Fourier transform

$$\hat{\mathbf{x}}_{\lambda}^{(p)}(f) = \frac{1}{L}\sum_{\tau=0}^{L-1} \mathbf{x}_{\lambda}^{(p)}(\tau) w(\tau) e^{\frac{-2\pi i \tau f}{L}}$$

is computed for each segment, $\{w(\tau)\}_{\tau=0,\ldots,L-1}$ being the Hamming window [13]. Then we defined

$$\mathbf{S}^{\mathbf{x}_\lambda}(f) = \frac{L}{PW}\sum_{p=1}^{P} \hat{\mathbf{x}}_{\lambda}^{p}(f)\hat{\mathbf{x}}_{\lambda}^{p}(f)^{H},$$

where $W = \frac{1}{L}\sum_{t=0}^{L-1} w(t)^2$.

Direct Estimation of the Cross–Power Spectrum of the Hidden Process. Fixed a frequency f, we propose to estimate the cross-power spectrum $\mathbf{S}^{\mathbf{X}}(f)$ through a least-squares approach with ℓ_1 regularization applied to the model in Eq. (4). Hence, after computing the Welch's estimator $\mathbf{S}^{\mathbf{y}}(f)$ of the cross–power spectrum of the observed data, we define [7, 29]

$$\begin{pmatrix} \widehat{\mathrm{Re}(\mathcal{S}^{\mathbf{X}})} \\ \widehat{\mathrm{Im}(\mathcal{S}^{\mathbf{X}})} \end{pmatrix} = \underset{\mathbf{s}}{\mathrm{argmin}} \left\{ \left\| \begin{pmatrix} \mathbf{G}\otimes\mathbf{G} & 0 \\ 0 & \mathbf{G}\otimes\mathbf{G} \end{pmatrix} \mathbf{s} - \begin{pmatrix} \mathrm{Re}(\mathcal{S}^{\mathbf{y}}) \\ \mathrm{Im}(\mathcal{S}^{\mathbf{y}}) \end{pmatrix} \right\|_2^2 + \lambda \|\mathbf{s}\|_1 \right\}, \quad (7)$$

where, for the sake of readability, we set $\mathbf{s} = (\mathbf{s}_2)$ and we omit the specification of the argument f.

In order to solve the optimization problem in (7) we used the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) [4] implemented as described in Algorithm 1. In particular we observe that in line 4 and 6 of the algorithm we have exploited the block-diagonal structure of the forward matrix defined in Eq. (5).

Following the original paper by Beck and Teboulle [4], in Algorithm 1 we used the shrinkage operator $\mathcal{T}_{\frac{\lambda}{L}} : \mathbb{R}^{n^2} \to \mathbb{R}^{n^2}$ defined as, for all $i = 1, \ldots, n^2$,

$$\mathcal{T}_{\frac{\lambda}{L}}(\mathbf{x})_i := \left(|x_i| - \frac{\lambda}{L}\right)^{+} \mathrm{sign}(x_i) = \begin{cases} x_i - \frac{\lambda}{L} & \text{if } x_i \geq \frac{\lambda}{L} \\ 0 & \text{if } |x_i| \leq \frac{\lambda}{L} \\ x_i + \frac{\lambda}{L} & \text{if } x_i \leq -\frac{\lambda}{L} \end{cases}, \quad (8)$$

where for all $z \in \mathbb{R}$, $(z)^+ := \max\{z, 0\}$ and $\mathrm{sign}(z)$ denote the ramp function (or positive part) and the sign function, respectively, and

$$L = 2\lambda_{\max}(\mathcal{G}^{\top}\mathcal{G}), \quad (9)$$

being $\lambda_{\max}(\mathcal{G}^{\top}\mathcal{G})$ the highest eigenvalue of the squared matrix $\mathcal{G}^{\top}\mathcal{G}$.

Remark 3.1 The value in Eq. (9) is a common choice for the constant L in optimization problem of the form (7) [4] . In fact, it corresponds to the Lipschitz constant of the gradient of the function $\mathbf{f} : \mathbb{R}^{2n^2} \to \mathbb{R}$, defined as

$$\mathbf{f}\begin{pmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \end{pmatrix} = \left\| \mathcal{G}\begin{pmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \end{pmatrix} - \begin{pmatrix} \mathrm{Re}(\mathcal{S}^{\mathbf{y}}) \\ \mathrm{Im}(\mathcal{S}^{\mathbf{y}}) \end{pmatrix} \right\|_2^2.$$

In our case, computing the Lipschitz constant L by means of Eq. (9) requires determining the eigenvalues of a matrix of size $2n^2 \times 2n^2$. However, a more efficient computation can be carried on by exploiting the block-diagonal structure of $\mathcal{G}$ and the properties of the Kronecker product:

$$\begin{aligned} L &= 2\lambda_{\max}\left(\begin{pmatrix} \mathbf{G}\otimes\mathbf{G} & 0 \\ 0 & \mathbf{G}\otimes\mathbf{G} \end{pmatrix}^{\top}\begin{pmatrix} \mathbf{G}\otimes\mathbf{G} & 0 \\ 0 & \mathbf{G}\otimes\mathbf{G} \end{pmatrix}\right) \\ &= 2\lambda_{\max}\begin{pmatrix} \mathbf{G}^{\top}\mathbf{G}\otimes\mathbf{G}^{\top}\mathbf{G} & 0 \\ 0 & \mathbf{G}^{\top}\mathbf{G}\otimes\mathbf{G}^{\top}\mathbf{G} \end{pmatrix} \\ &= 2\left[\lambda_{\max}(\mathbf{G}^{\top}\mathbf{G})\right]^2 . \end{aligned}$$

Hence, from a practical point of view, L can be computed as

$$L = 2\sigma_{\max}^4(\mathbf{G}) , \tag{10}$$

being $\sigma_{\max}(\mathbf{G})$ the highest singular value of $\mathbf{G}$.

Algorithm 1 FISTA for a direct estimation of the cross–power spectrum $\mathbf{S^x}$ in the frequency domain.

Input: $\mathbf{G} \in \mathbb{R}^{m\times n}$; λ, L, K, $\varepsilon > 0$; $\mathbf{S^y} \in \mathbb{C}^{m\times m}$ and $\mathbf{S}_0 \in \mathbb{C}^{n\times n}$ Hermitian matrices and corresponding vectorization $\mathcal{S}^{\mathbf{y}} := \mathrm{vec}(\mathbf{S^y}) \in \mathbb{C}^{m^2}$ and $\mathcal{S}_0 := \mathrm{vec}(\mathbf{S}_0) \in \mathbb{C}^{n^2}$.

