On the two-step estimation of the cross–power spectrum for dynamical linear inverse problems

Elisabetta Vallarino[1]**, Sara Sommariva**[1]**, Michele Piana**[1,2] **and Alberto Sorrentino**[1,2]

[1] Dipartimento di Matematica, Università di Genova, Italy
[2] CNR–SPIN, Genova, Italy

Abstract. We consider the problem of reconstructing the cross–power spectrum of an unobservable multivariate stochastic process from indirect measurements of a second multivariate stochastic process, related to the first one through a linear operator. In the two–step approach, one would first compute a regularized reconstruction of the unobservable signal, and then compute an estimate of its cross–power spectrum from the regularized solution. We investigate whether the optimal regularization parameter for reconstruction of the signal also gives the best estimate of the cross–power spectrum. We show that the answer depends on the regularization method, and specifically we prove that, under a white Gaussian assumption: (i) when regularizing with truncated SVD the optimal parameter is the same; (ii) when regularizing with the Tikhonov method, the optimal parameter for the cross–power spectrum is lower than half the optimal parameter for the signal. We also provide evidence that a one–step approach would likely have better mathematical properties than the two–step approach. Our results apply particularly to the brain connectivity estimation from magneto/electro-encephalographic recordings and provide a formal interpretation of recent empirical results.

1. Introduction

Dynamical inverse problems are typically concerned with two interplaying and in some sense still open issues. The first one is related to the reconstruction of the unobserved, multivariate stochastic process from the measured time series; the second one is the estimate of the statistical interdependence of the individual components of the multivariate stochastic process. A paradigmatic example of these issues is the estimate of brain functional connectivity from recordings of magneto/electro-encephalographic (M/EEG) data, currently a hot topic in neuroscience. Functional connectivity is systematically used to study both the healthy [10] and the pathological [38, 41] brain, either at rest [4] or during the execution of specific tasks [25, 43].

While there is no unique formal definition of functional connectivity, the term is generally used to identify various forms of statistical interdependence between the temporal waveforms of spatially distinct brain areas [34]. In the last couple of decades, such interdependence is increasingly studied in the frequency domain; this makes sense in light of the increasingly accepted model that neural interactions between brain regions are mediated by synchronization of their rhythmic activity in specific frequency

bands [13]. However, M/EEG only record the magnetic field/electric potential at the scalp; therefore functional connectivity between brain regions has to be estimated indirectly, using the scalp data and the physical model that relates neural currents to the recordings.

In this framework, the majority of connectivity studies employs a two–step approach [35]: first, an estimate of the source time courses is obtained using an inverse method; then, frequency–domain connectivity metrics are computed from the cross–spectrum of the reconstructed source time courses. Due to the multitude of available inverse methods [17, 42, 5, 9, 2, 26, 37, 19, 23] and connectivity metrics [1, 27, 6, 15, 33, 34], in the last decade there has been growing interest in validating and comparing different combinations of methods [12, 22, 7, 36, 30].

Recent empirical evidence suggests that the two–step approach might feature an unexpected parameter tuning issue. Indeed, the usual approach to overcome ill–posedness consists in searching a tradeoff between solution complexity and data fitting, and this tradeoff is realized by means of the selection of a proper regularization parameter. It would seem natural that the optimal estimate of the cross–spectrum can only be attained with the optimal reconstruction of the signal. Yet, in a recent study the authors in [21] have shown that the value of the regularization parameter that provides the best reconstruction of the source spectral power does not coincide with the value that provides the best reconstruction of the source-level functional connectivity quantified through coherence.

Motivated by this empirical result, in this work we investigate the following problem: let $\mathbf{Y}(t)$ be noisy and indirect measurements of a multivariate stochastic process $\mathbf{X}(t)$; let $\mathbf{x}_\lambda(t)$ be the reconstruction of the hidden signal, obtained by means of a regularization algorithm; finally, assume that the cross–spectrum of $\mathbf{X}(t)$, denoted as $\mathbf{S}^{\mathbf{X}}(f)$, is estimated from the reconstructed signal $\mathbf{x}_\lambda(t)$; under these conditions, does the optimal regularization parameter for reconstruction of the hidden signal coincide with the optimal regularization parameter for reconstruction of its cross–spectrum?

In particular, we will prove that the answer is "no" when the regularized solution is computed via Tikhonov regularization, thus confirming the empirical results of [21]. We will also prove that the answer is "yes" when the regularized solution is computed via truncated Singular Value Decomposition, thus showing that the answer to the question actually depends on the choice of the inverse method. In addition, we will show the potential of a one–step approach relying on a mathematical model directly relating the measured data to the unknown cross-spectrum. In particular, a preliminary analysis performed in this paper shows that the one–step approach enhances the filtering effectiveness of regularization with respect to the standard two-step approach.

The structure of the article is as follows: in Section 2 we provide the general definitions and formalize the main question of the paper. In Section 3 we express the reconstructions errors in terms of the filter factors and provide an interpretation. Section 4 contains the main results of our work: we show that the optimal regularization parameters for reconstruction of $\mathbf{x}(t)$ and $\mathbf{S}^{\mathbf{x}}(f)$ are generally different and that this difference depends on the inversion method. In Section 5 we show how the filter factors of the two–step approach have a jittering behaviour, while those of a possible one–step approach would be smooth. In Section 6 we show the results of a numerical simulation in which we remove the somewhat restrictive assumptions that are needed to prove the theorems of Section 4. Our conclusions will be offered in Section 6, together with the discussion of possible directions for future work.

2. Definition of the problem

Let $\mathbf{X}(t) = (X_1(t), \ldots, X_N(t))^\top$ be a multivariate, stationary stochastic process whose realizations $\mathbf{x}(t)$ cannot be observed directly; indirect information on $\mathbf{x}(t)$ can be obtained by observing the realizations $\mathbf{y}(t)$ of the process $\mathbf{Y}(t)$, which is a noisy linear mixture of $\mathbf{X}(t)$

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

where $\mathbf{G}$ is an $M \times N$ *forward* matrix, with $M = \dim(\mathbf{Y}(t))$, $N = \dim(\mathbf{X}(t))$, and $\mathbf{N}(t)$ is the measurement noise process, which is assumed to be a zero–mean Gaussian process independent from $\mathbf{X}(t)$. For ease of presentation, we further assume $M \leq N$ and $\mathbf{G}$ to be a full row rank matrix so that all its singular values are strictly positive; however, the results below can be easily extended to the general case.

We consider the case where one is interested in reconstructing the cross–spectrum of the process $\mathbf{X}(t)$, that contains information on the statistical dependencies between the different components of the signal. The cross–spectrum is a one–parameter family of $N \times N$ matrices $\mathbf{S}^{\mathbf{X}}(f)$, whose (j, k)–th element is defined as

$$S_{j,k}^{\mathbf{X}}(f) = \lim_{T \to +\infty} \frac{1}{T} E[\hat{X}_j(f, T)\hat{X}_k(f, T)^H] \tag{2}$$

where $\hat{X}_j(f, T)$ is the Fourier transform of $X_j(t)$ over the interval $[0, T]$, defined as

$$\hat{X}_j(f, T) = \int_0^T X_j(t)e^{-2\pi ift}dt \tag{3}$$

and X^H is the Hermitian transpose of X [3].

In this work we consider the case when the reconstruction of the cross–spectrum is done in a two–step process:

(i) First, a regularized estimate $\mathbf{x}_\lambda(t)$ of $\mathbf{x}(t)$ is computed as

$$\mathbf{x}_\lambda(t) = \mathbf{W}_\lambda \mathbf{y}(t) = \mathbf{V}\mathbf{\Phi}(\lambda)\mathbf{\Sigma}^\dagger \mathbf{U}^\top \mathbf{y}(t) \quad , \tag{4}$$

where: $\mathbf{G} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ is the singular value decomposition (SVD) of the forward matrix, being $\mathbf{U} \in R^{M \times M}$, $\mathbf{V} \in R^{N \times N}$, and $\mathbf{\Sigma} = \mathrm{diag}(\sigma_1, \ldots, \sigma_M) \in R^{M \times N}$ with $\sigma_1 \geq \ldots \geq \sigma_M > 0$; $\mathbf{\Sigma}^\dagger$ is the pseudo-inverse of $\mathbf{\Sigma}$; $\mathbf{\Phi}(\lambda) = \mathrm{diag}(\varphi_1(\lambda), \ldots, \varphi_M(\lambda), 0, \ldots, 0) \in R^{N \times N}$ are the *filter factors* [18], which are functions of one (or more) *regularization parameter(s)* λ.

