What and *where* in electromagnetic brain imaging

Alain de Cheveigné, Centre National de la Recherche Scientifique and Ecole normale supérieure.

Abstract

To understand the brain, we need to observe both the nature and dynamics of its activity (the "what"), and the location or distribution of its sources (the "where"). This paper proposes a new approach based on standard data-driven linear analysis, in which these two elements are derived in parallel from separate columns of the analysis matrix. A subset of columns enhances the activity of interest, and its complement defines spatial filters that suppress that activity (null filters). Each null filter is combined with an anatomy-dependent source model to estimate a subset of the sources space (zero set) in which the source might reside, and the final estimate is derived from the intersection of zero sets. There is no theoretical limit to the accuracy with which the location of a source can be estimated, but practical limits may arise from noise in the data, imperfect calibration, or an incomplete or inaccurate source model. The method is illustrated with simulated and real data from EEG, MEG and SEEG.

Introduction

In Plato's allegory of the cave, prisoners inferred the world from shadows shifting on a wall. According to Helmholtz's doctrine of unconscious inference, perception proceeds by inference from incomplete sensory evidence (Helmholtz 1967; Alhacen 1030; Hatfield 2002; Imbert 2020; Kawato 1993; Friston 2002). A similar idea applies also to the analysis of electromagnetic (EM) brain data, because the sparsely-sampled measurements from techniques such as electroencephalography (EEG), magnetoencephalography (MEG), or invasive electrode arrays, do not suffice to characterize the full activity of the brain (Baillet 2001, 2010). At best, we can infer that activity by combining sparse observations with priors,

and the question is how best to do so. In Plato's cave, the shadows offered weak clues as to *what* objects were, and strong cues as to *where* they were. Here, I transpose this what/where distinction (curiously observed also among pathways within the brain, Schneider 1969) to the problem of locating brain activity.

In electromagnetic (EM) brain imaging, observations (MEG, EEG, etc.) are combined with source and forward models to infer the dynamics of interesting processes in the brain, and their spatial distribution. Three main approaches to localization are*: dipole fitting, beamforming*, and *minimum current norm estimate* (MNE), and here I suggest a fourth, *zero-set imaging*. All approaches require a source model that constrains the nature and spatial characteristics of brain sources, and a forward model that relates them to observations at the sensor or electrode array. Source and forward models are informed by anatomical measurements (e.g. magnetic resonance imaging, MRI), priors (e.g. regions of interest, ROI), conductance information, and knowledge of sensor/electrode geometry. Sophisticated software tools are available to create such source and forward models and apply them to source localization and imaging (Oostenveld 2011; Tadel 2011; Gramfort et al 2014).

The principle

EM observations are related to brain activity via a mixing process that is assumed to be instantaneous and linear:

$$\mathbf{X} = \mathbf{SM} \qquad (1)$$

where matrices $\mathbf{X}$ and $\mathbf{S}$ represent the time series of observations and source activity respectively ($\mathbf{S}$ includes both brain and artifact or noise sources). There are I sources and J observation channels, with $I \gg J$. The mixing matrix $\mathbf{M}$ of size $I \times J$ is unknown. A useful viewpoint is that brain sources span a *vector space* of dimension I, containing all their linear combinations, and that observations span a *subspace* that space, of dimension J, upon which brain sources are projected. Multiple sources thus map to the same observation, explaining the indeterminacy of the inverse problem. Linear analysis involves applying a matrix to the data:

$$\mathbf{Y} = \mathbf{XU} \qquad (2)$$

where $\mathbf{Y}$ is a matrix of time series ("components") that hopefully reflect the brain activity in some useful fashion. For example, one column $\mathbf{y}_k$ might approximate the time course of a process or source of interest. In that case, the corresponding column $\mathbf{u}_k$ of the analysis matrix can be seen as a *spatial filter* (set of J coefficients) that enhances that source. In some situations, a subset of N columns of $\mathbf{U}$ may be useful to enhance a group of N sources collectively.

