Investigating complex networks with inverse models

Vincent Wens*
Laboratoire de Cartographie Fonctionnelle du Cerveau, UNI – ULB Neurosciences Institute, Université Libre de Bruxelles (ULB), Brussels, Belgium
(Dated: January 13, 2014)

Recent advances in neuroscience have motivated the study of network organization in spatially distributed dynamical systems from indirect measurements. However, the associated connectivity estimation, when combined with inverse modeling, is strongly affected by spatial leakage. We formulate this problem in a general framework and develop a new approach to model spatial leakage and limit its effects. It is analytically compared to existing regression-based methods used in electrophysiology, which are shown to yield biased estimates of amplitude and phase couplings.

Investigating the dynamical organization of spatially extended systems represents a challenging problem arising in many scientific disciplines, from physics to biology. In this context, network theory has emerged as a widely used phenomenological tool [1]. It is based upon a representation of complex systems as graphs [2], which are built by selecting the dynamical subunits defining nodes and a coupling measure defining edges. These modeling steps are known to affect the ensuing graph topology, which can lead to misinterpreting properties such as ‘small-world-ness’ [3, 4]. Node selection is especially difficult for distributed systems accessible through indirect or incomplete observations only. Directly using sensors as nodes may be of limited value, insofar as measurements miss or mix relevant subunits, leading to graphs that depend on the experimental set-up [4]. A natural way out is to reconstruct those subunits from observations, a problem that may not have a unique solution and thus require the introduction of an inverse model [5, 6]. Yet, model assumptions can strongly affect the resulting graph and generate spurious topological features [7].

To illustrate this, we shall hereafter consider a continuous system whose state is described by an unknown scalar field \( \Psi_0(r,t) \). The physics of the problem determines the direct operator \( L \) mapping this state to observed data \( \mu(t) \in \mathbb{R}^m \) via \( \mu(t) = (L \Psi_0)(t) + \varepsilon(t) \), where \( \varepsilon(t) \) denotes measurement noise. An inverse model then provides a reconstruction \( \Psi(r,t) = (W_r \mu)(t) \) of \( \Psi_0(r,t) \) by application of the inverse operator \( W_r \) on data. This typically yields a deformed representation of \( \Psi_0(r,t) \) [Figs. 1(a,c)] and spurious connectivity patterns [Figs. 1(b,d)]. We shall refer to this effect as spatial leakage [8].

In neuroscience, this issue affects the mapping of brain networks from scalp electrophysiological data [9]. Indeed, their limited spatial resolution typically leads to smooth reconstructions [Fig. 1(c)] of neural electric current flows, and connectivity maps between activity at \( r_0 \) and the rest of the brain display spurious coupling patterns, especially around \( r_0 \) [Fig. 1(d)]. In practice, the latter often dominates and hides true long-range interactions. Eliminating spurious local connectivity then allows to emphasize the large-scale structure of brain networks. Some techniques were introduced to that effect, such as imaginary coherence [10] or phase lag index [11, 12] for phase coupling, and orthogonalization [13, 14] for amplitude correlation. However, an approach valid for any type of connectivity indices has not yet been considered.

In this Letter, we discuss the concept of spatial leakage in the general framework of inverse problems. We explicitly model spatial leakage from a given location \( r_0 \) and introduce a general correction scheme suppressing spurious coupling around \( r_0 \) and emphasizing true long-range connectivity [Figs. 1(e,f)]. We analytically compare it to orthogonalization by deriving a simple formula connecting these two approaches. This also yields a comparison with imaginary coherence and phase lag index as well, which appear as orthogonalized phase coupling estimates.

Finally, we highlight the biases in connectivity estimation due to orthogonalization and absent in our new method, e.g., the effects of phase coupling on orthogonalized amplitude correlation.
Spatial structure of inverse operators. Spatial leakage refers to the spurious connectivity patterns induced by the spatial profile of inverse operators, independently of the system state. Therefore, it does not reflect dynamical features but merely structural properties depending on the direct model and reconstruction priors only, a basic fact that has seemingly not been used until now. This concept can be illustrated clearly using

\[(L\Psi_0)(t) = \int d^3 r \ell(r) \Psi_0(r,t). \tag{1}\]

The kernel \(\ell(r) \in \mathbb{R}^m\) is typically quite smooth in applications. For notational convenience, we shall hereafter use real scalar fields, the generalization to vector fields being straightforward. Let us consider prototypical inverse models based on minimizing the sum of the prediction error \(\frac{1}{2}||\mu(t) - (L\Psi)(t)||^2\) and a term \(\frac{1}{p} \int d^3 r |\Psi(r,t)|^p\) imposing a least \(L^p\)-norm constraint \((p > 1)\) [6]. Although an explicit analytical form can not be obtained generically, implying the need for numerical optimization, an implicit equation for the solution \(\Psi(r,t) = (W_r \mu)(t)\) can be derived (see Supp. Mat. [15]),

\[\Psi(r,t) = \frac{\ell(r)^T(\mu(t) - (L\Psi)(t))}{\ell(r)^T(\mu(t) - (L\Psi)(t))}^{1/p}, \tag{2}\]

where \(\alpha_p = (p - 2)/(p - 1)\). This shows that the spatial profile of the inverse operator is strongly controlled by the kernel \(\ell(r)\). In particular, when \(\ell(r)\) is continuous at \(r_0\), Eq. 2 shows that \(W_r\) shares this property. This leads to spurious local connectivity around \(r_0\) since we have \(\Psi(r,t) \approx \Psi(r_0,t)\), for \(r\) in some neighborhood of \(r_0\), independently of \(\mu(t)\). Only the precise shape and size of this neighborhood depends on data \(\mu(t)\) and parameter \(p\). Likewise, long-distance correlations in \(\ell(r)\) induce spurious couplings. For example, \(\ell(r) = R(r_0)\ell\) implies \(W_r = \kappa^{1/(p-1)} W_{r_0}\) by Eq. 2, so that \(\Psi(r,t)\) and \(\Psi(r_0,t)\) are temporally correlated whatever \(\mu(t)\).

