Contents lists available at ScienceDirect





Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth

Geometric classification of brain network dynamics via conic derivative discriminants



Matthew F. Singh^{a,b,c,*}, Todd S. Braver^b, ShiNung Ching^{c,d}

^a Department of Neuroscience, Washington University in St. Louis, St. Louis, MO, USA

^b Department of Psychology, Washington University in St. Louis, St. Louis, MO, USA

^c Department of Electrical and Systems Engineering, Washington University in St. Louis, St. Louis, MO, USA

^d Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, MO, USA

ARTICLE INFO

Keywords: Pattern classification

EEG

Dynamical systems

Neural dynamics

Machine learning

Directional statistics

ABSTRACT

Background: Over the past decade, pattern decoding techniques have granted neuroscientists improved anatomical specificity in mapping neural representations associated with function and cognition. Dynamical patterns are of particular interest, as evidenced by the proliferation and success of frequency domain methods that reveal structured spatiotemporal rhythmic brain activity. One drawback of such approaches, however, is the need to estimate spectral power, which limits the temporal resolution of classification.

New method: We propose an alternative method that enables classification of dynamical patterns with high temporal fidelity. The key feature of the method is a conversion of time-series data into temporal derivatives. By doing so, dynamically-coded information may be revealed in terms of geometric patterns in the phase space of the derivative signal.

Results: We derive a geometric classifier for this problem which simplifies into a straightforward calculation in terms of covariances. We demonstrate the relative advantages and disadvantages of the technique with simulated data and benchmark its performance with an EEG dataset of covert spatial attention. We reveal the timecourse of covert spatial attention and, by mapping the classifier weights anatomically, its retinotopic organization.

Comparison with existing method: We especially highlight the ability of the method to provide strong group-level classification performance compared to existing benchmarks, while providing information that is complementary with classical spectral-based techniques. The robustness and sensitivity of the method to noise is also examined relative to spectral-based techniques.

Conclusion: The proposed classification technique enables decoding of dynamic patterns with high temporal resolution, performs favorably to benchmark methods, and facilitates anatomical inference.

1. Introduction

Understanding how different brain regions interact in task-dependent ways is a key goal of cognitive neuroscience. In this regard, a frequent aim of neural data analysis is to characterize the spatiotemporal patterns present within brain activity and, subsequently, to enable the association of such patterns with specific cognitive states. Perhaps the most classic of such approaches, certainly within the domain of electrophysiological brain recordings, is time-frequency analysis which involves the projection of data into the Fourier domain in order to examine synchronization, spectral power and cross-channel (i.e., network) relationships.

Recently, increased attention has been directed toward

* Corresponding author at: 600 S. Euclid Ave., 63110 St. Louis, MO, USA. *E-mail address*: f.singh@wustl.edu (M.F. Singh).

https://doi.org/10.1016/j.jneumeth.2018.06.019

Received 10 October 2017; Received in revised form 22 June 2018; Accepted 25 June 2018 Available online 30 June 2018 0165-0270/ © 2018 Published by Elsevier B.V.

characterizing neural recordings in terms of their underlying dynamics via more holistic descriptions (Andrzejak et al., 2001; Adeli et al., 2007, 2008; Sitges et al., 2010). Such approaches include methods for characterizing chaos and instability reflected in neural recordings via estimation of Lyapunov exponents (Iasemidis et al., 1990; Güler et al., 2005) and manifold reconstruction methods such as Takens embedding (Takens, 1981; Sauer et al., 1991), which has been used to infer directed information flow between brain regions without overt analysis of rhythmicity (Tajima et al., 2015). As dynamical systems-based characterizations, these approaches are unified in that interest is placed on directly characterizing the underlying vector field that ultimately governs the time-evolution of the observed recordings. The vector field describes the instantaneous evolution of dynamical systems, i.e. the

time-derivatives of its state variables. In this spirit, here we propose a technique for elucidating properties of the vector field that involves analysis not of the observed time-series itself, but rather on the derivatives: the 'velocity' trajectories of neural activity. We use the term 'velocity' in the dynamical systems sense in which system states, such as a particular combination of cell voltages, are referred to as 'positions' within the state space and an observed sequence of 'positions' (states) forms a trajectory. The 'velocity' is thus considering how the system states change in time. It turns out that this rather straightforward transformation from considering system position to system velocity brings about several conceptual and practical advantages for classifying neural dynamics. Our approach is motivated in part by recent results in monotone dynamical systems theory that provide explicit characterizations of how the attractor landscape of a dynamical system relates to the geometry of its derivative phase flow (Smith, 1980; Sanchez, 2009, 2010; Feng et al., 2017). In other words, by analyzing the observed velocity trajectories we can make certain deductions regarding the properties of the underlying neural circuit dynamics. The spatial resolution of these deductions match the spatial resolution of the recordings themselves. Exploiting this notion, we describe a novel approach to decode different cognitive states by geometrically deducing and characterizing their corresponding velocity trajectories. As will be seen, this approach enjoys many advantageous statistical properties. In the particular case focused on in our paper, it reduces to an intuitive comparison of covariance geometry of the derivative time-series, which is amenable to classification using conic boundaries. We proceed to demonstrate the efficacy and ease of the proposed paradigm in a variety of examples, then apply it to an electroencephalogram (EEG) dataset (Treder et al., 2011) wherein we show its interpretability and performance in decoding covert spatial attention.

2. Classification for hypothesis testing in neuroscience

Classification techniques play two distinct roles within neuroscience: for developing basic scientific inference as with multivariate pattern analysis (MVPA); and for developing brain-based interfaces such as brain–computer interfaces (BCI). The technique that we now propose has been developed with the former objective (scientific inference) in mind, although BCI applications are also possible. MVPA is a pattern-based approach to studying brain function, in which investigators attempt to decode whether activation patterns contain significant information regarding task events or cognitive states. A common goal of MVPA studies is to determine the anatomical regions in the brain that contain such information (Haxby et al., 2001). Other studies have used MVPA-type analyses to test temporal rather than spatial hypotheses regarding brain function, as with studies of mnemonic periodic-replay (Fuentemilla et al., 2010; Jafarpour et al., 2013).

For all of these research questions, linear classification methods are by far the most common, and include logistic regression, linear discriminants, correlation, or support-vector machines (Haxby et al., 2014). However, a continual goal of methodological research on MVPA applications for neuroscience is to search for enhanced classification approaches. One potential direction is to leverage the recent advances with deep learning approaches (i.e., convolutional neural networks, CNNs) that have shown significant promise within machine learning applications. A few studies have successfully adopted such approaches to classify neural data, for example by using pre-trained image-classification networks (Spampinato et al., 2017; Wen et al., 2017). Another potential direction for MVPA methodological development that has been more widely explored is to improve upon the neural features that are utilized by classification algorithms, such as encoding models (Garcia et al., 2013; Haxby et al., 2014; Foster et al., 2016) or feature selection (Jafarpour et al., 2013).

In the current study, we also examine changes to the neural features utilized by classifiers, but adopt a qualitatively different approach, which instead considers a new form of information, namely the temporal evolution of brain activity. In particular, previous work using MVPA for EEG/MEG decoding has employed classification analysis using band-limited power, so that the classifier input is a periodic representation of brain activity. In contrast, the proposed technique uses a temporal derivative, which is a powerful transformation that can convert any spatiotemporal pattern (not just periodic ones) into another spatial feature. However, as we later discuss, this form of information is almost never separable with a linear boundary. Consequently, we develop an analogous classifier using a (quadratic) conic boundary. In practice, a positive classification using our method generates the following information: (1) a characterization of how the neural activity for the chosen EEG channels evolves differently in time based upon the task classes (the information is dynamically coded) and (2) a detailed temporal signature of the differential patterns of temporal evolution in the two classes.

In contrast to using complex formalisms such as deep learning, our classifier is explicit and employs only one form of information: the direction (covariance) of the temporal derivative. As a result, there is an explicit mapping between the classifier and the geometry of the system's temporal evolution. This feature aids the classifier's interpretability. In contrast, limited interpretability is seen as one of the greatest drawbacks of approaches such as convolutional neural networks and, to a lesser extent, linear decoders as well (Haufe et al., 2014). The primary contribution of the proposed technique for neuroscience investigations is that it provides information regarding how spatiotemporal patterns of neural activity relate to task states. This information is characterized in terms of the interaction of neural populations across time. Such a characterization promotes mechanistic inference regarding the nature of task-based interactions between neural generators. As opposed to simply linking task states with mathematical descriptions of brainwaves (i.e. spectral power) inferences are made in terms of the temporal interactions between neural populations that characterize a task state. Any experimental inferences drawn from applying the classifier are made at the spatial and temporal scales in which the data is recorded: e.g. interactions between large cortical populations as indexed by the EEG timeseries.

In the current study, we first provide simulated examples in which the system-states and corresponding inferences are made at the cellular scale. Next we provide analysis of the cortical populations that underlie task-states in humans using an experimental dataset from a benchmark EEG experiment (Treder et al., 2011). This benchmark dataset has subsequently been treated using conventional MVPA approaches involving extraction of rhythmic activity, in efforts to reveal circuit dynamics that underlie covert spatial attention (Samaha et al., 2016). Critically, we demonstrate the power of the ability of our approach to generate new neuroscientific insights regarding both the anatomical interactions and the precise temporal dynamics that give rise to covert spatial attention (see Section 3.4).

3. Results

3.1. Conic analysis reveals underlying system dynamics

We consider time-series (here, neural recordings) denoted as

$$x_{j}^{(i(t))}(t) = [x_{j,1}^{(i(t))}(t), x_{j,2}^{(i(t))}(t), ..., x_{j,n}^{(i(t))}(t)]^{T}]$$
(1)

where *n* denotes the number of observed variables, $j \in \mathbb{Z}^+$ is a trial index and i(t) = 1, 2, ..., K denotes a class label (e.g., a cognitive state). Our overall goal is to deduce the class label (at each time) by analyzing the derivatives of $x_i(t)$.