1: **Initialize:** $k = 0$, $t_0 = 0$, $\begin{pmatrix} \mathbf{w}_{0,1} \\ \mathbf{w}_{0,2} \end{pmatrix} = \begin{pmatrix} \mathbf{s}_{0,1} \\ \mathbf{s}_{0,2} \end{pmatrix} = \begin{pmatrix} \mathrm{Re}(\mathcal{S}_0) \\ \mathrm{Im}(\mathcal{S}_0) \end{pmatrix}$

2: **while** $k \le K$ **and** $e \le \varepsilon$ **do**

3: $\quad k = k + 1$

4: $\quad \begin{pmatrix} \mathbf{s}_{k,1} \\ \mathbf{s}_{k,2} \end{pmatrix} = \begin{pmatrix} \mathcal{T}_{\frac{\lambda}{L}}\left(\mathbf{w}_{k-1,1} - \frac{2}{L}(\mathbf{G}\otimes\mathbf{G})^{\top}((\mathbf{G}\otimes\mathbf{G})\mathbf{w}_{k-1,1} - \mathrm{Re}(\mathcal{S}^{\mathbf{y}})\right) \\ \mathcal{T}_{\frac{\lambda}{L}}\left(\mathbf{w}_{k-1,2} - \frac{2}{L}(\mathbf{G}\otimes\mathbf{G})^{\top}((\mathbf{G}\otimes\mathbf{G})\mathbf{w}_{k-1,2} - \mathrm{Im}(\mathcal{S}^{\mathbf{y}})\right) \end{pmatrix}$

5: $\quad t_k = \frac{1+\sqrt{1+4t_{k-1}^2}}{2}$

6: $\quad \begin{pmatrix} \mathbf{w}_{k,1} \\ \mathbf{w}_{k,2} \end{pmatrix} = \begin{pmatrix} \mathbf{s}_{k,1} \\ \mathbf{s}_{k,2} \end{pmatrix} + \frac{t_{k-1}-1}{t_k}\begin{pmatrix} \mathbf{s}_{k,1} - \mathbf{s}_{k-1,1} \\ \mathbf{s}_{k,2} - \mathbf{s}_{k-1,2} \end{pmatrix}$

7: $\quad e = \frac{||\mathbf{s}_{k,1}-\mathbf{s}_{k-1,1}||_1 + ||\mathbf{s}_{k,2}-\mathbf{s}_{k-1,2}||_1}{||\mathbf{s}_{k,1}||_1 + ||\mathbf{s}_{k,2}||_1}$

8: **end while**

Output: $\begin{pmatrix} \mathbf{s}_{k,1} \\ \mathbf{s}_{k,2} \end{pmatrix}$

Since we seek a solution to the optimization problem (7) to be interpreted as the cross-power spectrum of a stochastic process, $\mathbf{s}_1$ and $\mathbf{s}_2$ need to be the vectorization of a symmetric and antisymmetric matrix, respectively. The following results prove that this can be achieved by initializing FISTA through a random Hermitian matrix, as done in Algorithm 1.

Proposition 3.1 *Let $\mathcal{A} := \mathrm{vec}(\mathbf{A}) \in \mathbb{R}^{m^2}$ be the vectorization of a matrix $\mathbf{A} \in \mathbb{R}^{m\times m}$ and let $\mathcal{T}_\alpha$ the shrinkage operator defined as in Eq. (8) by replacing $\frac{\lambda}{L}$ with a generic parameter $\alpha > 0$. The following properties hold:*

(a) if $\mathbf{A}$ is symmetric, then $\mathcal{T}_\alpha(\mathcal{A})$ is the vectorization of a symmetric matrix;
(b) if $\mathbf{A}$ is antisymmetric, then $\mathcal{T}_\alpha(\mathcal{A})$ is the vectorization of an antisymmetric matrix.

Proof Since $\mathbf{A}_{ij} = \mathcal{A}_{m(j-1)+i}$ for all $i, j = 1, \ldots, m$, it can be shown that

$$\mathbf{A} \text{ is symmetric} \iff \mathcal{A}_{m(j-1)+i} = \mathcal{A}_{m(i-1)+j}, \text{ for all } i, j = 1, \ldots, m,$$

and

$$\mathbf{A} \text{ is antisymmetric} \iff \mathcal{A}_{m(j-1)+i} = -\mathcal{A}_{m(i-1)+j}, \text{ for all } i, j = 1, \ldots, m.$$

Hence, the thesis follows by the definition of $\mathcal{T}_\alpha$ observing that if $\mathbf{A}$ is symmetric then

$$\begin{aligned}\mathcal{T}_\alpha(\mathcal{A})_{m(j-1)+i} &= \left(|\mathcal{A}_{m(j-1)+i}| - \alpha\right)^+ \mathrm{sign}(\mathcal{A}_{m(j-1)+i}) \\ &= \left(|\mathcal{A}_{m(i-1)+j}| - \alpha\right)^+ \mathrm{sign}(\mathcal{A}_{m(i-1)+j}) = \mathcal{T}_\alpha(\mathcal{A})_{m(i-1)+j}.\end{aligned}$$

Similarly, if $\mathbf{A}$ is antisymmetric, then

$$\begin{aligned}\mathcal{T}_\alpha(\mathcal{A})_{m(j-1)+i} &= \left(|\mathcal{A}_{m(j-1)+i}| - \alpha\right)^+ \mathrm{sign}(\mathcal{A}_{m(j-1)+i}) \\ &= \left(|-\mathcal{A}_{m(i-1)+j}| - \alpha\right)^+ \mathrm{sign}(-\mathcal{A}_{m(i-1)+j}) = -\mathcal{T}_\alpha(\mathcal{A})_{m(i-1)+j}.\end{aligned}$$

□

Theorem 3.1 *Let $\left\{\begin{pmatrix}\mathbf{s}_{k,1}\\ \mathbf{s}_{k,2}\end{pmatrix}\right\}_{k\geq 0}$ and $\left\{\begin{pmatrix}\mathbf{w}_{k,1}\\ \mathbf{w}_{k,2}\end{pmatrix}\right\}_{k\geq 0}$ the sequences generated with Algorithm 1 with input the cross–power spectrum of the observed data $\mathbf{S}^{\mathbf{y}} \in \mathbb{C}^{m\times m}$ and a random Hermitian matrix $\mathbf{S}_0 \in \mathbb{C}^{n\times n}$ as initial point. Then $\{\mathbf{s}_{k,1}\}_{k\geq 0}$ and $\{\mathbf{w}_{k,1}\}_{k\geq 0}$ are vectorizations of symmetric matrices and $\{\mathbf{s}_{k,2}\}_{k\geq 0}$ and $\{\mathbf{w}_{k,2}\}_{k\geq 0}$ are vectorization of antisymmetric matrices.*

Proof To prove the theorem we proceed by induction.

For $k = 0$, $\begin{pmatrix}\mathbf{w}_{0,1}\\ \mathbf{w}_{0,2}\end{pmatrix} = \begin{pmatrix}\mathbf{s}_{0,1}\\ \mathbf{s}_{0,2}\end{pmatrix} = \begin{pmatrix}\mathrm{Re}(\mathcal{S}_0)\\ \mathrm{Im}(\mathcal{S}_0)\end{pmatrix}$. Hence the thesis follows from the fact that $\mathbf{S}_0$ is Hermitian.

Fixed $k > 0$, we assume that $\mathbf{s}_{k-1,1}$ and $\mathbf{w}_{k-1,1}$ are vectorization of two symmetric matrices denoted with $\mathbf{S}_{k-1,1}$ and $\mathbf{W}_{k-1,1}$, respectively. Analogously, we assume that $\mathbf{s}_{k-1,2}$ and $\mathbf{w}_{k-1,2}$ are vectorization of two antisymmetric matrices denoted with $\mathbf{S}_{k-1,2}$ and $\mathbf{W}_{k-1,2}$, respectively.

