A tutorial on generalized eigendecomposition for denoising, contrast enhancement, and dimension reduction in multichannel electrophysiology

Michael X Cohen[a]

[a]*Donders Centre for Medical Neuroscience, Radboud University Medical Center, the Netherlands,*

Abstract

The goal of this paper is to present a theoretical and practical introduction to generalized eigende-composition (GED), which is a robust and flexible framework used for dimension reduction and source separation in multichannel signal processing. In cognitive electrophysiology, GED is used to create spatial filters that maximize a researcher-specified contrast. For example, one may wish to exploit an assumption that different sources have different frequency content, or that sources vary in magnitude across experimental conditions. GED is fast and easy to compute, performs well in simulated and real data, and is easily adaptable to a variety of specific research goals. This paper introduces GED in a way that ties together myriad individual publications and applications of GED in electrophysiology, and provides sample MATLAB and Python code that can be tested and adapted. Practical considerations and issues that often arise in applications are discussed.

Keywords: EEG, MEG, LFP, oscillations, source separation, GED, eigendecomposition, components analysis, covariance matrix

1. Background and motivation

1.1. What are "sources" and (how) can they be separated?

The brain is an unfathomably complex and dynamic system, with countless neural operations that are characterized by both segregation and integration. Understanding the brain therefore requires a balance

between isolating neurocognitive elements while maintaining sufficient ecological validity to generalize from experimental control to real-world behavior. One of the primary difficulties in neuroscience research is that myriad neural and cognitive operations unfold simultaneously, and can overlap in time, frequency, and space. In other words, separating sources of cognitive and neural processes is one of the major challenges in neuroscience research.

The term "source" can have several interpretations: It can refer to a single physical location in the brain, a distributed set of locations, a neural ensemble, a single neuron, a synapse, a cognitive operation, a computational algorithm that a neural ensemble produces, a neurochemical modulation, etc. Therefore, the mechanism of source separation depends on the goals of the research. For example, at a behavioral level, careful experiment design can be used to separate attentional from sensory sources of reaction time variability. When studying neural oscillations, we can consider the Fourier transform to be a method of spectral source separation under the assumption that sources are mixed in the time domain but have non-overlapping frequency characteristics. FMRI is highly suitable for spatial source separation due to the small voxel sizes and relative locality of the BOLD response.

The type of source separation under consideration in this paper is a descriptive-statistical separation, where sources are isolated based on spatiotemporal patterns in a channel covariance matrix (a covariance matrix contains the pairwise linear interactions across channels). These sources can stem from a spatially restricted spatial location (e.g., modeled by a single dipole) or they can stem from an anatomically distributed but synchronous network. As explained later, GED has no anatomical constraints, and the sources are defined according to descriptive-statistical criteria, in particular, the information contained in covariance matrices.

Neural data are noisy and variable, and there are more sources in the brain than measurement points. This means that applying a source separation method does not guarantee that a single source has been isolated. (The same uncertainty holds for other methods as well; for example, an fMRI study may indicate a localized source while many other sources were involved but below the statistical threshold.) In this sense, source separation methods should be seen as *attempts* to isolate sources, and their accuracy depends on a variety of experimental and data-analytic factors. A more conservative interpretation of GED is that it acts as a contrast-enhancing filter by isolating relevant from irrelevant patterns in the data while simultaneously allowing for a reduced-dimensional signal to analyze.

1.2. Why multivariate analyses?

It is not contentious to write that electrophysiologists are uninterested in *electrodes*; instead, electro-physiologists are interested in using the numerical values that the electrodes provide to understand how the brain works. It is equally uncontentious to state that there is no one-to-one mapping of electrode to computational source. The neural computations underlying cognition are implemented by complex interactions across neural circuits comprising various types of cells, modulations by neurochemicals, etc.; these circuit interactions produce electromagnetic fields that can be quantified using LFP, EEG, or MEG (local field potential, electroencephalography, magnetoencephalogray), but these electrical fields propagate to multiple electrodes simultaneously. Furthermore, each electrode simultaneously measures the electrical fields from multiple distinct neural circuits, plus artifacts from muscles and heart beats, and from noise.

Thus, the manifest variables (a.k.a. observable data) that we measure (voltage fluctuations in electrodes) are indirect and comprise mixtures of the latent constructs we seek to understand. The mixing of sources at the electrodes motivates *multivariate* analyses that identify patterns distributed across electrodes. This can be contrasted with *univariate* analyses that consider each individual electrode to be a separate statistical unit (Figure 1).

A) Multiple sources, multiple electrodes
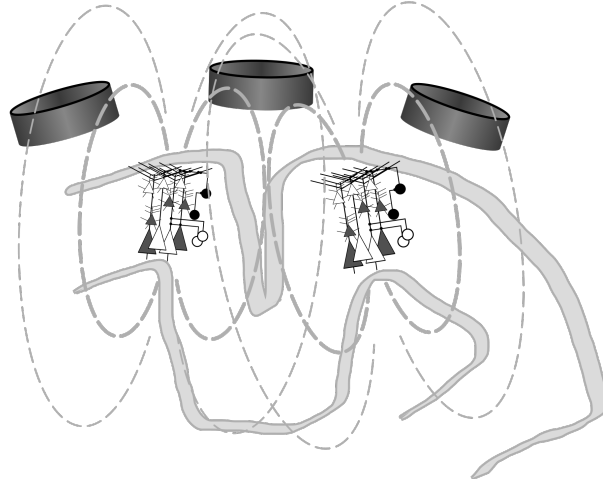
B1) Univariate approach
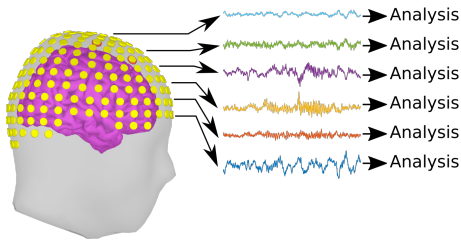
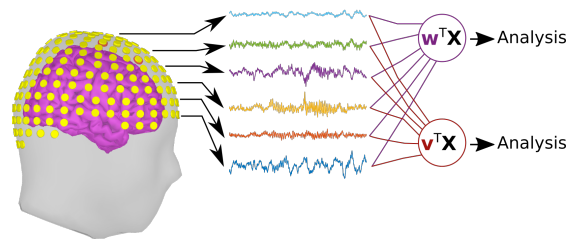**B2)** Multivariate approach

Figure 1: A) An anatomical interpretation of the source-separation problem. Electrophysiology involves measuring data from electrodes (black disks), which are manifest (observable) variables. But electrophysiologists are interested in understanding neural circuits (latent variables), which are complex combinations of neurons and neurochemicals, and which generate electrical fields that propagate to the electrodes. The challenge is to leverage the spatial correlations across electrodes to separate activity from different neural sources. The illustration here makes it seem like the problem is one of anatomical assignment, but because neural computations are spatially distributed, a single localized dipole is not physiologically plausible for all high-level sensory/cognitive/motor phenomena. Therefore, the goal of source separation can be treated as one of statistical assignment. B) The distinction between univariate (B1) and multivariate (B2) analyses lies in the conceptual and statistical use of the electrodes: The univariate approach treats each electrode as an independent measurement; the multivariate approach considers that the signals of interest are embedded in patterns that span many electrodes, and thus isolating those signals requires appropriately designed spatial filters (vectors $\mathbf{w}$ and $\mathbf{v}$ are spatial filters and $\mathbf{X}$ is the channels-by-time data matrix.

And yet, mass-univariate analyses have remained the most common analysis method for decades. The mass-univariate approach is nearly universally used because it is simple and has been successful throughout the history of neuroscience, and because electrode technology and computational power limited the types of analyses that were feasible in the past. However, these limitations no longer exist, and thus there is little justification for maintaining the position that analyses that dominated the literature decades

ago are *prima facie* optimal today.

Thus, the argument here is not that mass-univariate analyses are wrong or misleading; rather, the argument is that progress in neuroscience will be accelerated by shifting to conceptualizing and analyzing data with the goal of isolating and extracting information that is distributed across a set of electrodes. This argument has been made several times before (Makeig et al., 2002; Uhlhaas et al., 2009; Hebb, 1949; Buzsáki, 2010), although the increase in simultaneously recorded measurements, computational power, and analysis possibilities makes multivariate methods more attractive, feasible, and insightful now than in previous decades.

(Although the focus of this paper is on electrophysiology, the previous argument applies to any set of brain or behavioral manifest variables, including neurons, fMRI voxels, two-photon imaging pixels, reaction time, questionnaire items, etc.)

To be sure, there are myriad modern data analysis methods that leverage increased understanding of neurophysiology and computational power. Here I will focus on one family of spatial multivariate analyses, which are used as dimension-reducing, contrast-enhancing spatial filters. A spatial filter is a set of weights, such that the weighted combination of the manifest variables optimizes some criteria. That weighted combination results in a reduced number of virtual data channels (often called "components") compared to the original dataset, and that are designed such that each component isolates a pattern in the data that may be mixed or have low signal-to-noise quality in individual electrodes.

The purpose of this tutorial paper is to introduce a family of methods based on generalized eigendecomposition (GED). The GED seed can sprout many seemingly different multivariate applications, which makes it a powerful method to adapt to specific hypotheses, research goals, and dataset types. There are several other excellent general introductions to GED in the neuroscience literature (Parra et al., 2005; de Cheveigné and Parra, 2014; Haufe et al., 2014); the focus of the present paper is on practical issues that researchers might face when implementing GED.

1.3. A multitude of multivariate analyses

There are many multivariate analysis methods that have been introduced and validated on electrophysiology data. Below is a brief description of several commonly used methods. I focus on their limitations only to facilitate a contrast with the advantages of GED, without implying that these methods are flawed or should be avoided.

Principal components analysis (PCA) is a popular dimension-reduction method that finds a set of channel weights, such that the weighted combination of channels maximizes variance while keeping all components mutually orthogonal. PCA is an excellent tool for data compression, but has three limitations with regards to contrast enhancement, denoising, and source separation: It is descriptive as opposed to inferential; the PC vectors are constrained to be orthogonal in the channel space (which means that correlated sources in channel space remain correlated in PCA space); and maximizing variance does not necessarily maximize relevance (de Cheveigné and Parra, 2014).

Independent components analysis (ICA) is nearly ubiquitously used in M/EEG research to attenuate artifacts such as eye blinks and muscle activity, and is also used for data analysis (Debener et al., 2010). ICA is a blind separation method that relies on the assumption that sources are statistically independent and non-Gaussian distributed. It is often used as a descriptive measure, but cross-validation methods exist for evaluating statistical significance of individual components (Hyvärinen, 2011). In simulated EEG data, ICA has low accuracy at recovering ground-truth sources (Cohen, 2017a; Zuure and Cohen, 2021).

Decoding, a.k.a. multivariate pattern analysis, involves using machine-learning methods to classify experiment conditions (e.g., a particular motor response or visual stimulus) based on weighted combinations of brain signals (Cichy and Pantazis, 2017; King and Dehaene, 2014). Some linear classifiers (e.g., Fisher linear discriminant analysis) are built on GED, however, decoding methods generally threshold and binarize the data, thus discarding a considerable amount of rich and meaningful spectral and temporal variability.

Deep learning is a framework for mapping inputs to outputs, via myriad simple computational units, each of which implements a weighted sum of its inputs plus a nonlinearity (Schirrmeister et al., 2017). Deep

learning has been transformative in many computational fields including computer vision and language translation. Outside of the visual system, deep learning applications in neuroscience have not (yet) made a major impact. Part of the difficulty of deep learning in neuroscience is that its representations are complex, nonlinear, and difficult to interpret. In other words, deep learning has major applications in society and engineering, but is (so far) of limited value for providing mechanistic insights (the same can be said of decoders more generally) (Ritchie et al., 2019). On the other hand, deep learning and linear decompositions could be used synergistically, for example by using linear decompositions to reduce dimensionality and enhance the signal-to-noise characteristics of data that are categorized by a deep learning network (Lawhern et al., 2018).

There are several other multivariate components analyses that are less commonly used in neuroscience, such as factor analysis, Tucker decomposition, and nonnegative matrix factorization. An exhaustive list and comparison of multivariate components analyses is beyond the scope of this paper.

1.4. Motivation for and advantages of GED

GED as a tool for denoising, dimension reduction, and source separation of multichannel data has several advantages. First, it is based on specifying hypotheses. A cornerstone of experimental science is generating and testing null and alternative hypotheses. This manifests in statistical comparisons as a difference or a ratio between the means of two sample distributions. The spatial filters created by GED are designed to maximize a contrast between two features of the data — a feature to enhance, and a feature that acts as reference. Examples of feature-pairs include: an experiment condition and a control condition; a prestimulus period and a poststimulus period; the trial-average and the single-trial data; narrowband filtered and unfiltered data. When the two covariance matrices computed from those data features are equivalent, the GED returns a contrast ratio of 1, which is the expected null-hypothesis value. In this sense, GED is a supervised method, which can be contrasted with unsupervised methods like PCA or ICA.