Several data-driven methods exist to find such a matrix $\mathbf{U}$. Among them are *principal component analysis* (PCA), *independent component analysis* (ICA), *canonical correlation analysis* (CCA), *joint decorrelation* (JD) (de Cheveigné and Parra 2014), and others. Usually, the objective of the analysis is to observe the dynamics of a particular source of interest, using the column of the analysis matrix that isolates that source. Here, however, the focus is on the *other* columns of the analysis matrix, which form spatial filters that cancel the source, called *null filters* for that source.

Suppose that we have a *source model* to which we can map our source of interest, and a *forward model* that defines the source-to-channel gains $\mathbf{g}_{\tilde{\imath}}$ for each possible source location $\tilde{\imath}$. Multiplying the gain vector of a source by the null filter for that source yields zero, by definition, and there are usually other locations for which the product is zero, which collectively form the *zero set* of this filter. Given a null filter for a source i, the source is obviously located within its zero set and, if several such filters are available, within their *intersection*. This is the core idea of the zero-set method. To find this intersection, it is convenient to add the gain-filter products over filters $\mathbf{u}_k$ to obtain a cost function:

$$o(\tilde{\imath}) = \sum_k (\mathbf{g}_{\tilde{\imath}} \cdot \mathbf{u}_k)^2 / b_{\tilde{\imath}} \qquad (3)$$

that is zero when the index $\tilde{\imath}$ matches the source i. The location-dependent normalization factor $b_{\tilde{\imath}}$ is designed to counter systematic variations of gain over the source space (SI).

Localization thus derives from a combination of *data observations* encapsulated within the null filters, and *prior knowledge* encapsulated within the source model. Related to the source model, a rendering model may be used to translate the location within source parameter

space into a realistic image. Factors that determine realism (e.g. smoothness of the distribution of activity over cortical tissue) are encapsulated within that model.

If the search space is small enough, the location of the zero may be found by grid search over the source model parameter space. For a larger space, methods such as the Nelder-Mead simplex algorithm may allow more efficient search. If the real source does not fit any index within the source space, Eq. (3) cannot be satisfied. The investigator must then either expand the source model, or else tentatively use the position of the *minimum* of the cost function as a cue to location. Conversely, prior knowledge or interest may be used to prune the source space (e.g. ROI).

The localization and imaging pipeline

Suppose that we wish to localize the source of, say, a stimulus-evoked brain response. We gather two sets of data: EM recordings within an appropriate paradigm involving stimulation (Fig. 1, blue), and anatomical and co-registration data to inform source and forward models (Fig. 1 red). Based on the recordings, a data-driven linear analysis method produces an analysis matrix $\mathbf{U}$, such that one column $\mathbf{u}_k$ isolates the time course of the source of interest, and the other columns $\mathbf{u}_{k\prime\neq k}$ serve as null filters to estimate its location. The location is found by searching within the source model parameter space for the intersection of the zero sets of those null filters, or for the zero of the associated cost function. As a final step, a rendering model can be applied to produce a satisfying image of the neural activity, for example as distributed activity (Fig. 1).