More generally, various assumptions and parameters can affect the spatial properties of inverse operators. Yet, the issue of spatial leakage is of quite general concern and applies to other inverse models [16–18] as well [15]. For example, the whole class of models based on minimizing the sum of the prediction error and applying to other inverse models \([16–18]\) as well \([15]\).

Structural model. Let us first consider an ideal state localized at \(r_0\), i.e., \(\Psi_0(r,t) = \delta(r - r_0)f(t)\). Its reconstruction \(\Psi(r,t) = (W_r \mu)(t)\) from noiseless observations \(\mu(t) = (L\Psi_0)(t)\) is a functional of \(f(t)\), \(\Psi(r,t) = (R_{r_0}^{-1}f)(t)\), which defines the resolving operator \(R_{r_0}^{-1}\) [6] and describes spatial leakage from \(r_0\). For example, \(R_{r_0}^{-1} = W_r \ell(r_0)\) in the case of Eq. 1. Assuming \(R_{r_0,r_0}^{-1}\) invertible, the unknown function \(f(t)\) can be reformulated in terms of calculable quantities as \((R_{r_0,r_0}^{-1}\Psi(r_0,t))(t)\). This leads to a closed expression \(\Psi(r,t) = \Lambda(r,t)\) for this “pure leakage” field, where

\[\Lambda(r,t) = (R_{r,r_0} R_{r_0,r_0}^{-1}\Psi(r_0,t))(t). \tag{3}\]

When no assumptions are made about \(\Psi_0(r,t)\), we shall use Eq. 3 as a model for the contribution of spatial leakage from \(r_0\) to the reconstruction \(\Psi(r,t) = (W_r \mu)(t)\), and define a corrected field

\[\Phi(r,t) = \Psi(r,t) - \Lambda(r,t). \tag{4}\]

It can also be written as \(\Phi(r,t) = (W_r^{\text{new}} \mu)(t)\) in terms of a new inverse operator

\[W_r^{\text{new}} = W_r - R_{r,r_0} R_{r_0,r_0}^{-1} W_{r_0} \tag{5}\]

depending on the direct and inverse models only. Equations 4, 5 form the analytical basis of our new method.

Although it can be applied to any state and inverse model, its strict validity presents two restrictions. First, this subtractive scheme requires spatial leakage to linearly contribute to the reconstruction. This assumption holds locally when the inverse operator is continuous at \(r_0\), since then \(\Psi(r,t) \approx \Psi(r_0,t) = \Lambda(r_0,t) \approx \Lambda(r,t)\) for \(r\) close enough to \(r_0\). It is otherwise invalid, except when the resolving operator is linear. Second, the node \(r_0\) must be isolated from the other activated regions, i.e., \(\Psi_0(r,t)\) assumes the form \(\delta(r - r_0)f(t) + g(r,t)\) with \(g(r,t) \neq 0\) only where spatial leakage to \(r_0\) is negligible, \(R_{r_0,r_0} \approx 0\). Otherwise, \(g(r,t)\) would also contribute to the expression \((R_{r_0,r_0}^{-1}\Psi(r_0,t))(t)\) used for \(f(t)\) to derive Eq. 3. Explicitly, in the case of Eq. 1,

\[R_{r_0,r_0}^{-1}\Psi(r_0,t) = f + \int d^3 r R_{r_0,r_0}^{-1} R_{r_0,r_0} g(r,r_0). \tag{6}\]

The neighborhood of \(r_0\) where \(R_{r_0,r_0} g(r,r_0) \neq 0\) being typically of non-zero but limited size, local leakage is generally overcorrected, as illustrated by the identity \(W_{r_0}^{\text{new}} = 0\), whereas the reconstruction at sufficiently long distances remains unchanged, since \(W_r^{\text{new}} \approx W_r\) whenever leakage from \(r_0\) is negligible, \(R_{r_0,r_0} \approx 0\). Note also that spatial leakage between \(r\) and \(r'\) away from \(r_0\) is left uncorrected [Fig. 1(e)]. At the level of connectivity maps between \(\Psi(r_0,t)\) and \(\Phi(r,t)\), this means that local maps is suppressed whereas true long-distance interactions are preserved, although they are still deformed by spatial leakage [Fig. 1(f)]. Despite this limitation, the correction scheme given by Eqs. 4, 5 is useful to emphasize long-range connectivity. This makes it suitable, in particular, for studies of large-scale brain networks.

In this application, Eq. 1 is used and inverse models are taken of the form \((W_r \mu)(t) = w(r,t)^T \mu(t)\) with \(w(r,t) \in \mathbb{R}^m\). It then follows from the definitions that \((R_{r_0,r_0}^{-1}f)(t) = w(r,t)^T \ell(r_0)(t)\) and

\[\Phi(r,t) = \Psi(r,t) - \gamma(r,r_0,t) \Psi(r_0,t), \tag{7}\]

\[\gamma(r,r_0,t) = w(r,t)^T \ell(r_0) / w(r_0,t)^T \ell(r_0). \tag{8}\]
In the rest of this Letter, we shall use this particular case and compare this new method to alternative techniques used in electrophysiology.