Second, because of the inherent comparison of two covariance matrices, GED allows for inferential statistics to determine whether a component is significant. Methods for statistical inference are described in a later section.

Third, GED has only a few key researcher-guided analysis choices, which makes it easy to learn, apply, and adapt to new situations.

Fourth, and related to the previous point, there are no spatial or anatomical constraints. The order and relative locations of the physical data channels (electrodes, sensors, pixels, or voxels) does not affect the analysis. This means that the spatial maps can be physiologically interpreted without concern for biases imposed by an *a priori* anatomical or physical model.

Fifth, GED allows for individual differences in topographies. For example, alpha-band activity might be maximal at electrodes Pz, POz, Oz, PO7, etc., in different individuals, leading to possible difficulties and subjectivity with electrode selection. A GED that maximizes alpha-band activity allows for idiosyncratic functional-anatomical distributions for different individuals, while ensuring that the components across all individuals satisfy the same statistical criteria.

Sixth, GED is deterministic and non-iterative. This means that repeated decompositions of the same data give the same solution (this can be contrasted, for example, with ICA algorithms that are initialized with random weights). And it means that GED is fast. Indeed, the GED typically takes a few milliseconds to compute; most of the total analysis time comes from data preparation such as temporal filtering.

Finally, GED has a long history of applications in statistics, machine learning, engineering, and signal processing (Parra et al., 2005). Although not always called "GED," generalized eigendecomposition provides the mathematical underpinning of many analysis methods, including linear discriminant analysis, common-spatial pattern (used in brain-computer interface algorithms), blind-source separation (Tomé, 2006; Blankertz et al., 2007), and other methods discussed later. Although there are "tricks" for optimizing GED for specific applications, the general approach is well established in multiple areas of science (Parra and Sajda, 2003; Särelä and Valpola, 2005).

1.5. GED in the wild

GED is widely used in neuroscience, although the terminology differs. In brain-computer interfaces, GED is called common spatial pattern analysis (Blankertz et al., 2007) and is used to design spatial filters that facilitate neural control over a computer program (Rivet et al., 2011; Ai et al., 2018). Nikulin et al.

(2011) used GED to design a narrowband spatial filter, which they termed spatio-spectral decomposition. This method was extended to use broadband energy compared to narrowband energy over a successive range of center frequencies, and was termed spectral scanning (de Cheveigné and Arzounian, 2015) or, more generally, joint decorrelation (de Cheveigné and Parra, 2014). Dähne et al. (2014) used GED to design a spatial filter that maximizes a correlation between EEG and a behavioral measure like reaction time. Cohen (2017b) adapted GED to identify multivariate cross-frequency coupling. Several groups have used GED to optimize ERPs, which is particularly useful for single-trial analyses (Rivet and Souloumiac, 2013; Tanaka and Miyakoshi, 2019; Das et al., 2020). GED has been used to obtain components that have maximal signal-to-noise characteristics in steady-state evoked potential studies (Dmochowski et al., 2015; Cohen and Gulbinaite, 2017). GED may also provide an efficient method for spike-sorting in invasive multichannel recordings (Wouters et al., 2018), as well as online epileptic spike waveform detection (Dan et al., 2020).

There are many other applications of GED in electrophysiology research; the goal here is not to cite all of them, but instead to highlight that GED is a core part of the corpus of multivariate neuroscience analyses, even if its inclusion is not apparent from the titles of research papers.

1.6. Brief overview of GED

Before delving into the details, it will be useful to have the "bird's eye view" of GED (Figure 2). GED is a decomposition of two covariance matrices, here termed $\mathbf{S}$ and $\mathbf{R}$. These two covariance matrices come from different features of the data: an experiment condition and a control condition, narrowband filtered and unfiltered data, etc. The $\mathbf{S}$ matrix is the covariance of the "signal" data feature of interest, and the $\mathbf{R}$ matrix is the covariance of the "reference" data that provides a comparison.
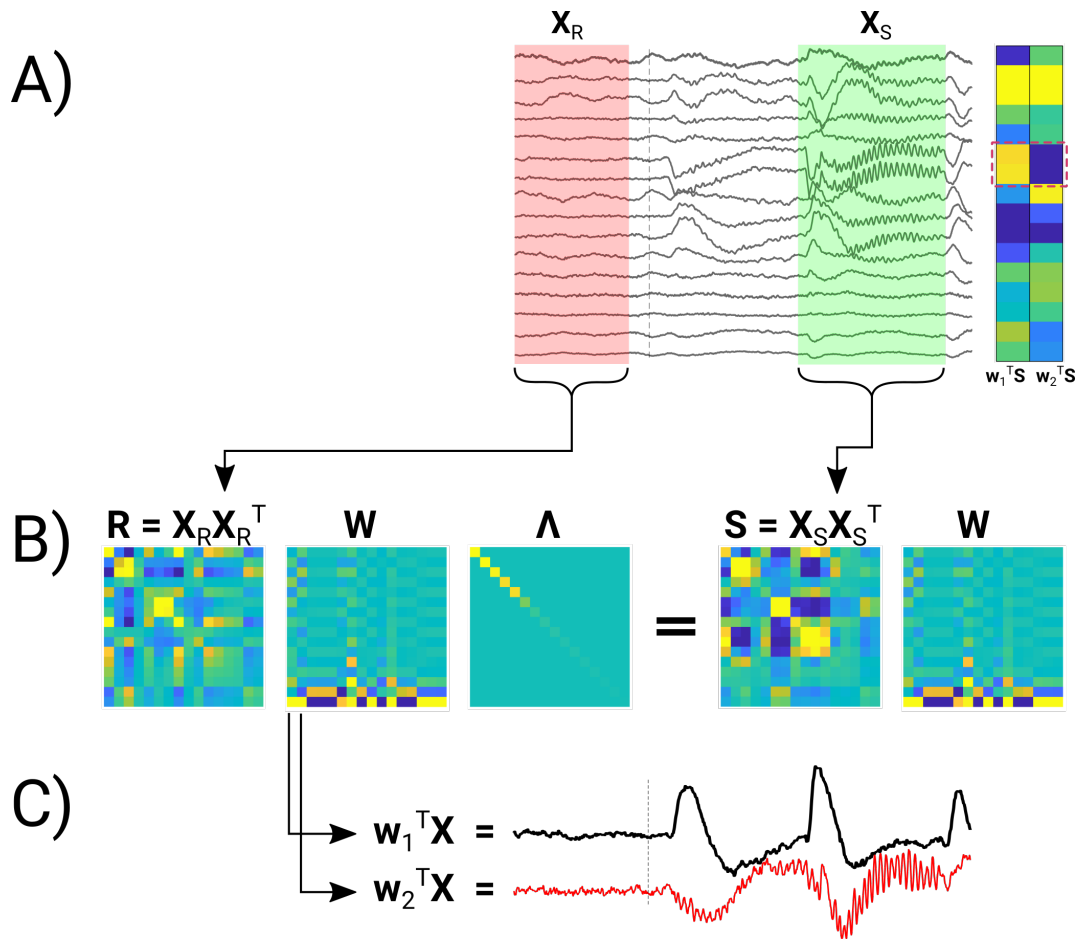
Figure 2: Graphical overview of GED. This example shows laminar recordings in mouse V1 after a visual stimulus onset. A) Two time windows are selected for the "reference" ($\mathbf{X_R}$) and "signal" ($\mathbf{X_S}$) features of the data. B) A generalized eigenvalue decomposition is performed on the two covariance matrices created from the two time windows. The resulting eigenvectors (matrix $\mathbf{W}$) are the spatial filters, and their corresponding eigenvalues (diagonal elements of matrix $\mathbf{\Lambda}$) encode the ratio of matrix $\mathbf{S}$ to $\mathbf{R}$ along each direction $\mathbf{w_i}$. C) Each eigenvector times the data produces a component, with an accompanying time series and spatial map (spatial maps are visualized to the right of the time series in panel A). In this example, the first component captured a low-frequency response to the visual stimulus, while the second component isolated the gamma-band response. The red dashed box in the spatial maps indicates the approximate location of cortical layer 4.

The GED finds a weighting of the data channels that maximizes a signal-to-noise ratio (SNR) that can be thought of as $\mathbf{S}/\mathbf{R}$ (division is not defined for matrices, but this conceptualization can be helpful). Inter-channel covariance patterns that are common between $\mathbf{S}$ and $\mathbf{R}$ are ignored.

The channel weight vector associated with the largest eigenvalue is the spatial filter, and the weighted sum of all channel time series is the component time series that maximizes the researcher-defined criteria

established by selecting data for the two covariance matrices. That component time series can be used in standard analyses such as ERP or time-frequency analysis, and its accompanying topography can be visualized for topographical or anatomical interpretation.

Figure 3 shows an example comparison of PCA and GED in a simulated 2D dataset. This illustration highlights several key features of GED, including the ability to separate sources that are correlated in the original channel space.
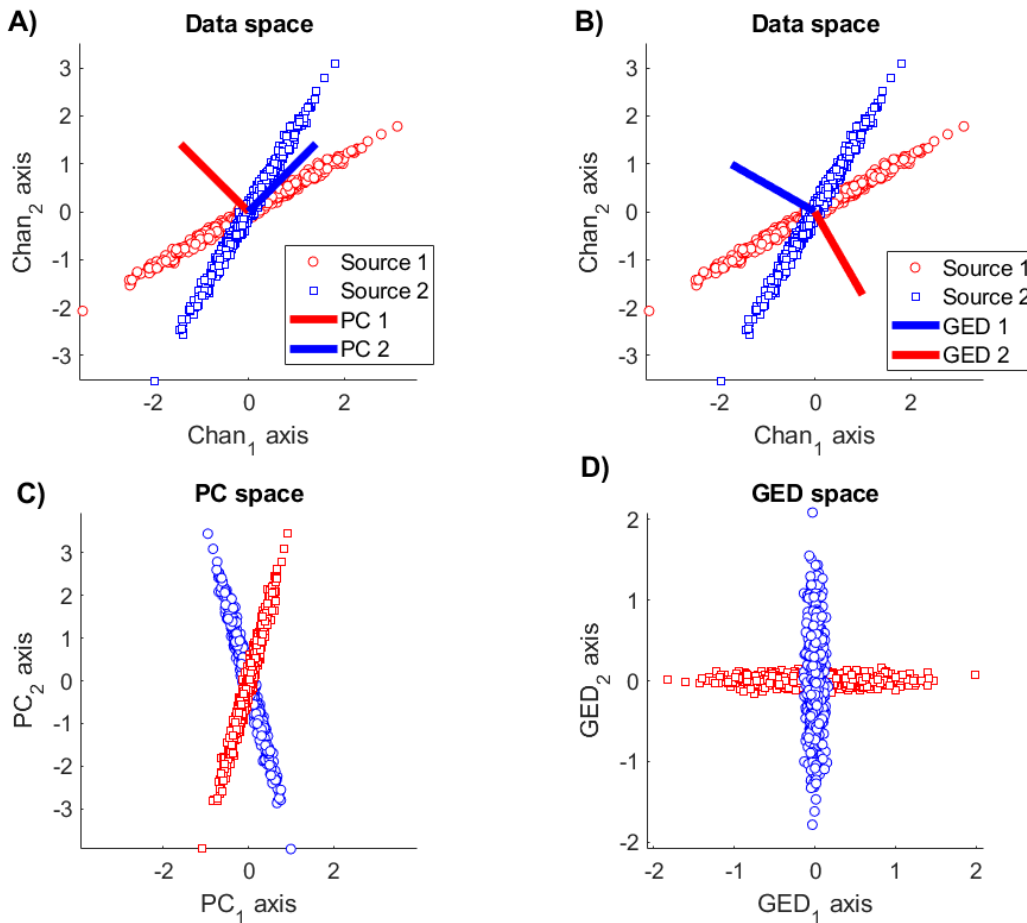

Figure 3: Simple example comparing PCA and GED. (A, B) The data were simulated as two "streams," indicated by the blue and red markers. Both plots show the same data. (C) A PCA on these data removes the correlation between variables, but does not separate the sources. The PC basis vectors, which form the axes PC_1 and PC_2, are shown in the channel space in panel A. (D) A GED on the channel covariance matrices for the red vs. blue streams was successful at source separation. The GED eigenvectors, which form axes GED_1 and GED_2, are shown in the channel space in panel B. Notice that the sources are correlated in the data space and in the PC space, and are orthogonal in the GED space.