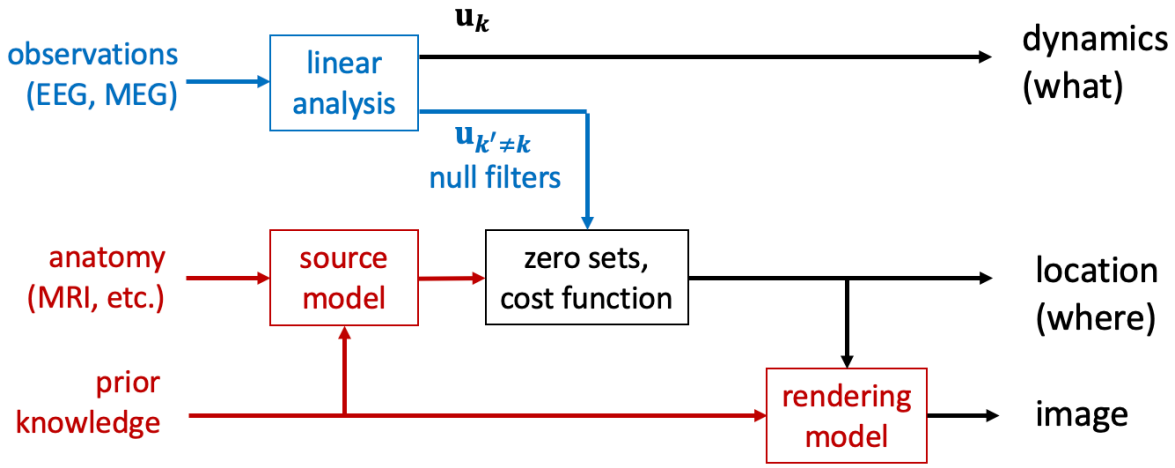

Fig. 1 Data analysis pipeline. Data-driven analysis of recordings (EEG, MEG, etc.), in blue, produces an analysis matrix $\mathbf{U}$ that is split into one column that isolates the dynamics of the source of interest, and remaining columns that serve as null filters for that source. In parallel, anatomical and sensor position data, in red, are acquired and combined with prior knowledge to produce a source/forward model. Combining both streams produces zero sets, the intersection of which indicates the location of the source within the parameter space of the source model. Optionally, a rendering model, also informed by anatomy and priors, produces a realistic image of the source.

A toy world

The reader might feel confused by the various "spaces" invoked so far: the *physical* space within the brain, the *signal* space of neural dynamics, and the *parameter* space of the source model. Here I introduce a simplified "toy world" that captures the essence of the full problem while allowing for intuitive visualization. Sources and sensors are both in a plane (the former within a disc, the latter on its periphery), and source-to-sensor gain varies as the inverse of the square of their distance, $1/d^2$ (Fig. 2 A). The planar source space is homeomorphic to the cortical surface, while the inverse quadratic dependency mimics Maxwell's equations. Real-world complexities such as three-dimensional location and orientation, anatomy, and inhomogeneous or anisotropic conduction, are conveniently ignored.

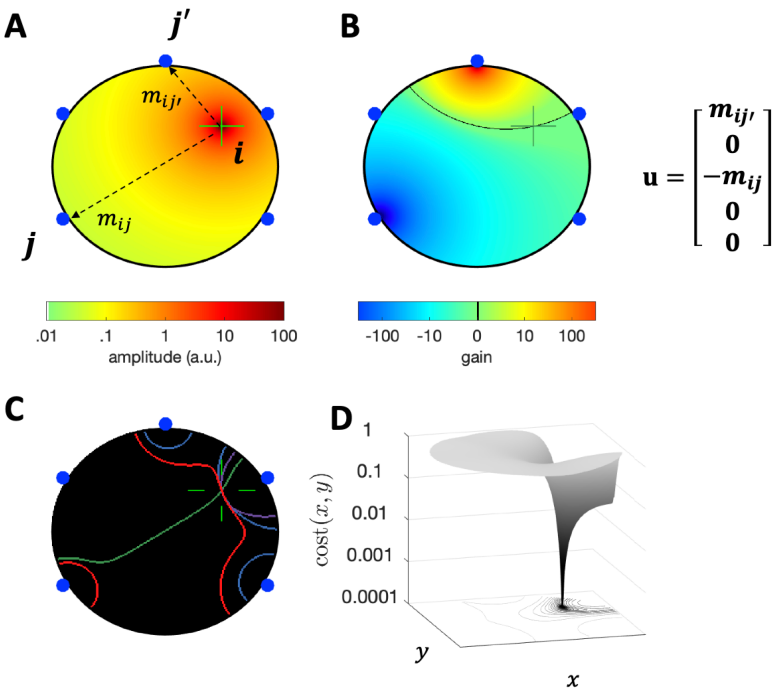

Fig. 2. "Toy world" consisting of an array of sensors (blue dots) placed at the periphery of a disc. A: A source within the disc (green cross) radiates within the plane with an amplitude that decreases with distance as $1/d^2$. B: Gain map for a linear combination of two sensors with weights chosen to cancel the source at the green cross (null filter). The weight vector (spatial filter) is shown on the right, with coefficients ordered clockwise from the lower right. The black line is the zero set of this filter. D: Zero sets of four null filters for this same source. The intersection of zero sets signals the source position (green cross). D: Cost function of position within the source space (Eq. 3). The global minimum (zero) indicates the source position.