By exploiting the linearity of the vec$(\cdot)$ operator and the properties of the Kronecker product, from line 4 of Algorithm 1 it follows:

$$\begin{aligned}\mathbf{s}_{k,1} &= \mathcal{T}_{\frac{\lambda}{L}}\left(\mathbf{w}_{k-1,1} - \frac{2}{L}\left(\mathbf{G}\otimes\mathbf{G}\right)^\top((\mathbf{G}\otimes\mathbf{G})\mathbf{w}_{k-1,1} - \mathrm{Re}(\mathcal{S}^y)\right)\right)\\ &= \mathcal{T}_{\frac{\lambda}{L}}\left(\mathrm{vec}(\mathbf{W}_{k-1,1}) - \frac{2}{L}\left(\mathbf{G}\otimes\mathbf{G}\right)^\top \mathrm{vec}\left(\mathbf{G}\mathbf{W}_{k-1,1}\mathbf{G}^\top - \mathrm{Re}(\mathbf{S}^{\mathbf{y}})\right)\right) \qquad (11)\\ &= \mathcal{T}_{\frac{\lambda}{L}}\left(\mathrm{vec}\left(\mathbf{W}_{k-1,1} - \frac{2}{L}\mathbf{G}^\top\mathbf{G}\mathbf{W}_{k-1,1}\mathbf{G}^\top\mathbf{G} + \frac{2}{L}\mathbf{G}^\top\mathrm{Re}(\mathbf{S}^{\mathbf{y}})\mathbf{G}\right)\right).\end{aligned}$$

Since $\mathbf{W}_{k-1,1}$ and $\mathrm{Re}(\mathbf{S}^{\mathbf{y}})$ are symmetric for the inductive hypothesis and the properties of the cross-power spectrum, respectively, the argument on the right side of the last equation in (11) results to be a symmetric matrix. Hence Proposition 3.1 ensures $\mathbf{s}_{k,1}$ is the vectorization of a symmetric matrix, $\mathbf{S}_{k,1}$.

Similarly from line 4 of Algorithm 1 it follows:

$$\begin{aligned}\mathbf{s}_{k,2} &= \mathcal{T}_{\frac{\lambda}{L}}\left(\mathbf{w}_{k-1,2} - \frac{2}{L}(\mathbf{G}\otimes\mathbf{G})^\top((\mathbf{G}\otimes\mathbf{G})\mathbf{w}_{k-1,2} - \mathrm{Im}(\mathcal{S}^{\mathbf{y}}))\right)\\ &= \mathcal{T}_{\frac{\lambda}{L}}\left(\mathrm{vec}\left(\mathbf{W}_{k-1,2} - \frac{2}{L}\mathbf{G}^\top\mathbf{G}\mathbf{W}_{k-1,2}\mathbf{G}^\top\mathbf{G} + \frac{2}{L}\mathbf{G}^\top\mathrm{Im}(\mathbf{S}^{\mathbf{y}})\mathbf{G}\right)\right),\end{aligned}$$

where the argument of the right side of the last equation is antisymmetric because $\mathbf{W}_{k-1,2}$ and $\mathrm{Im}(\mathbf{S}^{\mathbf{y}})$ are antisymmetric. Hence, Proposition 3.1 ensures $\mathbf{s}_{k,2}$ is the vectorization of a antisymmetric matrix, $\mathbf{S}_{k,2}$.

The fact that $\mathbf{w}_{k,1}$ and $\mathbf{w}_{k,2}$ are the vectorization of a symmetric and an antisymmetric matrix then follows from line 6 of Algorithm 1, that can be rewritten as follows

$$\begin{pmatrix}\mathbf{w}_{k,1}\\ \mathbf{w}_{k,2}\end{pmatrix} = \begin{pmatrix}\mathrm{vec}\left(\mathbf{S}_{k,1} + \frac{t_{k-1}-1}{t_k}(\mathbf{S}_{k,1} - \mathbf{S}_{k-1,1})\right)\\ \mathrm{vec}\left(\mathbf{S}_{k,2} + \frac{t_{k-1}-1}{t_k}(\mathbf{S}_{k,2} - \mathbf{S}_{k-1,2})\right)\end{pmatrix}.$$

□

4 Smart Product for an Efficient Computation of the FISTA Update

The FISTA update at line 4 of Algorithm 1 requires several matrix–vector multiplications involving the matrix $\mathbf{G}\otimes\mathbf{G}$ and its transpose $(\mathbf{G}\otimes\mathbf{G})^\top = \mathbf{G}^\top\otimes\mathbf{G}^\top$. Since $\mathbf{G}\in\mathbb{R}^{m\times n}$, $\mathbf{G}\otimes\mathbf{G}$ and $\mathbf{G}^\top\otimes\mathbf{G}^\top$ have size $m^2\times n^2$ and $n^2\times m^2$, respectively, the product of a vector by each one of these matrices has a cost proportional to $O(m^2n^2)$. In this section we show how the properties of the Kronecker product can be exploited to reduce both the computational cost and the memory requirements of such product. More specifically, our approach has two main advantages: on the one hand it reduces the computational cost to a value proportional to $O(\max(m,n)\,mn)$, on the other hand it avoids explicitly assembling the matrix $\mathbf{G}\otimes\mathbf{G}$, by directly employing the matrix $\mathbf{G}$.

First, we recall that the Kronecker product enjoys the mixed-product property [17], therefore it holds

$$\underbrace{\mathbf{G}\otimes\mathbf{G}}_{m^2\times n^2} = (\mathbf{G}\,\mathbf{I}_n)\otimes(\mathbf{I}_m\,\mathbf{G}) = \underbrace{(\mathbf{G}\otimes\mathbf{I}_m)}_{m^2\times nm}\underbrace{(\mathbf{I}_n\otimes\mathbf{G})}_{nm\times n^2}. \tag{12}$$

Now we present a result that establishes a connection between the elements of $\mathbf{G}\otimes\mathbf{I}_m$ and those of $\mathbf{I}_m\otimes\mathbf{G}$ through matrices that permute rows and columns. This relationship plays a crucial role in the procedure for efficiently computing the matrix-vector product with $(\mathbf{G}\otimes\mathbf{G})$ (or its transpose, $(\mathbf{G}\otimes\mathbf{G})^\top$).

Proposition 4.1 *Let $\mathbf{G}\in\mathbb{R}^{m\times n}$ be a rectangular matrix. Then there exist two matrices $\mathbf{P}_{m^2}\in\mathbb{R}^{m^2\times m^2}$ and $\mathbf{P}_{nm}\in\mathbb{R}^{nm\times nm}$ such that*

$$\mathbf{G}\otimes\mathbf{I}_m = \mathbf{P}_{m^2}(\mathbf{I}_m\otimes\mathbf{G})\mathbf{P}_{nm}. \tag{13}$$

Proof The thesis is equivalent to identify two permutation functions for the row and column indices that rearrange the elements of $\mathbf{I}_m\otimes\mathbf{G}$, making it identical to $\mathbf{G}\otimes\mathbf{I}_m$. Relation (13) is then established by defining $\mathbf{P}_{m^2}\in\mathbb{R}^{m^2\times m^2}$ and $\mathbf{P}_{nm}\in\mathbb{R}^{nm\times nm}$ as identity matrices of the corresponding sizes, with the specified permutations applied to their rows and columns [17].