2. Mathematical and statistical aspects

2.1. The math of GED

The goal of GED is to find a set of channel weights that multiply the channel time series, such that the weighted sum of channels maximizes a contrast between two data features. For example, consider that the to-be-maximized feature is a post-stimulus time window, and that the pre-stimulus time window is the reference period. If the post-stimulus data are contained in a channels-by-time matrix $\mathbf{X}_S$, the pre-stimulus data are contained in a channels-by-time matrix $\mathbf{X}_R$, and the set of weights are contained in column vector $\mathbf{w}$, then the goal of GED can be expressed as:

$$\lambda = \frac{\|\mathbf{w}^\mathsf{T}\mathbf{X}_S\|^2}{\|\mathbf{w}^\mathsf{T}\mathbf{X}_R\|^2} \tag{1}$$

$\|\cdot\|^2$ indicates the squared magnitude of the vector (the sum of its squared elements). Thus, λ is the ratio of the magnitude of the "signal" data filtered through $\mathbf{w}$, to the magnitude of the "reference" data filtered through the same $\mathbf{w}$. The squared magnitude of a vector can be expressed as the dot product of the row vector with itself. Thus:

$$\lambda = \frac{(\mathbf{w}^\mathsf{T}\mathbf{X}_S)(\mathbf{w}^\mathsf{T}\mathbf{X}_S)^\mathsf{T}}{(\mathbf{w}^\mathsf{T}\mathbf{X}_R)(\mathbf{w}^\mathsf{T}\mathbf{X}_R)^\mathsf{T}} \tag{2}$$

$$= \frac{\mathbf{w}^\mathsf{T}\mathbf{X}_S\mathbf{X}_S^\mathsf{T}\mathbf{w}}{\mathbf{w}^\mathsf{T}\mathbf{X}_R\mathbf{X}_R^\mathsf{T}\mathbf{w}} \tag{3}$$

$$= \frac{\mathbf{w}^\mathsf{T}(\mathbf{X}_S\mathbf{X}_S^\mathsf{T})\mathbf{w}}{\mathbf{w}^\mathsf{T}(\mathbf{X}_R\mathbf{X}_R^\mathsf{T})\mathbf{w}} \tag{4}$$

The parentheses in equation 4 reveal that the solution to equation 1 can be developed through covariance matrices.

A covariance matrix is an $M \times M$ matrix in which the element in row i and column j contains the covariance between channels M_i and M_j, defined as the sum of the element-wise multiplications of the mean-centered channel time series. Thus, the covariance matrix contains all linear pairwise interactions. A covariance is simply a non-normalized Pearson correlation coefficient, and thus a covariance matrix is a correlation matrix that retains the scale of the data (e.g., μV).

If the data are organized in a channels-by-time matrix $\mathbf{X}$ with n time points, then the covariance matrix

is given by

$$\mathbf{C} = \mathbf{X}\mathbf{X}^{\mathsf{T}}(n-1)^{-1} \tag{5}$$

In general, the size of the covariance matrix should be channels-by-channels, and so the multiplication would be expressed as $\mathbf{X}^{\mathsf{T}}\mathbf{X}$ if the data were organized as time-by-channels. The division by $n-1$ is a normalization factor that prevents the covariance from increasing simply by increasing the number of observations (time points).

The data must be mean-centered before the multiplication. Mean-offsets in the data will cause the GED solutions to point in the direction of the offsets instead of the direction that maximizes the desired optimization criterion. Mean-centering means that the average value of each channel, over the time window used to compute the covariance matrix, is zero.

Variance normalization is not necessary if all channels are in the same scale (e.g,. microvolts). Multi-modal data might need within-modality normalization, which is discussed in section 6.1.

MATLAB code for computing the covariance matrix follows (variable `data` is channels-by-time).

```
data = data-mean(data,2); % mean-center
S = data*data' / (size(data,2)-1);
```

As introduced in the previous section, GED requires two covariance matrices: the $\mathbf{S}$ matrix computed from features of the data to highlight, and the $\mathbf{R}$ matrix computed from features of the data to use as reference ($\mathbf{S}$ for *signal* and $\mathbf{R}$ for *reference*). The signal and reference data should be matched on as many features as possible, similar to how a control condition in an experiment should be matched to the experimental condition. In this sense, the GED can be thought of as maximizing a signal-to-*reference* ratio instead of a signal-to-*noise* ratio.

If $\mathbf{S}$ is computed from a time window during stimulus presentation and $\mathbf{R}$ is computed from a time window before stimulus presentation (the inter-trial interval), then the resulting GED will isolate the covariance patterns that maximally differentiate stimulus processing compared to the inter-trial-interval, while excluding any patterns in the data that are present before and during the stimulus period, such

as ongoing spontaneous activity or slower neurocognitive processes that are unrelated to stimulus processing.

Once the two covariance matrices are formed, the goal is to find an M-element vector of weights (called vector $\mathbf{w}$; each element w_i is the weight for the ith data channel), which acts as a *spatial filter* that reduces the dimensionality of the data from M channels to 1 component. The elements in $\mathbf{w}$ are constructed such that the linear weighted sum over all channels maximizes the ratio of "multivariate power ratio" in $\mathbf{S}$ to that in $\mathbf{R}$. Rewriting equation 4 using covariance matrices leads to the form $\mathbf{w}^\mathsf{T}\mathbf{Sw}$, which is known as the quadratic form. The quadratic form is a single number that encodes the variance in matrix $\mathbf{S}$ along direction $\mathbf{w}$. Therefore, the goal of GED is to maximize the ratio of the quadratic forms of the two matrices, where that ratio is encoded in λ.

$$\lambda = \frac{\mathbf{w}^\mathsf{T}\mathbf{Sw}}{\mathbf{w}^\mathsf{T}\mathbf{Rw}} \tag{6}$$

This expression is also known as the generalized Rayleigh quotient. Note that when the data covariance matrices are the same, $\lambda = 1$, which can be considered the null-hypothesis value ($\mathrm{H}_0 : \mathbf{S} = \mathbf{R}$).

The goal of GED is to find the vector $\mathbf{w}$ that maximizes λ. This is the objective function.

$$\underset{\mathbf{w}}{\arg\max} \frac{\mathbf{w}^\mathsf{T}\mathbf{Sw}}{\mathbf{w}^\mathsf{T}\mathbf{Rw}} \tag{7}$$

Equation 7 would produce only one spatial filter, however, it can be expanded to include additional vectors $\mathbf{w}_2$ up through $\mathbf{w}_M$, with each vector $\mathbf{w}_i$ subject to the constraint that it maximizes λ_i while being uncorrelated with previous components (that is, the components are orthogonal in the GED space, as shown in Figure 3). It can be shown that the expansion of one vector to a set of M vectors means that equation 6 can be rewritten as follows.

$$\mathbf{\Lambda} = (\mathbf{W}^\mathsf{T}\mathbf{RW})^{-1}\mathbf{W}^\mathsf{T}\mathbf{SW} \tag{8}$$

where each column of $\mathbf{W}$ is a spatial filter, and each diagonal element of $\mathbf{\Lambda}$ is the multivariate ratio in the direction of the corresponding column in $\mathbf{W}$ (additional details on the derivation of this solution can be found elsewhere, e.g. Ghojogh et al., 2019; Parra and Sajda, 2003). Some algebraic manipulations

brings us to the matrix solution to the optimization problem:

$$\mathbf{\Lambda} = (\mathbf{W}^\mathsf{T}\mathbf{R}\mathbf{W})^{-1}\mathbf{W}^\mathsf{T}\mathbf{S}\mathbf{W} \tag{9}$$

$$\mathbf{\Lambda} = \mathbf{W}^{-1}\mathbf{R}^{-1}\mathbf{W}^{-\mathsf{T}}\mathbf{W}^\mathsf{T}\mathbf{S}\mathbf{W} \tag{10}$$

$$\mathbf{\Lambda} = \mathbf{W}^{-1}\mathbf{R}^{-1}\mathbf{S}\mathbf{W} \tag{11}$$

$$\mathbf{R}\mathbf{W}\mathbf{\Lambda} = \mathbf{S}\mathbf{W} \tag{12}$$

Equation 12 is known as a generalized eigendecomposition on matrices $\mathbf{S}$ and $\mathbf{R}$. This means that the set of weights that maximizes the multivariate signal-to-noise ratio — the spatial filter — is an eigenvector, and the value of that ratio is the corresponding eigenvalue. It is also useful to see that equation 12 is conceptually equivalent to a "regular" eigendecomposition on a matrix product.

$$\mathbf{B} = \mathbf{R}^{-1}\mathbf{S} \tag{13}$$

$$\mathbf{W}\mathbf{\Lambda} = \mathbf{B}\mathbf{W} \tag{14}$$

One of the limitations of PCA is that the eigenvectors matrix is orthogonal, meaning that all components must be orthogonal. That constraint comes from the eigendecomposition on a symmetric matrix, and all covariance matrices are symmetric. This constraint is a limitation because sources that are correlated in the channel space cannot be separated by PCA (e.g., Figure 3). However, the product of two symmetric matrices is generally not symmetric, and so the eigenvectors of a GED may be correlated (though they are linearly independent for distinct eigenvalues). This is the reason that sources that are correlated in channel space can be separated by GED (and will be orthogonal in the GED space). Also note that matrix $\mathbf{R}$ is inverted "on paper"; in practice, the GED is solved without explicitly inverting matrices (Ghojogh et al., 2019), which means that GED can be performed on reduced-rank data.

The two matrices that GED returns contain pairs of eigenvectors and eigenvalues, with each eigenvector (spatial filter) having a corresponding eigenvalue (multivariate ratio). In this pairing, the eigenvector $\mathbf{w}_i$ points in a specific direction in the dataspace, but does not convey the importance of that direction (the $\mathbf{w}$'s are unit-length in the space of $\mathbf{R}$ because $\mathbf{W}^\mathsf{T}\mathbf{R}\mathbf{W} = \mathbf{I}$). In contrast, the corresponding eigenvalue λ_i encodes the importance of the direction, but is a scalar and therefore has no intrinsic direction. The implication of this is that the eigenvector associated with the largest eigenvalue is the spatial filter that maximizes the ratio $\mathbf{S} : \mathbf{R}$ along direction $\mathbf{w}_1$. The next-largest eigenvalue is paired with the eigenvector that maximizes that ratio while being $\mathbf{R}$-orthogonal to the first direction (that is,

satisfying the constraint that $\mathbf{w}_2^\mathsf{T}\mathbf{X}$ is orthogonal to $\mathbf{w}_1^\mathsf{T}\mathbf{X}$ (Parra et al., 2019)). And so on for all other directions.

It is not reasonable to expect that all components are meaningful and interpretable. Indeed, if the two covariance matrices are similar to each other (e.g., created from two different experiment conditions that are matched on many perceptual and motor features), there may be only one significant component — or possibly no significant components. The GED simply returns all solutions without a p-value or confidence interval that would indicate interpretability. Inferential statistics can be applied to the eigenvalue spectrum, which is described in section 2.4.

GED is easy to implement in MATLAB. The returned solutions are not guaranteed to be sorted by eigenvalue magnitude, and thus it is convenient to sort the solution matrices.

```
[W,L] = eig(S,R);
[eigvals,sidx] = sort(diag(L),'descend');
eigvecs = W(:,sidx);
```

After sorting, the component time series and component map are created by multiplying an eigenvector with, respectively, the multichannel data and the covariance matrix (discussed in more detail in section 3.6).

```
comp_ts  = eigvecs(:,1)' * data; % data are chansXtime
comp_map = eigvecs(:,1)' * S;
```

Trial-related data are often stored as 3D matrices (e.g., channels, time, trials), and so the code may need to be modified to reshape the data to 2D for the multiplication, and then back into 3D for further analyses.

The computation of a component time series is an important divergence from typical machine-learning classification or discriminant analyses: The goal of traditional classification analyses is to use $\mathbf{w}$ to binarize the data and predict whether the data were drawn from "class A" or "class B". However,

the application of GED described here results in a continuous time series that has considerably richer information than a binary class label.

2.2. Assumptions underlying GED

GED relies on several implicit and explicit assumptions.

1. Signals mix linearly in the physical data channels. This assumption is necessary because GED is a linear decomposition. For M/EEG/LFP, this assumption is readily feasible, because electrical fields propagate instantaneously and sum linearly (within measurement capabilities) (Nunez et al., 2006). For other measurement modalities like fMRI, multiunit activity, or 2-photon imaging, the linear-mixing assumption can be interpreted that spatially distributed manifestations of the neural computations are simultaneously active (within a reasonable tolerance given by the sampling rate) and sum linearly. This assumption appears to be met in fMRI data (van Dijk et al., 2020; Boynton et al., 2012).

2. The targeted features of the data are stable within the windows used to construct the covariance matrices. Stability here means that the characteristics of the signal remain consistent over the covariance matrix time window. If the underlying neural dynamics change qualitatively within that time window, for example if multiple distinct sensory/cognitive/motor processes are engaged, then the resulting GED decomposition may be less well-separated, more difficult to interpret, or less statistically robust.

3. Covariance is a meaningful basis for source separation. Covariance is a second-order statistical mixed moment, and captures all of the pairwise linear interactions. This is notably distinct from, for example, ICA, which begins by whitening the data and optimizes based on higher-order moments such as kurtosis or entropy. Validation studies in simulated data (described later) confirm the viability of this assumption.

4. The data features used to create the $\mathbf{S}$ and $\mathbf{R}$ matrices are physiologically or cognitively sufficiently different to produce an interpretable spatial filter. To take an extreme example, designing a spatial filter to separate 10.0 Hz from 10.1 Hz EEG activity is unlikely to produce meaningful results. The data features should be selected such that they are comparable in many aspects and differ in one or a small number of key aspects. Analogously, an experimental task condition and its associated control condition should be carefully designed to be similar but meaningfully differentiable.