(ii) Then, an estimate of the cross–spectrum is obtained from these reconstructed time–series using the Welch's method [44], which consists in partitioning the data in P overlapping segments $\{\mathbf{x}_\lambda^p(t)\}_{p=1,\ldots,P}$, computing $\hat{\mathbf{x}}_\lambda^p(f) = \frac{1}{L}\sum_{t=0}^{L-1}\mathbf{x}_\lambda^p(t)w(t)e^{-\frac{2\pi itf}{L}}$, the Discrete Fourier Transform of the signals multiplied by a window function $w(t)$, and then averaging these modified periodograms:

$$\mathbf{S}^{\mathbf{x}_\lambda}(f) = \frac{L}{PW}\sum_{p=1}^{P}\hat{\mathbf{x}}_\lambda^p(f)\hat{\mathbf{x}}_\lambda^p(f)^H, \qquad f = 0, \ldots, L - 1, \tag{5}$$

where L is the length of each segment and $W = \frac{1}{L}\sum_{t=0}^{L-1}w(t)^2$.

This two–step approach is largely used, e.g., in connectivity estimation from M/EEG data, where the estimated cross-spectrum is typically used to compute a large pool of connectivity metrics such as coherence [31], imaginary part of coherency [27], and phase slope index [29]. Naturally but crucially, this estimate depends on the choice of the regularization method, as well as on the choice of λ, which modulates the degree of regularization of the estimate $\mathbf{x}_\lambda$.

In this work we will mainly focus on two regularization methods, namely truncated SVD (tSVD) and Tikhonov regularization. The reason of this specific choice is as follows: the Tikhonov method is one of the more commonly employed methods for connectivity estimation in M/EEG, and it has been used by Hincapié and colleagues in the paper that motivated this study [21]; tSVD is a method which is easy to deal with analytically, and in addition it will provide a different result than the Tikhonov method, thus showing that the answer to the main question of this study is method–dependent.

Henceforth, $\mathbf{u}_i$ and $\mathbf{v}_i$ will denote the i-th column of matrices $\mathbf{U}$ and $\mathbf{V}$, respectively. tSVD relies on the 1–parameter family of regularized estimates

$$\mathbf{x}_\lambda(t) = \sum_{i=1}^{\lambda} \frac{\mathbf{u}_i^\top \mathbf{y}(t)}{\sigma_i} \mathbf{v}_i \qquad \lambda \in \{1, \ldots, M\} \quad , \tag{6}$$

which are obtained from equation (4) by setting

$$\varphi_i(\lambda) = \begin{cases} 1 & \text{if} \quad i \leq \lambda \\ 0 & \text{if} \quad i > \lambda \end{cases} \quad . \tag{7}$$

Tikhonov estimates are defined as

$$\mathbf{x}_\lambda(t) = \sum_{i=1}^{M} \frac{\sigma_i^2}{\sigma_i^2 + \lambda} \frac{\mathbf{u}_i^\top \mathbf{y}(t)}{\sigma_i} \mathbf{v}_i \qquad \lambda \geq 0 \quad , \tag{8}$$

which are obtained from equation (4) by setting $\varphi_i(\lambda) = \frac{\sigma_i^2}{\sigma_i^2 + \lambda}$. From now on, for simplicity, we omit the dependence of $\mathbf{\Phi}$ and φ_i on λ. Also note that in the two methods the parameter λ assumes values in different sets. In tSVD λ determines the number of retained SVD components, and therefore assumes integer values in $\{1, \ldots, M\}$, where a small λ value means few retained components and thus an high level of regularization. In Tikhonov regularization λ determines the strength with which each SVD component contributes to the solution; in this case λ assumes continuous values in $[0, +\infty)$ and the higher the value the higher the degree of regularization.

For the two mentioned methods, we consider the problem of the optimal choice of the regularization parameter λ for reconstruction of the cross–spectrum. We define optimality through the minimization of the norm of the discrepancy, specifically we define the two following optimal values for the parameter.

Definition 1. *Consider the regularized solution (4) and the cross–spectrum estimate (5) associated to a realization of equation (1); we define the optimal parameter for the reconstruction of* $\mathbf{x}(t)$

$$\lambda_\mathbf{x}^* = \arg\min_\lambda \varepsilon_\mathbf{x}(\lambda) \quad \text{with} \quad \varepsilon_\mathbf{x}(\lambda) = \sum_t \|\mathbf{x}_\lambda(t) - \mathbf{x}(t)\|_2^2 \quad , \tag{9}$$

and the optimal parameter for the reconstruction of $\mathbf{S}^\mathbf{x}(f)$

$$\lambda_\mathbf{S}^* = \arg\min_\lambda \varepsilon_\mathbf{S}(\lambda) \quad \text{with} \quad \varepsilon_\mathbf{S}(\lambda) = \sum_f \|\mathbf{S}^{\mathbf{x}_\lambda}(f) - \mathbf{S}^\mathbf{x}(f)\|_F^2 \quad , \tag{10}$$

where $\| \cdot \|_2$ and $\| \cdot \|_F$ are the L^2-norm and the Frobenius norm, respectively; $\varepsilon_{\mathbf{x}}(\lambda)$ and $\varepsilon_{\mathbf{S}}(\lambda)$ will be called reconstruction errors.

In the following sections we shall answer the following question: does the optimal regularization parameter for reconstruction of $\mathbf{x}(t)$, $\lambda_{\mathbf{x}}^*$, coincide with the optimal regularization parameter for reconstruction of $\mathbf{S}^{\mathbf{x}}(f)$, $\lambda_{\mathbf{S}}^*$?

3. Reconstruction errors with filter factors

In this section we aim at deriving an explicit formulation of $\varepsilon_{\mathbf{x}}(\lambda)$ and $\varepsilon_{\mathbf{S}}(\lambda)$ in terms of the filter factors $\mathbf{\Phi}$. To this end we observe that from equations (1) and (4) we can derive the following relationship between the true and the reconstructed signal:

$$\mathbf{x}_\lambda(t) = \mathbf{R}_\lambda \mathbf{x}(t) + \mathbf{W}_\lambda \mathbf{n}(t) \tag{11}$$

where $\mathbf{R}_\lambda = \mathbf{W}_\lambda \mathbf{G}$ is the resolution matrix [11, 18].

A similar relationship between the true and the estimated cross-spectrum can be derived by substituting equation (11) into definition (5) and by exploiting the linearity of the Discrete Fourier Transform:

$$\begin{aligned}\mathcal{S}^{\mathbf{x}_\lambda}(f) &= (\mathbf{R}_\lambda \otimes \mathbf{R}_\lambda)\, \mathcal{S}^{\mathbf{x}}(f) + (\mathbf{W}_\lambda \otimes \mathbf{W}_\lambda)\, \mathcal{S}^{\mathbf{n}}(f) \\ &\quad + (\mathbf{W}_\lambda \otimes \mathbf{R}_\lambda)\, \mathcal{S}^{\mathbf{x}\mathbf{n}}(f) + (\mathbf{R}_\lambda \otimes \mathbf{W}_\lambda)\, \mathcal{S}^{\mathbf{n}\mathbf{x}}(f)\end{aligned} \tag{12}$$

where $\mathcal{S}^{\mathbf{x}}(f)$ is the vector obtained by concatenating the columns of the matrix $\mathbf{S}^{\mathbf{x}}(f)$, $\otimes$ is the Kronecker product, and $\mathbf{S}^{\mathbf{x}\mathbf{n}}(f)$ is the cross–spectrum between $\mathbf{x}$ and $\mathbf{n}$, i.e., following the notation in equation (5), $\mathbf{S}^{\mathbf{x}\mathbf{n}}(f) = \frac{L}{PW}\sum_{p=1}^{P} \hat{\mathbf{x}}^p(f)\hat{\mathbf{n}}^p(f)^H$.

Since $\mathbf{X}(t)$ and $\mathbf{N}(t)$ are independent, $\mathcal{S}^{\mathbf{x}\mathbf{n}}(f)$ and $\mathcal{S}^{\mathbf{n}\mathbf{x}}(f)$ are negligible provided that enough data time-points are available. Hence from definition 1 it follows

$$\varepsilon_{\mathbf{x}}(\lambda) = \sum_t \| (\mathbf{R}_\lambda - \mathbf{I}_N)\, \mathbf{x}(t) + \mathbf{W}_\lambda \mathbf{n}(t) \|_2^2 \tag{13}$$

$$\varepsilon_{\mathbf{S}}(\lambda) = \sum_f \| (\mathbf{R}_\lambda \otimes \mathbf{R}_\lambda - \mathbf{I}_{N^2})\, \mathcal{S}^{\mathbf{x}}(f) + (\mathbf{W}_\lambda \otimes \mathbf{W}_\lambda)\, \mathcal{S}^{\mathbf{n}}(f) \|_2^2 \tag{14}$$

where $\mathbf{I}_N$ is the identity matrix of size $N \times N$.