A source (green cross) radiates within the disc as shown as a color scale (Fig. 2 A), impinging with a different amplitude on each sensor. Adding two sensors with appropriate weights produces a spatial filter with zero gain at the source position (a *null filter* for that source) (Fig. 2 B). The gain is also zero at other positions that together form the *zero set* of that filter (black line in Fig. 2 B). Null filters for a given source form a subspace of dimension $J - 1$ of the space of spatial filters (Supplementary Information, SI), and thus there may be multiple null filters for a source, with zero sets that intersect at the position of the source (Fig. 2 C). If the location of the source is unknown, it can be estimated from the intersection of their zero sets. Rather than dealing directly with zero sets, it is convenient to define a cost function as in Eq. 3, which is zero at the position of the source (Fig. 2 D).

Is the zero unique? The question is difficult to answer in full generality, but again the toy world can provide some insight. With J sensors and K null filters, the zero set for filter k consists of positions (x, y) such that:

$$\sum_j \frac{u_{jk}}{(x-x_j)^2+(y-y_j)^2} = 0. \qquad (4)$$

Multiplying each term by the denominators of the other terms to give all terms the same denominator, we have:

$$\sum_j u_{jk} \prod_{j' \neq j} \left[(x - x_{j'})^2 + (y - y_{j'})^2\right] = 0. \qquad (5)$$

This can be expanded into a multivariate polynomial equation in x, y of order $2(J-1)$. For two unknowns (x, y), we need at least two such equations (and thus at least two null filters). Bezout's theorem tells us that there may be up to $n = [2(J-1)]^2$ zeros for this set of equations (or $n = [d(J-1)]^d$ for a d-dimensional source space). Thus, there might indeed be multiple zeros of the cost function, albeit finite in number. Empirically, the zero is often unique (see SI for a more complete discussion), but it is worth keeping in mind the possibility that localization might sometimes be ambiguous.

Simulated MEG

Insights gained from the toy world need checking within more realistic configurations. A source model with 15002 dipoles, positioned on the cortical surface with orientations normal to that surface, was calculated using Brainstorm (Tadel et al. 2011), based on a data set from the Brainstorm tutorial (SI). A dipolar source position was chosen (Fig. 3 A), and a single-column matrix of Gaussian noise, representing simulated source activity, was multiplied by the corresponding gain vector to obtain 274-channel simulated MEG data. PCA was applied to these data, and columns 2-274 of the transform matrix were retained as null filters for the source. The cost function (Eq. 3), plotted in Fig. 3 B, shows a sharply-defined minimum at a position that matches that of the source, with a precision limited only by the spatial resolution of the source model. Other configurations and source models are explored in SI, as are the sensitivity of the method to noise in the data, jitter of the forward model coefficients, and mismatch between real and model sources. In a nutshell: the method appears tolerant to noise and jitter, but sensitive to a mismatch between real and model source, an issue that becomes apparent with real data.