A final, implicit, assumption of GED is that the components "look reasonable" to our intuition. Our training and experiences as neuroscientists have instilled in us a sense of what data "should" look like, in terms of spatial and temporal characteristics (e.g., smoothness or spectral energy distributions) and statistical effects sizes. Data or results that are inconsistent with our expectations are likely to be rejected, or the analysis redone using different parameters or data features. This is not unique to GED — it is nearly ubiquitous in science that our expectations, intuitions, and experiences leads us to evaluate data and results in certain ways. This is a subtle but important aspect of high-quality research: On the one hand, artifacts, noise, and confounds can be detected by experienced scientists while being missed by novices. But on the other hand, we must take care to avoid overfitting, researcher degrees of freedom, and p-hacking.

Careful experiment and analysis design, along with appropriate inferential statistics, can help ensure that results are valid and generalizable. But it is also important to be cognizant of the role — and potential bias — that researcher expectations can have on data analysis pipelines. The implications of researcher overfitting are discussed extensively elsewhere (Ioannidis, 2008; Head et al., 2015; Kriegeskorte et al., 2009).

2.3. Understanding and avoiding overfitting

Overfitting is a term in statistics and machine learning that refers to models being optimized for a desired result, at the expense of generalizability. Overfitting is potentially dangerous because the model parameters can be driven by noise or other non-reproducible patterns in the data.

On the other hand, overfitting is a powerful and useful approach when used appropriately. GED is based on fitting a spatial filter that maximizes the contrast between matrices $\mathbf{S}$ and $\mathbf{R}$. Thus, one can use GED to leverage overfitting in a beneficial way without introducing systematic biases into the analyses that could confound the results. This requires some additional considerations, compared to blind decomposition methods such as ICA or PCA (these methods are "blind" in that they do not involve specifying a contrast that guides the separation).

There are four approaches that will help evaluate whether the results reflect overfitting noise or sampling variability, and thus whether the model is likely to generalize to new data.

1. Apply statistical contrasts that are orthogonal to the maximization criteria. For example, use GED to create a spatial filter that maximizes gamma-band activity across all experiment conditions, and then statistically compare gamma activity between conditions. In this case, a statistical test of the presence of gamma activity *per se* is biased by the creation of the spatial filter. But a statistical test on condition differences in gamma is not biased by the filter construction. Note that this method does not guarantee an unbiased test in all cases. For example, a trial count imbalance could bias the GED towards the condition that contributes the most data.

2. Use cross-validation to evaluate generalization performance. Cross-validation is a commonly used method in machine learning, and involves fitting the model (the spatial filter) using part of the data and applying the model to a small portion of the data that was not used to train the model. Typical train/test splits are 80/20% or 90/10%. One cross-validation fold would mean including only 10% of the experimental data, which is unlikely to provide a sufficient amount of data in typical electrophysiology experiments. An alternative is to use k-fold cross-validation, where the analysis is repeated, each time using a different 90/10 split of the data. Thus, after 10 splits, all of the data are included, without any of the "test" trials included in the GED construction.

3. Create the spatial filter based on independent data. This is similar to cross-validation and is a technique that is sometimes used in fMRI localizer tasks (for example, having a separate experiment acquisition to identify the fusiform face area). Thus, the GED spatial filter would be created from a different set of trials, possibly in a different experiment block.

4. Apply inferential statistics (via permutation-testing) to evaluate the probability that a component would arise given overfitting of data when the null hypothesis is true. This is because overfitting noise will produce a maximum eigenvalue greater than 1 (the H_0 value), but that eigenvalue is unlikely to survive a significance test with an appropriate p-value. The mechanism for this is described below.

2.4. Inferential statistics

As mentioned earlier, the eigenvectors simply point in a particular direction in the data space, while the eigenvalues encode the importance of that direction for separating $\mathbf{S}$ from $\mathbf{R}$. Thus, the goal of inferential statistics of GED is to determine the probability that eigenvalues as large as observed could have been obtained by chance (due to noise and sampling variability), given that the null hypothesis is

true. The expected value of λ is 1 when the two covariance matrices are the same; but in real data, $\mathbf{S}$ and $\mathbf{R}$ can be expected to differ due to sampling variability and noise, even if they are drawn from the same population. Thus, the eigenvalues can be expected to be *distributed* around 1, and we can expect that roughly one-half of the eigenvalues of a GED on null-hypothesis data will be larger than 1.

Inferential statistical evaluation of GED solutions involves creating an empirical null-hypothesis distribution of eigenvalues, and comparing the empirical eigenvalue relative to that distribution. One might initially think of generating covariance matrices from random numbers. However, this approach would produce an inappropriately liberal statistical threshold, because real data have a considerable amount of spatiotemporal structure that must be incorporated into the null hypothesis distribution (Theiler et al., 1992). Thus, an appropriate approach is to randomize the mapping of data into covariance matrix, without generating fake data.

Consider an experiment with 100 trials, and the GED is based on the pre-trial to post-trial contrast. Each trial provides two covariance matrices, and thus there are 200 covariance matrices in total. To generate a null-hypothesis GED, each covariance matrix is randomly assigned to be averaged into $\mathbf{S}$ or $\mathbf{R}$, and a GED is performed. From the resulting collection of M eigenvalues, the largest is selected. This is the largest eigenvalue that arose under the null hypothesis. This shuffling procedure is then repeated many times (e.g., 1000), with each iteration having a new random assignment of data segment to covariance matrix. The resulting collection of 1000 pseudo-λ's creates a distribution of the largest eigenvalues expected under the null hypothesis that $\mathbf{S} = \mathbf{R}$. A p-value can be computed as the proportion of pseudo-λ values that are larger than λ. For example, if 4/1000 null-hypothesis maximum eigenvalues were larger than the observed λ, the p-value associated with that GED component would be p=.004. Note that these tests are one-tailed, as we are interested only in components that are larger than the null-hypothesis distribution. Using the largest eigenvalue from each null-hypothesis GED corrects for multiple comparisons over the M eigenvalues. This procedure is called extreme-value correction or maxT correction (Nichols and Holmes, 2002; Maris and Oostenveld, 2007).

Alternatively, one can define a critical threshold above which any observed λ can be considered statistically significant, as the 95% percentile of the pseudo-λ distribution. If that numerical value is, e.g., 1.7, then any GED component with an associated $\lambda > 1.7$ can be considered statistically significant at a 5% threshold.

Unfortunately, things are not always so simple. If $\mathbf{S}$ and $\mathbf{R}$ are not in the same scale, then the expected λ under the null hypothesis might be different from 1. This can occur, for example, if the $\mathbf{S}$ matrix comes from narrowband filtered data while the $\mathbf{R}$ matrix comes from broadband data. A narrowband signal will contain less variance than the broadband signal from which it was extracted. Randomly shuffling data segments into $\mathbf{S}$ or $\mathbf{R}$ might produce a pseudo-λ distribution around 1, whereas the empirical largest λ might be .1. If overall differences in covariance matrix magnitude are expected, the matrices can be normalized. Normalization of the covariance matrices is discussed in section 3.3.

The methods described above are for statistical evaluation of a GED within one individual. But neuroscience studies typically involve group-level analyses, in which the goal is to determine whether the features of the data are consistent across individuals, which provides support for generalizability to a population. In many cases, it is advantageous to take one component per subject (e.g., the component that maximizes the desired $\mathbf{S} : \mathbf{R}$ contrast; component selection is discussed in section 3.7), and N subjects would therefore produce N components. Group-level analyses would then be done on the results of the component time series analyses: ERP, power spectrum, time-frequency power, or whatever is the relevant result. Thus, procedures for group-level analyses are the same as those for analyzing data from one electrode per subject.

3. Practical aspects

3.1. Preparing data for GED

GED doesn't "know" what is real brain signal and what is noise or artifact; it simply identifies the patterns in the data that maximally separate two covariance matrices. Therefore, the data should be properly cleaned prior to GED. This may include rejecting noisy trials, temporal filtering, ICA for projecting out non-brain sources, and excluding bad channels. It is also possible to clean the covariance matrices by excluding any segment covariances that are "far away" from their mean (discussed in section 3.3). Data cleaning procedures are often equipment- and lab-specific, and so data cleaning for GED can follow the same data cleaning protocol already established in the lab. GED can produce good solutions with reduced-rank data (discussed in section 3.4), so there is no concern about removing artifact ICs.

Channel interpolation is not necessary, because the interpolated channels are linear combinations of other

channels, and thus do not contribute unique information to the decomposition. Removed channels can be interpolated in the component maps to facilitate averaging across subjects.

For EEG and LFP, the reference scheme does not impact the GED solution. Re-referencing is a linear operation, and GED is a linear decomposition. Of course, the channel weights will differ if a different reference is used, and thus the filter forward model will also change. Referencing is an oft-debated issue in EEG research (Yao et al., 2019), and average reference is often recommended. The only real restriction on re-referencing with GED is that the same reference must be used to compute the GED and the components. For example, one should not construct the GED using earlobe reference and then apply the spatial filter to average-referenced data.

3.2. Selecting data features for $\mathbf{S}$ and $\mathbf{R}$ covariance matrices

Selecting two data features for the GED to separate is the single most important decision that the researcher makes during a GED-based data analysis. It is also the reason why GED is such a flexible and versatile analytic backbone (de Cheveigné and Parra, 2014; Parra et al., 2005). This means that GED forces researchers to think carefully and critically about their hypotheses and analyses, which is likely to have positive effects on the quality of the research.

The main hard constraint is that the data channels must be the same and in the same order; it is not possible to separate covariance matrices of different sizes, nor is a GED on covariance matrices with channels in different orders interpretable.

There are many features of the data to select for the contrast-enhancing ($\mathbf{S}$) and reference ($\mathbf{R}$) matrices; below is a non-exhaustive list that illustrates the possibilities.

1. **Condition differences**. The data for the $\mathbf{S}$ covariance matrix come from the condition of interest (e.g., trials with an informative attentional cue) and the data for the $\mathbf{R}$ matrix come from a control condition (e.g., trials with an uninformative attentional cue) (Zuure and Cohen, 2021). One must be mindful that experiment confounds might bias the GED result. For example, if condition $\mathbf{S}$ has faster reaction times than condition $\mathbf{R}$, then the GED result might reflect motor rather than attentional processing.

2. **Task effects**. The data for the $\mathbf{S}$ covariance matrix come from a within-trial time window (e.g., 0 to 800 ms post-trial-onset) and the data for the $\mathbf{R}$ matrix come from a pre-trial baseline period (e.g., -500 to 0 ms) (Duprez et al., 2020). When data from all conditions are pooled together, this approach avoids the risk of overfitting to one condition.

3. **Spectral contrast**. The data for the $\mathbf{S}$ matrix are narrowband filtered in some range of interest (e.g., alpha, 10 Hz), and the data for the $\mathbf{R}$ matrix are broadband (de Cheveigné and Arzounian, 2015) or narrowband from neighboring frequencies (Nikulin et al., 2011).

4. **Trial average**. The $\mathbf{S}$ matrix is computed from the trial-averaged response such as an ERP, and the $\mathbf{R}$ matrix is computed from the average of the single-trial covariance matrices (Rivet et al., 2009; Tanaka and Miyakoshi, 2019; Dmochowski et al., 2015). This approach will enhance the phase-locked features of the signal.

3.3. Computing the covariance matrices

The quality of the GED result depends entirely on the quality of the covariance matrices, so it is important to make sure that the covariance matrices are made from a sufficient amount of clean data.

One way to increase the stability of a covariance matrix is to increase the number of time points in the data segment. However, the size of the time window presents a trade-off between cognitive specificity vs. covariance quality: Shorter time windows (e.g., 100 ms) better isolate phasic sensory/cognitive/motor events, but risk a noisier covariance matrix. On the other hand, longer time windows (e.g., 1000 ms) may increase the quality of the covariance matrix but may span multiple distinct task events associated with distinct neural/cognitive systems. Keep in mind that the temporal resolution of the component time series is that of the data, and is not limited by the time windows used to create the covariance matrices. This is because the spatial filter can be applied to the entire time series data, not only the data within the covariance time window.

If the covariance matrix is computed from temporally narrowband filtered data, then the time windows to compute the covariance matrices should be at least one cycle, and preferably longer. For example, if the channel data are filtered at 4 Hz, then the time window to compute the covariance matrix should be at least 250 ms. To avoid edge effects, the temporal filter should be applied to the epoched or continuous data, not only to the data inside the covariance time window.

Ideally, the data for the two covariance matrices are of comparable quality. Matching the amount of data used for both covariance matrices (that is, the same number of trials or time points) can help ensure equal data quality. However, simply matching the amount of data does not guarantee matching covariance quality if the data used to create $\mathbf{R}$ are noisier than the data used to create $\mathbf{S}$.