Proposition 1. *The reconstruction errors defined in (9) and (10) are given by:*

$$\varepsilon_{\mathbf{x}}(\lambda) = \sum_t \sum_{i=M+1}^{N} \left(\mathbf{v}_i^\top \mathbf{x}(t)\right)^2 + \sum_t \sum_{i=1}^{M} \left[(\varphi_i - 1)^2 \left(\mathbf{v}_i^\top \mathbf{x}(t)\right)^2 + \varphi_i^2 \frac{\left(\mathbf{u}_i^\top \mathbf{n}(t)\right)^2}{\sigma_i^2} \right] \tag{15}$$

and

$$\begin{aligned}\varepsilon_{\mathbf{S}}(\lambda) &= \sum_f \sum_{\substack{i \geq M+1 \ or \\ j \geq M+1}} \left| (\mathbf{v}_i \otimes \mathbf{v}_j)^\top \mathcal{S}^{\mathbf{x}}(f) \right|^2 + \sum_f \sum_{i,j=1}^{M} \left[(\varphi_i \varphi_j - 1)^2 \left| (\mathbf{v}_i \otimes \mathbf{v}_j)^\top \mathcal{S}^{\mathbf{x}}(f) \right|^2 \right. \\ &\quad \left. + \left(\frac{\varphi_i \varphi_j}{\sigma_i \sigma_j} \right)^2 \left| (\mathbf{u}_i \otimes \mathbf{u}_j)^\top \mathcal{S}^{\mathbf{n}}(f) \right|^2 + 2 \left(\varphi_i \varphi_j - 1 \right) \frac{\varphi_i \varphi_j}{\sigma_i \sigma_j} Re \left(\overline{(\mathbf{v}_i \otimes \mathbf{v}_j)^\top \mathcal{S}^{\mathbf{x}}(f)} (\mathbf{u}_i \otimes \mathbf{u}_j)^\top \mathcal{S}^{\mathbf{n}}(f) \right) \right]\end{aligned} \tag{16}$$

Proof. To prove equation (15) we observe that

$$\mathbf{W}_\lambda = \mathbf{V}\mathbf{\Phi}\mathbf{\Sigma}^\dagger \mathbf{U}^\top = \sum_{i=1}^{M} \mathbf{v}_i \frac{\varphi_i}{\sigma_i} \mathbf{u}_i^\top$$

and

$$\mathbf{R}_\lambda - \mathbf{I}_N = \mathbf{V}\boldsymbol{\Phi}\boldsymbol{\Sigma}^\dagger\boldsymbol{\Sigma}\mathbf{V}^\top - \mathbf{I}_N = \sum_{i=1}^{M} \mathbf{v}_i(\varphi_i - 1)\mathbf{v}_i^\top - \sum_{i=M+1}^{N} \mathbf{v}_i\mathbf{v}_i^\top.$$

Then the thesis follows from equation (13) by exploiting the orthonormality of $\mathbf{V}$ and the independence between processes $\mathbf{X}(t)$ and $\mathbf{N}(t)$.

Analogously, equation (16) follows from equation (14) by observing

$$\mathbf{W}_\lambda \otimes \mathbf{W}_\lambda = (\mathbf{V} \otimes \mathbf{V})\left(\boldsymbol{\Phi}\boldsymbol{\Sigma}^\dagger \otimes \boldsymbol{\Phi}\boldsymbol{\Sigma}^\dagger\right)(\mathbf{U} \otimes \mathbf{U})^\top = \sum_{i,j=1}^{M} (\mathbf{v}_i \otimes \mathbf{v}_j)\frac{\varphi_i\varphi_j}{\sigma_i\sigma_j}(\mathbf{u}_i \otimes \mathbf{u}_j)^\top$$

and

$$\mathbf{R}_\lambda \otimes \mathbf{R}_\lambda - \mathbf{I}_{N^2} = (\mathbf{V} \otimes \mathbf{V})\left(\boldsymbol{\Phi}\boldsymbol{\Sigma}^\dagger\boldsymbol{\Sigma} \otimes \boldsymbol{\Phi}\boldsymbol{\Sigma}^\dagger\boldsymbol{\Sigma} - \mathbf{I}_{N^2}\right)(\mathbf{V} \otimes \mathbf{V})^\top$$

$$= \sum_{i,j=1}^{M} (\mathbf{v}_i \otimes \mathbf{v}_j)(\varphi_i\varphi_j - 1)(\mathbf{v}_i \otimes \mathbf{v}_j)^\top - \sum_{\substack{i \geq M+1 \ or \\ j \geq M+1}} (\mathbf{v}_i \otimes \mathbf{v}_j)(\mathbf{v}_i \otimes \mathbf{v}_j)^\top$$

$\square$

Remark 1. The expression in (15) is a classical result in regularization theory [18], in which the reconstruction error is expressed in terms of three distinct components. The first component is the norm of the projection of the original signal onto the kernel of $\mathbf{G}$, i.e. the part of the signal that cannot be reconstructed. The second term is the *regularization error*, i.e. the error introduced by regularization itself; indeed, this term vanishes when the value of all the filters is one. The last term is the *perturbation error*, i.e. the backprojection of stochastic noise components onto the reconstructed signal, that regularization tries to reduce.

Remark 2. Expression (16) is the analogue of (15) for the cross–spectrum estimated with the two–step approach. To the best of our knowledge this expression is novel and has never been studied. The reconstruction error $\varepsilon_\mathbf{S}(\lambda)$ here is made of four distinct components: three of them have the same interpretation of those appearing in $\varepsilon_\mathbf{x}(\lambda)$; the fourth term is a non–vanishing mixed term, that depends on both the signal and the noise spectra; as we shall see below, this term turns out to be negative at least in some special cases.

4. The relationship between the optimal regularization parameters: two case studies

We will now address the main question posed in the introduction: does the optimal regularization parameter for reconstruction of the time–series coincide with the optimal regularization parameter for reconstruction of the cross–spectrum? As we shall see, the answer depends on the specific choice of the inverse method, i.e. of the form of the filter factors. Here we study first the case of tSVD, and then the case of the Tikhonov method.

In order to proceed analytically, in this section we make the further assumption that both the signal and the noise are white–noise Gaussian processes, with covariance matrices $\omega^2\mathbf{I}_N$ and $\alpha^2\mathbf{I}_M$, respectively. The Gaussian assumption is often not too far fetched; in M/EEG, particularly, it is widely used and, even though perhaps the data

distribution is not exactly Gaussian, the Gaussian assumption is implicit (when not explicit) in the vast majority of connectivity studies [28]. The white–noise assumption, on the other hand, is stronger, as it implies that there is no temporal structure in the signal: we will come back to this point in the Discussion.

4.1. Truncated SVD

When tSVD is employed, by substituting the values of the corresponding filter factors into equations (15) and (16) we get the following corollary of Proposition 1.

Corollary 1. *Consider the tSVD estimate* $\mathbf{x}_\lambda(t)$ *given by equation (6), with regularization parameter* $\lambda \in \{1, \ldots, M\}$. *Then*

$$\varepsilon_{\mathbf{x}}(\lambda) = \sum_t \sum_{i=\lambda+1}^N \left(\mathbf{v}_i^\top \mathbf{x}(t)\right)^2 + \sum_t \sum_{i=1}^\lambda \frac{\left(\mathbf{u}_i^\top \mathbf{n}(t)\right)^2}{\sigma_i^2} \tag{17}$$

and

$$\varepsilon_{\mathbf{S}}(\lambda) = \sum_f \sum_{\substack{i \geq \lambda+1 \ or \\ j \geq \lambda+1}} \left|(\mathbf{v}_i \otimes \mathbf{v}_j)^\top \mathcal{S}^{\mathbf{x}}(f)\right|^2 + \sum_f \sum_{i,j=1}^\lambda \frac{\left|(\mathbf{u}_i \otimes \mathbf{u}_j)^\top \mathcal{S}^{\mathbf{n}}(f)\right|^2}{\sigma_i^2 \sigma_j^2} \tag{18}$$

Remark 3. When regularization is accomplished through tSVD, the mixed term in $\varepsilon_{\mathbf{S}}(\lambda)$ vanishes; this allows us to compute the optimal regularization parameter explicitly.