The proposed method for neural data analysis and classification in this context consists of two steps. First, for each class (*i*), we obtain an estimate of the derivative time-series, $\dot{x}^{(i)}(t)$. Second, we use generalized cones as a geometric classifier for individual points of the derivative time series. A generalized cone can be understood as an object in

Euclidean space that is invariant to scaling: if a vector v is an element of a cone, then so is αv for any non-negative $\alpha \in \mathbb{R}^+$. Thus, the cone is a fundamentally directional object. Our approach will involve using such cones to partition Euclidean space into sets of derivatives attainable only by specific classes. This procedure can reveal how a system transitions between states (but not necessarily the speed at which these transitions occur).

Our subsequent results will demonstrate that this approach complements spectral analysis (e.g., assessment of oscillatory power), which, while an assay of dynamics, does not explicitly include directional information. More specifically, unlike many existing methods for spatiotemporal classification, we do not operate upon summary parameters such as power within a specific frequency band (Fuentemilla et al., 2010; Jafarpour et al., 2013). Instead, it is the individual data points of the derivative time series that are used as the basis for classification since these contain inherent spatiotemporal information. In this way, our classification approach can capture the instantaneous rules, (i.e., the dynamics), governing how the (class-specific) timeseries evolve. This approach is particularly favorable as it reduces a spatio-temporal classification problem in the original time series to a spatial-only problem in the derivative time-series. As we demonstrate later, this reduction significantly aids analysis and interpretation.

The motivation for such an approach stems from recent results in the theory of smooth monotone dynamical systems. For our purposes, a smooth dynamical system consists of a state space (here, we consider this to be \mathbb{R}^n) coupled with a phase flow, or vector field that describes, for each state, how the system evolves forwards (and, backwards) in time. Recent research has linked geometric knowledge about the vector field (i.e., the derivatives along system trajectories or paths) to the asymptotic behavior of the system state variables. Such characterizations have been made possible by formally studying the extent to which the derivatives along a path may be confined within generalized cones (Smith, 1980; Sanchez, 2009, 2010; Feng et al., 2017). Although a detailed discussion of monotone dynamics is beyond our current scope (see e.g., Smith, 2008), these recent advances motivate our use of cones for classification in derivative space.

To the extent that spatial patterns are useful for classification, the derivatives can provide us crucial information regarding the underlying system dynamics (spatiotemporal patterns). Perhaps an intuitive analogy is that of vertical ascent for two sorts of aircraft: (fixed-wing) airplanes and helicopters. While helicopters can make a direct vertical translation, airplanes require strong lateral velocity before any vertical motion is obtainable. Thus, for an airplane, reaching a point just above the current one requires a complex (spiral-like) velocity trajectory. Observing this velocity trajectory is thus highly informative in deducing the vehicle in question. Aerodynamic constraints also greatly constrain the turning radius of an airplane. As a result, the set of obtainable velocities for an airplane relative to its body-frame differs greatly from that of a helicopter, despite both being able to reach any point in airspace. Similarly, neural circuits are seemingly able to produce a wide variety of spatial activity patterns, although the set of attainable derivatives may be more limited. For example, in Fig. 1 we consider two different toy networks, each composed of three continuous, recurrent Hopfield-model neurons (Hopfield, 1984, see Methods; Fig. 1A and B). The two networks (systems) feature a common connection scheme, but differ in their internal dynamics. In the second system ("System B") every cell exhibits greater decay (in neuroscience parlance, leak conductance) (Fig. 1B). The systems produce qualitatively similar trajectories wherein all initial conditions lead to a damped oscillatory response for both cases (Fig. 1C and E). Despite this similarity, the modeled systems differ in the oscillatory envelope of their response (Fig. 1F). In the ideal case of noise-free uninterrupted observations, the trajectories of these systems are easily discriminable in the original phase space (Fig. 1E). However, trial-based designs that are the primary workhorse of experimental neuroscience recordings (e.g., single- or multi-unit electrophysiology, EEG, fMRI, etc.) generate large numbers

of short duration observations, in which the variability of initial states present at the trial start can result in significant entanglement of activation trajectories. When we replicate these conditions for the two simulated networks, their recordings become inextricable in the phasespace of cell activity (Fig. 1C). However, the systems remain easily separable in the phase space of their derivative signals (Fig. 1D). As the derivative space is indicative of instantaneous changes in activity, dynamical features of the system are captured in individual observations, so the systems may be differentiated regardless of observation length or initial conditions. Of course, this classification can be performed to a certain degree via direct methods (e.g., linearization) if the full analytical model is already known. Our goal, in contrast, is to make the same dynamical inferences (i.e., about the structure and equations governing the system) based only upon observation of the activity. Using our proposed approach, we observe that derivatives are invariant within a particular directional region (Fig. 1D). Based upon this invariance, we can directly surmise that System A (Fig. 1A) cannot directly move across the Cell 1 axis as its derivatives can never point in that direction Fig. (1). Instead, its derivative space suggests System A could only traverse the space by moving indirectly in that direction, just as physical constraints prevent vertical translation of a fixed-wing aircraft. One possible solution, is for System A to generate a series of oscillations leading to traversal of the space along the Cell 2 axis, which is of course consistent with the actual dynamics (Fig. 1E and F). In contrast, derivatives of System B are rarely orthogonal to the Cell 2 axis (Fig. 1D) and the corresponding prediction holds that System B traverses the Cell 2 axis more directly (with fewer oscillations) than System A (Fig. 1E and F).

To operationalize the above analysis, we propose a classifier to discriminate systems based upon their directional derivatives. Most classifiers used to analyze neural data, such as linear support vector machines (Boser et al., 1992), assume that the classes differ in some combination of their univariate means and thus form linear boundaries (Norman et al., 2006). For the case of derivatives, this assumption is not valid. In fact, a well-known result in dynamical systems theory, which we rely upon, is that the derivatives of a smooth, bounded dynamical system must be balanced (e.g. Sanchez, 2009): over long scales the mean derivative is always zero. Otherwise, the system would grow infinitely along the direction of the mean derivative. In the case of a neural system, the mean change in the measured activity variable (e.g., voltage) must be zero over a sufficiently long recording, lest the activity grow infinitely. As the mean derivative is thus irrelevant, the motivating results in dynamical systems theory concern invariance along cones. Due to scale-invariance, conic boundaries discriminate between classes based upon the directional components of the data, rather than magnitude features (such as speed of evolution), and are always centered at zero. With very mild assumptions upon the way derivatives are distributed for each class, we can derive a Bayes-optimal classifier for directional data (see Materials and Methods) which reduces to a comparison of covariance matrices. The boundaries of this classifier form a quadratic cone.

For an unlabeled time-series $x^{(i(t))}(t)$ we define our quadratic-cone classifier as:

$$i(t) = \arg\min_{k} \left(\frac{\dot{x}(t)^{T} \Sigma_{(\dot{x}_{k})}^{-1} \dot{x}(t)}{\det |\Sigma_{(\dot{x}_{k})}^{-1}|^{1/n}} \right)$$
(2)

Here, $\Sigma_{(\dot{x}_k)} \in \mathbb{R}^{n \times n}$ is the covariance matrix associated with the derivative time-series of class *k*. At an implementation level, this covariance can be estimated numerically by collecting labeled training data from each class. Thus, the classifier operates on a point-wise basis: assigning a class label to each individual data-point of the derivative time-series. In the two class case with class labels ± 1 we describe the cone in terms of a single matrix *P*:

$$\dot{i}(t) = \operatorname{sgn}\left(\dot{x}(t)^T P \dot{x}(t)\right)$$
(3)



Fig. 1. The two variants of a recurrent continuous neural network (A, B). Line widths indicate connection strength and green/black indicates excitatory/inhibitory. The red markers in network B indicate a greater decay rate. Observed trajectories do not differentiate between generative systems in the original state space (C.1, C.2), but are readily distinguishable in the derivative space and its projection onto a spherical surface (D.1, D.2). Random intervals of each orbit are plotted for 150 initial conditions and two views (view 1: C.1, D.1; view 2: C.2, D.2) with derivatives projected onto ∂S^{n-1} . In the original phase space (E) full trajectories of both systems approach a focus, however, they differ in the approach envelope which separates derivatives (F). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$$P \coloneqq \frac{\sum_{(\dot{x}_{l-1})}^{-1}}{\det |\Sigma_{(\dot{x}_{l-1})}^{-1}|^{l/n}} - \frac{\sum_{(\dot{x}_{l+1})}^{-1}}{\det |\Sigma_{(\dot{x}_{l+1})}^{-1}|^{l/n}}$$
(4)

We denote the induced cone as $C(P) := \{y \in \mathbb{R}^n | y^T \text{Py} \le 0\}$. The decoding rule may be restated in terms of the cone C(P) as $\dot{x}(t) \in \text{Int}(C(P)) \Leftrightarrow i$ (t) = -1 with Int indicating the interior (i.e. when the inequality is strict). We present further information concerning practical and efficient calculations of these terms and the case of singular covariances (i.e. fewer samples than channels) in Materials and Methods. Formally, we use the term 'derivative' in the sense of backward approximation by Newton's difference quotient, however the scale-invariance property of cones makes such formalism unnecessary and the backward derivative may be replaced by the difference time series: $\dot{x}(t) \simeq x(t) - x(t - \Delta t)$. In general, we have had success implementing the derivative in this form without further preprocessing, even though an extensive literature exists concerning more sophisticated methods of numerical differentiation (i.e. Ahnert and Abel, 2007).