When working with data that are segmented into multiple trials, the covariance matrix of each trial can be computed and then averaged over trials. Covariance-averaging can also be used for continuous (e.g., resting-state) data that are epoched into, e.g., 2-second epochs (Allen et al., 2004). In this case, each individual data segment must be mean-centered before its covariance is computed. Illustrative MATLAB code shows how this can be implemented (data is a variable with dimensions: channel, time, trials).

```
covmat = zeros(nbchans);
for triali=1:ntrials
    seg = data(:,:,triali); % extract one trial
    seg = seg - mean(seg,2); % mean-center
    covmat = covmat + seg*seg'/(size(seg,2)-1);
end
covmat = covmat / triali;
```

Computing N covariance matrices also allows for an additional data cleaning step based on distance. In particular: The average covariance matrix $\overline{\mathbf{S}}$ is computed, and then the Euclidean distance (or any other distance metric) between each segment's covariance matrix $\mathbf{S}_n$ and $\overline{\mathbf{S}}$ is computed. Those N distances can be z-scored, any covariance matrices with an excessive distance (e.g., $z > 2.31$ corresponding to $p < .01$) are excluded, and a new average $\overline{\mathbf{S}}$ is re-computed. The justification of this procedure is that large-distance covariance segments are multivariate outliers that may bias the results.

Because covariance matrices retain the scale of the data, normalization is not necessary. Even if $\mathbf{S}$ and $\mathbf{R}$ are in different scales, normalization may be unnecessary. This is because in most cases, the relative eigenvalues are important (e.g., for sorting), not the absolute numerical values. Normalization is necessary only when (1) permutation testing is used while $\mathbf{S}$ and $\mathbf{R}$ are in different scales, or (2) data are combined from different modalities that have very different numerical ranges (e.g., EEG and MEG).

It is possible to z-normalize each data channel, however, this will change the covariance matrix, and therefore will make the filter forward model difficult to interpret, because z-normalizing each channel separately alters the magnitude of the between-channel covariances. In other words, channels with low variance are inflated whereas channels with high variance are deflated. Of course, this is the goal of z-normalizing, but some aspect of channel differences in variance are meaningful, for example alpha-band variance is higher at posterior-central channels compared to lateral temporal channels.

An alternative is to mean-center each channel separately, and then divide all channels by their pooled standard deviation. This approach preserves the relative covariance magnitudes within modality, while simultaneously ensuring that the total dataset has a pooled standard deviation of 1.

Covariance matrices are typically computed via equation 5 ($\mathbf{X}\mathbf{X}^\mathsf{T}$). However, alternative covariance estimators have been proposed to address certain situations, such as excessive noise or outliers. Such methods include median absolute deviation and Riemannian geometric averaging (Miah et al., 2020; Blum et al., 2019; Barachant et al., 2012). Covariance matrices can also be constructed using time points as features instead of, or in addition to, channels. Such time-delay embedding would turn the GED into a spatiotemporal filter, which is not discussed here. Rigorous comparisons of these estimators is beyond the scope of this paper.

3.4. Regularization

Regularization involves adding some constant to the cost function of an optimization algorithm. Regularization has several benefits in machine learning, including "smoothing" the solution to reduce overfitting and increasing numerical stability, particularly for reduced-rank datasets. Regularization is not necessary for GED, but can improve its performance when there are numerical stability issues.

There are several forms of regularization, including L1 (a.k.a. Lasso), L2 (ridge), Tikhonov, shrinkage, and others. Although there are mathematical differences between different regularization techniques, it is often the case that various methods produce comparable benefits (e.g., Wong et al., 2018).

Here I focus on shrinkage regularization, because it is simple and effective, and commonly used in GED applications (Lotte and Guan, 2010). Shrinkage regularization involves adding a small number to the

diagonal of the $\mathbf{R}$ matrix (and, thus, $\widetilde{\mathbf{R}}$ replaces $\mathbf{R}$ in equation 12). That small number is some fraction of the average of $\mathbf{R}$'s eigenvalues.

$$\widetilde{\mathbf{R}} = \mathbf{R}(1 - \gamma) + \gamma\alpha\mathbf{I}_M \tag{15}$$

$$\alpha = \sum_{i=1}^{M} \lambda_i \tag{16}$$

$\mathbf{I}_M$ is the $M \times M$ identity matrix, α is the average of all eigenvalues of $\mathbf{R}$, and γ is the regularization amount. Scaling down the $\mathbf{R}$ matrix by 1-γ ensures that $\widetilde{\mathbf{R}}$ and $\mathbf{R}$ have the same trace. This is useful because the trace of a matrix equals the sum of its eigenvalues, and thus the total "energy" of the eigenspectrum is preserved before and after regularization.

The MATLAB code is a direct implementation of equation 15.

```
gamma = .01; % 1% regularization
Rr = R*(1-gamma) + gamma*mean(eig(R))*eye(length(R));
```

The effect of shrinkage regularization on GED can be understood by noting that when $\gamma = 1$, the GED turns into a PCA on covariance matrix $\mathbf{S}$. Thus, regularization biases the GED towards high-variance solutions at the potential cost of reduced separability between $\mathbf{S}$ and $\mathbf{R}$. Because maximizing variance does not necessarily maximize relevance, excessive regularization may lead to results that are numerically stable but that are less sensitive to the desired contrast. For this reason, one should use as little regularization as possible but as much as necessary.

There are two ways to determine the amount of regularization to use. One is to pre-select a value for γ and apply that value to all datasets. In our experience, for example, $\gamma = 1\%$ improves the solution for matrices that are reduced-rank, noisy, or are difficult to separate, while having little or no noticeable effect on the GED solution for clean and easily separable matrices. A fixed level of regularization is not guaranteed to be an optimal amount for all datasets, but it has the advantage of being simple, unbiased, and applied equally to all of the data.

A second approach is to use cross-validation to empirically identify a γ value (Parra et al., 2019). The idea is to re-compute the GED multiple times using a range of γ values, and select the amount of

shrinkage that maximizes a data feature or experiment effect of interest (e.g., the ratio of narrowband to broadband power). A potential risk of cross-validation is that the γ parameter that maximizes one data feature (e.g., narrowband power) may reduce sensitivity to a different data feature (e.g., condition difference).

3.5. Sign uncertainty

Eigenvectors have a fundamental sign uncertainty. That is, because eigenvectors point along a 1D subspace (which is an infinitely long line that passes through the origin of the eigenspace), eigenvector $\mathbf{w}$ is the same as $-\mathbf{w}$. Sign uncertainties do not affect spectral or time-frequency analyses, but they do affect time-domain (ERP) analyses and topographical maps. This can cause interpretational and averaging difficulties, because, for example, a P3 ERP component can manifest as a negative deflection.

There are two principled methods for adjusting the sign of the eigenvector. One is to ensure that the electrode with the largest absolute value in the component map is positive. Thus, if the largest-magnitude electrode has a negative value, the entire eigenvector is multiplied by -1. (This method could be adapted to the ERP, for example ensuring that the component ERP deflection at 300 ms is positive.)

A second method is to assume that the eigenvector sign will be consistent for a majority of subjects, and then flip the signs of the subjects with the "minority sign." A group-averaged ERP or topographical map is computed (without any sign-flipping), and each individual subject's data are correlated with the group average. The eigenvector from subjects with a negative correlation to the group-average can be flipped, and the ERP or topographical map can be recomputed. This second method facilitates group-level data aggregation, but one must be cautious to avoid a biased statistical result if the group-level polarity is tested against zero.

3.6. Applying the spatial filter to the data

Each eigenvector (each column of the $\mathbf{W}$ matrix) is a set of weights for computing the weighted average of all data channel time series. This is the reason that the eigenvector is called a spatial filter. Each vector-data multiplication reduces the dimensionality of the data from $M \times T$ to $1 \times T$ (for M channels

and T time points). The filter must be applied to data with the same channels in the same order as was used to create the covariance matrices.

The spatial filter is created from one feature of the data, but can be applied to other data. For example, a spatial filter created by contrasting 10 Hz activity ($\mathbf{S}$ matrix) to broadband activity ($\mathbf{R}$ matrix) can be applied to the broadband signal (this is done in a validation simulation shown later). In this case, the component is designed to optimize alpha-band activity, but it is not constrained to pass through only alpha-band activity; activity in lower or higher frequencies may also be observed in the component time series. The interpretation of, for example, 40 Hz energy in the alpha-band filter would be that the higher-frequency activity, though spectrally separate from alpha, has a similar topography as alpha, and thus passes through the alpha-optimized spatial filter.

Applying the spatial filter to data beyond what was used to construct the $\mathbf{S}$ matrix also reduces the potential for bias or overfitting.

The component time series, computed as $\mathbf{w}^{\mathsf{T}}\mathbf{X}$, does not necessarily have the same units as the data $\mathbf{X}$ (e.g., μV or fT). This is because eigenvectors have arbitrary scaling. For some applications, the units don't matter. Indeed, units do not affect qualitative or statistical comparisons across conditions or over time. Furthermore, spectral and time-frequency analyses often involve normalization to decibel or percent change, meaning that the original units don't matter. Finally, measures of phase dynamics, synchronization, and correlation are unitless, and thus data scaling has no effect.

Nonetheless, it may be useful to scale the component time series. This can be done by z-normalizing the component time series, or unit-norming the eigenvector. The formula and MATLAB code for vector normalization are shown below.

$$\widetilde{\mathbf{w}} = \mathbf{w}/\|\mathbf{w}\| \tag{17}$$

```
w = evecs(:,1); % first eigenvector
w = w/norm(w); % scale to unit norm
```

In addition to the component time series, GED provides component maps that can be visually interpreted and averaged across individuals. These maps, however, are not the eigenvectors themselves; indeed, the

eigenvectors are generally not interpretable in a physiological sense, because they simultaneous encode weights to boost channels that maximize the contrast, and suppress channels that are irrelevant or noisy (Haufe et al., 2014). Instead, the component maps are obtained from the columns of $\mathbf{W}^{-\mathsf{T}}$. For GED, this can be simplified without the inverse as the covariance between the component time series and the channel data $((\mathbf{w}^{\mathsf{T}}\mathbf{X})\mathbf{X}^{\mathsf{T}})$ (Parra et al., 2005), or as the eigenvector times the covariance matrix $(\mathbf{w}^{\mathsf{T}}\mathbf{S})$ (Haufe et al., 2014) (a scaling factor is omitted for simplicity, as discussed above). This latter form provides a nice parallel of using the eigenvector to compute the component time series $(\mathbf{w}^{\mathsf{T}}\mathbf{X})$ and the component map $(\mathbf{w}^{\mathsf{T}}\mathbf{S})$.

A conceptualization is that the eigenvectors represent how to contort the channel data from "outside" to see the underlying source, whereas the component maps represent the underlying source projecting outwards onto the electrodes. That said, although the eigenvectors are not directly physiologically interpretable, they contain rich and high-spatial-frequency information, and have been used to identify empirical frequency band boundaries based on changes in spatial correlation patterns of eigenvectors across neighboring frequencies (Cohen, 2021).

3.7. Which component to use

Theoretically, the component with the largest eigenvalue has the best separability. However, this should be visually confirmed before applying the spatial filter to data and interpreting the results, because the component that mathematically best separates two data features is not guaranteed to be physiologically interpretable. For example, the top component in a GED that maximizes low-frequency activity may reflect eye-blink artifacts not entirely removed by ICA. In my lab, we typically produce a MATLAB figure for each dataset that shows the eigenspectrum, topoplots, and ERPs from the first 5 components (Figure 4). Our default choice is the largest component, but we sometimes select later components based on topography or ERP. Picking components must be done carefully to avoid biased data selection. One possibility is to create a topographical template of the "ideal" topography (e.g., a Gaussian around electrode FCz for a midfrontal component), and then pick the GED component that has the strongest spatial correlation with that template. It is also important that any method of component selection is orthogonal to the statistical analyses performed, as discussed earlier.
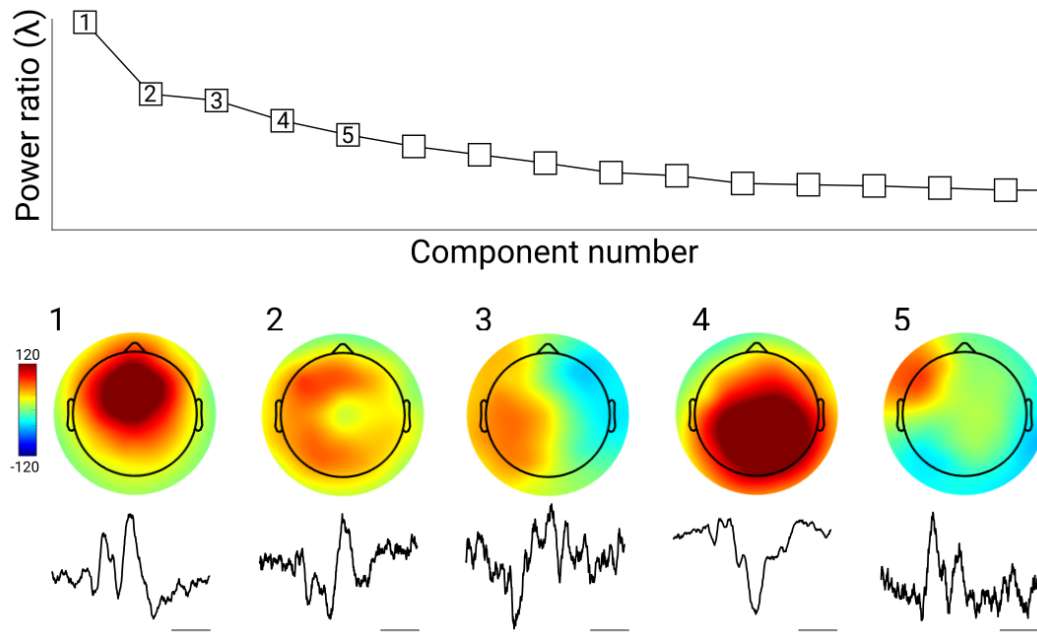
Figure 4: Picking a component based on visual inspection. A GED was performed on 64-channel EEG with the **S** matrix computed from 6-Hz filtered activity and the **R** matrix computed from broadband activity. Both matrices were created from data in a 100-800 ms window relative to trial onset. The top row shows the eigenspectum (also called a scree plot), with 15/64 components shown. The middle row shows the component maps (all maps have the same color scale shown on the left), and the bottom row shows the ERP (the spatial filter was created from narrowband data but was applied back to the broadband signal). Horizontal bar indicates 300 ms. In this example, the first component isolated midfrontal theta, while the fourth component reflects parietal theta.