Then we first highlight that the structures of $\mathbf{G}\otimes\mathbf{I}_m$ and $\mathbf{I}_m\otimes\mathbf{G}$ are given by

$$\mathbf{G}\otimes\mathbf{I}_m = \underbrace{\begin{bmatrix} g_{1,1}\mathbf{I}_m & g_{1,2}\mathbf{I}_m & \cdots & g_{1,n}\mathbf{I}_m \\ g_{2,1}\mathbf{I}_m & g_{2,2}\mathbf{I}_m & \cdots & g_{2,n}\mathbf{I}_m \\ \vdots & \vdots & \ddots & \vdots \\ g_{m,1}\mathbf{I}_m & g_{m,2}\mathbf{I}_m & \cdots & g_{m,n}\mathbf{I}_m \end{bmatrix}}_{nm \text{ columns}}, \quad \mathbf{I}_m\otimes\mathbf{G} = \left.\begin{bmatrix} \mathbf{G} & O_{m,n} & \cdots & O_{m,n} \\ O_{m,n} & \mathbf{G} & \cdots & O_{m,n} \\ \vdots & \vdots & \ddots & \vdots \\ O_{m,n} & O_{m,n} & \cdots & \mathbf{G} \end{bmatrix}\right\} m^2 \text{ rows}$$

Consequently, we obtain the elements of $\mathbf{G}\otimes\mathbf{I}_m$ from those of $\mathbf{I}_m\otimes\mathbf{G}$ with a permutation that brings the $i-$th row in position

$$\mathrm{mod}(i-1,m)\,m + \left\lfloor \frac{i-1}{m} \right\rfloor + 1, \quad \text{for } i = 1\ldots m^2; \tag{14}$$

and the $j-$th column in position

$$\mathrm{mod}(j-1,n)\,m + \left\lfloor \frac{j-1}{n} \right\rfloor + 1, \quad \text{for } j = 1\ldots nm, \tag{15}$$

where, for all $a, b\in\mathbb{N}$ with $\mathrm{mod}(a,b)$ we denote the remainder after the division of a by b and with $\lfloor a\rfloor$ the floor of a.

Indeed, note that, $\mathrm{mod}(i-1,m)$ (resp. $\mathrm{mod}(j-1,n)$) is the quantity that serves to group the row indices (resp. coloumn) into blocks of m rows (resp. coloumn), determining the relative position within each block. The value $\left\lfloor \frac{i-1}{m} \right\rfloor$ (resp. $\left\lfloor \frac{j-1}{n} \right\rfloor$)

determines the block number we are considering. See [3, Section 2.5] and [6, Remarks 2-3] for further visualizations and definitions on the block rectangular permutations.

Let $\mathbf{e}_\ell$ be the $\ell-$th canonical vector with 1 in position ℓ and zeros elsewhere, that is $(\mathbf{e}_\ell)_j = \delta_{j\ell}$. We conclude the proof choosing $\mathbf{P}_{m^2}$ and $\mathbf{P}_{nm}$ the two permutation matrices where the one entry equal to 1 is given exploiting the two permutation functions (14) and (15). In detail,

$$\mathbf{P}_{m^2} = \left[e_{\xi_r(1)} \mid e_{\xi_r(2)} \mid \cdots \mid e_{\xi_r(m^2)}\right], \quad \mathbf{P}_{nm} = \left[e_{\xi_c(1)} \mid e_{\xi_c(2)} \mid \cdots \mid e_{\xi_c(nm)}\right],$$

where

$$\begin{aligned} \xi_r(i) &= \mathrm{mod}(i-1, m)\, m + \left\lfloor \frac{i-1}{m} \right\rfloor + 1, \quad i = 1 \ldots m^2, \\ \xi_c(j) &= \mathrm{mod}(j-1, n)\, m + \left\lfloor \frac{j-1}{n} \right\rfloor + 1, \quad j = 1 \ldots nm. \end{aligned} \tag{16}$$

□

Theorem 4.1 *Let $\mathbf{G} \in \mathbb{R}^{m\times n}$ be a rectangular matrix and $\mathbf{x} \in \mathbb{R}^{n^2}$. The matrix–vector product $(\mathbf{G}\otimes\mathbf{G})\mathbf{x}$ can be performed in $O(\max(m,n)\,mn)$ operations with a matrix–free approach (denoted by Algorithm 2).*

Proof Eq. (12) and Proposition 4.1 imply

$$(\mathbf{G}\otimes\mathbf{G})\mathbf{x} = (\mathbf{G}\otimes\mathbf{I}_m)(\mathbf{I}_n\otimes\mathbf{G})\mathbf{x} = \mathbf{P}_{m^2}(\mathbf{I}_m\otimes\mathbf{G})\mathbf{P}_{nm}(\mathbf{I}_n\otimes\mathbf{G})\mathbf{x},$$

where $\mathbf{P}_{m^2}$ and $\mathbf{P}_{nm}$ are the permutation matrices defined by functions ξ_r and ξ_c in Eq. (16) in Proposition 4.1. Hence the product $(\mathbf{G}\otimes\mathbf{G})\mathbf{x}$ can be decomposed in few steps that form Algorithm 2. We now analyze the proposed procedure and demonstrate that it does not require assembling any of the matrix tensor products. Moreover, we prove that the total computational cost of each step is proportional to $\max(m,n)\cdot mn$.

The cost of step 2 and 4 of Algorithm 2 is 0 since they only require the index exchange suggested by ξ_r and ξ_c in Eq. (16). The vector $\mathbf{x} \in \mathbb{R}^{n^2}$ can be split into n vectors $\hat{x}_\ell$ of size n. In particular,

$$\hat{x}_\ell = \left[x_{(\ell-1)n+k}\right]_{k=1}^{n}, \quad \ell = 1, \ldots, n.$$

Similarly, the vector $\widehat{\mathbf{y}}$ can be split into m vectors of size n.

The latter and the block diagonal structures of $\mathbf{I}_n\otimes\mathbf{G}$ and $\mathbf{I}_m\otimes\mathbf{G}$, imply that the multiplications in steps 1 and 3 of Algorithm 2 avoid assembling the matrix $\mathbf{G}\otimes\mathbf{G}$ (or any other matrix tensor product). Moreover, they consist of n and m matrix-vector products with the matrix $\mathbf{G}$, respectively. Therefore the total computational cost is proportional to $n\cdot nm + m\cdot nm$, that is $O(\max(n,m)\cdot mn)$. □

Remark 4.1 Since the results presented in this section hold true for any generic rectangular matrix, Algorithm 2 can be straightforwardly applied for computing also the

Algorithm 2 Efficient computation of the matrix–vector multiplication $(\mathbf{G} \otimes \mathbf{G})\mathbf{x}$

Input: $\mathbf{G} \in \mathbb{R}^{m \times n}$; $\mathbf{x} \in \mathbb{R}^{n^2}$; $\mathbf{P}_{nm}$ and $\mathbf{P}_{m^2}$ as in Proposition 4.1.
1: Compute $\mathbf{y} := (\mathbf{I}_n \otimes \mathbf{G})\,\mathbf{x} \in \mathbb{R}^{nm}$;
2: Permute the element of $\mathbf{y}$ to define $\widehat{\mathbf{y}} := \mathbf{P}_{nm}\mathbf{y}$;
3: Compute $\mathbf{z} := (\mathbf{I}_m \otimes \mathbf{G})\,\widehat{\mathbf{y}} \in \mathbb{R}^{m^2}$;
4: Permute the element of $\mathbf{z}$ to define $\widehat{\mathbf{z}} := \mathbf{P}_{m^2}\mathbf{z}$;
Output: $\widehat{\mathbf{z}}$

matrix–vector product $\left(\mathbf{G}^\top \otimes \mathbf{G}^\top\right)\mathbf{x}$. Moreover, in connectivity studies, $m \propto 10^2$ is the number of M/EEG sensors and $n \propto 10^3$ is the number of cortical locations and in general $n >> m$. Then, the estimation cost of the Theorem 4.1 becomes $O(mn^2)$.

Remark 4.2 The previous result allows us to handle the natural increase in the complexity of the optimization problem in Eq. (7) and of the associated resolution scheme based on FISTA. We stress that, in terms of computational efficiency, in the classical two–step approach based on Tikhonov regularization, the inverse problem reduces to the solution of a smooth least-squares problem, which can be handled efficiently. The latter is not a computational bottleneck that our method is designed to address. However, in the context of M/EEG connectivity studies, efficiency should be understood in terms of the trade-off between computational effort and reconstruction accuracy. With our method we seek a substantially higher estimation accuracy than the classical approach, at the price of solving a more involved linear inverse problem. The results of Theorem 4.1 and Algorithm 2, together with the complexity estimate discussed above, ensure that this increased algebraic complexity does not translate into a prohibitive per-iteration cost, so that the FISTA-based scheme remains computationally viable in realistic large-scale scenarios.