3.2. Conic geometry blindly identifies upstream network states

To test the degree to which conic-invariance describes changes in functional structure (Honey et al., 2009), we simulated the case in which an experimenter tries to study a complex neural network but may only access a small number of outputs (Fig. 2A). We formalized this scenario by simulating a chaotic (Lorenz, 1963) system's effects on a downstream recurrent network of four bursting Hindmarsch-Rose model neurons (Hindmarsh and Rose, 1984) with recordings only available for the first three cells (Fig. 2B and C). Cells are either excitatory (green) or inhibitory (red). The fourth, unobserved, inhibitory neuron forms the main recipient of upstream input and is normally quiescent save when activated by the upstream component. When the upstream component is in the "off" state, the network acts as a delayed negative feedback loop. When the upstream component is in the "on" state, a new indirect inhibitory path is opened via Cell 4. Thus, by altering the activity of a mediating cell, the upstream component's activation dictates possible interactions in the downstream network.

In order to blindly determine the upstream component's activity state we performed unsupervised clustering of downstream neural voltage. Using the three observed neurons we performed 2-means clustering (Lloyd, 1982) for either the spectral power of the original voltage trace (V(t)) or, for conic analysis, the covariance of the derivative voltage traces ($\Sigma_{(V)}$). Spectral power and covariances were calculated over equal-length non-overlapping time windows. Using the k-means algorithm (Lloyd, 1982) produces both a set of centroids (cluster means) and labels assigning each data point to the cluster to which it is closest. For the spectral classification we directly used the labels assigned to each time bin by k-means as the training class. For conic classification we discarded the original labels for each time bin and instead generated new labels through our proposed conic method. The two covariance centroids were treated as class covariances and we then classified each time point according to (2). Because the conic classifier provides instantaneous class labels, we assigned whole-bin classes post hoc in a winner-take-all paradigm based on the sum:

$$\arg\min_{k}\left(\sum_{t\in T_{l}}\left(\frac{\dot{x}(t)^{T}\Sigma_{(\dot{x}_{k})}^{-1}\dot{x}(t)}{\det|\Sigma_{(\dot{x}_{k})}^{-1}|^{1/n}}\right)\right),\tag{5}$$

where T_l denotes the *l*th bin.

We repeated this simulation/analysis sequence while varying simulation parameters for a total of 136 simulations per case (see Materials and Methods). The parameters of interest were the variance of intrinsic system noise (a Brownian motion), measurement noise, coupling strength with the upstream component, and the window length for each bin. Results demonstrate a general advantage of conic classification over spectral clustering (Fig. 2D). This relationship was most pronounced for strong coupling strength or low measurement noise. Unlike conic decoding, spectral decoding was largely unaffected by the white-noise added to the recorded voltages. Spectral performance exceeded conic decoding in extreme cases of measurement noise, which is well known to degrade estimation of derivative time series. However, both methods were resistant to intrinsic noise within the system. These two noise varieties (intrinsic vs. measurement) had equal variance ranges, but unlike measurement noise, intrinsic noise interacts with system states. For instance, Brownian motion in the voltage variable has less impact during the fall of an action potential than near the firing threshold, so this differential sensitivity may preserve those portions of dynamics to which the conic method is most sensitive. We conclude that for most parameter choices within this simulation, the proposed conic decoding method exceeded spectral classification of upstream system states but is susceptible to idealized measurement noise. However, we later demonstrate that the conic method is actually resistant to many (realistic) sources of experimental noise and artifact (see Discussion).



Fig. 2. Simulated comparison of blind conic vs. spectral classification. (A) The system consists of an upstream chaotic attractor which provides binary downstream input to a two-layer recurrent system. Green cells are excitatory and red inhibitory. (B) Example simulated voltages for the three cells in the bottom layer using median model parameters. Voltage traces are in black and blue bars indicate when the upstream input was "On". (C) The state of the upstream system is easily determinable by plotting the voltages in the normalized derivative space. (D) Percent correct by decoding method and simulation parameters. Neither method was affected by the level of intrinsic system noise (IntVar), but conic decoding suffered with the addition of measurement noise (ExtVar). Conic decoding accuracy continued to increase with the upstream coupling strength (Stim), while spectral decoding showed a concave relationship. Both methods performed slightly better for moderate bin sizes (nWin) over very small ones but further increases did not improve performance. The shaded regions give standard deviations (n = 136). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.3. The conic method produces instantaneous decoding of covert spatial attention

In this section, we transition from simulated to real experimental data. Specifically, we consider the method's application to EEG data recorded during a spatial covert-orienting task, taking the data from a publicly available database (Schmidt et al., 2011; Treder et al., 2011). During the task a briefly presented central cue indicated where a subsequent task-relevant target would appear with 80% accuracy. The central cues had near identical visuo-spatial features and subjects were instructed to focus upon the cue (see Materials and Methods). This design feature implies that the putative neural dynamics associated with particular orientations correspond to covert spatial attention rather than simply the direction of gaze. The original results using this dataset have demonstrated that the spatial distribution of alpha power (across channels) may be used to determine to which attended location subjects are covertly orienting. These previous analyses (Schmidt et al., 2011) were performed pair-wise using L1-regularized logistic regression for alpha-band power in parieto-occipital electrodes and the authors presented the best pair accuracy for each subject as well as the number of significantly classified pairs per subject out of a total of 15 (see Table 1). In a first set of analyses, we use all trials with delays of at least 1300 ms and only consider data after the first 250 ms post-cue. We consider the same metrics to those previously reported for comparison

Table 1

Group-level comparisons between conic and alpha-band decoding as reported in Treder et al. (2011) and Samaha et al. (2016). The second column ("Measure") indicates the parameter tested, namely the number of significant pairs (locations) decoded, the maximal decoding accuracy among all possible pairs, and the mean accuracy when performing 6-way classification (all locations at once). The original and conic results are given in Mean (SD) format. We used independent samples *t*-tests and all *p*-values reported are two-tailed. Due to the much greater variance of conic decoding for the 6-way classification relative to Samaha et al. (2016), we corrected for unequal variance in the reported *t*-value. We use the ** superscript to denote uncorrected significance at p < .01.

	Measure	Original	Conic	t	р	df
Treder 11	Sig. pairs Max accuracy (2- way)	3.5 (2.7) 74.6 (2.3)	10.3 (1.8) 76.4 (3.2)	3.558 ^{**} .572	.004 ^{**} .577	15 15
Samaha 16	Mean accuracy (6-way)	23 (4)	32.1 (7)	3.119**	.0088**	12.1

(Treder et al., 2011) and all reported accuracies for the current analyses are based upon leave-one-out cross-validation. Accuracies are averaged between conditions to prevent spurious decoding from unequal class sizes.

3.3.1. Group-level conic performance exceeds the limits of alpha-band classification

In the case of pairwise classification, current results appear favorably (Fig. 3A and B) to those previously published (Treder et al., 2011). In their analysis, Schmidt and colleagues (Treder et al., 2011) provide two metrics: the number of significantly discriminated pairs and the maximal pairwise accuracy. We emphasize comparisons in terms of number of significant pairs as these take into account more subtle spatial comparisons whereas the previously reported maximal pairwise accuracy corresponded to spatially opposite locations for seven out of eight participants in Treder et al. (2011). Results demonstrate that the conic classifier is, on average, superior to α -band based decoding in terms of number of significant pairwise classifications (Table 1). However, the conic advantage in maximal pairwise accuracy was insignificant (Table 1).

3.3.2. Conic and spectral decoding form complementary techniques

As with traditional approaches applied to the dataset (Treder et al., 2011), conic decoding exhibits variable performance across subjects. Interestingly, however, conic and alpha-band decoding performance across individuals tended to be negatively correlated in terms of the number of significant pairs per subject (Fig. 3C). Due to the small sample size and suspected outlier (Subject 8), we performed a rankorder test with the outlier and parametric test both with and without the outlier. Using the full data with rank-order correlation demonstrated a negative relationship between conic and spectral performance (Table 2). The parametric relationship with the full data did not reach significance due to the low sample size, but this relationship greatly increases when removing subject 8 for whom no pairs were significantly classified by the conic method (Table 2). However, the low sample size makes it difficult to determine whether the suspected outlier does in fact represent anomalous data or just the extreme poor end of conic classification ability. Thus, while the small sample size interferes with statistical inferences, results suggest that not only are these decoding methods complementary, but also that many subjects exhibiting poor spectral-based decoding may actually possess strong conic classification. In summary, the proposed conic method and spectral



Table 2

The relationship between conic and spectral (Treder et al., 2011) decoding (number of significant pairs) trends negative. Statistics correspond to non-parametric correlation (Kendall), parametric correlation (Pearson), and parametric with the high leverage subject (#8) removed. All *p*-values are two-tailed.

Statistic	Outlier	Correlation	р	df
Kendall (ĸ)	Yes	5879	.0738	6
Pearson (r)	Yes	355	.3882	6
Pearson (r)	No	940	.0016	5

decoding may form complementary approaches, which is to say that they may be decoding different forms of information. Consequently, subjects insensitive to one technique may benefit from the other. Although conic performance was significantly better on average, spectral decoding was superior for two subjects (#1 and 8) so the conic classifier may still be missing some information that is present in the spectral domain and using the two methods in conjunction might provide superior results.