Because there is theoretical motivation and mathematical justification for selecting the component with the largest eigenvalue, it is not necessary to statistically evaluate the top component, unless there is a cause for concern about overfitting. For example, if a component is created based on maximizing alpha-band activity over all conditions, the top component can be computed and the inferential statistical analyses can be done on the differences in alpha power or phase dynamics across the experiment conditions, and not on the significance of the component *per se*. Analogously, in univariate analyses, one typically selects a data channel and then applies a statistical comparison across conditions, without necessarily testing whether that channel has a significant amount of activity *per se*.

3.8. Complex GED solutions

The GED may return complex solutions, which means complex-valued eigenvalues and eigenvectors. When there are complex solutions to an eigenvalue problem, they appear in conjugate pairs. Complex solutions are often found in components with smaller eigenvalues, while the larger components have zero-valued imaginary parts. The eigenvalues (and therefore also the eigenvectors) of symmetric matrices are guaranteed to be real-valued, but $\mathbf{R}^{-1}\mathbf{S}$ is not symmetric, and therefore its eigenvalues may be complex-valued.

Complex solutions can arise if the covariance matrices are reduced-rank or ill-conditioned, and usually indicate that the covariance matrices have poor signal-to-noise characteristics, reduced rank, or that $\mathbf{S}$ and $\mathbf{R}$ are difficult to separate. Thus, complex solutions may indicate problems with the data that should be addressed. On the other hand, reduced-rank matrices do not necessarily produce complex solutions if the data are high quality and the covariance matrices are separable.

There are several approaches to dealing with complex solutions: First, use more data to create the covariance matrices (e.g., longer time windows or wider spectral bands, or more trials); Second, redefine the GED contrast so that the matrices are more separable (e.g., all conditions against the inter-trial-interval instead of one condition against another); Third, if the data are reduced-rank, compress the data from M (channels) to r (matrix rank) dimensions, e.g., using PCA, and then run GED on the compressed-data covariance matrices; Fourth, apply regularization to fill in null dimensions and thus force the rank to M.

3.9. Subject-specific or group-level decomposition

If the same electrodes are placed in the same locations in different individuals, then the researcher has the choice to perform the GED separately on each individual, or to perform one GED on covariance matrices that are averaged across individuals. This is analogous to group-ICA, where data from all individuals are pooled to derive a single set of components that is based on data from all subjects (Calhoun et al., 2001, 2009). It is also possible to utilize GED to maximize consistency across individuals (Parra et al., 2019), and simultaneously within- and across-subject variance (Tanaka, 2020).

Needless to say, the purpose of acquiring data from multiple individuals is to make inferences about a population, which motivates group-level statistical analyses. However, we must consider the features of the data we expect to be consistent across individuals, versus the features of the data we allow to vary across individuals. Performing the GED on each subject individually is based on the expectation that the statistical contrast ($\mathbf{S}$ vs. $\mathbf{R}$) is consistent across individuals, whereas the exact topography can vary due to factors such as cortical folding, functional localization, and electrode placement. Allowing for individual topographical variability also opens possibilities to explore individual differences in age, genetics, task performance, disease state, or any other subgroup analysis.

3.10. Two-stage compression and separation

A two-stage GED involves (1) data compression via PCA and then (2) source separation via GED. This is useful when there are many data channels, severely reduced-rank covariance matrices, or complex-valued GED solutions. The initial compression stage should be added to the analysis pipeline only when necessary, e.g., when GED returns unsatisfactory results while the data matrices are very large (e.g., hundreds of channels).

The goal of the first stage is to produce an $N \times T$ dataset, where N is the number of principal components with $N < M$. This is obtained as the eigenvectors matrix $\mathbf{V}$ (of all the data, not from the GED) times the data matrix, using only the top N components.

$$\mathbf{Y} = \mathbf{V}_{1:N}^{\mathsf{T}} \mathbf{X} \tag{18}$$

The subscript $1:N$ indicates to use the first N columns in the eigenvectors matrix. The number of PCA components to retain (N) can be based on one of two factors. First, one can use the rank of the data matrix as the dimensionality. The advantage of this approach is that it prevents information loss in the compression. In other words, the original and compressed data have the same information, but the post-PCA data matrix has fewer dimensions. This guarantees full-rank matrices in the GED and thus should improve numerical stability of the solution.

A second method is to select the number of compressed dimensions based on a statistical criterion. If the eigenvalues of the PCA are converted to percent total variance explained, then a variance threshold of, e.g., 0.1% can be applied. The compressed signal thus comprises only principal components that

contribute more than 0.1% of the multivariate signal power. In this case, the data matrix used for GED contains less information than the original data. The motivation for this approach is that principal components that account for a tiny amount of variance might reflect noise or unrelated activity. On the other hand, the PCA is driven by total variance — and variance does not equal relevance (de Cheveigné and Parra, 2014). Thus, there is a risk that some of the rejected information is important for separating $\mathbf{S}$ from $\mathbf{R}$.

The GED is then performed on the reduced-dimensional data ($\mathbf{Y}$ instead of $\mathbf{X}$). The resulting component maps, however, should be interpreted in the channel space for visualization and cross-subject averaging. Therefore, the spatial filter can be right-multiplied by the first N PC vectors (the first N columns in $\mathbf{V}$), which will "undo" the first-stage compression: $\mathbf{w}^\mathsf{T}\mathbf{S}\mathbf{V}_{1:N}^\mathsf{T}$.

4. Validation in simulations

We and others have done many simulations over the years to demonstrate that GED is highly accurate at reconstructing simulated activity. Some simulation protocols and results are published (e.g., Cohen, 2017a; Barzegaran et al., 2019; Sabbagh et al., 2020; Nikulin et al., 2011); many others are done during piloting, testing, exploring, and teaching. Simulations have the advantages of allowing full control over variables such as the amount and color of noise, the signal strength and characteristics, the number of trials, electrodes, and sources, and so on. On the other hand, simulations rarely capture the full complexity of EEG data and accompanying artifacts, and thus one should not assume that GED (or any other method) always identifies The Truth in empirical data simply because it performed well in a simulation.

Protocols for simulating EEG data range in biological and physical plausibility (Næss et al., 2021; Barzegaran et al., 2019; Aznan et al., 2019; Anzolin et al., 2021). For example, a pure sine wave mixed with Gaussian noise can be useful to evaluate the performance of an algorithm, but is not a neurophysiologically plausible time series. More sophisticated models incorporate morphological and physiological details of neurons, realistic geometry, and electrical forward models (e.g., Næss et al., 2021), and therefore can be used to test more precise hypotheses. The approach taken in this paper is to balance physiological plausibility (anatomical forward model, narrowband signals) with simplicity

(e.g., narrowband-filtered noise to mimic a neural oscillation).

The simulation approach presented here involves generating time series signals and noise at thousands of dipole locations in the brain, passing those time series through an anatomical forward model (a.k.a. a leadfield matrix) to generate EEG activity at simulated electrodes, and then analyzing the electrode-level activity. The results can be qualitatively and quantitatively compared to the signals that were generated in the dipoles. This pipeline is useful because it is fast and efficient, allows the researcher to control the strength, nature, and locations of the signal and noise, while also providing physiologically plausible EEG topographical characteristics.

MATLAB and Python code that accompanies this tutorial are available at github.com/mikexcohen/GED_tutorial (no additional toolboxes are required), and includes two simulations. The first is very simple (1000 ms of data, of which 500 ms is a pure sine wave) and illustrates the high reconstruction accuracy of GED and the poor performance of PCA. This provides an introduction to implementing simulated data and performing GED, and is not shown here.

The second simulation shows a better use-case: isolating an alpha-band component embedded in noise during simulated resting-state data. Figure 5A shows the projection of the "alpha dipole" activation onto the scalp. A GED was computed on narrowband-filtered signals ($\mathbf{S}$) compared to broadband ($\mathbf{R}$). Figure 5B shows the scree plot, which clearly indicates one component that dominates the multivariate data. For reference, a traditional univariate analysis was applied (Welch's method on each electrode). At this SNR level, the true alpha dynamics are barely visible in the scalp and power spectrum. At higher signal gain values, the electrode data more accurately recovered the ground truth activity (not shown here, but easy to demonstrate with a minor adjustment to the code).
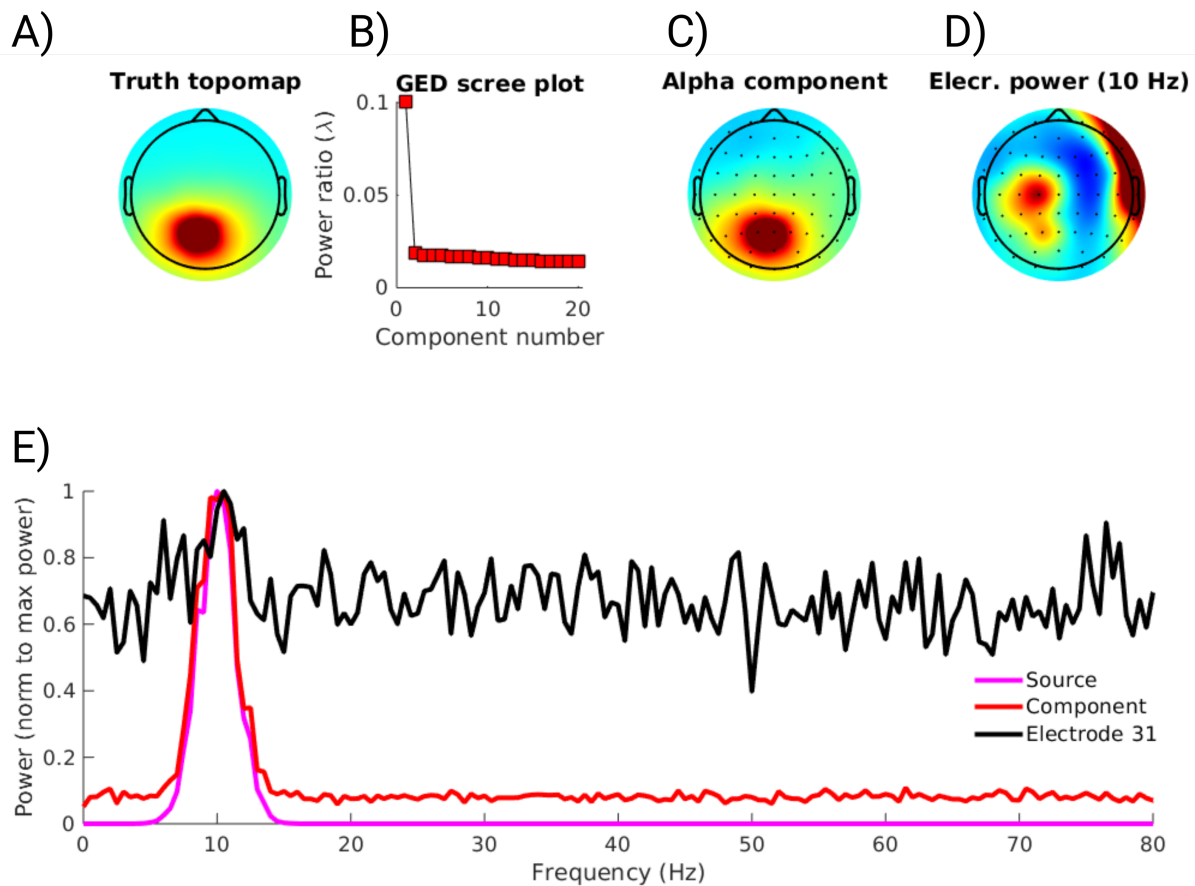
Figure 5: EEG simulation demonstrates the advantage of GED over electrode-level analyses. Data were generated in 2004 brain dipoles and projected to 64 scalp electrodes (black dots). One dipole in parietal cortex was simulated as a non-stationary alpha oscillation; all other dipoles contained Gaussian random noise with a $\sigma = 1$ standard deviation. The source dipole projection is visualized in panel A. 100 segments of 2 seconds each were generated to simulate resting-state EEG data. A GED was run on 10 Hz filtered signal ($\mathbf{S}$) vs. broadband ($\mathbf{R}$). The eigenspectrum is shown in panel B. Note that all component eigenvalues are much smaller than 1, because there is more energy in the entire spectrum compared to only 10 Hz. The top component had a topography (panel C) that closely matched the ground truth. Power spectra were computed at each of 64 electrodes, and the 10 Hz power is shown in panel D. Panel E shows the power spectra from electrode 31 (closest to the ground truth dipole maximal projection), the 10 Hz GED component, and the dipole source activity (ground truth). Each topographical map has a different numerical range, so they should be interpreted qualitatively.