Theorem 1. *Let* $x_\lambda(t)$ *be the tSVD estimate as given by equation (6), with regularization parameter* $\lambda \in \{1, \ldots, M\}$; *assume* $\mathbf{X}(t)$ *and* $\mathbf{N}(t)$ *to be white–noise Gaussian processes with covariance matrices* $\omega^2 \mathbf{I}_N$ *and* $\alpha^2 \mathbf{I}_M$, *respectively. Then*

$$\lambda_{\mathbf{x}}^* = \lambda_{\mathbf{S}}^* = \max\left\{\lambda \in \{1, \ldots, M\} \ s.t \ \sigma_\lambda \geq \frac{\alpha}{\omega}\right\} \tag{19}$$

Proof. As $\mathbf{X}(t)$ and $\mathbf{N}(t)$ are white–noise Gaussian processes with covariance matrices $\omega^2 \mathbf{I}_N$ and $\alpha^2 \mathbf{I}_M$, we have $\mathbf{S}^{\mathbf{x}}(f) = \omega^2 \mathbf{I}$ and $\mathbf{S}^{\mathbf{n}}(f) = \alpha^2 \mathbf{I}$. Provided that enough data time-points are available‡, these imply that

$$\sum_t \left(\mathbf{v}_i^\top \mathbf{x}(t)\right)^2 = T\omega^2 \qquad \sum_t \left(\mathbf{u}_i^\top \mathbf{n}(t)\right)^2 = T\alpha^2 \tag{20}$$

and

$$\begin{aligned}
\left|(\mathbf{v}_i \otimes \mathbf{v}_j)^\top \mathcal{S}^{\mathbf{x}}(f)\right|^2 &= \omega^4 \left|\mathbf{v}_j^\top \mathbf{v}_i\right|^2 = \omega^4 \delta_{ij} \\
\left|(\mathbf{u}_i \otimes \mathbf{u}_j)^\top \mathcal{S}^{\mathbf{n}}(f)\right|^2 &= \alpha^4 \left|\mathbf{u}_j^\top \mathbf{u}_i\right|^2 = \alpha^4 \delta_{ij}
\end{aligned} \tag{21}$$

where δ_{ij} is the Kronecker delta. By substituting equalities (20)-(21) in equations (17) and (18) we get

$$\varepsilon_{\mathbf{x}}(\lambda) = (N - \lambda) T\omega^2 + T\alpha^2 \sum_{i=1}^\lambda \frac{1}{\sigma_i^2}$$

‡ The sufficient number of time points is of course dependent on the size of the problem, and in particular it grows with the dimensions of the data and of the unknown.

$$\varepsilon_{\mathbf{S}}(\lambda) = (N - \lambda) L\omega^4 + L\alpha^4 \sum_{i=1}^{\lambda} \frac{1}{\sigma_i^4}$$

The thesis follows by observing that the increments

$$\varepsilon_{\mathbf{x}}(\lambda) - \varepsilon_{\mathbf{x}}(\lambda - 1) = -T\omega^2 + \frac{T\alpha^2}{\sigma_\lambda^2}$$

and

$$\varepsilon_{\mathbf{S}}(\lambda) - \varepsilon_{\mathbf{S}}(\lambda - 1) = -L\omega^4 + \frac{L\alpha^4}{\sigma_\lambda^4}$$

are non–decreasing functions of λ and thus $\varepsilon_{\mathbf{x}}(\lambda)$ and $\varepsilon_{\mathbf{S}}(\lambda)$ have a unique minimum at the biggest λ for which such increments are negative. $\qquad\square$

4.2. Tikhonov method

We now consider the case when regularization is performed by means of the standard Tikhonov formula.

Corollary 2. *Let $x_\lambda(t)$ be the Tikhonov estimate as given by equation (8), with regularization parameter $\lambda \geq 0$. Then*

$$\varepsilon_{\mathbf{x}}(\lambda) = \sum_t \sum_{i=M+1}^{N} \left(\mathbf{v}_i^\top \mathbf{x}(t)\right)^2$$
$$+ \sum_t \sum_{i=1}^{M} \left[\frac{\lambda^2}{(\sigma_i^2 + \lambda)^2} \left(\mathbf{v}_i^\top \mathbf{x}(t)\right)^2 + \frac{\sigma_i^2}{(\sigma_i^2 + \lambda)^2} \left(\mathbf{u}_i^\top \mathbf{n}(t)\right)^2 \right] \tag{22}$$

and

$$\varepsilon_{\mathbf{S}}(\lambda) = \sum_f \sum_{\substack{i \geq M+1 \ or \\ j \geq M+1}} \left|(\mathbf{v}_i \otimes \mathbf{v}_j)^\top \mathcal{S}^{\mathbf{x}}(f)\right|^2$$
$$+ \sum_f \sum_{i,j=1}^{M} \left[\left(\frac{\sigma_i^2 \sigma_j^2}{(\sigma_i^2 + \lambda)(\sigma_j^2 + \lambda)} - 1 \right)^2 \left|(\mathbf{v}_i \otimes \mathbf{v}_j)^\top \mathcal{S}^{\mathbf{x}}(f)\right|^2 \right.$$
$$+ \frac{\sigma_i^2 \sigma_j^2}{(\sigma_i^2 + \lambda)^2 (\sigma_j^2 + \lambda)^2} \left|(\mathbf{u}_i \otimes \mathbf{u}_j)^\top \mathcal{S}^{\mathbf{n}}(f)\right|^2 \tag{23}$$
$$\left. + 2 \left(\frac{\sigma_i^2 \sigma_j^2}{(\sigma_i^2 + \lambda)(\sigma_j^2 + \lambda)} - 1 \right) \frac{\sigma_i \sigma_j}{(\sigma_i^2 + \lambda)(\sigma_j^2 + \lambda)} Re\left(\overline{(\mathbf{v}_i \otimes \mathbf{v}_j)^\top \mathcal{S}^{\mathbf{x}}(f)} (\mathbf{u}_i \otimes \mathbf{u}_j)^\top \mathcal{S}^{\mathbf{n}}(f) \right) \right]$$

Again we assume that $\mathbf{X}(t)$ and $\mathbf{N}(t)$ are white–noise Gaussian processes with covariance matrices $\omega^2 \mathbf{I}_N$ and $\alpha^2 \mathbf{I}_M$. Under this assumption equations (22) and (23) become

$$\varepsilon_{\mathbf{x}}(\lambda) = T(N - M)\omega^2 + T\omega^2 \sum_{i=1}^{M} \frac{\lambda^2}{(\sigma_i^2 + \lambda)^2} + T\alpha^2 \sum_{i=1}^{M} \frac{\sigma_i^2}{(\sigma_i^2 + \lambda)^2} \tag{24}$$

and

$$\varepsilon_{\mathbf{S}}(\lambda) = L(N-M)\omega^4 + L\omega^4 \sum_{i=1}^{M} \left(\frac{\sigma_i^4}{(\sigma_i^2 + \lambda)^2} - 1 \right)^2 + L\alpha^4 \sum_{i=1}^{M} \frac{\sigma_i^4}{(\sigma_i^2 + \lambda)^4}$$
$$+ 2L\omega^2\alpha^2 \sum_{i=1}^{M} \left(\frac{\sigma_i^4}{(\sigma_i^2 + \lambda)^2} - 1 \right) \frac{\sigma_i^2}{(\sigma_i^2 + \lambda)^2} \tag{25}$$

where we notice that, as anticipated in the previous section, the fourth addend is negative; this fact suggests that, to the extent that the other terms are comparable to those in the corresponding expression for the tSVD (18), the reconstruction error generated by the Tikhonov method is smaller than the one generated by tSVD.

By differentiating equations (24) and (25) we have

$$\frac{d}{d\lambda}\varepsilon_{\mathbf{x}}(\lambda) = 2T\left(\omega^2\lambda - \alpha^2\right) \sum_{i=1}^{M} \frac{\sigma_i^2}{(\sigma_i^2 + \lambda)^3} \tag{26}$$

and

$$\frac{d}{d\lambda}\varepsilon_{\mathbf{S}}(\lambda) = 4L\omega^2 \sum_{i=1}^{M} \frac{\sigma_i^2}{(\sigma_i^2 + \lambda)^5} (\alpha^2 + \sigma_i^2\omega^2)$$
$$\cdot \left(\lambda + \sigma_i^2 + \sqrt{\sigma_i^4 + \sigma_i^2 \frac{\alpha^2}{\omega^2}} \right) \left(\lambda + \sigma_i^2 - \sqrt{\sigma_i^4 + \sigma_i^2 \frac{\alpha^2}{\omega^2}} \right) \tag{27}$$

We are now able to prove the following theorem.