**Comparison with orthogonalization.** The philosophy behind these techniques is to consider connectivity estimates insensitive to any instantaneous linear coupling, of which spatial leakage (Eqs. 7, 8) is a particular case. To implement this idea, we still consider Eq. 7 for complex scalar fields but now determine the real coefficient \(\gamma(r, r_0, t)\) via linear regression. This yields

\[
\tilde{\Phi}(r, t) = \Phi(r, t) - \gamma(r, r_0, t) \Psi(r_0, t),
\]

\[
\gamma(r, r_0, t) = \text{Re}[\langle \Phi(r, t) \Psi(r_0, t)^* \rangle / |\Psi(r_0, t)|^2].
\] (10)

The orthogonalization method [14] uses these equations to remove spatial leakage from \(r_0\). Geometrically, \(\Phi(r, t)\) coincides with the orthogonal projection of \(\Psi(r, t)\) onto the direction perpendicular to \(\Psi(r_0, t)\). This implies that \(\Phi(r, t)\) is left invariant by the addition to \(\Psi(r, t)\) of any complex number co-linear with \(\Psi(r_0, t)\). It thus follows from Eqs. 7, 8 that \(\Psi(r, t)\) can be replaced by \(\Phi(r, t)\) in the definition of \(\tilde{\Phi}(r, t)\), which yields the useful formula

\[
\tilde{\Phi}(r, t) = \frac{1}{2} \left( \Phi(r, t) - \frac{\Psi(r_0, t)^2}{|\Psi(r_0, t)|^2} \Phi(r, t)^* \right),
\] (11)

which directly relates the two correction schemes.

Orthogonalization uses no information about the structure of spatial leakage beyond Eq. 7, but requires complex fields since otherwise \(\tilde{\Phi}(r, t)\) vanishes identically. It is driven by data and is applicable to any pair of signals. Equation 11 overcorrects spatial leakage from \(r_0\) at the expense of true interactions, since all instantaneous linear relations between \(\Psi(r, t)\) and \(\Psi(r_0, t)\) are suppressed. In particular, the reconstruction is affected even when no leakage is present, since then \(\Phi(r, t) = \Psi(r, t)\) but \(\tilde{\Phi}(r, t) \neq \Psi(r, t)\). Regression can be made slightly less conservative by adapting its hypotheses to the structure of spatial leakage. For example, under the assumption of a stationary inverse model, \(\partial \omega(r, t)/\partial t = 0\), a time-independent coefficient \(\gamma(r, r_0, t)\) was derived in [13] by minimizing the mean squared error \(\langle |\tilde{\Phi}(r, \cdot)|^2 \rangle\), where brackets denote time averaging. In any case, regression always requires linearity of the resolving operator, as for Eqs. 4, 5, and modifies \(\Psi(r, t)\) even when spatial leakage is negligible, contrary to the structural model correction.

To further explore the limitations of orthogonalization, we shall consider some explicit connectivity indices and show that it can non-trivially bias their estimation.

Brain network analyses are often based upon cerebral rhythms and concentrate on phase-locking and amplitude co-modulation [19], which appear more generally as long-range communication mechanisms within distributed oscillatory networks in the mean field approximation. We now study phase and amplitude couplings between \(\Psi(r_0, t)\) at a given \(r_0\) and \(\Phi(r, t)\) at \(r \neq r_0\), to those obtained between \(\Psi(r_0, t)\) and \(\tilde{\Phi}(r, t)\). To that aim, it is useful to rewrite Eq. 11 geometrically in terms of phases and amplitudes:

\[
\tilde{\theta}(r, r_0, t) = \frac{\pi}{2} \times \text{sign}[\sin \theta(r, r_0, t)],
\]

\[
|\tilde{\Phi}(r, t)| = |\Phi(r, t)| \times |\sin \theta(r, r_0, t)|,
\] (13)

where \(\theta(r, r_0, t)\) denotes the phase lag between \(\Phi(r, t)\) and \(\Psi(r_0, t)\).

Let us first focus on phase connectivity and consider two widely used coupling measures between \(\Phi(r, t)\) and \(\Psi(r_0, t)\): cross-density \((\Phi(r, \cdot) \Psi(r_0, \cdot))^*\) and phase coherence (exp \(i \theta(r, r_0, \cdot)\)). The effect of orthogonalization on cross-density is directly obtained using Eq. 11,

\[
(\Phi(r, \cdot) \Psi(r_0, \cdot))^* = i \text{Im} [(\Phi(r, \cdot) \Psi(r_0, \cdot))^*].
\] (14)

Orthogonalization removes the real part of cross-density, and thus generally underestimates its modulus compared to our new method. Likewise, it forces phase coherence to be purely imaginary. More precisely, Eq. 12 yields

\[
\langle \exp [i \theta(r, r_0, \cdot)] \rangle = i (f_+ - f_-),
\] (15)

where \(f_+\) and \(f_-\) denote the fraction of time spent by \(\theta(r, r_0, t)\) in the ranges \([0, \pi]\) and \([\pi, 2\pi]\), respectively. Orthogonalized phase coherence thus depends in a coarse way on the phase relations obtained with the structural correction, and is generally biased. For example, the phase-locking \(|\langle \exp [i \theta(r, r_0, \cdot)] \rangle| \approx 1\) occurring when \(\theta(r, r_0, t)\) narrowly varies around 0 or \(\pi\) symmetrically \((f_+ = f_-)\) is suppressed after orthogonalization, \(|\langle \exp [i \theta(r, r_0, \cdot)] \rangle| = 0\). Interestingly, the right-hand sides of Eqs. 14 and 15 respectively determine imaginary coherence [10] and phase lag index [11, 12]. Notably, the phase lag index \([f_+ - f_-]\) was introduced in [11] on heuristic grounds and conjectured to be insensitive to instantaneous linear couplings. The proof follows here as consequence of our derivation from orthogonalization.