5. Section 5. Interpretation

5.1. Does one component mean one "source"?

As written at the outset of this paper, the term "source" has multiple interpretations, so the title of this subsection is itself an ill-posed question.

It is important to keep in mind that GED is a contrast-enhancing procedure that is based on statistical characteristics of the data. A GED component may correspond to a physiological or cognitive source, but this is not trivially guaranteed. Indeed, a GED component is simply a weighted sum of voltage values recorded by a collection of electrodes, which itself is not a "source." In this sense, GED, along with other multivariate methods including ICA, can be seen as tools that, along with careful experiment design and data analyses, facilitate an attempt at source separation. Analogously, phase synchronization between brain regions does not directly indicate functional connectivity; but synchronization, along with careful experiment design and theoretical justification, can be interpreted as reflecting functional connectivity.

Another consideration is that GED provides linear basis vectors for characterizing the data. On the one hand, this is appropriate given that electrical fields mix linearly at the electrodes. However, a linear decomposition does not mean that the data are optimally described by linear basis vectors, as illustrated in Figure 6.
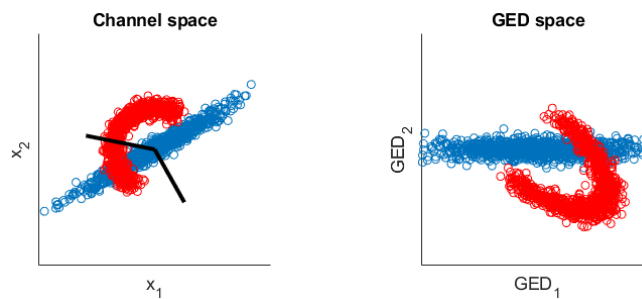

Figure 6: Because GED is a linear decomposition, nonlinear-distributed sources may not be suitably captured by GED. The black lines in the left panel show the GED eigenvectors.

Another issue that impedes interpreting a component as a source is that one source can manifest in multiple components, which is discussed more below.

5.2. Multiple significant components

Although the component with the largest eigenvalue is theoretically the most relevant (that is, it maximally enhances the desired contrast), it is possible that multiple linearly separable dimensions separate $\mathbf{S}$ from $\mathbf{R}$ (e.g., Figure 2). For example, in a GED comparing task vs. baseline, multiple perceptual and motor operations will separate task performance from resting state. When appropriate, permutation testing can be used to determine the number of statistically significant components from a GED.

The exact interpretation of multiple significant components is not entirely clear. The simple interpretation is that they reflect distinct brain features that separate the two covariance matrices. In other words, that there is a multidimensional subspace that separates $\mathbf{S}$ and $\mathbf{R}$, and the number of significant components is the dimensionality of that subspace. Whether each dimension in this space maps onto a neural or cognitive source, however, is not guaranteed.

Furthermore, a single source can be appear in multiple components, for example, due to changes in covariance structure resulting from traveling waves, rotating dipoles, or head movements that change the locations of the EEG electrodes or placement inside the MEG. A matrix with repeated eigenvalues (or, in the case of noise, numerically distinct but statistically indistinguishable) can indicate the presence of an eigenplane. In this case, the eigenvectors associated with the repeated eigenvalues are basis vectors for that subspace, and any linear combination of the vectors is also part of the same "source."

5.3. Do components reflect dipoles?

GED components are a linear weighted combination of data channels, where the weights are defined by maximizing the difference between two covariance matrices. There are no constraints or assumptions about spatial smoothness or anatomical localization. This means that GED components do not necessarily correspond to a single anatomical generator that can be described by a dipole. And there is no need for them to be dipolar. This contrasts with a common view of ICA: Indeed, independent components are often evaluated in terms of their "dipolarity" (that is, their fit to a single dipole projection) (Delorme et al., 2012) (although ICs are not constrained to find only dipolar components). Although dipolarity may indicate physiological plausibility for sensory or motor processes that are likely

to be driven by a spatially restricted population of neurons, neuroscience is increasingly moving towards the idea that many cognitive processes are implemented by spatially distributed networks (Bassett and Sporns, 2017). This suggests that the fit to a single dipole is not necessarily an important consideration when evaluating a spatial filter.

We have observed that GED components are sometimes dipole-like and other times spatially distributed (e.g., Figure 4). In simulations, non-dipolar topographies can arise from noisy data, although the component time series accurately reflects the simulated ground truth. Therefore, the component topography is one indicator of the spatial filter quality, but it should not be the sole basis for selecting or rejecting a GED component. It is possible to project the component map onto the brain given a suitable forward model (Cohen and Gulbinaite, 2017; Hild and Nagarajan, 2009), but this is simply a useful visualization, not a guarantee of an anatomical origin.

The recommendation here is to interpret the GED components in a conservative manner corresponding to how the components were created, which is maximizing a statistical contrast in multivariate data. The physiological interpretation is that each GED component captures a functionally cohesive and temporally coherent network, which could be spatially restricted or spatially distributed. That said, a strongly dipolar component map suggests a spatially restricted anatomical source (Nunez et al., 2006), which can facilitate a discussion of the spatial extent of the component generator.

5.4. There is no escaping the ill-posed problem

Finally, it is important to keep in mind that all source separation problems — be they statistical or anatomical — are fundamentally *ill-posed*. This means that there are more unknown factors than observable measurements. Just because GED gives a solution does not mean that it gives the *correct* solution, or even the only solution. There is an infinite number of possible brain generator configurations that could produce an observed EEG signal; as researchers, we must define an objective function to satisfy based on assumptions about the data.

Ultimately, the ill-posed problem means that our interpretation of GED (or any other multivariate method) should be appropriately conservative: We cannot know if we are uncovering true sources in the brain, but we can be confident that the components reflect patterns in the data that boost

an hypothesis-relevant signal while also minimizing noise and individual variability that contaminate traditional univariate data selection methods.

6. Future directions and conclusions

6.1. Multimodal datasets

GED does not "know" or "care" about the origin or type of data. It is possible to include multiple sources of data into the same data matrix, for example, EEG+MEG (Zuure et al., 2020), LFP+single units (Cohen et al., 2021), EEG+FMRI, or any other combinations of brain, behavior, and body read-outs (Dähne et al., 2014). Any continuous signals that are relevant for an hypothesis can be included in the covariance matrices. Multimodal datasets therefore offer a powerful means to investigate how dynamics are integrated across multiple spatial and temporal scales, and across multiple neural and behavioral observable variables.

Temporal up/downsampling may be necessarly to have data values at each time point. And data in very different scales should be suitably normalized. For example, EEG data in microvolts and MEG data in Tesla differ in raw numerical values by roughly 14 orders of magnitude.

6.2. Better neurobiological and computational interpretations

As a statistical method to facilitate practical data analysis, GED (and many other multivariate techniques) is unambiguously useful. Its mathematical foundations are clear, and uncountable simulation and empirical studies demonstrate its flexibility and power.

However, as neuroscientists, we are not interested in data methods *per se*; we are interested in using the results of data analyses to understand brain function and its role in behavior and disease. Therefore, it would be ideal if GED components could be interpreted in a more physiological manner. It is currently premature to make claims about neural circuit configurations purely on the basis of GED topographies or eigenspectra, but there are three paths to being able to link GED to neural circuit configurations.

The first avenue is empirical multimodal studies. The idea would be to use combined EEG+fMRI, or combined EEG/LFP+spikes. One could then empirically evaluate the sensitivity of GED to distinct spatial or neural functional configurations.

A second promising avenue for linking GED to neural circuits comes from developments in computational modeling that are increasingly biophysically and morphologically accurate (e.g., Næss et al., 2021; Neymotin et al., 2020). Such generative forward models allow researchers to simulate data at the neural level (including specifying different types of neurons across different cortical layers, density of inter-neuronal connectivity, etc.) and then produce scalp-level MEEG data. One could then apply GED to the biophysically simulated EEG data after manipulating neurophysiologically meaningful circuit or connectivity parameters.

Finally, the component topography could be inspected for evidence of a dipolar vs. distributed pattern. Although (as written earlier), GED results should not be evaluated purely based on topographical distributions, the topography can be used to provide evidence for an anatomically restricted vs. distributed set of generators. (On the other hand, a seemingly distributed topography may also reflect noisy data.)

6.3. Conclusions

The goal of this paper was to present an approachable tutorial on the use of GED for multivariate source separation. There are several other lucid introductions to the theory and advantages of GED (Parra et al., 2005; de Cheveigné and Parra, 2014; Haufe et al., 2014; Tomé, 2006; Blankertz et al., 2007); the contribution of this paper is to focus on intuition and practical aspects that researchers might encounter while implementing GED. Although the discussion centered on electrophysiology, these methods are generic and can be applied to any type of multichannel time series data, including fMRI or calcium imaging.

Arguably, progress in neuroscience will require moving beyond traditional mass-univariate analyses, and incorporating multivariate source separation methods. GED is certainly not the only useful multivariate method, nor is it suitable for all situations and datasets. But it has many statistical and practical advantages, and its flexibility and high SNR should make it a tool in the toolbox of every neuroscientist collecting multichannel time series data.

Acknowledgements

References

Ai, Q., Liu, Q., Meng, W., Xie, S.Q., 2018. Chapter 6 - EEG-Based Brain Intention Recognition, in: Advanced Rehabilitative Technology. Academic Press, Cambridge, MA, USA, pp. 135–166. doi:10.1016/B978-0-12-814597-5.00006-0.

Allen, J.J.B., Coan, J.A., Nazarian, M., 2004. Issues and assumptions on the road from raw signals to metrics of frontal EEG asymmetry in emotion. Biol. Psychol. 67, 183–218. doi:10.1016/j.biopsycho.2004.03.007, arXiv:15130531.

Anzolin, A., Toppi, J., Petti, M., Cincotti, F., Astolfi, L., 2021. SEED-G: Simulated EEG Data Generator for Testing Connectivity Algorithms. Sensors 21, 3632. doi:10.3390/s21113632.

Aznan, N.K.N., Atapour-Abarghouei, A., Bonner, S., Connolly, J., Moubayed, N.A., Breckon, T., 2019. Simulating Brain Signals: Creating Synthetic EEG Data via Neural-Based Generative Models for Improved SSVEP Classification. arXiv doi:10.1109/IJCNN.2019.8852227, arXiv:1901.07429.

Barachant, A., Bonnet, S., Congedo, M., Jutten, C., 2012. Multiclass brain-computer interface classification by Riemannian geometry. IEEE Trans. Biomed. Eng. 59, 920–928. doi:10.1109/TBME.2011.2172210, arXiv:22010143.

Barzegaran, E., Bosse, S., Kohler, P.J., Norcia, A.M., 2019. EEGSourceSim: A framework for realistic simulation of EEG scalp data using MRI-based forward models and biologically plausible signals and noise. J. Neurosci. Methods 328, 108377. doi:10.1016/j.jneumeth.2019.108377.

Bassett, D.S., Sporns, O., 2017. Network neuroscience. Nat. Neurosci. 20, 353–364. doi:10.1038/nn.4502, arXiv:28230844.

Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., Muller, K.R., 2007. Optimizing spatial filters for robust eeg single-trial analysis. IEEE Signal processing magazine 25, 41–56.

Blum, S., Jacobsen, N.S.J., Bleichner, M.G., Debener, S., 2019. A Riemannian Modification of Artifact Subspace Reconstruction for EEG Artifact Handling. Front. Hum. Neurosci. 0. doi:10.3389/fnhum.2019.00141.

Boynton, G.M., Engel, S.A., Heeger, D.J., 2012. Linear Systems Analysis of the fMRI Signal. Neuroimage 62, 975. doi:10.1016/j.neuroimage.2012.01.082.

Buzsáki, G., 2010. Neural syntax: cell assemblies, synapsembles and readers. Neuron 68, 362. doi:10.1016/j.neuron.2010.09.023.

Calhoun, V.D., Adali, T., Pearlson, G.D., Pekar, J.J., 2001. A method for making group inferences from functional MRI data using independent component analysis. Hum. Brain Mapp. 14, 140–151. doi:10.1002/hbm.1048, arXiv:11559959.