Theorem 2. *Let $x_\lambda(t)$ be the Tikhonov estimate as given by equation (8), with regularization parameter $\lambda \geq 0$; assume $\mathbf{X}(t)$ and $\mathbf{N}(t)$ to be white–noise Gaussian processes with covariance matrices $\omega^2\mathbf{I}_N$ and $\alpha^2\mathbf{I}_M$, respectively. Then*

$$\lambda_{\mathbf{x}}^* = \frac{\alpha^2}{\omega^2} \tag{28}$$

and

$$\lambda_{\mathbf{S}}^* < \frac{\lambda_{\mathbf{x}}^*}{2} \tag{29}$$

Proof. The first statement simply follows from equation (26) by observing that $\frac{d}{d\lambda}\varepsilon_{\mathbf{x}}(\lambda) \geq 0$ if and only if $\lambda \geq \frac{\alpha^2}{\omega^2}$.
Instead, equation (27) implies that $\frac{d}{d\lambda}\varepsilon_{\mathbf{S}}(\lambda) > 0$ if

$$\lambda > -\sigma_i^2 + \sqrt{\sigma_i^4 + \sigma_i^2 \frac{\alpha^2}{\omega^2}} \tag{30}$$

Consider the function $h : [0, +\infty) \ni z \to -z^2 + \sqrt{z^4 + z^2 \frac{\alpha^2}{\omega^2}}$. As schematically shown in Figure 1, h is strictly increasing and bounded above by $\frac{\lambda_{\mathbf{x}}^*}{2} = \frac{\alpha^2}{2\omega^2}$. As a consequence, the condition (30) is satisfied if $\lambda \geq \frac{\lambda_{\mathbf{x}}^*}{2}$, that means $\varepsilon_{\mathbf{S}}(\lambda)$ is strictly increasing in $[\frac{\lambda_{\mathbf{x}}^*}{2}, +\infty)$ and thus inequality (29) holds. $\square$

The main interest of Theorem 2 is that it provides a simple relationship between $\lambda_{\mathbf{S}}^*$ and $\lambda_{\mathbf{x}}^*$. However, expression (27) contains more information about the values of $\lambda_{\mathbf{S}}^*$, as stated in the following Theorem.
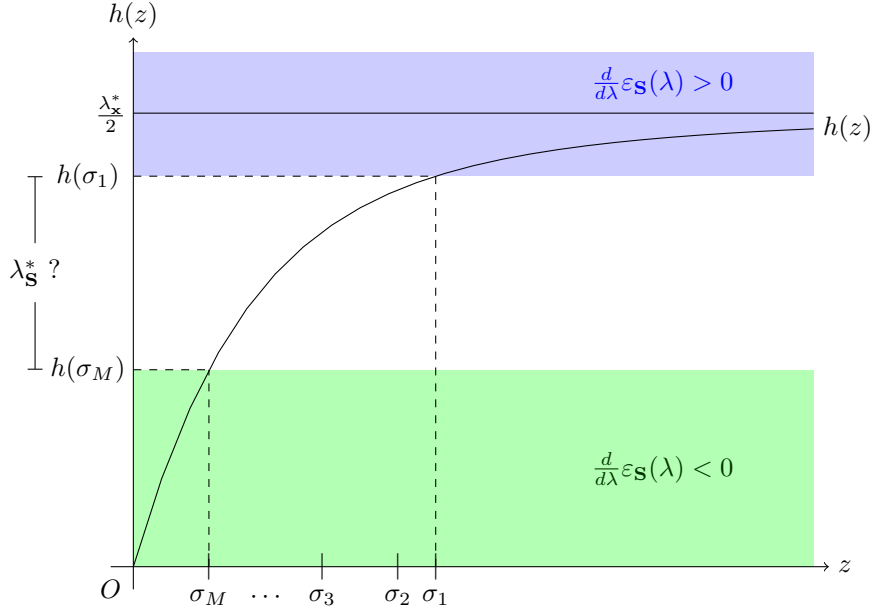
Figure 1. Plot of the function h(z) defined in the proof of Theorem 2. h(z) is related to the sign of the addends at the right hand side of (27). If $\lambda < h(\sigma_M)$ all the addends in (27) are negative therefore $\varepsilon_{\mathbf{S}}(\lambda)$ is decreasing (green area), whereas if $\lambda > h(\sigma_1)$ all the addends in (27) are positive therefore $\varepsilon_{\mathbf{S}}(\lambda)$ is increasing (blue area); it follows that the optimal regularization parameter $\lambda_{\mathbf{S}}^*$ lies in the interval $[h(\sigma_M); h(\sigma_1)]$. Moreover, for $\lambda \geq \frac{\lambda_{\mathbf{x}}^*}{2}$ all the addends in (27) are positive regardless of the singular values, and therefore $\varepsilon_{\mathbf{S}}(\lambda)$ is increasing; this fact leads to the inequality $\lambda_{\mathbf{S}}^* < \frac{\lambda_{\mathbf{x}}^*}{2}$

Theorem 3. *Under the same hypotheses of Theorem 2, the value of $\lambda_{\mathbf{S}}^*$ belongs to the interval $[h(\sigma_M), h(\sigma_1)]$, where $h(z) = -z^2 + \sqrt{z^4 + z^2 \frac{\alpha^2}{\omega^2}}$.*

Proof. As schematically shown in Figure 1, when $\lambda > h(\sigma_1)$, all the addends in (27) are positive and thus $\frac{d}{d\lambda}\varepsilon_{\mathbf{x}}(\lambda)$ is positive; on the other hand, when $\lambda < h(\sigma_M)$ the derivative $\frac{d}{d\lambda}\varepsilon_{\mathbf{x}}(\lambda)$ is negative as all the addends are negative. □

Remark 4. Theorem 3 also gives information on the limiting behaviour of $\lambda_{\mathbf{S}}^*$ as $\lambda_{\mathbf{x}}^* = \frac{\alpha^2}{\omega^2}$ approaches very small or very large values. In the no–noise scenario, when $\lambda_{\mathbf{x}}^* \sim 0$, $\lambda_{\mathbf{S}}^*$ grows approximately linearly with $\lambda_{\mathbf{x}}^*$. The other boundary is however more interesting. Indeed, when $\lambda_{\mathbf{x}}^* \to \infty$ the extremes of the interval $h(\sigma_1)$ and $h(\sigma_M)$ grow with the same order of $\sqrt{\lambda_{\mathbf{x}}^*}$. Therefore, when noise gets larger not only $\lambda_{\mathbf{S}}^*$ is smaller than $\lambda_{\mathbf{x}}^*$, but it also grows more slowly.

Remark 5. Theorems 2 and 3 imply that, when regularization is accomplished through the Tikhonov method, $\lambda_{\mathbf{x}}^*$ does not depend on the forward matrix $\mathbf{G}$, while $\lambda_{\mathbf{S}}^*$ does. The fact that $\lambda_{\mathbf{x}}^*$ does not depend on $\mathbf{G}$ may appear counter–intuitive: if the singular values grows, also the effective SNR of the data grow, and then the regularization parameter should become smaller. In fact, the regularization parameter does become smaller *with respect to the data*; in other words, this is the classical

behaviour of the optimal regularization parameter, where we are changing the SNR by increasing the strength of the exact signal, rather than decreasing the variance of the noise.

Remark 6. When $M = N$ and $\sigma_1 = \ldots = \sigma_M = 1$, the forward matrix $\mathbf{G}$ is orthogonal and the inverse problem in equation (1) is well–posed. Theorems 2 and 3 imply that $\lambda_{\mathbf{S}}^*$ and $\lambda_{\mathbf{x}}^*$ are different also under these conditions, as

$$h(\sigma_M) = h(\sigma_1) = -1 + \sqrt{1 + \frac{\alpha^2}{\omega^2}} < \frac{\alpha^2}{\omega^2}$$

Although unrealistic, this case is of particular interest in M/EEG functional connectivity because it corresponds to the ideal case where there is no *cross–talk* or *source–leakage* between sources [20].

Indeed, in this case the resolution matrix is proportional to the identity matrix $(\mathbf{R}_\lambda = (1 + \lambda)^{-1} \mathbf{I})$, i.e. the estimate at one location is not influenced by neural activity at different locations. Our result shows that also in this ideal case the optimal values of the regularization parameters are different.

5. Beyond the two–step approach: Filter factor for a direct estimation of $\mathbf{S}^{\mathbf{x}}(f)$ from $\mathbf{S}^{\mathbf{y}}(f)$

As an alternative to the two–step approach described so far, one may directly estimate the cross–power spectrum of the unknown $\mathbf{S}^{\mathbf{x}}(f)$ from that of the data $\mathbf{S}^{\mathbf{y}}(f)$. Indeed, from equation (1) and from the linearity of the Fourier Transform it follows

$$\mathcal{S}^{\mathbf{y}}(f) = (\mathbf{G} \otimes \mathbf{G})\mathcal{S}^{\mathbf{x}}(f) + \mathcal{S}^{\mathbf{n}}(f) \quad , \tag{31}$$

which describes a linear inverse problem.