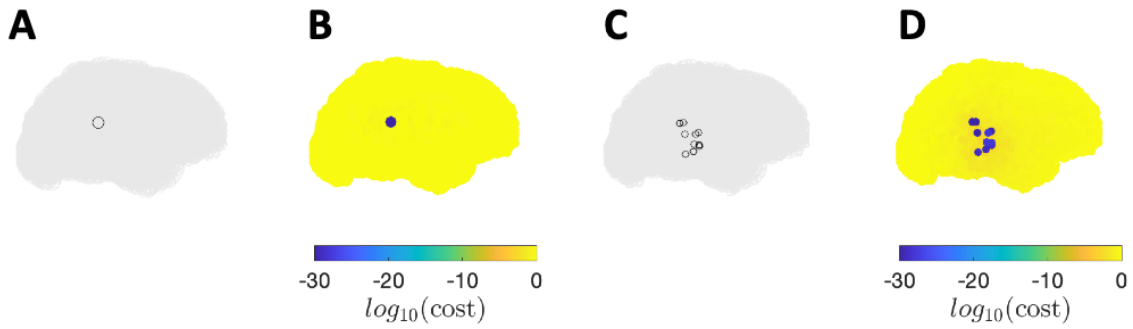

Figure 3. Localization of simulated MEG souces. A: Location of simulated source. B: Cost function, assuming a constrained orientation cortical surface model. C: Locations of 10 simulated sources with independent activities. D: Cost function.

An interesting case is *multiple sources* with independent activity. The time courses of $N = 10$ such sources (Fig. 3 C) were synthesized as a 10-column matrix of Gaussian-distributed values, which was multiplied by the corresponding columns of the model gain matrix to obtain 274-channel simulated MEG. Columns 1-10 of the PCA matrix capture all the variance, and thus columns 11-274 constitute null filters for all 10 sources. The cost function that is plotted in Fig. 3 D. Sharply-defined minima are visible at positions that match those of the sources, a result that holds even if the source waveforms are correlated. However, if the source waveforms are *linearly dependent*, in particular identical, localization breaks down (SI). The multiple-source case is interesting because it often happens that linear analysis isolates a *group* of components from background activity, but cannot distinguish between them.

Real MEG and EEG

The method was applied to 274-channel MEG data recorded in response to repeated auditory stimulation in an odd-ball detection task (publicly-available data used by the Brainstorm tutorial, SI). The JD algorithm was applied repeatedly: first to isolate stimulus-locked activity, second to isolate button-press-locked activity, and third to isolate narrowband activity near 10 Hz (alpha). In each case, null filters were chosen from columns of the transform matrix excluding the first N ($N = 1$ for auditory and button press, $N = 10$ for alpha). A cost function was calculated over locations of a volume source model (2 mm grid, arbitrary dipole orientation). For the auditory response, sensors and source locations were restricted to left or right hemispheres and analyzed separately.

Figure 4 A and B show the cost function for the left and right hemispheres respectively for the auditory response. Each pixel represents the minimum cost over a stack of voxels perpendicular to the viewing plane, thus deep sources are visible. Small values of the cost function are clustered in temporal regions of both hemispheres. Compared to the simulation (Fig. 3), the cost does not reach small values (note the range on the colorbar), implying that the real source does *not* fit the model of a single dipole. Here, the analysis uses one component of the JD analysis ($N = 1$), but the auditory evoked response extends over additional JD components. If several components are included ($N > 1$), the patterns are wider (SI).

Visually, these patterns are roughly consistent with those found with techniques such as dipole fitting, beamforming, and minimum norm estimate (SI). An important difference is that those methods estimate the *distribution of activity* over source space, whereas the plots in Fig. 4 reflect the *distribution of fit* between the real source and the source model as a function of position in its parameter space. This difference (fit vs activity) suggests that the present method may offer insights that are complementary to other techniques. Note that the model parameter space here also includes orientation, not visible in these plots.