The case of amplitude connectivity is more intricate because it is affected by phase relations, as follows from Eq. 13. In particular, the existence of an interaction between phase lag \(\theta(r, r_0, t)\) and amplitude \(|\Psi(r_0, t)|\) introduces spurious coupling between \(|\Phi(r, t)|\) and \(|\Psi(r_0, t)|\), which may lead to overestimation. This effect is an artifact of orthogonalization and is independent of spatial leakage itself. To illustrate this, let us compare the temporal correlation \(\rho(r, r_0)\) between \(|\Phi(r, t)|\) and \(|\Psi(r_0, t)|\) to its orthogonalized equivalent \(\tilde{\rho}(r, r_0)\) between \(|\tilde{\Phi}(r, t)|\) and \(|\Psi(r_0, t)|\). We first assume independence of amplitudes and phase lag, so that the expectation values \(|\langle \sin \theta(r, r_0, \cdot) \rangle|^k |\Phi(r, \cdot)|^l |\Psi(r_0, \cdot)|^m \rangle\) factorize into \(|\langle \sin \theta(r, r_0, \cdot) \rangle|^k |\langle \Phi(r, \cdot) \rangle|^l |\langle \Psi(r_0, \cdot) \rangle|^m\rangle\) for integers \(k, l, m \geq 0\). Using the definition of the Pearson correlation coefficient, it is direct to show that

\[
\tilde{\rho}(r, r_0) = \frac{M[\sin \theta(r, r_0, \cdot)\]}{\sqrt{1 + M[\sin \theta(r, r_0, \cdot)]^2 + M[\Phi(r, \cdot)]^2}},
\] (16)
where $M[x]$ denotes the mean of $|x(t)|$ divided by its standard deviation. This implies that amplitude correlation is underestimated, $\hat{\rho}(r, r_0) \leq \rho(r, r_0)$, with equality in the limit of perfect phase-locking where $\theta(r, r_0, t)$ converges to a constant $\neq 0$ or $\pi$, i.e., $M[\sin \theta(r, r_0, t)] \to \infty$. However, this bound can be violated in the presence of coupling between amplitudes and phase lag. This situation was investigated in [14] using a model of coherent sources, which involves amplitudes $a_0(t)$ and $a_1(t)$ with identical Rayleigh distribution, and phases $\phi_0(t)$ and $\phi_1(t)$ uniformly distributed on the circle, all four signals being independent of each other. This model assumes that $\Psi(r_0, t) = a_0(t) \exp(i\phi_0(t))$ and

$$\Phi(r, t) = c a_0(t) e^{i(\phi_0(t)+\psi)} + \sqrt{1 - c^2} a_1(t) e^{i\phi_1(t)},$$

(17)

The parameter $0 \leq c \leq 1$ quantifies the strength of the linear relation between $\Psi(r_0, t)$ and $\Phi(r, t)$, and $\psi$ controls their mean phase lag. The relation between $c$, $\psi$, $\rho(r, r_0)$ and $\hat{\rho}(r, r_0)$ was studied numerically in [14], but can be derived analytically using an appropriate approximation scheme. In [15], we show that

$$\rho(r, r_0) \approx c^2,$$

(18)

$$\hat{\rho}(r, r_0) \approx \frac{c^2}{\sqrt{c^4 + 2c^2 + 1/2}},$$

(19)

where $\hat{c} = c \sin \psi/\sqrt{1 - c^2}$. Eliminating $c$ from Eq. 19 exhibits the non-linear dependence of orthogonalized amplitude correlation $\hat{\rho}(r, r_0)$ on $\rho(r, r_0)$ and phase lag $\psi$. Let us notice that, depending on the values of $c$ and $\psi$, orthogonalization may underestimate (e.g., for $\psi = 0$) or overestimate (e.g., for $\psi = \pi/2$) amplitude correlation. In the latter case, the ratio $\hat{\rho}(r, r_0)/\rho(r, r_0)$ always lies between 1 and $\sqrt{2}$, showing the existence of spurious amplitude coupling. These conclusions based on approximations qualitatively agree with numerical simulations [15]. Let us also note that the overestimation of amplitude correlation may in general be much more dramatic than in the above model. For example, the case where $\theta(r, r_0, t)$ is given by $\arcsin(\kappa |\Psi(r_0, t)/\Phi(r, t)|)$ with positive $\kappa$ implies $\hat{\rho}(r, r_0) = 1$ independently of $\rho(r, r_0) \leq 1$, showing that $\hat{\rho}(r, r_0)/\rho(r, r_0)$ is unbounded.

Orthogonalization in its full generality has been applied with some success to investigations of brain connectivity through electrophysiology [10–14]. However, our analysis highlights some of its shortcomings, especially the counter-intuitive introduction of spurious amplitude coupling. In this Letter, we introduced a new method specifically correcting for spatial leakage and based on a direct characterization of its structure rather than on generic regression arguments. Although this method has some limitations which can be clearly identified, it offers a simple, general and principled way to analyze the large-scale network organization of spatially extended systems. This theory will be directly relevant to investigations of electrophysiological brain networks taking place at the level of reconstructed sources. More generally, it should find applications in other scientific fields as well, since it applies to any real-world complex system for which measurements are indirect or incomplete.

I would like to thank X. De Tiége and N. Goldman for inspiring discussions and their valuable comments on this manuscript. This work was supported by a research grant from the belgian Fonds de la Recherche Scientifique (convention 3.4611.08, FRS-F.N.R.S.).
Appendix A: Examples of inverse models

We showed in the main text that the kernel \( \ell(r) \) of the direct operator is a prime ingredient controlling the spatial structure of the minimum \( \mathcal{L}_p \)-norm inverse model. In this Appendix, we extend our discussion and indicate how reconstruction priors of the inverse model may affect the structure of spatial leakage. We also consider a fundamentally different type of inverse model based on spatial filtering. In the following, the direct operator \( L \) is always assumed to be given by Eq. 1.