Analogously to what we did in the previous Sections for the forward operator $\mathbf{G}$, we can introduce the SVD of the forward operator $\mathbf{G} \otimes \mathbf{G} = (\mathbf{U} \otimes \mathbf{U})(\mathbf{\Sigma} \otimes \mathbf{\Sigma})(\mathbf{V} \otimes \mathbf{V})^\top$, up to reordering the elements of $\mathbf{\Sigma} \otimes \mathbf{\Sigma}$ and the corresponding columns of $\mathbf{U} \otimes \mathbf{U}$ and $\mathbf{V} \otimes \mathbf{V}$. We can then express a *one–step* regularized estimate of the cross–spectrum in terms of the SVD and of the filter factors

$$\mathcal{S}^{\mathbf{x}}_\lambda(f) = (\mathbf{V} \otimes \mathbf{V}) \, \widetilde{\mathbf{\Phi}}(\lambda) \, (\mathbf{\Sigma} \otimes \mathbf{\Sigma})^\dagger \, (\mathbf{U} \otimes \mathbf{U})^\top \, \mathcal{S}^{\mathbf{y}}(f)$$

$$= \sum_{i,j}^{M} \widetilde{\varphi}_{i,j}(\lambda) \frac{(\mathbf{u}_i \otimes \mathbf{u}_j)^\top \, \mathcal{S}^{\mathbf{y}}(f)}{\sigma_i \sigma_j} (\mathbf{v}_i \otimes \mathbf{v}_j) \quad . \tag{32}$$

In particular, if Tikhonov regularization is employed, the filter factors read

$$\widetilde{\varphi}_{i,j}(\lambda) = \frac{\sigma_i^2 \sigma_j^2}{\sigma_i^2 \sigma_j^2 + \lambda} \tag{33}$$

while in tSVD the components such that the product $\sigma_i \sigma_j$ is below the threshold defined by λ are filtered out. Instead, in the classical two–step approach the filter factors for the estimated cross–spectrum are simply given by the product of the filter factors for the estimated source time–courses, that means each of the singular value σ_i and σ_j is individually filtered, instead of their product $\sigma_i \sigma_j$. Indeed, the cross–spectrum of the regularized estimate $\mathbf{x}_\lambda(t)$ in equation (4) is

$$\mathcal{S}^{\mathbf{x}_\lambda}(f) = (\mathbf{V} \otimes \mathbf{V}) \left(\mathbf{\Phi}\mathbf{\Sigma}^\dagger \otimes \mathbf{\Phi}\mathbf{\Sigma}^\dagger \right) (\mathbf{U} \otimes \mathbf{U})^\top \, \mathcal{S}^{\mathbf{y}}(f)$$

$$= \sum_{i,j}^{M} \varphi_i(\lambda)\varphi_j(\lambda) \frac{(\mathbf{u}_i \otimes \mathbf{u}_j)^\top \, \mathcal{S}^{\mathbf{y}}(f)}{\sigma_i \sigma_j} (\mathbf{v}_i \otimes \mathbf{v}_j) \quad . \tag{34}$$
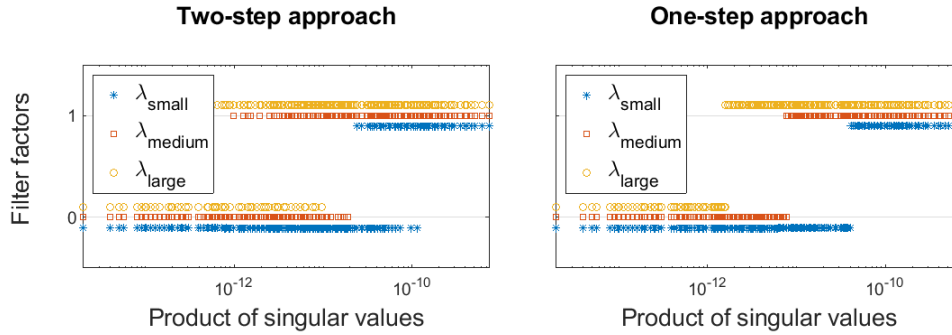
Figure 2. Filter factors for the tSVD method. On the x–axis the product of the singular values $\sigma_i\sigma_j$, on the y–axis the corresponding values of the filter factors $\varphi_i(\lambda)\varphi_j(\lambda)$ for the two–step approach (left) and $\widetilde{\varphi}_{i,j}(\lambda)$ for the one–step approach (right). The three different colors correspond to three different values of the regularization parameter, as illustrated in the legend. Please notice that the filter factors for tSVD are either zero or one, but different colors are plotted at slightly different levels for the sake of clarity.

As a comparison in Figures 2 and 3 we plotted the filter factors $\widetilde{\varphi}_{i,j}(\lambda)$ and $\varphi_i(\lambda)\varphi_j(\lambda)$ for the tSVD and Tikhonov method. The forward matrix was obtained by randomly selecting (uniform sampling) $M = 20$ sensors and $N = 25$ source locations from a standard MEG forward operator based on a realistic, three–layer boundary element method (BEM) head model, publicly available within the mne–python software [16]. Such selection of a subset of rows and columns allows to produce more readable plots than those obtained by using the whole leadfield, that would be qualitatively similar but with a denser cloud of points. The specific choice of source and sensor locations does not modify the results: we tried several random configurations and they all provided similar plots. Figure 2 and 3 highlight the potential advantages of the one–step approach over the two–step approach. In the case of tSVD, the filter factors of the two–step approach are zero whenever either $i < \lambda$ or $j < \lambda$, which implies a jittering behaviour when plotted as a function of the product $\sigma_i\sigma_j$. In the one–step approach this issue is not present, because filtering is applied directly to the product of the singular values. In the case of the Tikhonov method we observe a similar behaviour, where in the one–step approach the filter factors increase smoothly when the product $\sigma_i\sigma_j$ increases, while in the two–step approach also higher values of such product may be severely filtered because of the effect of the regularization parameter on the individual singular values.

6. A numerical simulation

In this Section we use a numerical simulation to show what happens when the rather restrictive assumption of a white Gaussian signal, that was needed to prove the results in Section 4, no longer holds. We exploit the same simulation to show that in this case the one–step approach described by equation (32) enables to estimate the cross–power spectrum with a lower reconstruction error. We remark, however, that this simulation is just an example and is not meant to be a full validation, for which a more thorough set of simulations would be needed and which is beyond the scope of this article.

Following Hincapié et al. [21], we simulated two interacting oscillatory sources.
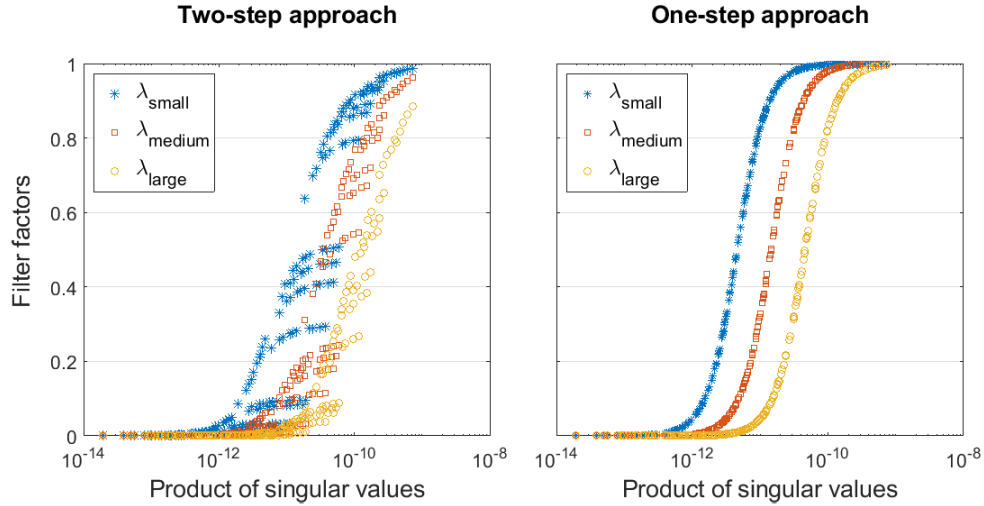
Figure 3. Filter factors for the Thikonov method. On the x–axis the product of the singular values $\sigma_i\sigma_j$, on the y–axis the corresponding values of the filter factors $\varphi_i(\lambda)\varphi_j(\lambda)$ for the two–step approach (left) and $\widetilde{\varphi}_{i,j}(\lambda)$ for the one–step approach (right). The three different colors correspond to three different values of the regularization parameter, as illustrated in the legend.

The first source was placed in the temporal lobe of the left hemisphere, the second source was placed in the occipital lobe of the right hemisphere at a distance of 11.8 cm from the first one. The time courses of the two sources were simulated with a coherence level of 0.4, as follows: first the base frequency was set to 12 Hz for both sources; then the instantaneous frequency was randomly drawn independently around the base frequency, causing fluctuations in the phase relationship between the two signals. A total number of $T = 30,000$ time points was used. The process was repeated until the desired level of coherence (0.4) was attained.