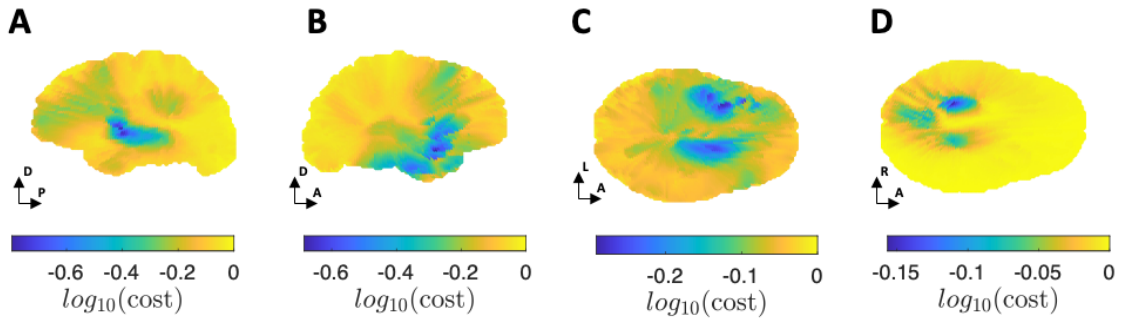

Fig. 4. Localization of sources in MEG. A: Cost function associated with the auditory response, left hemisphere. B: Same, right hemisphere. C: Cost function associated with the button press. D: Cost function associated with 10 Hz narrowband activity. Each pixel represents the minimum cost over a stack of voxels perpendicular to the viewing plane (minimum-take-all transparent brain).

Similarly, Fig. 4 C shows the cost function for the response preceding the button press. Figure 4 D shows the cost function associated with the first 10 components of the alpha response. Again, the relatively large values suggest that the cortical source is not a single dipole.

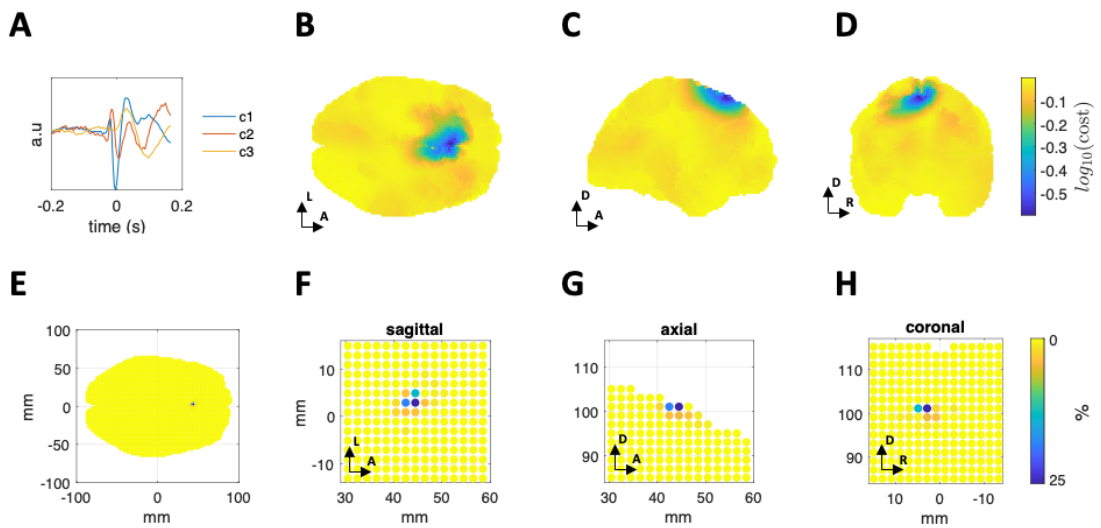

Fig. 5. Localization of interictal spikes in EEG. A: Waveforms of three components of a JD analysis time locked to interictal spikes. B-D: Cost function associated with the first (earliest) component, based on a volume source model on a 2mm grid, projected on sagittal, axial, and coronal planes (minimum-take-all). E-H: empirical probability distribution of the position of the minimum cost, based on bootstrap resampling (n=100).