**Minimum \( \mathcal{L}_p \)-norm estimates.** The prototypical example of inverse models consists in optimizing the fit between observed data \( \mu(t) \) and prediction \( (L\Psi)(t) \), with constraints on \( \Psi(r,t) \) imposing prior model assumptions. We minimize the strictly convex functional \( V = V_{\text{mismatch}} + V_{\text{prior}} \) over real scalar fields \( \Psi(r,t) \), where

\[
V_{\text{mismatch}} = \frac{1}{2} \int dt \| \mu(t) - (L\Psi)(t) \|^2, \quad (S1)
\]

\[
V_{\text{prior}} = \frac{\lambda}{p} \int dt d^3r |\Psi(r,t)|^p, \quad (S2)
\]

with \( |x|^2 = x^T x \) and \( \lambda > 0, p > 1 \) [6]. The solution is unique, and, when expressed as a functional of data, \( \Psi(r,t) = (W_r \mu)(t) \), defines the inverse operator \( W_r \). The extremization equation \( \delta V / \delta \Psi(r,t) = 0 \) reads

\[
\ell(r)^T (\mu(t) - (L\Psi)(t)) = \lambda |\Psi(r,t)|^{p-2} \Psi(r,t). \quad (S3)
\]

Taking the absolute value and eliminating \( |\Psi(r,t)| \),

\[
\Psi(r,t) = \lambda^{\alpha_p-1} \times \frac{\ell(r)^T (\mu(t) - (L\Psi)(t))}{|\ell(r)^T (\mu(t) - (L\Psi)(t))|^{\alpha_p}}, \quad (S4)
\]

where \( \alpha_p = (p-2)/(p-1) < 1 \). Equation 2 is recovered by setting \( \lambda = 1 \). Note that this derivation does not directly apply to the important case \( p = 1 \) where \( V_{\text{prior}} \) is not differentiable [6].

The qualitative spatial properties of \( W_r \), described in the main text are unaffected by certain model parameters but are by others. For example, changing Eq. S1, which embodies the physical assumption that measurement noise \( \varepsilon(t) = \mu(t) - (L\Psi)(t) \) is a gaussian white process, into \( V_{\text{mismatch}} = \frac{1}{2} \int dt dt' \varepsilon(t)^T M(t,t') \varepsilon(t') \), which allows to incorporate a more realistic description of noise, replaces \( \ell(r)^T \varepsilon(t) \) by \( \ell(r)^T \int dt' M(t,t') \varepsilon(t') \) in Eq. S4.

This leaves the structure of spatial leakage, i.e. data-independent spatial properties of the reconstructed field, qualitatively unmodified. On the other hand, changing the \( \mathcal{L}_p \)-norm affects spatial leakage. Indeed, using \( V_{\text{prior}} = \frac{\lambda}{p} \int dt d^3r \omega(r) |\Psi(r,t)|^p \), where \( \omega(r) > 0 \) is a weight function, instead of Eq. S2, incorporates an extra factor \( \omega(r)^{\alpha_p-1} \) into the right-hand side of Eq. S4. This can strongly modify the structure of spatial leakage. For example, a discontinuity in the weight function forces the reconstruction to be discontinuous too, even when \( \ell(r) \) is smooth.

**The linear case.** The situation simplifies in the widely used case \( p = 2 \), for which Eq. S4 is linear since \( \alpha_2 = 0 \). The inverse operator now takes the form

\[
W_r = \ell(r)^T A, \quad (S5)
\]

where \( A \) is an \( r \)-independent operator. In this case, \( W_r \) is seen to be at least as smooth as \( \ell(r) \), while mere continuity holds in general. Interestingly, Eq. S5 actually follows from the unitary symmetry of \( V_{\text{prior}} \) with \( p = 2 \), as we show in Appendix B. The qualitative structure of spatial leakage for the minimum \( \mathcal{L}_2 \)-norm estimate thus applies to all inverse models possessing this symmetry, may they be linear or non-linear. Only \( A \) depends on the details of the model. An explicit expression for this operator can not be obtained in general, unlike for the minimum \( \mathcal{L}_2 \)-norm model where [6]

\[
(A\mu)(t) = \left( \lambda + \int d^3r \ell(r) \ell(r)^T \right)^{-1} \mu(t). \quad (S6)
\]

This result can be derived by rearranging Eq. S4 with \( \alpha_p = 0 \) as

\[
\int d^3r' a(r,r') \Psi(r',t) = \ell(r)^T \mu(t), \quad (S7)
\]

where \( a(r,r') = \lambda \delta(r-r') + \ell(r)^T (r'') \ell(r') \). The linear operator on the left-hand side can be inverted using an expansion for large \( \lambda \),

\[
\Psi(r,t) = \int d^3r' \frac{1}{\lambda} \left[ \delta(r-r') - \frac{1}{\lambda} \ell(r)^T \ell(r') + \frac{1}{\lambda^2} \int d^3r'' \ell(r'')^T \ell(r'') \ell(r') - \ldots \right] \ell(r')^T \mu(t)
\]

\[
= \frac{1}{\lambda} \ell(r)^T \left[ 1 - \frac{1}{\lambda} \int d^3r' \ell(r')^T \ell(r')^T + \frac{1}{\lambda^2} \left( \int d^3r' \ell(r') \ell(r')^T \right)^2 - \ldots \right] \mu(t), \quad (S8)
\]
which leads to Eqs. S5, S6 after re-summation of this geometric series and analytic continuation in $\lambda$.