5 Numerical Validation

The numerical Section is organized as follows. In Subsection 5.1 and 5.2 we describe how MEG synthetic data were simulated an analyzed using both the classical two–step approach and the proposed one, referred as one-step approach. The results and comparison of the two methods are summarized in Subsection 5.3. We highlight that our approach may still be regarded as a two-step procedure as it requires the estimation of the cross-power spectrum of the recorded time-series before the optimization step. However, since the optimization is carried out in the frequency domain and thus regularization is applied directly to the variable to be estimated, the use of the term one-step is justified for the sake of conciseness in the current numerical section.

5.1 Brain Connectivity Configurations

To test the performance of the proposed approach under different experimental conditions, we considered the following two brain configurations.

Configuration 1: three active sources with unidirectional coupling from source 1 to source 2. Inspired by previous works [14, 27, 32], the time–courses of the active sources were simulated by filtering in the band [8, 12]Hz (α band) a multivariate autoregressive (MVAR) process of order $P = 5$ defined as

$$\begin{pmatrix} z_1(t) \\ z_2(t) \\ z_3(t) \end{pmatrix} = \sum_{k=1}^{P} \begin{pmatrix} a_{1,1}(k) & 0 & 0 \\ a_{2,1}(k) & a_{2,2}(k) & 0 \\ 0 & 0 & a_{3,3}(k) \end{pmatrix} \begin{pmatrix} z_1(t-k) \\ z_2(t-k) \\ z_3(t-k) \end{pmatrix} + \begin{pmatrix} \epsilon_1(t) \\ \epsilon_2(t) \\ \epsilon_3(t) \end{pmatrix}. \tag{17}$$

The non-zero elements $a_{i,j}(k)$ of the coefficient matrix were drawn from a normal distribution of zero mean and standard deviation 0.9. We retained only coefficients resulting in (i) a stable MVAR process [19] and (ii) triplets of signals, $(z_1(t), z_2(t), z_3(t))^T$, such that the ℓ_2 norm of the strongest one is less than 3 times the ℓ_2 norm of the weakest one and such that the average over the range [8, 12]Hz of the sum of their power spectra was at least 1.2 times the average over the entire frequency range [32].

Configuration 2: three sources with unidirectional coupling from source 1 to source 2 and 3. The time–courses of the active sources were generated as for Configuration 1, but substituting the model in Eq. (17) with

$$\begin{pmatrix} z_1(t) \\ z_2(t) \\ z_3(t) \end{pmatrix} = \sum_{k=1}^{P} \begin{pmatrix} a_{1,1}(k) & 0 & 0 \\ a_{2,1}(k) & a_{2,2}(k) & 0 \\ a_{3,1}(k) & 0 & a_{3,3}(k) \end{pmatrix} \begin{pmatrix} z_1(t-k) \\ z_2(t-k) \\ z_3(t-k) \end{pmatrix} + \begin{pmatrix} \epsilon_1(t) \\ \epsilon_2(t) \\ \epsilon_3(t) \end{pmatrix}. \tag{18}$$

5.2 Simulation and Analysis of the Observed MEG Time–Series.

We exploited the model in Eq. (1) for generating 50 realisations, $\{\mathbf{y}(t)\}_{t=1}^{T}$, of the observable process for each configuration defined in the previous section. Specifically, we fixed $T = 10,000$, and for all $t = 1, \ldots, T$, we set

$$\mathbf{y}(t) = \mathbf{G}\mathbf{x}(t) + \mathbf{e}(t), \tag{19}$$

where

- we extracted the forward operator $\mathbf{G}$ from the sample dataset within the MNE Python package [11] by considering only magnetometers, and by downsampling the available source space to $n = 6940$ points. Hence, in our numerical experiment $\mathbf{G}$ has size 102×6940 and each column $\mathbf{g}_i$, $i = 1, \ldots, n$, represents the magnetic field generated by a point-like unit source placed at the i–th point of the source-space with orientation normal to the local cortical surface;
- we defined $\mathbf{x}(t)$ by randomly selecting three points of the source–space so that the pairwise source distances were greater than 4 cm and the pairwise ratios of the ℓ_2 norms of the corresponding columns of $\mathbf{G}$ were close to one. The components of $\mathbf{x}(t)$ corresponding to the drawn location were set equal to the time–courses defined by Eq. (17) or (18) while the value of the remaining $n - 3$ components was kept equal to 0;

– we sampled $\mathbf{e}(t)$ from a multivariate Gaussian distribution $\mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_m)$ where we chose σ^2 so as to obtain a signal-to-noise ratio (SNR) equal to 5 dB.

To test the proposed method, for each one of the simulated data, $\{\mathbf{y}(t)\}_{t=1}^T$, we then computed the Welch's estimator, $\mathbf{S}^{\mathbf{y}}(f)$, and we applied Algorithm 1 by setting f equal to the frequency in the range [8, 12]Hz where the component $S_{12}^{\mathbf{y}}(f)$ peaks. In our experiments, we set the maximum number of iterations K equal to 5,000, the tolerance ε equal to 10^{-5}, and we computed the Lipschitz constant L as in Eq. (10). To avoid inverse crime and mimic real-life scenarios where the active brain sources seldom match points of the source-space, the matrix $\mathbf{G}$ used within Algorithm 1 is obtained from that employed in Eq. (19) for simulating the MEG data by further reducing the source-space to $n = 644$ points. Finally, we tested four different values of the regularization parameters, by choosing four different scaling factors κ evenly spaced in log-space in the range $[10^{-2}, 10^{-1}]$. Then we set $\lambda = \kappa\,\lambda^*$, where $\lambda^* = 2\left\|\mathcal{G}\begin{pmatrix}\mathrm{Re}(\mathcal{S}^{\mathbf{y}})\\ \mathrm{Im}(\mathcal{S}^{\mathbf{y}})\end{pmatrix}\right\|_\infty^2$, being $\mathcal{S}^{\mathbf{y}} = \mathrm{vec}(\mathbf{S}^{\mathbf{y}}(f))$. The value λ^* has been shown to be an upper bound for the optimization problem (7) to admit a non-null solution [10], while the lower bound $10^{-2}\lambda^*$ has been chosen to guarantee enough sparsity in the obtained estimates.

The performance of the proposed approach was compared to that of the classical two–step approach described in Section 3.1. In detail, we first computed the Tikhonov estimator $\{\mathbf{x}_\lambda(t)\}_{t=1}^T$ of the neural sources as in Eq. (6) by using the coarse operator $\mathbf{G}$ including $n = 644$ source locations used also within Algorithm 1. Four different regularization parameters were tested, namely $\lambda = \xi\ 10^{-SNR/10}$, with $\xi \in \{0.1, 1, 10, 100\}$. This range has been chosen based on previous literature showing that the value $10^{-SNR/10}$ corresponds to the regularization parameter that provides the best possible estimate of the brain activity in the case of uncorrelated Gaussian signals [33]. For each value of the parameter, we then computed the Welch's estimator, $\mathbf{S}^{\mathbf{x}_\lambda}(f)$, of the time–series $\{\mathbf{x}_\lambda(t)\}_{t\in T}$.