3.3.3. Conic weights generate retinotopic maps of covert attention

The conic method also appears to replicate previously published results showing that spatial aspects covert-orienting of covert orienting are most prominent of posterior electrodes. In fact, spatial location is highly prominent in the conic weights (see Materials and Methods) and, in this case, provides anatomical localization of dynamical information (Fig. 4) without requiring search-light type analyses (Etzel et al., 2013). To determine whether conic weights provide anatomically meaningful information we performed a spatial mapping of the weights for each cone-generating matrix. We defined the map for each condition in terms of the contrast 'location x' vs. the average hemifield (3 locations) opposite 'x'. We averaged the covariances across subjects for viewing purposes (group level), but in most cases similar results held at the



Fig. 3. Comparison of conic and alpha-band based decoding methods in terms of number of significantly decoded locationpairs. (A) Both decoding methods possess large inter-subject variation in decoding performance. (B) At the group-level conic decoding outperforms spectral. (C) At the individual-level the techniques are negatively correlated and thus form complementary methods. For subject 8, the conic classifier did not significantly discriminate any locational pairs for the full uncensored data. However, censoring trials using the methods for eyeblink and motion artifact rejection adopted in Samaha et al. (2016) improved performance (see Materials and Methods).

individual subject level as well. As a matrix, the conic classifier is most naturally visualized as a weighted graph. However, to produce a spatial map we assigned each channel a weight based upon its contribution to each of the matrix's eigenvectors (see Materials and Methods). We find that the regions representing the covertly attended location experience decreases in the amplitude of derivatives (negative weights) while regions representing the opposite spatial location experience relative increases in the amplitude of derivatives (positive weights) (Fig. 4). Thus, covert attention may be associated with neural activity whose variability (as distinct from variance) is more restricted in regions representing the covertly attended location.

3.3.4. Instantaneous decoding tracks the evolution of neural states

The conic-geometry of these instantaneous changes is easily visible for the 2-class case (i.e. lower left vs. upper locations) when projecting the derivative time series into the eigen-coordinates (see Materials and Methods) with the largest eigenvalues (Fig. 5A and B). These may be thought of as 'principal components', but as they relate the differential activation between classes they also possess sign information. Further, the absolute magnitude of each eigenvalue indicates how strongly selective it is for the class associated with its sign (Fig. 5B). The timeseries of these components indicate that the information differentiating between classes is associated with increased dynamics of certain spatial patterns (the eigenvectors). Critically however, these dynamics are not necessarily associated with consistent increases in particular frequency bands (Fig. 5B). Point-wise classifier performance may be visualized by projecting data into the plane of positive and negative eigenlengths (see Materials and Methods) which indicates the combined 'magnitude' of dynamics associated with each class (Fig. 5C). Data is displayed for the median-performing subject in terms of number of significantly classified pairs (number four). Analogous plots for the best and worst-performing subjects are included in Materials and Methods.

Fig. 4. Anatomical weights distributions. Inverting conic weights to anatomically map classifier features produces a retinotopic map of covert spatial attention in posterior channels. Center panel indicates the corresponding region of head space (posterior coronal) while radial panels denote the weight mapping for subjects attending to that location vs. opposite locations with hot colors indicating increased prevalence in the discriminating spatio-temporal pattern.



Fig. 5. Conic geometry for the median subject (#4). (A) Plotting the first 2 negative components and the first positive component in derivative space reveals conic geometry. Data is displayed from a variety of viewpoints. (B) Plotting the primary positive and negative eigencomponents of the cone for a 2-location contrast demonstrates that these components increase derivative amplitude during the corresponding cognitive state. (C) Two dimensional projections onto the length in the "positive" and "negative" eigenspaces for the discriminative cone reveals robust classification of individual observations lasting only a few milliseconds.

3.4. Conic decoding reveals distributed interactions at high temporal resolution

In a recent paper, Samaha and colleagues (Samaha et al., 2016) used the same dataset to test the hypothesis that the objects of spatial covert attention are differentially encoded in phase and power of the alpha band using both decoding classifiers and encoding models. The authors found an immediate phase-coupling response between frontal and posterior regions during the early post-cue interval (0-400 ms) that was not spatially tuned, but conversely identified a sustained spatiallytuned response in alpha-power that began at roughly 450 ms post-cue. The authors interpreted these results to indicate that previously reported observations of increased anterior-posterior phase coupling during spatial covert attention reflects top-down attentional modulation of posterior sites. In contrast, they argue that the results rule out phase coupling as a mechanism for encoding the contents of spatial covert attention. In contrast to phase coupling, the conic classifier enables investigation of more general interactions between channels. Therefore, we tested whether the same conclusions - namely a lack of spatial information in anterior-posterior interactions and an onset of spatial tuning at 450 ms (in our case significant decoding), would hold for the conic classification approach.

To distinguish between local and distributed features, we compared the full conic classifier to reduced versions that only utilized univariate (variance) or multivariate (correlation) information (see Materials and Methods). The variance-only version of the conic classifier results in a weighted sum of univariate classification rules similar to signal power while having no sensitivity to information such as signal phase. In contrast, the correlation-only version ensures that the classifier can only consider the relationship between channels but not channel-specific information such as power. We then compared 6-way classification accuracy (all 6 locations at once) for the three conic classifiers (full/ covariance, variance, and correlation) with the results recently reported by Samaha and colleagues (Samaha et al., 2016).

Compared to the previously reported results by Samaha and colleagues with this dataset (Samaha et al., 2016), we found that the full and reduced conic classifiers performed favorably when applied to the most posterior channels. The full conic classifier achieved a mean accuracy of $32.1 \pm 7\%$ which was not only significantly better than chance levels, but also significantly better than the classification performance achieved by Samaha and colleagues $23.1 \pm 4\%$ based upon alpha-power (Table 1). The reduced conic classifiers corresponding to

Table 3

Statistical comparisons of the reduced cones (correlation/variance only) to chance, alpha-band decoding (Samaha et al., 2016), and the full conic classifier as well as to each other. Testing compared to chance was performed via one-sample *t*-test and comparisons between conic classifiers and alpha-band decoding were performed with independent samples, unequal variance *t*-tests. Comparisons between full/reduced conic classifiers were performed with paired *t*-tests. All *p*-values are two-tailed and uncorrected for multiple comparisons. * = p < .05, ** = p < .01.

	Mean	Corr	Corr			Var		
		t	р	df	t	р	df	
Chance	16.67	3.343*	.0124*	7	6.855**	.00024**	7	
Samaha 16	23 (4)	.529	.6058	12.69	.672	.5141	12.19	
Full cone	32.1 (7)	-7.985^{**}	.00009**	7	-3.815^{**}	.0066**	7	
Var	24.13 (3.08)	.081	.938	7		-		
Corr	24.3 (6.46)		-			-		

just correlation or just variance also demonstrated significant classification ability, relative to chance accuracy (Table 3). Their performance was worse than that of the full conic classifier and the two versions were equivalent to each other in classification accuracy as well as to the alpha-power classifier used by Samaha and colleagues (Table 3). Therefore, significant information is present in both single-channel measures (variance/power) as well as in cross-channel measures (correlation) for posterior channels. Moreover the combination of these variables (covariance) in the full cone provides far more information than either the correlation or variance alone. We found no additional information in the anterior channels above what was present in posterior channels alone (see Materials and Methods) replicating the findings of Samaha and colleagues (Samaha et al., 2016) that anterior-posterior phase coupling is not spatially tuned.

Next, we tested whether the conic classifier produces a similar temporal profile of covert spatial attention as the inverted encoding models utilized by Samaha and colleagues (Samaha et al., 2016). Their reported results indicated that significant channel tuning for alphapower to the attended location began at roughly 450 ms post-cue. In contrast, the conic classifier identifies significant and sustained information regarding the attended location at a substantially earlier post-cue time period. Significance over the time course was assessed by permutation testing n = 350,000 (see Materials and Methods). For the desired levels of significance thresholds did not vary significantly over time or classification procedure (all σ^2 < .005%) so we plot them as static. Mean classification accuracy for the full conic method exceeded chance (p < .05) starting at roughly 310 ms post-cue (Fig. 6 A), thus preceding the identified onset of alpha-band tuning (Samaha et al., 2016) by nearly 150 ms. We note that this initial period was not included in the classifier training and significant decoding is attained from that point onwards possibly reflecting similarly sustained information. Thus, the decoding pattern meets the temporal persistency required of sustained attentional markers. Still, one possibility for the earlier onset of significant decoding is that the discrepant interval reflects cue-driven responses. However, in the very early trial interval (before 250 ms) classification accuracy was consistently at or below chance. We therefore consider it unlikely that residual activity from visual responses drives the earlier onset of conic decoding (see Materials and Methods). Results from central and parietal electrodes suggest that the earlier decoding ability stems from posteriorly generated neural signals (see Materials and Methods). Lastly, we considered whether the earlier onset of significant decoding with the conic classifier could be attributed to any preprocessing steps, yet this seems unlikely. Our only preprocessing step consisted of a conservative wavelet smoothing (see Materials and Methods). In particular, the kernel we employed is far shorter than the temporal discrepancy between methods (cones vs. alpha-power tuning) and wavelet convolution moves information forward in time so any biases would actually diminish the effect.

In summary, the conic decoding method is able to identify the contents of covert spatial attention from approximately 310 ms post-cue from posterior sites, roughly 150 ms prior to inverted encoding models (Samaha et al., 2016; Fig. 6A). Follow-up analyses suggest that this discrepancy is unlikely to be caused by visual responses to the cue, transient cue-related responses in neighboring regions, or preprocessing steps (see Materials and Methods). Moreover, the time-course of the deconstructed cone (Fig. 6B) indicated that this initial period of significant decoding is more attributed to local changes (the variance component) than distributed changes (the correlation component; see Materials and Methods). Moreover, in this period and the surrounding \pm 100 ms were marked by a general anti-correlation between decoding accuracy based upon variance vs. correlation (see Materials and Methods; Fig. 6B). To be clear, we do not claim that this result alone is sufficient to support an alternative account of the timecourse of spatial attention, but we use it to illustrate how the proposed method may contribute to neuroscience investigations through its improved temporal sensitivity and flexibility (i.e. the ability to construct reduced models that enable selective probing of univariate versus multivariate effects).