**Spatial filtering.** A widely used alternative to the above constrained least-squares criterion is the beamformer [16]. It relies on fundamentally different ideas, yet the discussion on spatial leakage remains similar. The inverse operator is defined by a linear spatial filter

$$ (W_r \mu)(t) = w(r)^T \mu(t) $$

(S9)

that focuses on the signal originating from location $r$. In the *linearly constrained minimum variance* beamformer [16], this filtering is achieved by imposing the unit-gain constraint $w(r)^T \ell(r) = 1$, which ensures that field activity at $r$ passes through the filter undeformed, and by minimizing the mean square $(W_r \mu)^2$ to dampen the contribution from elsewhere. Introducing the data covariance matrix $\Sigma = (\mu \mu^\dagger)$ and a Lagrange multiplier $\lambda(r)$ enforcing the constraint, this problem amounts to minimize the function

$$ w(r)^T \Sigma w(r) + \lambda(r) (w(r)^T \ell(r) - 1) $$

(S10)

over $w(r) \in \mathbb{R}^m$. A standard calculation yields [16]

$$ w(r) = \Sigma^{-1} \ell(r) $$

(S11)

The analytical dependence of the inverse operator in $\ell(r)$ is qualitatively similar to that of Eq. S4 in the formal case $p = 0$, and the discussion of spatial leakage properties made in the main text holds unchanged, except that the spatial filter $w(r)$ is at least as smooth as the kernel $\ell(r)$.

These comments also apply to some generalizations of the beamformer because the qualitative structure of spatial leakage in Eq. S11 does not depend on the data covariance matrix $\Sigma$. For example, the non-stationary beamformer of [17] uses a time-dependent data covariance $\Sigma(t)$, making the spatial filter weights given by Eq. S11 vary in time but with the same qualitative dependence in $\ell(r)$. They also apply to frequency-domain colored beamformers where $\Sigma$ is replaced by the cross-density matrix $\Sigma(\nu) = (\mu(\nu) \mu(\nu)^\dagger)$ of the data Fourier transform $\mu(\nu)$ at frequency $\nu$ [18]. It is noteworthy that in this example, the orthogonalization method of [14] can not be applied, contrary to the new structural leakage correction scheme presented in the main text. Indeed, the structural model is analogous to Eqs. 7, 8 but uses the complex coefficient

$$ \gamma(r, r_0, \nu) = \frac{\ell(r)^T \Sigma^{-1}(\nu) \ell(r_0)}{\ell(r)^T \Sigma^{-1}(\nu) \ell(r)} $$

(S12)

obtained using Eqs. 8 and S11 with $\Sigma$ replaced by $\Sigma(\nu)$. Orthogonalization must therefore be adapted to this case by computing the complex-valued regression coefficient, which yields a trivial corrected field $\hat{\Phi}(r, \nu) = 0$.

**Appendix B: Unitary symmetry and spatial structure**

In this Appendix, we show that Eq. S5 derives from the unitary invariance of the $L_2$-norm. More generally, we investigate the consequences of unitary symmetry on the analytical dependence of inverse operators in $\ell(r)$. To the best of the author’s knowledge, this question has not been considered in the literature. We first explain what is meant by unitary symmetry and then present two elementary derivations of Eq. S5.

**Unitary symmetry.** It is convenient to introduce the standard inner product

$$ (\Psi_1 | \Psi_2) = \int d^3 r \Psi_1(r) \Psi_2(r), $$

(S13)

which allows to write a compact expression for the direct operator defined by Eq. 1,

$$ L\Psi = (\ell|\Psi). $$

(S14)

It is then quite natural to consider unitary transformations $\Psi \rightarrow U\Psi$ acting on fields $\Psi(r)$ in the fundamental representation

$$ (U\Psi)(r) = \int d^3 r' U(r, r') \Psi(r') $$

(S15)

and leaving the inner product invariant,

$$ (U\Psi_1 | U\Psi_2) = (\Psi_1 | \Psi_2). $$

(S16)

Since we consider here real-valued fields, this imposes the orthogonality constraint

$$ \int d^3 r U(r, r') U(r, r'') = \delta(r' - r''). $$

(S17)

The presence of the parameter $\ell(r)$ explicitly forbids unitary transformations to be symmetries of the direct and inverse problems. A standard trick is to promote $\ell(r)$ to a background field transforming in the same representation, hence restoring symmetry in the direct model since $(U\ell(U\Psi)) = (\ell|\Psi)$.

We now consider inverse models that are invariant under unitary transformations $\Psi \rightarrow U\Psi, \ell \rightarrow U\ell$. We also assume that, among its parameters, $\ell(r)$ is the only function of $r$.

**Extremizing invariant functionals.** Let us suppose that the reconstruction $\Psi(r, t) = (W_r \mu)(t)$ is defined as the extremum of some unitary invariant functional

$$ V = \mathcal{V}[m, s] $$

(S18)

in which $\Psi(r, t)$ and $\ell(r)$ appear through the invariants $m(t) = (\ell|\Psi(\cdot, t))$ and $s(t, t') = (\Psi(\cdot, t)|\Psi(\cdot, t'))$ only. The extremization equation $\delta V / \delta m(t) = 0$ reads

$$ \ell(r)^T \frac{\delta V}{\delta m(t)} + 2 \int dt' \frac{\delta V}{\delta s(t, t')} \Psi(r, t') = 0. $$

(S19)
Applying the inverse of the linear operator in the second term, if it exists, we obtain the implicit equation

$$\Psi(r, t) = \ell(r)^T B[t, m, s],$$  \hspace{1cm} (S20)

for some coefficients $B[t, m, s]$ depending on the derivatives of $V$, as well as on all other $r$-independent parameters such as data $\mu(t)$. Since this holds whatever $\mu(t)$, we obtain Eq. S5 with $(A\mu)(t) = B[t, m, s]$, which is independent of $r$, as claimed. This argument represents a direct generalization of the derivation of Eq. S4 for $p = 2$. It shows in particular that physical assumptions about measurement noise $\varepsilon(t) = \mu(t) - (L\Psi)(t)$ are irrelevant to this argument, since $V_{\text{minit}}$ in Eq. S1 can be replaced by any functional $\mu(t)$ and $(L\Psi)(t)$. Likewise, the gaussian white noise prior assumption on $\Psi(r, t)$ embodied in the $L_2$-norm term, Eq. S2 with $p = 2$, can be waived by using any functional of $s(t, t')$ for $V_{\text{prior}}$.