Source time courses were then projected to the sensor level through a forward operator obtained by downsampling the same MEG leadfield used in the previous section; the final forward operator has $M = 102$ sensors (magnetometers) and $N = 274$ source points guaranteeing a uniform coverage of the brain. With the exception of the two oscillatory sources, all other source time courses were set to zero. White Gaussian noise was added at the sensor level to reach five different values of Signal–to–Noise Ratio (SNR) defined as

$$\text{SNR} = 10\log_{10}\left(\frac{\sum_{t=0}^{T-1}||\mathbf{Gx}(t)||^2}{\sum_{t=0}^{T-1}||\mathbf{n}(t)||^2}\right) \quad, \tag{35}$$

where $\mathbf{n}(t) = \sigma\tilde{\mathbf{n}}(t)$, being $\tilde{\mathbf{n}}(t) \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, and σ was defined in order to attain the desired SNR. The values of SNR were evenly selected in the range $[-10\text{dB}, 10\text{dB}]$.

For each simulated data we numerically computed the two optimal parameters $\lambda_{\mathbf{x}}^*$ and $\lambda_{\mathbf{S}}^*$, as defined in Definition 1, for tSVD and for the Tikhonov method. The optimal values are reported in Table 1. For both methods, optimal reconstruction of the cross–spectrum requires less regularization than optimal reconstruction of the signal. For tSVD this result is qualitatively different from the white noise case, where

we proved that the two optimal parameters are equal; for the Tikhonov method, on the other hand, the optimal values obey the same inequality, and the ratio between the optimal values is very similar to the one reported in [21]. This is expected, since the numerical simulation here was constructed following the same scheme.

Finally, we numerically computed the optimal regularization parameter for the one–step approach defined as

$$\tilde{\lambda}_{\mathbf{S}}^* = \arg\min_\lambda \tilde{\varepsilon}_{\mathbf{S}}(\lambda) \quad \text{with} \quad \tilde{\varepsilon}_{\mathbf{S}}(\lambda) = \sum_f \|\mathbf{S}_\lambda^{\mathbf{x}}(f) - \mathbf{S}^{\mathbf{x}}(f)\|_F^2 \qquad (36)$$

and we compared the reconstruction error reached by the one–step and the two–step approach when the corresponding optimal regularization parameters are employed, i.e. $\varepsilon_{\mathbf{S}}(\lambda_{\mathbf{S}}^*)$ and $\tilde{\varepsilon}_{\mathbf{S}}(\tilde{\lambda}_{\mathbf{S}}^*)$. Figure 4 shows the reconstruction errors for the cross–power spectrum for tSVD and Tikhonov method for both the two–step and the one–step approach as a function of the SNR. The line corresponding to the two-step approach is always above the one corresponding to the one-step approach, showing that the latter provides a better estimation for the cross-power spectrum.

	tSVD		Tikhonov		
SNR	$\lambda_{\mathbf{x}}^*$	$\lambda_{\mathbf{S}}^*$	$\lambda_{\mathbf{x}}^*$	$\lambda_{\mathbf{S}}^*$	$\lambda_{\mathbf{S}}^*/\lambda_{\mathbf{x}}^*$
-10	7	25	110	$9.36 \cdot 10^{-1}$	$8.5 \cdot 10^{-3}$
-5	8	35	35.0	$2.93 \cdot 10^{-1}$	$8.4 \cdot 10^{-3}$
0	8	45	9.86	$9.53 \cdot 10^{-2}$	$9.7 \cdot 10^{-3}$
5	25	53	3.15	$3.00 \cdot 10^{-2}$	$9.5 \cdot 10^{-3}$
10	28	69	1.12	$9.40 \cdot 10^{-3}$	$8.4 \cdot 10^{-3}$

Table 1. $\lambda_{\mathbf{x}}^*$ and $\lambda_{\mathbf{S}}^*$ for the different values of SNR. In tSVD, the number of retained SVD components for the reconstruction of the cross–power spectrum is always higher than for the time–series reconstruction, showing that less regularization is needed for the first one. For the Tikhonov method, $\lambda_{\mathbf{S}}^*$ is always smaller than $\lambda_{\mathbf{x}}^*$, showing that less regularization is needed for the cross–power spectrum reconstruction.

7. Discussion and future work

Motivated by an analysis pipeline which is largely used for connectivity studies in the M/EEG community, in this article we have considered the problem of whether, in a two–step approach to the reconstruction of the cross–power spectrum of an unobservable signal, one should set the regularization parameter differently than what one would do for the reconstruction of the signal itself.

First, making use of filter factor analysis, we obtained an explicit expression for the reconstruction error for the cross–power spectrum under the two–step approach. This formula is the analogous of the well–known formula for the reconstruction error in linear inverse problems, and holds in general. Then, under additional hypotheses of a white Gaussian signal and white Gaussian noise, we proved that the optimal values coincide for tSVD, while in the Tikhonov method the optimal value for the cross–spectrum is at most half the optimal value for the signal, thus proving also that the answer actually depends on the inverse method. We speculate that such difference may be partly due to the fact that, with white Gaussian signals, the error in estimating
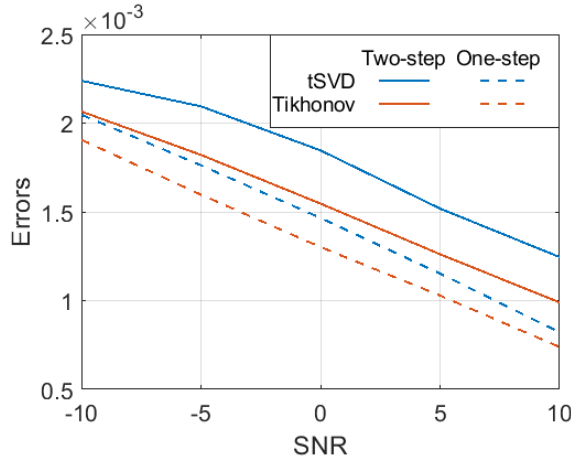
Figure 4. Reconstruction errors for the cross–spectrum reconstruction as a function of the SNR for tSVD (blue) and for the Tikhonov method (red), for the two–step (solid line) and for the one–step (dotted line) approach. For both methods the one–step approach provides a better reconstruction of the cross–power spectrum.

the cross–power spectrum involves the square of the filter factors; for tSVD, where filter factors can only be 0 or 1, the components corresponding to the largest singular values will possibly be weighted by a 1, and when the filter is squared for the cross–spectrum, the largest component will still be weighted by a 1. For Tikhonov, where filter factors range in the interval $[0, 1]$, the components corresponding to the largest singular values will be weighted by a factor lower than one, and when such factor is squared it becomes even smaller; therefore, in order not to filter excessively the largest components, a smaller regularization parameter is needed.

The results of Section 4 are in line with the results of [21], which showed empirically that the optimal estimate of connectivity is obtained with a regularization parameter smaller than the one providing the optimal estimate of the power spectrum, i.e. of the signal strength. Quantitatively, the recommendation in [21] was to use a parameter two orders of magnitude lower, while our main theorem for the Tikhonov method guarantees $\lambda_{\mathbf{S}}^* < \lambda_{\mathbf{x}}^*/2$.

Theorems 1, 2 and 3 have been obtained under the somewhat unrealistic assumption that the signal is a white–noise Gaussian process. While this is an important limitation with respect to the applications, preliminary numerical results indicate that the optimal value for the cross–spectrum is further reduced by the presence of a temporal structure in the signal: in the numerical simulation in Section 6 we observed that, with interacting, oscillatory signals, the optimal values in tSVD no longer coincide, and $\lambda_{\mathbf{S}}^*$ in the Tikhonov method is approximately two orders of magnitude smaller than $\lambda_{\mathbf{X}}^*$, in line with the mentioned results in [21]. In any case, future work will be devoted to investigating in detail the effect of a more plausible temporal structure of the input waveforms.

In addition, our results so far only concern the cross–power spectrum; future work will investigate the impact of the regularization parameter on the estimated value of the connectivity measure, such as Imaginary part of Coherency, Partial Directed Coherence and Granger causality.

Finally, as we point out in Section 5, our results suggest that the two–step approach to estimation of the cross–power spectrum, and more in general of brain functional connectivity, might be sub–optimal. This idea is in line with literature on the topic [24, 8, 14, 32, 39, 40]. Indeed, by looking at the filter factors obtained by the two–step approach, and comparing them to the filter factors one would get with a one–step approach to estimation of the power spectrum, we expect a better behaviour for this second option. Newly presented methods such as PSIICOS [32] present one–step approaches to the estimation of connectivity that benefit from this fact. Future work will be devoted to investigating more thoroughly this alternative approach.