Calhoun, V.D., Liu, J., Adali, T., 2009. A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. Neuroimage 45, Suppl. doi:10.1016/j.neuroimage.2008.10.057, arXiv:19059344.

de Cheveigné, A., Arzounian, D., 2015. Scanning for oscillations. Journal of neural engineering 12, 066020.

de Cheveigné, A., Parra, L.C., 2014. Joint decorrelation, a versatile tool for multichannel data analysis. Neuroimage 98, 487–505.

Cichy, R.M., Pantazis, D., 2017. Multivariate pattern analysis of MEG and EEG: A comparison of representational structure in time and space. Neuroimage 158, 441–454. doi:10.1016/j.neuroimage.2017.07.023, arXiv:28716718.

Cohen, M.X., 2017a. Comparison of linear spatial filters for identifying oscillatory activity in multichannel data. Journal of neuroscience methods 278, 1–12.

Cohen, M.X., 2017b. Multivariate cross-frequency coupling via generalized eigendecomposition. ELife 6, e21792.

Cohen, M.X., 2021. A data-driven method to identify frequency boundaries in multichannel electro-physiology data. Journal of Neuroscience Methods 347, 108949.

Cohen, M.X., Englitz, B., França, A.S.C., 2021. Large- and multi-scale networks in the rodent brain during novelty exploration. eNeuro doi:10.1523/ENEURO.0494-20.2021.

Cohen, M.X., Gulbinaite, R., 2017. Rhythmic entrainment source separation: Optimizing analyses of neural responses to rhythmic sensory stimulation. Neuroimage 147, 43–56. doi:10.1016/j.neuroimage.2016.11.036, arXiv:27916666.

Dähne, S., Meinecke, F.C., Haufe, S., Höhne, J., Tangermann, M., Müller, K.R., Nikulin, V.V., 2014. Spoc: a novel framework for relating the amplitude of neuronal oscillations to behaviorally relevant parameters. NeuroImage 86, 111–122.

Dan, J., Vandendriessche, B., Van Paesschen, W., Weckhuysen, D., Bertrand, A., 2020. Computationally-Efficient Algorithm for Real-Time Absence Seizure Detection in Wearable Electroencephalography. Int. J. Neural Syst. 30, 2050035. doi:10.1142/S0129065720500355, arXiv:32808854.

Das, N., Vanthornhout, J., Francart, T., Bertrand, A., 2020. Stimulus-aware spatial filtering for single-trial neural response and temporal response function estimation in high-density EEG with applications in auditory research. Neuroimage 204, 116211. doi:10.1016/j.neuroimage.2019.116211.

Debener, S., Thorne, J., Schneider, T.R., Viola, F.C., 2010. Using ICA for the Analysis of Multi-Channel EEG Data, in: Simultaneous EEG and fMRI. Oxford University Press, Oxford, England, UK, pp. 121–135. doi:10.1093/acprof:oso/9780195372731.003.0008.

Delorme, A., Palmer, J., Onton, J., Oostenveld, R., Makeig, S., 2012. Independent EEG Sources Are Dipolar. PLoS One 7, e30135. doi:10.1371/journal.pone.0030135.

van Dijk, J.A., Fracasso, A., Petridou, N., Dumoulin, S.O., 2020. Linear systems analysis for laminar fMRI: Evaluating BOLD amplitude scaling for luminance contrast manipulations. Sci. Rep. 10, 1–15. doi:10.1038/s41598-020-62165-x.

Dmochowski, J.P., Greaves, A.S., Norcia, A.M., 2015. Maximally reliable spatial filtering of steady state visual evoked potentials. Neuroimage 109, 63. doi:10.1016/j.neuroimage.2014.12.078.

Duprez, J., Gulbinaite, R., Cohen, M.X., 2020. Midfrontal theta phase coordinates behaviorally relevant brain computations during cognitive control. NeuroImage 207, 116340.

Ghojogh, B., Karray, F., Crowley, M., 2019. Eigenvalue and Generalized Eigenvalue Problems: Tutorial. arXiv URL: https://arxiv.org/abs/1903.11240v1, arXiv:1903.11240.

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.D., Blankertz, B., Bießmann, F., 2014. On the interpretation of weight vectors of linear models in multivariate neuroimaging. Neuroimage 87, 96–110.

Head, M.L., Holman, L., Lanfear, R., Kahn, A.T., Jennions, M.D., 2015. The Extent and Consequences of P-Hacking in Science. PLoS Biol. 13, e1002106. doi:10.1371/journal.pbio.1002106.

Hebb, D.O., 1949. The Organization of Behavior: A Neuropsychological Theory. Wiley, Chichester, England, UK.

Hild, K.E., Nagarajan, S.S., 2009. Source localization of EEG/MEG data by correlating columns of ICA and lead field matrices. IEEE Trans. Biomed. Eng. 56, 2619–2626. doi:10.1109/TBME.2009.2028615, `arXiv:19695993`.

Hyvärinen, A., 2011. Testing the ica mixing matrix based on inter-subject or inter-session consistency. NeuroImage 58, 122–136.

Ioannidis, J.P.A., 2008. Why most discovered true associations are inflated. Epidemiology 19, 640–648. doi:10.1097/EDE.0b013e31818131e7, `arXiv:18633328`.

King, J.R., Dehaene, S., 2014. Characterizing the dynamics of mental representations: the temporal generalization method. Trends in cognitive sciences 18, 203. doi:10.1016/j.tics.2014.01.002.

Kriegeskorte, N., Simmons, W.K., Bellgowan, P.S.F., Baker, C.I., 2009. Circular analysis in systems neuroscience: the dangers of double dipping - Nature Neuroscience. Nat. Neurosci. 12, 535–540. doi:10.1038/nn.2303.

Lawhern, V.J., Solon, A.J., Waytowich, N.R., Gordon, S.M., Hung, C.P., Lance, B.J., 2018. EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces. J. Neural Eng. 15, 056013. doi:10.1088/1741-2552/aace8c, `arXiv:29932424`.

Lotte, F., Guan, C., 2010. Regularizing common spatial patterns to improve bci designs: unified theory and new algorithms. IEEE Transactions on biomedical Engineering 58, 355–362.

Makeig, S., Westerfield, M., Jung, T.P., Enghoff, S., Townsend, J., Courchesne, E., Sejnowski, T.J., 2002. Dynamic brain sources of visual evoked responses. Science 295, 690–694. doi:10.1126/science.1066168, `arXiv:11809976`.

Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods 164, 177–190. doi:10.1016/j.jneumeth.2007.03.024, `arXiv:17517438`.

Miah, A.S.M., Rahim, M.A., Shin, J., 2020. Motor-Imagery Classification Using Riemannian Geometry with Median Absolute Deviation. Electronics 9, 1584. doi:10.3390/electronics9101584.

Næss, S., Halnes, G., Hagen, E., Hagler, Jr., D.J., Dale, A.M., Einevoll, G.T., Ness, T.V., 2021. Biophysically detailed forward modeling of the neural origin of EEG and MEG signals. Neuroimage 225, 117467. doi:10.1016/j.neuroimage.2020.117467, arXiv:33075556.

Neymotin, S.A., Daniels, D.S., Caldwell, B., McDougal, R.A., Carnevale, N.T., Jas, M., Moore, C.I., Hines, M.L., Hämäläinen, M., Jones, S.R., 2020. Human Neocortical Neurosolver (HNN), a new software tool for interpreting the cellular and network origin of human MEG/EEG data. eLife doi:10.7554/eLife.51214.

Nichols, T.E., Holmes, A.P., 2002. Nonparametric permutation tests for functional neuroimaging: A primer with examples. Hum. Brain Mapp. 15, 1–25. doi:10.1002/hbm.1058.

Nikulin, V.V., Nolte, G., Curio, G., 2011. A novel method for reliable and fast extraction of neuronal eeg/meg oscillations on the basis of spatio-spectral decomposition. NeuroImage 55, 1528–1535.

Nunez, P.L., Nunez, E.P.B.E.P.L., Srinivasan, R., Press, O.U., Srinivasan, A.P.C.S.R., 2006. Electric Fields of the Brain. Oxford University Press, Oxford, England, UK.

Parra, L., Sajda, P., 2003. Blind source separation via generalized eigenvalue decomposition. The Journal of Machine Learning Research 4, 1261–1269.

Parra, L.C., Haufe, S., Dmochowski, J.P., 2019. Correlated components analysis - extracting reliable dimensions in multivariate data. arXiv:1801.08881.

Parra, L.C., Spence, C.D., Gerson, A.D., Sajda, P., 2005. Recipes for the linear analysis of EEG. Neuroimage 28, 326–341. doi:10.1016/j.neuroimage.2005.05.032, arXiv:16084117.

Ritchie, J.B., Kaplan, D.M., Klein, C., 2019. Decoding the Brain: Neural Representation and the Limits of Multivariate Pattern Analysis in Cognitive Neuroscience. British J. Philos. Sci. 70, 581–607. doi:10.1093/bjps/axx023, arXiv:31086423.

Rivet, B., Cecotti, H., Souloumiac, A., Maby, E., Mattout, J., 2011. Theoretical analysis of xDAWN algorithm: Application to an efficient sensor selection in a p300 BCI,

in: 2011 19th European Signal Processing Conference. IEEE, pp. 1382–1386. URL: `https://ieeexplore.ieee.org/document/7073970`.

Rivet, B., Souloumiac, A., 2013. Optimal linear spatial filters for event-related potentials based on a spatio-temporal model: Asymptotical performance analysis. Signal Processing 93, 387–398.

Rivet, B., Souloumiac, A., Attina, V., Gibert, G., 2009. xDAWN algorithm to enhance evoked potentials: application to brain-computer interface. IEEE Trans. Biomed. Eng. 56, 2035–2043. doi:10.1109/TBME.2009.2012869, `arXiv:19174332`.

Sabbagh, D., Ablin, P., Varoquaux, G., Gramfort, A., Engemann, D.A., 2020. Predictive regression modeling with MEG/EEG: from source power to signals and cognitive states. Neuroimage 222, 116893. doi:10.1016/j.neuroimage.2020.116893.

Schirrmeister, R.T., Springenberg, J.T., Fiederer, L.D.J., Glasstetter, M., Eggensperger, K., Tangermann, M., Hutter, F., Burgard, W., Ball, T., 2017. Deep learning with convolutional neural networks for EEG decoding and visualization. Hum. Brain Mapp. 38, 5391. doi:10.1002/hbm.23730.

Särelä, J., Valpola, H., 2005. Denoising Source Separation. Journal of Machine Learning Research 6, 233–272.

Tanaka, H., 2020. Group task-related component analysis (gTRCA): a multivariate method for inter-trial reproducibility and inter-subject similarity maximization for EEG data analysis - Scientific Reports. Sci. Rep. 10, 1–17. doi:10.1038/s41598-019-56962-2.

Tanaka, H., Miyakoshi, M., 2019. Cross-correlation task-related component analysis (xTRCA) for enhancing evoked and induced responses of event-related potentials. Neuroimage 197, 177–190. doi:10.1016/j.neuroimage.2019.04.049, `arXiv:31034968`.

Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., Doyne Farmer, J., 1992. Testing for nonlinearity in time series: the method of surrogate data. Physica D 58, 77–94. doi:10.1016/0167-2789(92)90102-S.

Tomé, A.M., 2006. The generalized eigendecomposition approach to the blind source separation problem. Digital Signal Processing 16, 288–302.

Uhlhaas, P.J., Pipa, G., Lima, B., Melloni, L., Neuenschwander, S., Nikolić, D., Singer, W., 2009. Neural synchrony in cortical networks: history, concept and current status. Front. Integr. Neurosci. 3, 17. doi:10.3389/neuro.07.017.2009, `arXiv:19668703`.

Wong, D.D.E., Fuglsang, S.A., Hjortkjær, J., Ceolini, E., Slaney, M., de Cheveigné, A., 2018. A Comparison of Regularization Methods in Forward and Backward Models for Auditory Attention Decoding. Front. Neurosci. 12. doi:10.3389/fnins.2018.00531.

Wouters, J., Kloosterman, F., Bertrand, A., 2018. Towards online spike sorting for high-density neural probes using discriminative template matching with suppression of interfering spikes. J. Neural Eng. 15, 056005. doi:10.1088/1741-2552/aace8a, `arXiv:29932426`.

Yao, D., Qin, Y., Hu, S., Dong, L., Vega, M.L.B., Sosa, P.A.V., 2019. Which Reference Should We Use for EEG and ERP practice? Brain Topogr. 32, 530–549. doi:10.1007/s10548-019-00707-x.

Zuure, M.B., Cohen, M.X., 2021. Narrowband multivariate source separation for semi-blind discovery of experiment contrasts. Journal of Neuroscience Methods 350, 109063.

Zuure, M.B., Hinkley, L.B., Tiesinga, P.H.E., Nagarajan, S.S., Cohen, M.X., 2020. Multiple Midfrontal Thetas Revealed by Source Separation of Simultaneous MEG and EEG. J. Neurosci. 40, 7702–7713. doi:10.1523/JNEUROSCI.0321-20.2020.