3.5. Invariant properties promote robust classification

The proposed approach yields a number of fruitful properties that minimize the contaminating effects of variance due to both 'noise'/artifact and within-subject variance. We conducted a number of simulations using the EEG dataset to demonstrate that the conic classifier is robust to three sorts of noise: (slow) scalar multiplication (Fig. 7 A), temporal warping (Fig. 7 B), and multivariate box noise (Fig. 7C) (see Materials and Methods). The first two relate to the scale-invariant properties of cones. The first case (Fig. 7A) results from simply differentiating the product of the original time series $x_k(t)$ and a scalar function f(t). When f is slow, we have $d/dt[f(t)x_k(t)] = f(t)\dot{x}_k(t) + \dot{f}(t)x_k(t) \approx f(t)\dot{x}_k(t)$. Due to the scale invariance of cones, $\dot{x}_k(t) \in \text{Int } C(P) \Leftrightarrow f(t)\dot{x}(t) \in \text{Int } C(P) \text{ when } f(t) \neq 0.$ This case is especially relevant for multiplicative noise in which a multivariate signal is multiplied by an unknown scalar function, as could be the case for amplifier noise or changes in reference electrode conductance for EEG recordings. However, while the scaling will not affect how datapoints are labeled for a fixed classifier, it is possible that bias in the scaling function's magnitude can affect the calculated covariance, hence the corresponding classifier. Normalizing individual data points of $\dot{x}(t)$ can remove this effect, although this procedure also can change the calculated covariance.

When sampling rates are high so that the time-series are relatively smooth, the conic method is also invariant to temporal scaling (Fig. 7B). More formally, we define a class invariant interval as a continuous interval $\tau := [T_1, T_2] \subset \mathbb{R}^+$ with $t_1, t_2 \in \tau \Rightarrow i(t_1) = i(t_2)$. Let $y(t \in \tau) = x(f(t))$ for any strictly positive monotone $f: \tau \rightarrow \tau$ in which τ is a nonnegative class-invariant section of the time series. Then $\dot{x}(f(t)) \in Int C(P)$ $\Leftrightarrow \dot{v}(t) \in \text{Int } C(P)$. This case follows immediately by applying the chain rule and noticing that strict monotonicity implies $\dot{f}(t) \neq 0$. As the conic method is invariant to scalar multiplication, so too is it invariant to temporal "stretching" which corresponds to a conserved order of events despite irregularity in the precise timing. This problem of temporal "stretching" is referred to as dynamic time warping (DTW), and has been discussed by many previous authors, including in the use of DTW on derivative time series for classification (Górecki and Łuczak, 2013; Górecki and Łuczak, 2015; Łuczak, 2016). In general, dynamic time warping algorithms give a similarity measure between time series evolving at different rates by comparing the order of events between time-series rather than on a point-by-point basis. The recently proposed derivative-DTW variants operate similarly, but on the time series of first and second derivatives. Thus, both DTW and the currently proposed method are able to perform classification on warped time series. Unlike DTW, however, our approach does not explicitly consider the order of events, but rather operates on the set of realizable changes (the derivative geometry) and is thus applicable to systems in which different initial conditions also produce different orders of events. For example, in the Lorenz attractor (Lorenz, 1963), different initial conditions not only affect how long a trajectory stays within one of the attractor's two "loops", but also how many cycles it completes before exiting. The resultant time series from very similar initial conditions within the same system thus produce time series whose order of events differ greatly, despite generating identical state-space geometries after sufficient time. As with the previous case of multiplication by a scalar function, temporal warping will not affect how data points are classified for a fixed cone, but it can affect the calculated covariance, hence the classifier. As before, this effect can be removed by normalizing individual data points of the derivative time series.

The method is also resistant to additive multivariate box noise (Fig. 7C). This follows simply from the fact that the box function has derivative zero everywhere save at a finite number of points



Fig. 6. Decoding time series and performance for 6-way classification (all possible locations at once) using the conic method. (A) The "instantaneous accuracy" time series produced by applying a classifier trained starting 750 ms post-cue (vertical dashed line) to the entire delay time series. Cone-like classifiers using the full covariance (red), the correlation-only (black), or the variance-only (blue) reveal an earlier onset of significant decoding than previously reported with the same data. Horizontal lines indicate the significance threshold for the permutation *p*-value indicated. Accuracy is for prediction over 30 ms windows and was smoothed with a Gaussian kernel. The shaded green bar indicates the early interval in which conic decoding was significant but spectral decoding was not (Samaha et al., 2016). (B) Decoding time series (same as A) but plotted to highlight the early anti-correlated behavior (outer bar) between information decodable by variance and information decodable by correlation. The inner bar is the same as in (A) and indicates that early decoding was primarily due to within-channel variance. Shaded regions indicate the standard error of mean accuracy across subjects. (C) The conic classifier performed significantly better at the full 6-way spatial contrast than any other classification method. For support vector machines (SVM) using spectral power (cyan), only the alpha-band (8–13 Hz) was significantly different from chance (indicated by the dashed line) (Samaha et al., 2016). All conic classifiers were significant and the restriction to either only variance or only correlation produced performance indistinguishable from that of alpha-power SVM. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



Fig. 7. Invariant properties of the conic classifier with respect to different noise forms. Left column displays three different transformations applied to the same segment of EEG data from one trial type. (Middle column) The same forms of noise applied to a constant EEG segment from a different trial type. (Right column) Data for each trial projected onto a cone derived from the original (prenoise) data (blue = lower-right condition, orange = upper-left). Line color indicates channel number. (A) The combination of derivative filtering and conic scale invariance leads to robust performance for scalar multiplicative noise. (B) Scale invariance of cones leads to classifier invariance with respect to dynamic temporal warping (DTW). (C) Filtering properties of derivatives lead to robust performance in the face of multivariate box noise. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(corresponding to steps) and thus has minimal influence on the derivative time series. When the sampling rate is sufficiently high, many slow signals such as those induced by motion artifact may be approximated via a series of box functions, making conic decoding robust to artifactual noise. Thus the filtering of slow signal by derivatives combined with the scale-invariant properties of cones make the

classifier robust to various sources of artifact and within-subject variability. Lastly, we will show that in addition to robustness to noise, our conic method is especially tuned towards task-related signals as opposed to contaminant spontaneous activity.

3.6. Sensitivity to fast-slow interactions

In general, the current literature suggests that very slow changes in the EEG amplitude are particularly prone to drift and motion artifact (i.e. Masterton et al., 2007). In contrast, increasing evidence suggests that fast neural activity is associated with task performance (Jerbi et al., 2009). However, this does not necessarily indicate that slow components are always artifactual, but simply that fast components, when present, may be more informative. There are also many cases in which neural activity manifests meaningful interactions between fast and slow signals (Canolty et al., 2006; Monto et al., 2008; Torta et al., 2009). Thus, a method is needed that (1) can analyze slow signals when fast signals are weak/absent. (2) penalizes slow signals in the presence of faster ones, and (3) prioritizes slow components that exhibit slow-fast coupling over those that do not. The derivative covariance matrix exhibits all of these properties, especially when the components have similar variance (related to spectral power). Formally, consider two multivariate processes: X(t), Y(t): $\mathbb{R}^+ \to \mathbb{R}^n$ and class labels $i(t) \in \{k\}$. As before, we will denote the class-dependent derivative covariances as $\Sigma_{\dot{x}_{k}}$ and $\Sigma_{\dot{y}_{k}}$. Now consider a second, slower process B(t) and a constant 0 < c < 1 defined such that for any class-invariant interval τ (as defined above) and non-negative scalars $t, \epsilon: [t, t + \epsilon] \subseteq \tau$ we have B $(t + c\epsilon) - B(t) = Y(t + \epsilon) - Y(t)$. In other words, B(t) evolves c times the speed of Y(t). The covariance matrix for the combined signal S (t) := X(t) + B(t) is:

$$\Sigma_{\dot{S}_{k}} = \Sigma_{\dot{X}_{k}} + c^{2}\Sigma_{\dot{Y}_{k}} + c\left(\operatorname{cov}(\dot{X}_{k}, \dot{Y}_{k}) + \operatorname{cov}(\dot{Y}_{k}, \dot{X}_{k})\right)$$
(6)

The notation cov(A, B) indicates the covariance between two multivariate processes, whereas we have previously indicated single-process covariance Σ_A as shorthand for cov(A, A). Clearly $c^2 < c$ and if \dot{X}_k , \dot{Y}_k are independent, $cov(\dot{X}_k, \dot{Y}_k) = 0$. When \dot{X}_k is sufficiently small (with the frequency profile held constant), the classifier's inner term $\Sigma_{\hat{S}_{k}}^{-1}/\det |\Sigma_{\hat{S}_{k}}^{-1}|^{1/n}$ approaches $\Sigma_{\hat{Y}_{k}}^{-1}/\det |\Sigma_{\hat{Y}_{k}}^{-1}|^{1/n}$ which satisfies the first condition: sensitivity to slow components in the absence of fast ones. Likewise as *c* becomes small (for $\Sigma_{\dot{x}_k} \neq 0$), the classifier's inner term is biased towards the higher frequency X term even if the components have similar amplitude thus satisfying our second condition. Lastly, when X, Y have amplitude, the classifier is more sensitive to the interaction between fast and slow components via $cov(\dot{X}_k, \dot{Y}_k)$ than those that are independent of *X* (hence restricted to Σ_{Y_k}) as $c^2 < c$. Thus our classifier satisfies all three conditions to balance the influence of fast/ slow signal components in a manner that emphasizes task-related components. If all three scenarios are mixed throughout a single class, direct calculation of the covariance based upon $\dot{S}_k(t)$ will be biased in favor of periods featuring fast components due to their greater derivative magnitude. Again, this bias can be removed by normalizing individual data-points of the derivative time series (using any norm). This step will change the calculated covariance for each class, but will not affect the data's classifiability as $\dot{x}(t) \in \text{Int}(C(P)) \Leftrightarrow \alpha \dot{x}(t) \in \text{Int}(C(P))$ for any nonnegative α . The classifier's performance regarding the three criteria is exemplified by the continued sensitivity in the bursting network example (Fig. 2) which featured fast-slow coupling of downstream/upstream dynamics and within-cell fast-slow coupling during burst activity (Hindmarsh and Rose, 1984).