The method was also applied to 29-channel EEG data recorded from an epileptic patient between seizures (Brainstorm tutorial). The data were analyzed with the JD method to isolate activity time locked to markers aligned with interictal spikes. Figure 5 A shows the time-averaged waveforms of the first three JD components, and Fig. 5 B-D shows the cost function associated with the first (earliest component), viewed in sagittal, axial and coronal planes. As in Fig. 4, each pixel represents the minimum over a stack of voxels (transparent brain). The cost function is smallest at a superficial position on the left margin of the longitudinal fissure, roughly consistent with results from dipole fitting, beamforming, and minimum norm estimation (SI). Figure 5 E-H shows the empirical distribution of the position of the minimum, based on a bootstrap resampling (n=1000). The distribution is tight: 95% of trials fall within

4 mm of its centroid, however in the absence of ground truth one cannot say if that location coincides with the true source. Accurate localization of interictal activity is clinically important. Another example based on SEEG is described in SI.

Caveats. The theory assumes a perfect fit between the real source, on one hand, and the source model for some value of its parameters on the other. Here, instead, we found no perfect match to the source model, so we are left to interpret the pattern of imperfect fit to the inexact model (Fig. 4). That pattern seems informative, and consistent with other methods, but the theoretical basis to interpret it is incomplete. Hopefully, progress in understanding underlying brain processes will lead to more complete models that allow a perfect fit, obviating the need to interpret the pattern of imperfect fit. The question is further explored in the SI.

Discussion

Zero-set imaging makes a clean distinction between *what* (source activity waveforms) and *where* (their locations), that are associated with distinct columns of the data-driven analysis matrix. There is also a clear separation between the *observations* (embodied in the data-derived null filters) and *priors* (embodied in the source model). The data-driven analysis that isolates a source provides, as a side-product, the null filters that play the role of "sufficient statistics" for the purpose of localization. Linear analysis methods (PCA, ICA, etc.) are central stage, in contrast to pipelines in which they play a secondary role (such as interference rejection).

Zero-set imaging bears a superficial resemblance to dipole fitting, in that a source model is scanned for a fit to the data. For dipole fitting, the fit is made directly to data values measured at the sensors, for zero-set imaging the fit is indirect after multiplication by the null filters. Zero-set imaging differs from beamforming in that localization follows, rather than precedes, source separation: source separation (spatial filtering) is performed by the location-agnostic linear analysis stage, rather than the data-informed beamformer solution. Zero-set imaging differs from minimum norm estimation in that the prior is entirely encapsulated in the source model, rather than enforced by a minimum norm constraint in the calculation of an inverse.

Beamforming and minimum norm require an estimate of a noise covariance matrix. For zero-set imaging that step is subsumed by the linear analysis stage, for example JD (de Cheveigné and Parra 2014).

An appeal of zero-set imaging is that it allows, in ideal situations, sources to be localized with *arbitrary resolution and accuracy*, limited only by the resolution and accuracy of the source and forward models. There is no equivalent of a "point-spread volume" (Barnes et al 2004; Hauk et al 2022; Jaiswal 2020). This holds if the source model is complete, so that a real source fits a model source for some value of its parameters, but unfortunately such was not the case for the real MEG and EEG data used for evaluation. At this point we lack principled insights as to how to interpret the pattern of imperfect fit to an incomplete model (e.g. Fig. 4). This is a major weakness of the approach. On a brighter note, progress in understanding the brain may lead to more complete models that can be tested using this approach, a zero value of the cost indicating a perfect fit of the model to the real source. This paper focused on MEG and EEG, but the same principles are applicable to other recording modalities such as electrode arrays, in scientific and possibly also industrial or clinical contexts (see SI for an example of localization of interictal spikes from EEG). These questions and applications remain to be investigated.

Returning to Plato's allegory, if the shadow were scaled and placed *between* the light source and the shadow-casting object, it would perfectly mask that object, an almost perfect metaphor for a null filter. Just as multiple shadows from multiple light sources provide multiple constraints on an object's volume in space, multiple null filters constrain the location of a brain source. Looking at an illuminated object tells us *what* that object is, and looking at its shadows tells us accurately *where* it is. A similar idea works also for the brain.

Acknowledgments

Previous drafts benefitted from comments of Israel Nelken, Lucas Parra, and Malcolm Slaney.

Code availability

https://github.com/adc2/WhatVSWhere

Supplementary Information

Available from the BioRxiv site. See this for important details.