Let $\widehat{\mathbf{S}}$ be an estimate of the cross-power spectrum obtained with either the proposed method or the classical two–step approach. In order to quantitatively compare the two approaches, we separated the real and the imaginary part of $\widehat{\mathbf{S}}$. For both of them, we then computed a weighted sum over all the estimated pairwise interactions exceeding a given threshold of the euclidean distance between the interacting-source locations and the closest pair of sources truly connected in the sense of the Wasserstein 2-distance. More formally, denoted with $\mathcal{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_{6940}\}$ and $\mathcal{W} = \{\mathbf{w}_1, \ldots, \mathbf{w}_{644}\}$ the source spaces associated to the forward operators used for simulating the data and within the inverse procedures, respectively, we defined

$$\mathrm{Err}^{\mathrm{Re}} = \sum_{(\mathbf{w}_i,\mathbf{w}_j)\in\mathcal{E}^{rec}} \frac{\left|\mathrm{Re}\left(\widehat{S}_{ij}\right)\right|}{\max\limits_{i<j}\left|\mathrm{Re}\left(\widehat{S}_{ij}\right)\right|} \min_{(\mathbf{v}_p,\mathbf{v}_q)\in\mathcal{E}^{true}} d\left((\mathbf{w}_i,\mathbf{w}_j),(\mathbf{v}_p,\mathbf{v}_q)\right), \tag{20}$$

where $\mathcal{E}^{rec} = \left\{(\mathbf{w}_i,\mathbf{w}_j) \in \mathcal{W}\times\mathcal{W} \mid i<j \text{ and } \left|\mathrm{Re}\left(\widehat{S}_{ij}\right)\right| \geq \tau\right\}$ collects the pairs of source locations between which the estimated cross spectrum exceeds a given threshold, namely $\tau = 0.5 \max\limits_{i<j}\left\{\left|\mathrm{Re}\left(\widehat{S}_{ij}\right)\right|\right\}$; $\mathcal{E}^{true} \subset \mathcal{V}\times\mathcal{V}$ is the set of truly connected

Table 1 Level of sparsity in the solution provided by the proposed approach for decreasing values of the regularization parameters. For each configuration, each cell of the table shows in the first row the percentage of simulated data where the real (first column) or the imaginary (second column) part of the estimated cross-power spectrum shows at least one non-null interaction; in the second row the minimum and maximum number of supra-threshold connections across these data; and in the third row the mean number of supra-threshold interactions

	Conf 1		Conf 2	
	Real part	**Imag. part**	**Real part**	**Imag. part**
λ_4	98.0%	58.0%	100.0%	78.0%
	(1, 87)	(1, 41)	(1, 292)	(1, 284)
	4	4	17	17
λ_3	98.0%	76.0%	100.0%	88.0%
	(1, 156)	(1, 318)	(1, 401)	(1, 221)
	6	18	19	14
λ_2	100.0%	86.0%	100.0%	92.0%
	(1, 281)	(1, 462)	(1, 1013)	(1, 526)
	20	31	44	30
λ_1	100.0%	96.0%	100.0%	96.0%
	(1, 260)	(1, 325)	(1, 4303)	(1, 2942)
	20	42	144	120

sources, and

$$d\left((\mathbf{w}_i, \mathbf{w}_j), (\mathbf{v}_p, \mathbf{v}_q)\right) = \sqrt{\frac{1}{2}\min\left\{||(\mathbf{w}_i, \mathbf{w}_j) - (\mathbf{v}_p, \mathbf{v}_q)||_2^2, ||(\mathbf{w}_i, \mathbf{w}_j) - (\mathbf{v}_q, \mathbf{v}_p)||_2^2\right\}}$$

is the Wasserstein 2-distance [16].

Similarly, $\mathrm{Err}^{\mathrm{Im}}$ was defined as in (20) by replacing $\mathrm{Re}\left(\widehat{S}_{ij}\right)$ with $\mathrm{Im}\left(\widehat{S}_{ij}\right)$.

For both the proposed method and the classical two-step approach we chose the value of the regularization parameter that provides the lowest total error $\mathrm{Err}^{\mathrm{Re}} + \mathrm{Err}^{\mathrm{Im}}$. The corresponding estimated cross-power spectra were compared in order to asses the advantages of the proposed method.

5.3 Results

Table 1 illustrated the behavior of the proposed approach when varying the amount of regularization. As expected, the higher the value of the regularization parameter the sparser the resulting estimation of the cross–power spectrum. In detail, for the highest value of the parameter, namely $\lambda_4 := 0.1\,\lambda^*$, for many simulated data the resulting estimation of the cross-power spectrum does not show any non-null interaction. On the other hand, for the lowest value of the parameter, namely $\lambda_1 := 0.01\,\lambda^*$, only in two datasets for both the configurations the imaginary part of the estimated cross-power spectrum was equal to zero, but the number of supra-threshold connections increased up to few hundreds for Configuration 1 and few thousands for Configuration 2. We recall that the threshold τ was set equal to half the value of the strongest connection.

Figure 1 shows the main advantages of the proposed one-step method over the classical two-step approach. In this example, both methods identified connected sources nearby the truly interacting ones, however the number of false positives is much higher for the two–step approach. Further, for the two–step approach the value of the cross-

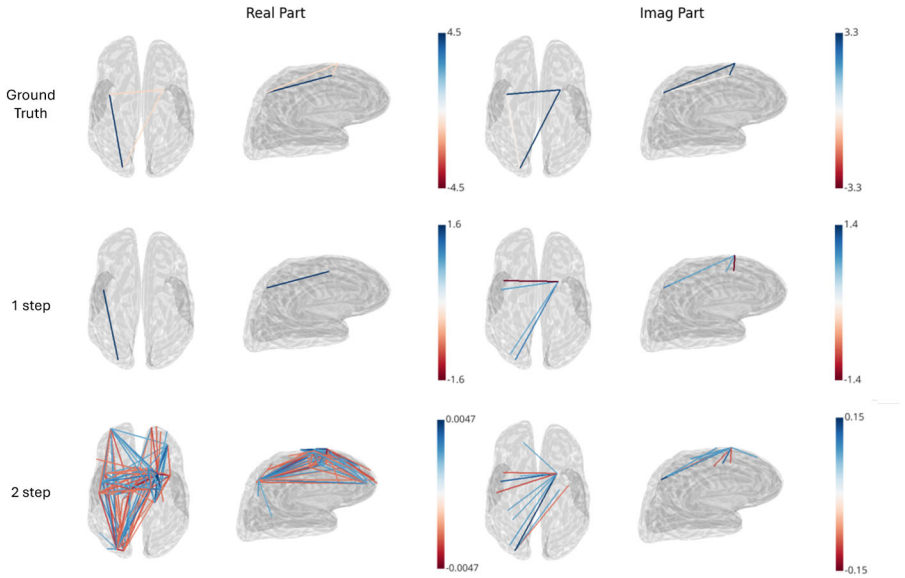

Fig. 1 Original and estimated cross-power spectrum for one MEG data simulated by using Configuration 2. In this simulation $\mathrm{Err}^{\mathrm{Re}}$ is equal to 0.077 for the one-step method and to 0.521 for the two–step approach, while $\mathrm{Err}^{\mathrm{Im}}$ is equal to 0.019 and 0.086 respectively

power spectrum is underestimated by several orders of magnitude. As expected, this improved estimation accuracy comes at the price of a higher computational cost: as shown by Figure 2 for this simulation more than 3000 iterations of the FISTA algorithm were required before the objective function in Eq. (7) and the total estimation error $\mathrm{Err}^{\mathrm{Re}} + \mathrm{Err}^{\mathrm{Im}}$ reached a plateau. As a consequence, the running time for analyzing each one of the simulated MEG data with the proposed one-step approach is about 30 minutes, while the classical two-step method takes around 2 minutes.