4. Discussion

We have introduced a novel method for performing classification of spatiotemporal signals. The key innovative feature of the approach is the use of the derivative rather than the original time-series as the basis for classification. By using the derivative time-series, our conic approach to classification is sensitive to temporal information, even though the classification method does not explicitly consider time as a variable. After computing the derivative time series, the method proceeds identically to static classification methods, with each time point of the derivative time series corresponding to one sample. From a dynamical-perspective the conic approach emphasizes decoding based upon the vector field of a system rather than the time course of trajectories.

We have given one example of an explicit form for conic classification which bears some similarity to a special case of quadratic discriminant analysis. However, unlike conventional quadratic discriminants, the balancing properties of derivatives and scale-invariant property of cones enable a natural explicit form for the *n*-class boundaries. These properties also lead to a number of invariance properties that cancel certain forms of noise such as scalar-multiplicative noise (i.e. amplifier noise), multivariate box-wave noise and low frequency noise (as elaborated on below). The form we provide relies exclusively upon the covariance matrices for each class's derivative time series and is essentially a comparison of the degree to which their derivative covariance matrices differ.

4.1. Continuous-time classification reveals the evolution of neural states

As illustrated in Fig. 5 the conic approach not only provides an accurate prediction for the trial, but also generates a prediction for each time point. Together, these predictions form a time series describing the evolution of the system between states with high temporal resolution. The key step of our conic classification technique is to capture the dynamic properties of states by converting to the derivative time series. Previous dynamic approaches have generated state time-series using parameter estimates at fixed or sliding windows (Fuentemilla et al., 2010; Jafarpour et al., 2013; van de Nieuwenhuijzen et al., 2013; King and Dehaene, 2014). In contrast, the current approach generates a time series of individual predictions from a single classifier. Users have additional flexibility in choosing the desired resolution via smoothing etc., or application of secondary classification techniques to the cone-generated time series for higher-order analyses.

4.2. Extension to hemodynamic signals

Presently, we have only presented results of conic decoding for data with relatively high sampling rates, namely neuronal simulations and EEG. However, the potential exists for applications involving much slower modalities provided the events of interest are observable on a similarly long timescale. Task-related dynamics have long been observed in the BOLD signal (e.g. Cohen et al., 1997) and more recent studies have also found dynamic relations between regions (e.g. Chang and Glover, 2010; Kafashan et al., 2016; Riehl et al., 2017). Although the low sampling rate of fMRI is non-ideal for derivative estimation, the aforementioned work suggests BOLD dynamics may be observable over longer intervals. The temporal derivative of interpolated data might then prove a suitable, smoother proxy. As such, the proposed method of conic classification has potential application to fMRI studies of temporally extended cognitive states.

4.3. The conic method extends classification to dynamic features

The special relevance of our proposed derivative-based classification technique to neuroscience is that it specifically decodes dynamic features (versus spatial patterns). The mechanism of the classifier is to decode how a system evolves across time, which we believe is particularly salient to neuroscience. Indeed, an influential view of brain function is to view the brain as a dynamical system. Consequently, the most productive approach would be to map cognitive constructs not only onto regions of the brain, but also the "patterns" of how those regions interact in time. For example, a primary interest of the cognitive neuroscience community is to determine the mapping from characteristic waveforms (event-related potentials or spectral components) to cognitive states. Classical characterizations have been used to inform dynamic models, which then predict anatomical regions of interest. Indeed, several modeling and empirical results now concern how specific waveform features, such as frequency, are formed. However, primary results derived from these characterizations still ultimately link cognitive states to specific mathematical properties of signals (e.g. patterns of frequency decomposition). For example, previous analyses of the covert spatial orienting dataset (Treder et al., 2011; Samaha et al., 2016) relate band-limited power (in the alpha band) to covert attention. In contrast, our method does not rely on deriving secondary measures from data, but rather examines dynamics in their most primitive form: the temporal derivatives of the signals themselves.

5. Materials and methods

5.1. Derivation of the conic classification criteria (2)

To derive our equation for conic classification we assume that the distributions for the two classes are elliptical in the sense that they may be described in terms of $f_i \sim g_i(x^T \Sigma_i^{-1} x)$. Unlike the traditional definition of an elliptic function such as those described in standard quadratic discriminant analysis we do not assume that the function f is non-increasing. In fact, this must not be the case from the perspective of autonomous dynamical systems as the origin/'mean' in derivative space corresponds to a fixed point and should thus have low density in the case of neural signals which are highly transient. Thus we assume nothing regarding the distributions f_i save that they are non-degenerate. Instead of restricting the distribution, we restrict our set of classifiers to be scale-invariant (conic) and thus we may instead consider the distribution projections on the surface of the (n - 1)-dimensional sphere (S^{n-1}) without loss of generality. What follows involves deriving the probability density of an elliptic distribution onto ∂S^{n-1} . These densities immediately lead to the derivation of a maximum likelihood classifier of the form (2).

We impose zero density at the origin as it has no direction and hence is never classifiable. From a practical standpoint this assumption should be insignificant for continuous valued data with a finite number of datapoints. Our treatment of the general elliptic case is in essence the same as the Gaussian case with mean zero long-considered in directional statistics (Wang and Gelfand, 2013). We then apply simple monotone transformations to the distribution to ease computation in high dimensional settings.

Proposition: Consider a set of elliptical probability distributions $\{f\}_{i=1,...,m}$ over \mathbb{R}^n with all means equal zero and $f_i(0) = 0 \forall i \in \{1, ..., m\}$. Denote the corresponding covariance matrices for each elliptical distribution as $\{\Sigma\}_{i=1,...,m}$, and a random variable *X* valued over \mathbb{R}^n . Define the projection operator $\Pi_S : \mathbb{R}^n \{0\} \to \partial S^{n-1}$ and its derived distributions (h_i) as follows:

$$\Pi_{S}(x) = \frac{x}{\|x\|_{2}} \quad X \sim f_{i} \Rightarrow \Pi_{S}(X) \sim h_{i}$$
(7)

Then the following two assignment functions are equivalent:

$$\arg\max_{i}h_{i}(\Pi_{S}(X)) = \arg\min_{i}\frac{X^{T}\Sigma_{i}^{-1}X}{\det|\Sigma_{i}^{-1}|^{1/n}}$$
(8)

Proof. Consider the elliptic distribution f_i and covariance matrix Σ_i . Thus, by the definition of an elliptic function, there is a function $g_i: \mathbb{R} \to \mathbb{R}$ and a functional $\kappa: L^1 \to \mathbb{R}$ such that:

$$f_i(X) \coloneqq \kappa(g_i) \det |\Sigma_i^{1/2}| g_i(X^T \Sigma_i^{-1} X)$$
(9)

Here κ_{g_i} denotes the scaling function which depends upon g_i but not Σ_i .

$$\Theta_i(\phi(x)) \coloneqq (\Pi_S(x))^I \Sigma_i^{-1}(\Pi_S(x)) \tag{10}$$

The distribution of $\Pi_S(x) \in \partial S^{n-1}$ is found by integrating Θ_i along the radial component of x; $r := ||x||_2$:

$$f_i(\Pi_S(x)) = \kappa(g_i) \det |\Sigma_i^{1/2}| \int_{r=0}^{\infty} r^{n-1} g_i(r^2 \Theta_i(\phi(x))) dr \wedge d\Omega$$
(11)

In which $d\Omega$ is the volume element of the (n - 1)-spherical shell. Using a simple change of variables $(u = r^2 \Theta(\phi))$ we remove the angular component $\Theta(\phi)$:

$$f_{i}(\Pi_{S}(x)) = \kappa(g_{i}) \frac{\det |\Sigma_{i}^{1/2}|}{2\Theta(\phi)} \int_{u=0}^{\infty} g_{i}(u) \left[\frac{u}{\Theta(\phi)}\right]^{\frac{n}{2}-1} \mathrm{d}u \wedge \mathrm{d}\Omega$$
(12)

$$=\kappa(g_i)\frac{\det|\Sigma_i^{1/2}|}{2\Theta(\phi)^{n/2}}\int_{u=0}^{\infty}g_i(u)u^{\frac{n}{2}-1}\mathrm{d}u\wedge\mathrm{d}\Omega\tag{13}$$

We will refer to the remaining integral as $\xi(g_i)$. Now consider a second distribution f_0 characterized by a separate scalar function $g_0 \neq g_i$, but identical covariance: $\Sigma_0 = \Sigma_i$, hence $\Theta_0(\phi(x)) = \Theta_1(\phi(x))$. The distributions may then be written:

$$f_{i}(\Pi_{S}(x)) = \kappa(g_{i}) \frac{\det |\Sigma_{i}^{1/2}|}{2\Theta_{i}(\phi)^{n/2}} \xi(g_{i})$$
(14)

$$f_0(\Pi_S(x)) = \kappa(g_0) \frac{\det |\Sigma_i^{1/2}|}{2\Theta_i(\phi)^{n/2}} \xi(g_0)$$
(15)

As both functions are probability distributions they must sum to 1. As the distributions are defined over ∂S^{n-1} they are integrated over $d\Omega$. Factoring out the scalar terms produces integrals dependent only upon the angular component $\Theta_i(\phi)$.