References

al Haytham I. 1030 (2002) Book of optics (in Hatfield).

Baillet, S., Mosher, J.C., and Leahy, R.M. (2001) Electromagnetic brain mapping. IEEE Signal Processing Magazine, 18(6):14–30.

Baillet, S. (2010) The Dowser in the Fields: Searching for MEG Sources, in MEG, an Introduction to Methods, edited by Hansen, P.C., Kringelbach, M.L., Salmelin, R., Oxford University Press, 83-123.

Barnes, G. R., Hillebrand, A., Fawcett, I. P., & Singh, K. D. (2004). Realistic spatial sampling for MEG beamformer images. *Human Brain Mapping*, *23*(2), 120–127. https://doi.org/10.1002/hbm.20047

de Cheveigné, A., Parra, L. (2014) Joint decorrelation: a versatile tool for multichannel data analysis. *NeuroImage*, 98, 487-505, doi: 10.1016/j.neuroimage.2014.05.068.

de Cheveigné, A. (2020) ZapLine: A simple and effective method to remove power line artifacts, *Neuroimage*, 207, 116356, doi: 10.1016/j.neuroimage.2019.116356.

de Cheveigné, A (2021). Shared Component Analysis. *NeuroImage*, 226, 117614.

Frauscher, B., Von Ellenrieder, N., Zelmann, R., Doležalová, I., Minotti, L., Olivier, A., Hall, J., Hoffmann, D., Nguyen, D. K., Kahane, P., Dubeau, F., & Gotman, J. (2018). Atlas of the normal intracranial electroencephalogram: Neurophysiological awake activity in different cortical areas. *Brain*, *141*(4), 1130–1144. https://doi.org/10.1093/brain/awy035

Friston, K. (2002). Functional integration and inference in the brain. *Progress in Neurobiology*, *68*(2), 113–143. https://doi.org/10.1016/S0301-0082(02)00076-X.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Parkkonen, L., & Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data. *NeuroImage*, *86*, 446–460. https://doi.org/10.1016/j.neuroimage.2013.10.027

Hatfield G. (2002) Perception as unconscious inference. In Heyer D., & Mausfeld R. (Eds) Perception and the physical world: Psychological and philosophical issues in perception (pp. 113–143). John Wiley and Sons.

Hauk, O., Stenroos, M., & Treder, M. S. (2022). Towards an objective evaluation of EEG/MEG source estimation methods – The linear approach. *NeuroImage*, *255*, 119177. https://doi.org/10.1016/j.neuroimage.2022.119177.

Helmholtz H. (1867). Handbuch der Physiologischen Optik (English tranl.: 1924 JPC Southall as Treatise on Physiological Optics) Voss.

Imbert M (2020). La fin du regard éclairant. Une révolution dans les sciences de la vision au XIe siècle. Ibn al-Haythan, Vrin, Paris.

Jaiswal, A., Nenonen, J., Stenroos, M., Gramfort, A., Dalal, S. S., Westner, B. U., Litvak, V., Mosher, J. C., Schoffelen, J.-M., Witton, C., Oostenveld, R., & Parkkonen, L. (2020). Comparison of beamformer implementations for MEG source localization. *NeuroImage*, *216*, 116797. https://doi.org/10.1016/j.neuroimage.2020.116797

Kawato, M., Hayakawa, H., & Inui, T. (1993). A forward-inverse optics model of reciprocal connections between visual cortical areas. *Network: Computation in Neural Systems*, *4*(4), 415–422. https://doi.org/10.1088/0954-898X_4_4_001

Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.-M. (2011) FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. Computational Intelligence and Neuroscience, 2011:156869.

Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., & Leahy, R. M. (2011). Brainstorm: A User-Friendly Application for MEG/EEG Analysis. Computational Intelligence and Neuroscience, 2011, 1–13. https://doi.org/10.1155/2011/879716

Schneider, G. E. (1969). Two Visual Systems. *Science*, *163*, 895–902.