More in general, in our experiments, the one-step approach outperformed the classical two-step method as demonstrated by Table 2 and Figure 3. More specifically, for both the approaches we selected for each simulated data the best regularization parameter among those tested as the one minimizing the sum $\mathrm{Err}^{\mathrm{Re}} + \mathrm{Err}^{\mathrm{Im}}$. When using these optimal parameters, the average error of the one–step approach is systematically lower than that of the two–step approach for both the real and the imaginary part of the cross–power spectrum in both the configurations. More quantitatively, Table 2 shows that on average the total error $\mathrm{Err}^{\mathrm{Re}} + \mathrm{Err}^{\mathrm{Im}}$ of the one-step approach is almost one order of magnitude lower than the total error of the two-step approach. This is due to the fact that the number of spurious interactions is much higher for the two–step approach than for the proposed method, as shown in the second row of Figure 3.

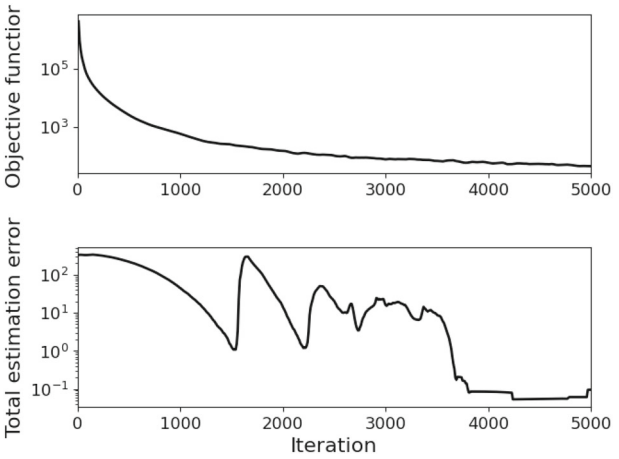

Fig. 2 Objective function and total estimation error, $\mathrm{Err}^{\mathrm{Re}} + \mathrm{Err}^{\mathrm{Im}}$, with respect to the iteration number when the one-step approach is used for analyzing the same simulated MEG data considered in Figure 1

Table 2 Comparison of the total error $\mathrm{Err}^{\mathrm{Re}} + \mathrm{Err}^{\mathrm{Im}}$ for the best estimates of the one-step and the two-step approach. Each cell shows mean (minimum, maximum) value across the 50 simulated data.

	Conf 1	**Conf 2**
1step	0.28 (0.006, 2.63)	0.39 (0.020, 3.04)
2step	1.50 (0.23, 6.93)	1.65 (0.26, 8.86)

6 Conclusions

The estimation of the brain cortical cross-power spectrum from M/EEG data is typically achieved in two steps: first the Tikhonov's least square estimator $\{\mathbf{x}_\lambda(t)\}_{t\in\mathbb{R}}$ of the brain cortical activity is computed, and then the cross-power spectrum of $\{\mathbf{x}_\lambda(t)\}_{t\in\mathbb{R}}$ is estimated through e.g. Welch's method. However, this approach often results in a large number of false positives. This issue is partially overcome by deriving from the cross-power spectrum metrics, such as the imaginary part of coherency, that are insensitive to linear mixing of the truly interacting brain source. However, such metrics fail in identifying instantaneously correlated source. In this work we overcome this issue by suggesting a novel approach that directly estimates the cross-power spectrum of the cortical sources from that of the recorded M/EEG time–series. The number of false positives is controlled by constructing an ℓ_1–regularized least square estimator of the cross-power spectrum computed through FISTA.

The proposed method leverages the tensorial structure of the global coefficient matrix. Therefore, we developed an efficient algorithm where the computational cost is primarily driven by operations involving the forward matrix $\mathbf{G}$. The advantages of our method over the classical two-step approach have been demonstrated on a large number of synthetic data mimicking different brain configurations.

From a methodological point of view, two are the main developments we are planning. The first one concerns the implementation of an automatic procedure to choose the regularization parameter by extensively exploring the space of admissible values.

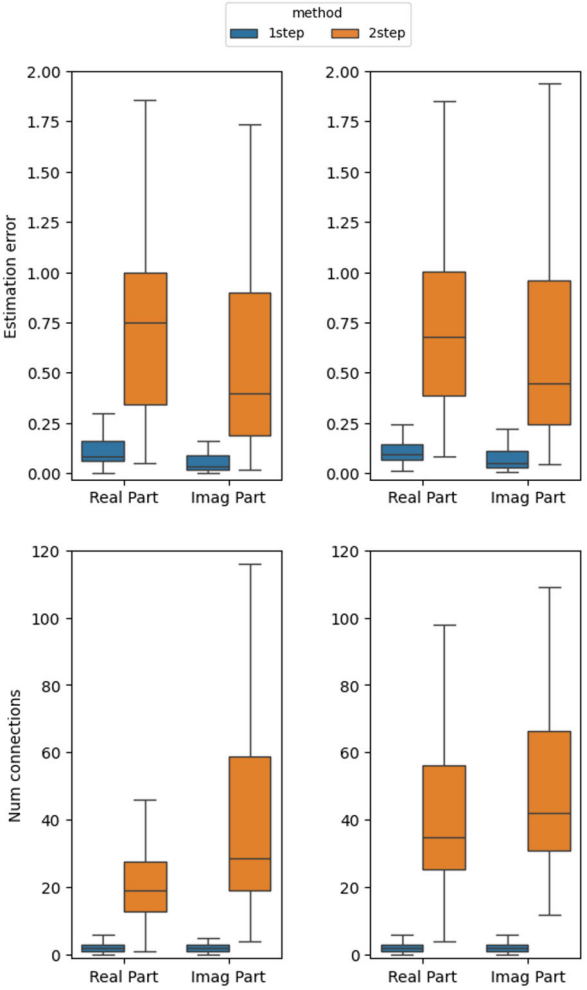

Fig. 3 Quantitative comparison of the performance of the one-step and the two-step approach. First row: error in estimating the real and the imaginary part of the cortical cross-power spectrum. Second row: number of supra-threshold connections in the real and imaginary part of the estimated cross-power spectrum. In each row the left-side panel refers to the first simulated brain configuration where only 1 pair of truly connected sources are present, while the right-side panel refers to the second brain configuration involving two pairs of interacting sources. Boxplots summarize results across 50 different simulated data. For the ease of visualization outliers were omitted

Moreover, advanced variants of FISTA, such as schemes with restarts or adaptive step sizes, possibly employing back-tracking strategies [24] or adaptive estimation of the Lipschitz constant L [1], could be employed to further improve the computational efficiency of the proposed method, and to enable a more systematic investigation of the trade-off between accuracy and computational cost.

The second one concerns implementing different approaches for solving the inverse problem in Eq. (3), such as Bayesian Monte Carlo approaches [26]. The latter would have the advantage of also quantifying the uncertainty of the provided estimation even though at the price of a higher computational cost.

Finally, we observe that the current implementation of the proposed approach provides estimates of the cross-power spectrum at fixed frequencies. Future work will be devoted to investigate how to combine the information from multiple frequencies and/or frequencies ranges.