Acknowledgments

References

[1] Baccalá L A and Sameshima K 2001 Partial directed coherence: a new concept in neural structure determination *Biological cybernetics* **84** 463–474

[2] Bekhti Y, Lucka F, Salmon J and Gramfort A 2018 A hierarchical bayesian perspective on majorization-minimization for non-convex sparse regression: application to M/EEG source imaging *Inverse Problems* **34** 085010

[3] Bendat J S and Piersol A G 2011 *Random data: analysis and measurement procedures* vol 729 (John Wiley & Sons)

[4] Brookes M J, Woolrich M, Luckhoo H, Price D, Hale J R, Stephenson M C, Barnes G R, Smith S M and Morris P G 2011 Investigating the electrophysiological basis of resting state networks using magnetoencephalography *Proceedings of the National Academy of Sciences* **108** 16783–16788

[5] Calvetti D, Pascarella A, Pitolli F, Somersalo E and Vantaggi B 2015 A hierarchical krylov–bayes iterative inverse solver for MEG with physiological preconditioning *Inverse Problems* **31** 125005

[6] Chella F, Marzetti L, Pizzella V, Zappasodi F and Nolte G 2014 Third order spectral analysis robust to mixing artifacts for mapping cross-frequency interactions in EEG/MEG *Neuroimage* **91** 146–161

[7] Chella F, Marzetti L, Stenroos M, Parkkonen L, Ilmoniemi R J, Romani G L and Pizzella V 2019 The impact of improved MEG–MRI co-registration on MEG connectivity analysis *NeuroImage* **197** 354–367

[8] Cheung B L P, Riedner B A, Tononi G and Van Veen B D 2010 Estimation of cortical connectivity from EEG using state-space models *IEEE Transactions on Biomedical engineering* **57** 2122–2134

[9] Costa F, Batatia H, Oberlin T, D'Giano C and Tourneret J Y 2017 Bayesian EEG source localization using a structured sparsity prior *NeuroImage* **144** 142–152

[10] De Pasquale F, Della Penna S, Snyder A Z, Lewis C, Mantini D, Marzetti L, Belardinelli P, Ciancetta L, Pizzella V, Romani G L *et al.* 2010 Temporal dynamics of spontaneous MEG activity in brain networks *Proceedings of the National Academy of Sciences* **107** 6040–6045

[11] de Peralta Menendez R G, Andino S L G and Lütkenhöner B 1996 Figures of merit to compare distributed linear inverse solutions *Brain Topography* **9** 117–124

[12] Fraschini M, Demuru M, Crobe A, Marrosu F, Stam C J and Hillebrand A 2016 The effect of epoch length on estimated EEG functional connectivity and brain network organisation *Journal of neural engineering* **13** 036015

[13] Fries P 2005 A mechanism for cognitive dynamics: neuronal communication through neuronal coherence *Trends in cognitive sciences* **9** 474–480

[14] Fukushima M, Yamashita O, Knösche T R and Sato M 2015 MEG source reconstruction based on identification of directed source interactions on whole-brain anatomical networks *NeuroImage* **105** 408–427

[15] Geweke J 1982 Measurement of linear dependence and feedback between multiple time series *Journal of the American statistical association* **77** 304–313

[16] Gramfort A, Luessi M, Larson E, Engemann D A, Strohmeier D, Brodbeck C, Parkkonen L and Hämäläinen M S 2014 MNE software for processing MEG and EEG data *Neuroimage* **86** 446–460

[17] Hämäläinen M S and Ilmoniemi R J 1994 Interpreting magnetic fields of the brain: minimum norm estimates *Medical & biological engineering & computing* **32** 35–42

[18] Hansen P C 2005 *Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion* vol 4 (Siam)

[19] Hari R and Puce A 2017 *MEG-EEG Primer* (Oxford University Press)

[20] Hauk O, Stenroos M and Treder M 2019 EEG/MEG source estimation and spatial filtering: The linear toolkit *Magnetoencephalography: From Signals to Dynamic Cortical Networks* 1–37

[21] Hincapié A S, Kujala J, Mattout J, Daligault S, Delpuech C, Mery D, Cosmelli D and Jerbi K 2016 MEG connectivity and power detections with minimum norm estimates require different regularization parameters *Computational Intelligence and Neuroscience* **2016** 19

[22] Hincapié A S, Kujala J, Mattout J, Pascarella A, Daligault S, Delpuech C, Mery D, Cosmelli D and Jerbi K 2017 The impact of MEG source reconstruction method on source-space connectivity estimation: a comparison between minimum-norm solution and beamforming *Neuroimage* **156** 29–42

[23] Ilmoniemi R J and Sarvas J 2019 *Brain Signals: Physics and Mathematics of MEG and EEG* (Mit Press)

[24] Kiebel S J, Garrido M I, Moran R J and Friston K J 2008 Dynamic causal modelling for EEG and MEG *Cognitive neurodynamics* **2** 121

[25] Liljeström M, Stevenson C, Kujala J and Salmelin R 2015 Task-and stimulus-related cortical networks in language production: Exploring similarity of MEG- and fMRI-derived functional connectivity *Neuroimage* **120** 75–87

[26] Luria G, Duran D, Visani E, Sommariva S, Rotondi F, Sebastiano D R, Panzica F, Piana M and Sorrentino A 2019 Bayesian multi-dipole modelling in the frequency domain *Journal of neuroscience methods* **312** 27–36

[27] Nolte G, Bai O, Wheaton L, Mari Z, Vorbach S and Hallett M 2004 Identifying true brain interaction from EEG data using the imaginary part of coherency *Clinical neurophysiology* **115** 2292–2307

[28] Nolte G, Galindo-Leon E, Li Z, Liu X and Engel A K 2019 Mathematical relations between measures of brain connectivity estimated from electrophysiological recordings for gaussian distributed data *bioRxiv* 680678

[29] Nolte G, Ziehe A, Nikulin V V, Schlögl A, Krämer N, Brismar T and Müller K R 2008 Robustly estimating the flow direction of information in complex physical systems *Physical review letters* **100** 234101

[30] Nunes R V, Reyes M B and De Camargo R Y 2019 Evaluation of connectivity estimates using spiking neuronal network models *Biological cybernetics* **113** 309–320

[31] Nunez P L, Silberstein R B, Shi Z, Carpenter M R, Srinivasan R, Tucker D M, Doran S M, Cadusch P J and Wijesinghe R S 1999 EEG coherency II: experimental comparisons of multiple measures *Clinical Neurophysiology* **110** 469–486

[32] Ossadtchi A, Altukhov D and Jerbi K 2018 Phase shift invariant imaging of coherent sources (PSIICOS) from MEG data *NeuroImage* **183** 950–971

[33] Pereda E, Quiroga R Q and Bhattacharya J 2005 Nonlinear multivariate analysis of neurophysiological signals *Progress in neurobiology* **77** 1–37

[34] Sakkalis V 2011 Review of advanced techniques for the estimation of brain connectivity measured with EEG/MEG *Computers in biology and medicine* **41** 1110–1117

[35] Schoffelen J M and Gross J 2019 Studying dynamic neural interactions with MEG *Magnetoencephalography: from signals to dynamic cortical networks* 1–23

[36] Sommariva S, Sorrentino A, Piana M, Pizzella V and Marzetti L 2019 A comparative study of the robustness of frequency-domain connectivity measures to finite data length *Brain topography* **32** 675–695

[37] Sorrentino A and Piana M 2017 Inverse modeling for MEG/EEG data *Mathematical and Theoretical Neuroscience* (Springer, Cham) pp 239–253

[38] Stam C J 2010 Use of magnetoencephalography (MEG) to study functional brain networks in neurodegenerative disorders *Journal of the Neurological Sciences* **289** 128–234

[39] Subramaniyam N P, Tronarp F, Särkkä S and Parkkonen L 2017 Expectation–maximization algorithm with a nonlinear Kalman smoother for MEG/EEG connectivity estimation *EMBEC & NBC 2017* (Springer) pp 763–766

[40] Tronarp F, Subramaniyam N P, Särkkä S and Parkkonen L 2018 Tracking of dynamic functional connectivity from MEG data with Kalman filtering *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (IEEE) pp 1003–1006

[41] Van Mierlo P, Höller Y, Focke N K and Vulliemoz S 2019 Network perspectives on epilepsy using EEG/MEG source connectivity *Frontiers in Neurology* **10** 721

[42] Van Veen B D, Van Drongelen W, Yuchtman M and Suzuki A 1997 Localization of brain electrical activity via linearly constrained minimum variance spatial filtering *IEEE Transactions on biomedical engineering* **44** 867–880

[43] Wakeman D G and Henson R N 2015 A multi-subject, multi-modal human neuroimaging dataset *Scientific data* **2** 150001

[44] Welch P 1967 The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms *IEEE Transactions on audio and electroacoustics* **15** 70–73