$$\xi(g_0)\kappa(g_0)\frac{\det|\Sigma_i^{1/2}|}{2}\int_{\partial S^{n-1}}\Theta_i(\phi)^{-\frac{n}{2}}\mathrm{d}\Omega\tag{16}$$

$$=\xi(g_i)\kappa(g_i)\frac{\det|\Sigma_i^{1/2}|}{2}\int_{\partial S^{n-1}}\Theta_i(\phi)^{-\frac{n}{2}}d\Omega = 1$$
(17)

$$\operatorname{rg} \operatorname{max}_{i} f_{i}(X) = \operatorname{arg} \operatorname{min}_{i} \left(\left(f_{i}(X) / \eta_{n} \right)^{-\frac{2}{n}} \right)$$
$$= \operatorname{arg} \operatorname{min}_{i} \frac{X^{T} \Sigma_{i}^{-1} X}{\det |\Sigma_{i}^{-1}|^{1/n}}$$
(18)

which amounts to our classifier (2). \Box

a

To recover a full maximum likelihood classification which considers initial population proportions (p_i) we simply multiply the derived probability density proportion of the probability density $f_i(X)$ by p_i before raising solutions to -2/n hence:

$$\arg\max_{i} f_{i}(X)p_{i} = \arg\min_{i} \left((f_{i}(X)p_{i}/\eta_{n})^{-\frac{2}{n}} \right) = \arg\min_{i} \frac{X^{T}\Sigma_{i}^{-1}X}{(p_{i}^{2}\det|\Sigma_{i}^{-1}|)^{1/n}}$$
(19)

5.2. Computational considerations and limitations

A key computational step is calculation of the determinant. To avoid numerical instability associated with matrix inversion, we calculate the determinants via:

$$\det |M|^{\frac{1}{n}} = \prod_{i}^{n} (\sigma_i(M))^{\frac{1}{n}}$$
(20)

where $\sigma_{\{1,...,n\}}$ indicate the singular values of *M*. Nevertheless, retrieving the eigenvalues may be computationally expensive $(O(n^{k > 2}))$ as compared with linear classification algorithms and this remains a limitation of the proposed method. However, we have used the default MATLAB eigenvalue algorithm for conic classification of 2000 dimensional data on a midrange laptop (x64, dual Intel(R) Core(TM) i7-6500U @ 2.50 GHz).

Another limitation of the current method concerns when the covariance matrices are ill-conditioned. In some potential applications (e.g. fMRI) the dimension of the data (number of channels) may exceed the number of datapoints per class. The simplest approach to this issue is dimensionality reduction or to (artificially) increase the number of datapoints through interpolation. However, an alternative method is to retain the original number of dimensions and use sparse-estimation for the population inverse matrix. Previous authors (e.g. Friedman et al., 2008) have developed computationally efficient methods to perform this calculation. Many of these approaches rely upon LASSO or regularization methods so their accuracy may rely upon at least some channels being independent.

In practice, some eigenvalues calculated for the covariance matrices may be near zero due to dependency. To maintain stable estimates of the classifier we recommend setting these eigenvalues equal to zero in the pseudoinversion and removing them from the calculation of determinants. However, we did not need to threshold eigenvalues for any of the simulated classifications. For the empirical data we used a relative eigenvalue threshold of 10,000.

5.3. Simulation 1

In Fig. 1 we used a sigmoidal 3-neuron recurrent network to illustrate conic dynamics in a low dimensional space. The connection scheme is depicted schematically in Fig. 1A and B and features two coupled negative feedback loops. A common hub (neuron 3) inhibits excitatory cells 1, 2. Connections between neurons 2, 3 were stronger than those with neuron 1. The model was as follows:

$$\frac{\mathrm{dx}}{\mathrm{dt}} = W \left(1 + \tanh(s \circ (x - b)) \right) - D \circ x + c \tag{21}$$

We use the notation \circ to indicate the Hadamard product (element-wise multiplication). We simulated this network twice using the same slope (*s*) and connection weights (*W*) for each case:

$$W = \begin{bmatrix} 0 & 0 & -3 \\ 0 & 0 & -25 \\ 3 & 25 & 0 \end{bmatrix} \quad s = \begin{bmatrix} 1.25 \\ 2.5 \\ 3 \end{bmatrix}$$
(22)

The two cases differed, however, in the values for decay (*D*), baseline (*c*), and threshold (*b*):

First Sim:
$$c = \begin{bmatrix} .7 \\ .584 \\ -2.407 \end{bmatrix}, \quad b = \begin{bmatrix} 7.909 \\ .597 \\ 3.374 \end{bmatrix}, \quad D = \begin{bmatrix} .11 \\ .14 \\ .13 \end{bmatrix}$$
 (23)

Second Sim:
$$c = \begin{bmatrix} 3.04 \\ 25.3 \\ -36.26 \end{bmatrix}, \quad b = \begin{bmatrix} 5.37 \\ -.194 \\ -.04 \end{bmatrix}, \quad D = \begin{bmatrix} 5.5 \\ 7 \\ 6.5 \end{bmatrix}$$
 (24)

Simulations were performed with Euler integration dt = .0005 and a run time of t = 3 for each iteration and then down-sampled by a factor of 15. For each case 150 iterations were performed with starting positions selected pseudo-randomly. Segments of each time-series were also pseudo-randomly selected using a random box-function (B(t)). The box function was formed by choosing 200 box-lengths distributed according to $(1 + r)^2$ with r distributed according to the standard normal distribution (r N(0, 1)). The resulting box-lengths were then rescaled so that they summed to the total window size (t = 3). The amplitude for each box was generated from a standard normal distribution and the inclusion rule for displaying a point x(t) was B(t) > 1.5. Thus, roughly 6.9% of the total points were displayed for each trajectory/initial condition.

All simulations were carried out in MATLAB2016b.

5.4. Simulation 2

For this simulation we considered a network of 4 bursting Hindmarsh–Rose neurons with single-exponential synapses. A chaotic Lorenz attractor provides input to the network as a representation of upstream network activity. The Lorenz attractor was given a standard parameterization with the addition of a Gaussian white-noise term (η_i) with standard deviation ϵ ("Internal Noise"):

$$\frac{\mathrm{d}x}{\mathrm{d}t} = 10(y-x) + \epsilon \eta_1(t) \tag{25}$$

$$\frac{\mathrm{d}y}{\mathrm{d}t} = x(28-z) - y + \epsilon \eta_2(t) \tag{26}$$

$$\frac{\mathrm{d}z}{\mathrm{d}t} = xy - 8/3z + \epsilon \eta_3(t) \tag{27}$$

The Lorenz attractor was Euler integrated with dt = .0025 for duration t = 50. Each time point was then upsampled (repeated) by a factor of 100 to ensure that the system did not fluctuate too quickly. The upstream system was considered 'on' when the first Lorenz variable (*x*) was greater than 0. Initial conditions for each iteration were normally distributed with mean $[-10 - 10 \ 27]$ and standard deviation.25.

The output current from the Lorenz attractor was a binary function of the first Lorenz variable's sign:

$$I_{\text{Lorenz}} = \kappa L(t) \begin{bmatrix} -.18 \\ .24 \\ .06 \\ 1.56 \end{bmatrix}, \quad L(t) = \begin{cases} 1 & x(t) > 0 \\ 0 & x(t) \le 0 \end{cases}$$
(28)

The term κ is a scaling factor ("Stimulus Strength") which we manipulated. Each neuron's internal dynamics were governed by the standard Hindmarsh–Rose equations (Hindmarsh and Rose, 1984):

$$\frac{du}{dt} = v - u^3 + 3u^2 + .5 + I(t) + \epsilon \gamma(t)$$
(29)

$$\frac{\mathrm{d}v}{\mathrm{d}t} = 1 - 5u^2 - v \tag{30}$$

$$\frac{dw}{dt} = .001 \left(4(u+1.6) - w \right)$$
(31)

Here I(t) denotes the combined input to each cell and $\gamma(t)$ denotes a Gaussian white-noise process with standard deviation ϵ ("Internal Noise"). Input consisted of a constant baseline level of input I_{base} , the synaptic currents I_{Syn} and the currents from the Lorenz attractor I_{Lorenz} . The synaptic currents were the product of a synaptic weight and a single dynamic synapse variable per neuron.

$$I_{\rm Syn}(t) = Wg(t) \quad W = \begin{bmatrix} 0 & 8 & -4 & -16.8\\ 12 & 0 & 0 & -14\\ 0 & 7 & 0 & -16\\ 16 & 0 & -20 & 0 \end{bmatrix}$$
(32)

$$\frac{dg}{dt} = -5g + .5(1 + \tanh(4u - 2))$$
(33)

$$I_{\text{base}} = \begin{bmatrix} 1.8\\ 1.5\\ 1.9\\ -1.6 \end{bmatrix}$$
(34)

The Hindmarsh-Rose system was Euler integrated with step size.05 and end time t = 5000. Initial conditions for all neuron variables (u, v, w) were drawn from a standard normal Gaussian distribution. All synapses were given initial condition zero. The initialization period from t = 0 to t = 1000 was removed to allow the system time to stabilize. The variables of interest consisted of the voltages (*u*) for the first three neurons only, which were down-sampled by a factor of 10. Gaussian noise was added with standard deviation ξ ("Measurement Noise"). We performed blind classification with clustering to assign system states (Lorenz input on/off) based upon the first three neurons' voltages. We performed classification using the covariance of the normalized derivative time series $(\Delta x_t / \| \Delta x_t \|_2)$ for each time bin as input to the k-means algorithm. Clustering for covariances was performed using k-means for 2 clusters for the upper triangle of the covariance (six elements). Similar clustering for spectral analyses were based upon 2cluster k-means for the vector composed of the separate real and imaginary components of the spectrum evaluated at 6 equi-spaced points spanning half the smallest window size. The conic decoding was then performed using the two centroids derived from the k-means clustering as the covariance matrix for each class. The prediction weights for all

points within a window were averaged to assign a class to the full window. For spectral clustering, labels were assigned directly based upon the centroid cluster to which each bin belonged. As the analysis involved blind clustering, cluster labels were arbitrary. In assigning cluster numbers to system states (i.e. does cluster 1 or 2 indicate input 'on') we used the assignment that provided maximal accuracy for each simulated trial. As such, the minimum accuracy for each trial is.5 and the expected accuracy is greater than.5 using this assignment. Therefore, we derive the null distribution for this case to evaluate bias. As the accuracy for a random assignment (random mapping between cluster number and 'on'/'off' state) is uniform U[0, 1], their mean is Bates distributed. For larger numbers of samples (as was the case for each trial), the Bates distribution for mean.5 over [0, 1] (as there are two classes) limits to the Gaussian N(.5, 1/(12n)). Thus the chance accuracy for random label assignment limits to N(.5, 1/(12n)). For our biased case (using the assignment that provides greatest accuracy), the assignment function is $max(\mu(A), 1 - \mu(A))$ with A the accuracy obtained by arbitrary assignment. For several hundred samples, as was our case for each trial, the expected values of this distribution is essentially.5 (limiting to $.5 + \frac{E[N(0,1)]}{\sqrt{12n}}$).