**Equivariance constraint.** A more explicit result can be obtained by directly solving the general constraints implied by unitary symmetry. The following argument is also slightly more general, in that it does not assume the inverse model to be defined as an extremization problem. The reconstruction is a functional

$$\Psi = F[\ell_1, \ldots, \ell_m]$$  \hspace{1cm} (S21)

of the $m$ components $\ell_k(r)$ of the kernel $\ell(r)$. Unitary symmetry imposes that, if $\Psi$ is the reconstruction for given functions $\ell_k$, then the reconstruction for the functions $U\ell_k$ must be $U\Psi$. Mathematically, this translates into the property that $F$ is unitary *equivariant*, i.e., for any linear operator $U$ satisfying Eqs. S15, S17, we have

$$F[U\ell_1, \ldots, U\ell_m] = UF[\ell_1, \ldots, \ell_m].$$  \hspace{1cm} (S22)

This assumption imposes strong constraints on the functional $F$, which in consequence must be of the form

$$F[\ell_1, \ldots, \ell_m] = \sum_{k=1}^m \ell_k C_k((\ell))^T),$$  \hspace{1cm} (S23)

as can be shown using classical theorems from invariant theory. Functions $C_k$ of the matrix $(\ell)^T$ of inner products $(\ell_j | \ell_k)$ are the most general unitary invariant functions of the $\ell_k$, whereas the factors $\ell_k$ are necessary for this sum to transform in the fundamental representation. All $r$-independent model parameters, including time $t$ and data $\mu(t)$, enter Eq. S23 through the $C_k$s only. We thus recover Eq. S5 again, with the $k^\text{th}$ component of $(A\mu)(t)$ given by $C_k$. The operator $A$ is indeed independent of $r$, since $\ell(r)$ only appears via the inner products $(\ell)^T$, as was explicit in Eq. S6.

Let us notice that the above arguments can be generalized to the case where other parameters transform non-trivially, by promoting them to background fields. Direct models involving other inner products and isometry groups or subgroups can also be considered along these lines.

**Appendix C: Derivations for amplitude correlation between coherent sources**

The effect of phase coupling on orthogonalized amplitude correlation was investigated in [14] using simulations of the coherent sources model presented in the main text. We claimed that Eqs. 18, 19 yield a good approximation of their results. In this Appendix, we make the approximation scheme used explicit and then derive these equations using elementary calculations. We also check the validity of this approximation by reproducing simulations of [14]. All probabilities considered here are based on ergodic densities, i.e., temporal averaging.

Let us recall that $a_0(t)$ and $a_1(t)$ are two positive signals with Rayleigh distribution, i.e., their fraction of time spent between values $a$ and $a + da$ is $-da \exp(-a^2/2\sigma^2)$, $\sigma > 0$ (Fig. S1), that $\phi_0(t)$ and $\phi_1(t)$ are uniformly distributed over $[0, 2\pi)$, and that all four variables are independent of each other. These assumptions imply in particular that $a_k(t) \exp i\phi_k(t)$ is distributed according to a complex gaussian with zero mean and variance $\sigma^2$.

Our goal is to compute the correlations

$$\rho(r, r_0) = \text{corr}[[\Phi(r, \cdot), [\Psi(r_0, \cdot)]],$$  \hspace{1cm} (S24)

between the amplitudes of $\Psi(r_0, t) = a_0(t) \exp i\phi_0(t)$ and $\Phi(r, t)$ given by Eq. 17 or $\tilde{\Phi}(r, t)$ given by Eq. 11.

**The approximation.** Computing these correlations as functions of $c$ and $\psi$ is a priori difficult. However, the following approximation scheme can be used. All amplitudes having finite variance, e.g., $\sigma^2 = (2 - \pi/2)\alpha^2$ for the Rayleigh distribution, they are effectively supported on a region where the amplitude squared can be reasonably approximated by a linear function, as exemplified in Fig. S1. Since correlation is invariant under linear transformations, this yields

$$\rho(r, r_0) \approx \text{corr}[[|\Phi(r, \cdot)|^2, |\Psi(r_0, \cdot)|^2],$$  \hspace{1cm} (S26)

between the magnitudes of $\Psi(r_0, t) = a_0(t) \exp i\phi_0(t)$ and $\Phi(r, t)$ given by Eq. 17 or $\tilde{\Phi}(r, t)$ given by Eq. 11.