Acknowledgements The authors are thankful to Dr. Martina Amerighi and Dr. Elisabetta Vallarino for their insightful discussions and suggestions. L.C., I.F., and S.S. are members of "Gruppo Nazionale per il Calcolo Scientifico" (INdAM-GNCS) and their work is partially supported by INdAM - GNCS Project "Analisi e applicazioni di matrici strutturate (a blocchi)" CUP E53C23001670001. The work of I.F. is supported by #NEXTGENERATIONEU (NGEU) and funded by the Ministry of University and Research (MUR), National Recovery and Resilience Plan (NRRP), project MNESYS (PE0000006) – A Multiscale integrated approach to the study of the nervous system in health and disease (DN. 1553 11.10.2022) for the part concerning the analysis of system structure for computation optimization, and its application to brain connectivity. S.S. acknowledges the support of the PRIN PNRR 2022 Project 'Computational mEthods for Medical Imaging (CEMI)' 2022FHCNY3, cup: D53D23005830006 for the conceptualization of the core optimization method in the context of multivariate statistical analysis.

Funding Open access funding provided by Università degli Studi di Genova within the CRUI-CARE Agreement.

Data Availability The codes for generating and analysing the used datasets during the current study are available in the GitHub repository, https://github.com/theMIDAgroup/fista_cps_conn

References

1. Aujol, J.F., Calatroni, L., Dossal, C., Labarrière, H., Rondepierre, A.: Parameter-free fista by adaptive restart and backtracking. SIAM J. Optim. **34**(4), 3259–3285 (2024)
2. Baillet, S., Mosher, J.C., Leahy, R.M.: Electromagnetic brain mapping. IEEE Signal Process. Mag. **18**(6), 14–30 (2001)
3. Barbarino, G., Garoni, C., Mazza, M., Serra-Capizzano, S.: Rectangular GLT sequences. Electron. Trans. Numer. Anal. **55**, 585–617 (2022)
4. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. **2**(1), 183–202 (2009)
5. Bendat, J.S., Piersol, A.G.: Random Data: Analysis and Measurement Procedures, vol. 729. John Wiley & Sons, New York (2011)
6. Bolten, M., Donatelli, M., Ferrari, P., Furci, I.: Symbol based convergence analysis in block multigrid methods with applications for Stokes problems. Appl. Numer. Math. **193**, 109–130 (2023)
7. Figueiredo, M.A., Nowak, R.D., Wright, S.J.: Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems. IEEE J. Sel. Top. Signal Process. **1**(4), 586–597 (2007)
8. Fries, P.: A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. Trends Cogn. Sci. **9**(10), 474–480 (2005)
9. Fukushima, M., Yamashita, O., Knösche, T.R., Sato, M.: MEG source reconstruction based on identification of directed source interactions on whole-brain anatomical networks. Neuroimage **105**, 408–427 (2015)
10. Gerstoft, P., Xenaki, A., Mecklenbräuker, C.F.: Multiple and single snapshot compressive beamforming. J. Acoust. Soc. Am. **138**(4), 2003–2014 (2015)
11. Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., Hämäläinen, M.S.: MEG and EEG data analysis with MNE-Python. Front. Neurosci. **7**(267), 1–13 (2013)

12. Hämäläinen, M.S., Ilmoniemi, R.J.: Interpreting magnetic fields of the brain: minimum norm estimates. Med. Biol. Eng. Comput. **32**(1), 35–42 (1994)
13. Hamming, R.W.: Digital Filters, 2nd edn. Prentice-Hall, Englewood Cliffs, New Jersey (1983)
14. Haufe, S., Ewald, A.: A simulation framework for benchmarking EEG-based brain connectivity estimation methodologies. Brain Topogr. **32**(4), 625–642 (2019)
15. Hincapié, A.S., Kujala, J., Mattout, J., Daligault, S., Delpuech, C., Mery, D., Cosmelli, D., Jerbi, K.: MEG connectivity and power detections with minimum norm estimates require different regularization parameters. Comput. Intell. Neurosci. **2016**(1), 3979547 (2016)
16. Hoffman, J.R., Mahler, R.P.: Multitarget miss distance via optimal assignment. IEEE Trans. Syst. Man Cybern. A: Syst. Hum. **34**(3), 327–336 (2004)
17. Horn, R.A., Johnson, C.R.: Matrix Analysis. Cambridge University Press, Cambridge (1985)
18. Ilmoniemi, R.J., Sarvas, J.: Brain Signals: Physics and Mathematics of MEG and EEG. The Mit Press (2019)
19. Lütkepohl, H.: New Introduction to Multiple Time Series Analysis. Springer-Verlag, Berlin (2005)
20. Nolte, G., Bai, O., Wheaton, L., Mari, Z., Vorbach, S., Hallett, M.: Identifying true brain interaction from EEG data using the imaginary part of coherency. Clin. Neurophysiol. **115**(10), 2292–2307 (2004)
21. Ossadtchi, A., Altukhov, D., Jerbi, K.: Phase shift invariant imaging of coherent sources (PSIICOS) from MEG data. Neuroimage **183**, 950–971 (2018)
22. Pereda, E., Quiroga, R.Q., Bhattacharya, J.: Nonlinear multivariate analysis of neurophysiological signals. Prog. Neurobiol. **77**(1), 1–37 (2005)
23. Sakkalis, V.: Review of advanced techniques for the estimation of brain connectivity measured with EEG/MEG. Comput. Biol. Med. **41**(12), 1110–1117 (2011)
24. Scheinberg, K., Goldfarb, D., Bai, X.: Fast first-order methods for composite convex optimization with backtracking. Found. Comput. Math. **14**(3), 389–417 (2014)
25. Schoffelen, J.M., Gross, J.: Studying dynamic neural interactions with MEG. In: Magnetoencephalography: From Signals to Dynamic Cortical Networks: 2nd edn. Springer International Publishing (2019)
26. Sommariva, S., Sorrentino, A.: Sequential Monte Carlo samplers for semilinear inverse problems and application to magnetocephalography. Inverse Probl. **30**(11), 114020 (2014)
27. Sommariva, S., Sorrentino, A., Piana, M., Pizzella, V., Marzetti, L.: A comparative study of the robustness of frequency-domain connectivity measures to finite data length. Brain Topogr. **32**(4), 675–695 (2019)
28. Stam, C.J., Nolte, G., Daffertshofer, A.: Phase lag index: assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources. Hum. Brain Mapp. **28**(11), 1178–1193 (2007)
29. Tibshirani, R.: Regression shrinkage and selection via the lasso. J. Roy. Statist. Soc. Ser. B **58**(1), 267–288 (1996)
30. Tihonov, A.N.: On the solution of ill-posed problems and the method of regularization. Dokl. Akad. Nauk SSSR **151**, 501–504 (1963)
31. Tronarp, F., Subramaniyam, N.P., Särkkä, S., Parkkonen, L.: Tracking of dynamic functional connectivity from MEG data with Kalman filtering. In: 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (2018)
32. Vallarino, E., Hincapié, A.S., Jerbi, K., Leahy, R.M., Pascarella, A., Sorrentino, A., Sommariva, S.: Tuning minimum-norm regularization parameters for optimal MEG connectivity estimation. Neuroimage **281**, 120356 (2023)
33. Vallarino, E., Sommariva, S., Piana, M., Sorrentino, A.: On the two-step estimation of the cross-power spectrum for dynamical linear inverse problems. Inverse Probl. **36**(4), 045010 (2020)
34. Vallarino, E., Sorrentino, A., Piana, M., Sommariva, S.: The role of spectral complexity in connectivity estimation. Axioms **10**(1), 35 (2021)
35. Vinck, M., Oostenveld, R., Van Wingerden, M., Battaglia, F., Pennartz, C.M.: An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias. Neuroimage **55**(4), 1548–1565 (2011)
36. Welch, P.: The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. IEEE Trans. Audio Electroacoust. **15**(2), 70–73 (1967)