**Derivation of Eq. 18.** It will prove useful to observe that $\Phi(r, t)$ is gaussian itself with zero mean, since it is given by a linear combination of $a_k(t) \exp i\phi_k(t)$, see Eq. 17. Using

$$|\Phi(r, t)| = \sqrt{1 - \alpha^2} a_1(t) e^{i(\phi_1(t) - \phi_0(t) - \psi)},$$  \hspace{1cm} (S28)

$$\langle a_0^2 \rangle = \langle a_1^2 \rangle$$ and $\langle \exp i(\phi_1 - \phi_0) \rangle = 0$, we also see that

$$\langle |\Phi(r, \cdot)|^2 \rangle = \langle a_1^2 \rangle,$$  \hspace{1cm} (S29)

We conclude that $\Phi(r, t)$ actually has same variance, and thus same distribution, as $a_1(t) \exp i\phi_0(t)$. In particular,

$$\langle |\Phi(r, \cdot)|^2 \rangle = \langle a_0^2 \rangle = \langle a_1^2 \rangle,$$  \hspace{1cm} (S30)
FIG. S1. Linear approximation of squared amplitude. The Rayleigh distribution of $a(t)$ with maximum reached at value $\alpha = 1$ is shown together with the function $a^2$. The Taylor expansion $a^2 \approx 2a_0 - \alpha^2$ around $a = \alpha$ is seen to be a good approximation in the $\alpha \pm \sigma$ interval (green shaded region), which contains about 80% of samples. Over a larger interval, linear regression still yields a reasonable approximation.

for $n \geq 0$. Now, using Eq. S28, the mutual independence between $a_0(t)$, $a_1(t)$ and $\phi_1(t) - \phi_0(t)$, and the identities $\langle a_0^2 \rangle = \langle a_1^2 \rangle$ and $\langle \cos(\phi_1 - \phi_0 - \psi) \rangle = 0$, we find

$$\langle |\Phi(\cdot,r)|^2 a_0^2 \rangle = \langle a_0^2 \rangle \cdot c^2 \times \langle a_1^2 \rangle,$$

(S31)

where $\sigma^2$ denotes the standard deviation of $a_0(t)^2$. Since $|\Psi(r_0,t)| = a_0(t)$ and $|\Phi(r,t)|$ have the same moments according to Eq. S30, this shows that the right-hand side of Eq. S26 equals $c^2$. We thus recover Eq. 18.

Let us notice that the independence in $\psi$ actually holds exactly. This is seen from Eq. S28, which shows that $\phi_0(t)$, $\phi_1(t)$ and $\psi$ only appear as $\phi_1(t) - \phi_0(t) - \psi$, and from the trivial property

$$\int d\phi_0 \, d\phi_1 \, f(\phi_1 - \phi_0 - \psi) = \int d\phi_0 \, d\phi_1 \, f(\phi_1 - \phi_0),$$

(S32)

for any function $f$ on the circle. To check the validity of our approximation for $\rho(r,r_0)$ as a function of $c$, results obtained from Eq. 18 and from direct simulations are plotted in Fig. S2(a). Comparison shows fair agreement and thus indicates that the approximation in Eq. S26 is quite good. Actually, fitting a power law $\rho(r,r_0) = c^k$ on the simulation curve using a log-log linear regression yields $k \approx 2.1$.

**Derivation of Eq. 19.** Let us now focus on Eq. S27. The orthogonalized field $\tilde{\Phi}(r,t)$ does not follow the same distribution $\Psi(r_0,t)$ and $\Phi(r,t)$. This is seen from its explicit expression

$$\tilde{\Phi}(r,t) = e^{i\phi_0(t)} \times \left[ c \sin \psi a_0(t) + \sqrt{1 - c^2} a_1(t) \sin(\phi_1(t) - \phi_0(t)) \right],$$

(S33)

which is obtained by direct application of Eqs. 11 and 17. Since the rescaling $|\Phi(r,t)| \to |\Phi(r,t)|/\sqrt{1 - c^2}$ leaves the correlation in Eq. S25 unchanged, and since $|\tilde{\Phi}(r,t)|/\sqrt{1 - c^2}$ depends on $c$ and $\psi$ through the combination $\tilde{c} = c \sin \psi/\sqrt{1 - c^2}$, we conclude that $\tilde{\rho}(r,r_0)$ must be a function of $\tilde{c}$ only. Its approximate dependence in $\tilde{c}$ can be obtained by direct computation of the right-hand side of Eq. S27. Calculations are similar to those presented above, although they are slightly more cumbersome. We obtain

$$\langle |\tilde{\Phi}(\cdot,r)|^2 a_0^2 \rangle - \langle |\tilde{\Phi}(\cdot,r)|^2 \rangle \langle a_0^2 \rangle = c^2 \sin^2 \psi \sigma^2 \quad (S34)$$

for the covariance between squared amplitudes $a_0(t)^2$ and $|\tilde{\Phi}(r,t)|^2$. Correlation is then obtained upon division by their standard deviations $\sigma_a^2$ and $\sigma_{\tilde{\Phi}(r,t)}^2$. For the latter, we find

$$\sigma_{|\tilde{\Phi}(r,t)|^2}^2 = (1 - c^2) \times \left[ (c^4 + 3/8) \sigma^2 + (2c^2 + 1/8) \langle a_0^2 \rangle \right].$$

(S35)

Plugged into Eq. S27 and using the definition of $\tilde{c}$, these results yield

$$\tilde{\rho}(r,r_0) \approx \frac{c^2}{\sqrt{c^4 + 3/8 + (2c^2 + 1/8)M[a_0^2]}},$$

(S36)

where $M[x]$ was defined in the main text after Eq. 16. Here, we have $M[a_0^2] = \langle a_0^2 \rangle / \sigma^2 = 1$ for the Rayleigh distribution, from which Eq. 19 finally follows.

This analytical result is compared to the simulation in Fig. S2(b), and good agreement is again observed. In particular, the fact that $\tilde{\rho}(r,r_0) \geq c \geq \rho(r,r_0)$ when $\psi = \pi/2$, whose interpretation is given in the main text, is seen to hold both for Eq. 19 and for the simulation.

FIG. S2. Analytical results versus simulations. Comparison of Eqs. 18 (a) and 19 (b) to simulations, in which the parameters $0 \leq c \leq 1$ and $0 \leq \psi \leq \pi/2$ were varied, show very good consistency.