To analyze the effect of parameterization we manipulated each parameter (intrinsic/measurement noise, stimulus strength, bin length) independently while keeping all others at their median value. There were a total of 136 trials per condition (parameter level) and 9 equispaced conditions per variable. As demonstrated in Fig. 2D, long run times ensured that despite the obvious noise/chaos within each trial (Fig. 2B) standard deviations for decoding accuracy were relatively small. Combined with the large sample size, these factors generated highly significant results for nearly every statistical comparison. As such we focused upon displaying performance. Note that the shading in Fig. 2D indicates standard deviation as opposed to standard error (which would not be visible).

5.5. Experiment 3

The data and task were obtained from an open repository associated with (Schmidt et al., 2011; Treder et al., 2011). Briefly, during the task, a multicolored hexagonal cue was presented for 200 ms and participants were instructed to focus upon the edge corresponding to a given color. The orientation of this edge predicted in which of 6 radially located discs the upcoming target would be displayed after a delay of 500–2000 ms. In 80% of trials the target appeared in the cued location. Targets were presented for 200 ms after which participants had to indicate the target's shape (either '+' or 'x'). The dataset from these studies is publicly available (BNCI-Horizon-2020.eu) and consists of eight subjects' pre-processed recordings for 60 EEG channels, 2 EOG channels. Data was recorded at 1 kHz but the published dataset has been down-sampled to 200 Hz. Our only further modification to the data was a single wavelet reconstruction to estimate derivatives given noisy data. For this step, we simply used a 1-D Daubechies 7-tap decomposition and level 3 reconstruction, although more specialized methods exist (Wang, 2007). Statistical analyses between previous and current results are based upon paired t-tests for each case. The significance of pair-wise accuracy was assessed with 1-sample t-tests for accuracy greater than chance (50%) and Bonferoni corrections which mimicked the original analyses with this dataset (Treder et al., 2011). For the 15 possible pairwise-comparisons this approach typically resulted in corrected significance thresholds of roughly 65% for p < .05, although the precise threshold varied with the number of trials (Table 4).

5.5.1. Visualizing conic projections

Plots in Fig. 5 display conic classification of subject 4's data for the lower upper right vs. lower left contrast for the 21st, 24th, 27th, and 30th trial of each of the two locations. Time series in Fig. 5B are the

Table 4

Number of trials for each subject meeting the trial length criteria by class (cued location).

Location		1	2	3	4	5	6
Subject	1	70	71	72	66	58	64
	2	70	66	72	73	59	67
	3	66	72	63	64	64	62
	4	68	74	69	68	64	69
	5	72	73	62	77	64	61
	6	47	53	49	51	49	48
	7	70	67	70	73	57	60
	8	73	67	59	73	64	69

concatenated time series of the largest eigencomponent for each location across the four classification periods each. We chose subject 4 to illustrate conic projections as the median-performing subject (in terms of conic-decoding performance). For context, we also include identical plots for the best-performing (#3; Fig. 8) and worst-performing subjects (#8; Fig. 9).

5.5.2. Individualized results for n-class decoding

In the main text we compare performance of the conic classifier with that of other methods on the same benchmark dataset. Namely, we presented pairwise classification accuracy for comparison to the original analyses by Treder and colleagues (Treder et al., 2011) and full 6-way classification for comparison to the more recent work by Samaha and colleagues (Samaha et al., 2016). However, one of the main advantages of the current classifier is that it is inherently multi-class as opposed to most linear or even quadratic classifiers which are generally pairwise, necessitating adhoc methods such as voting schemes to perform multi-class decoding. To better illustrate the ability of the conic classifier on multi-class problems we therefore include results comparing the full range of possible comparisons (2-way through 6-way) for each individual subject (Fig. 10). Results are plotted for the mean performance across all possible combinations of *n*-locations.

5.6. Effects of artifact rejection and eigenvalue estimation

For results displayed in the main text we included all trials meeting the trial-length criteria without consideration of possible eyeblink and motion artifact. Moreover, we censored eigenvalues with relative magnitudes less than 1/10,000 as calculation of very small eigenvalues can become numerically unstable. To ensure that these choices did not substantially affect conclusions we present the results corresponding to Fig. 3 for the cases of either motion artifact removed based upon the exclusion criteria adopted by Samaha and colleagues (Samaha et al., 2016) or with alternative conditions for eigenvalue censoring. We considered two alternative schemes for calculating the eigenvalues: either including all eigenvalues or, rather than removing small components, to remove the largest posterior component via Principal Component Analysis (PCA). PCA was performed upon the full data (all time points) in the native space (before derivative calculation/wavelet smoothing) and the largest principal component was removed. The eigenvalue and artifact conditions were fully crossed (Fig. 11) and we computed the number of significant pairs (Fig. 11A and B) and pairwise accuracies (Fig. 11C) for each subject. Group-level performance was very similar regardless of these preprocessing conditions (Fig. 11B). Likewise, individual performance as measured by decoding accuracy for each pair of locations was generally consistent across conditions (Fig. 11C). Classification accuracy for subject 8 slightly increased when removing artifact (averaged across eigenvalue conditions: paired -t $(14) = 2.16, p = .0485, \Delta \mu = .027 \pm .0485)$ while accuracy for subject slightly 6 decreased (paired - t(14) = -4.635, p = .0004, $\Delta \mu = -.021 \pm .0184$). We did not observe any significant changes due to eigenvalue conditions. These changes in pairwise accuracy had a



Fig. 8. Conic geometry for the best-decoding subject (#3). (A) Plotting the first 2 negative components and the first positive component in derivative space reveals conic geometry. Data is displayed from a variety of viewpoints. (B) Plotting the primary positive and negative eigencomponents of the cone for a 2-location contrast demonstrates that these components increase derivative amplitude during the corresponding cognitive state. (C) Two dimensional projections onto the length in the "positive" and "negative" eigenspaces for the discriminative cone reveals robust classification of individual observations lasting only a few milliseconds.

1

Fig. 9. Conic geometry for the worst-decoding subject (#8). (A) Plotting the first 2 negative components and the first positive component in derivative space reveals conic geometry. Data is displayed from a variety of viewpoints. (B) Plotting the primary positive and negative eigencomponents of the cone for a 2-location contrast demonstrates that these components increase derivative amplitude during the corresponding cognitive state. (C) Two dimensional projections onto the length in the "positive" and "negative" eigenspaces for the discriminative cone reveals robust classification of individual observations lasting only a few milliseconds. Note the large spike in this subject's data that is only visible in one of the eigenvectors.

corresponding increase/decrease in the number of significantly decoded pairs (Fig. 11A). In contrast, several high-performing subjects actually had the number of significantly classified pairs decrease when artifactual trials were removed (Fig. 11A). However, these decrements do not necessarily imply that performance worsened: since the significance threshold (in terms of accuracy) decreases with the number of classifications, removing trials due to artifact effectively increased the accuracy threshold for significance. Thus, although the changes in pairwise decoding accuracy due to preprocessing conditions were minor for most subjects, artifact-based censoring can lead to fewer significantly classified pairs by decreasing the sample size. The main take-away from this analysis is therefore that artifact had relatively little impact upon decoding performance for most subjects (Fig. 11C) but removing

artifact did increase performance for the worst-performing subject (#8) and negated the previously observed spectral advantage in decoding for this subject.

5.6.1. Decomposing the classifier into variance and correlation

Just as spectral relationships may be decomposed into power and phase, the covariance matrix may be decomposed into contributions of variance and correlation to assess local (variance) vs. distributed (correlation) contributions to task-related activity. Simply taking the variance of each channel as input to a standard classifier does not remove multivariate information as the classifier will still determine boundaries based upon the covariance of observations (the covariance of variances). However, by computing the cone using only the diagonal



Fig. 10. Individualized decoding accuracy for multi-class decoding. The *x*-axis ("*n*") denotes the number of classes and performance was averaged over all possible combinations of "*n*" spatial locations using the same windows as the pairwise comparisons (starting 250 ms post-cue) and including all trials lasting atleast 1300 ms (without artifact rejection).

(variance) terms all multivariate information is gone and the resultant classifier is actually identical to a weighted sum of univariate classification rules:

upon correlation alone, we used the regular conic classifier ((2)) but *z*-scored the data channel-wise for every channel. Thus, for every trial, each transformed channel had a variance of one so that the covariance was identical to the correlation.

$$\arg\min_{k} \left(\sum_{t \in T_{l}} \left(\frac{\sum_{c=1}^{n} \dot{x}_{c}(t)^{2} \sigma_{\dot{x}k_{c}}^{-2}}{\prod_{c=1}^{n} \sigma_{\dot{x}k_{c}}^{-2/n}} \right) \right)$$
(35)

where T_l denotes the *l*th bin and $\sigma_{xk_c}^2$ denotes the variance of the derivative signal for class *k* at channel *c*. To generate the classifier based

5.6.2. Testing the influence of anterior channels

In addition to the 13 most posterior channels we also tested the contribution of anterior channels. Using the first 11 anterior channels (first two rows) produced decoding that was weak ($\mu = .201 \pm .0275$),



Fig. 11. Effects of preprocessing steps (artifact-censoring and eigenvalue censoring) on classification performance. (A) The number of significantly decoded pairs for some subjects varied with the pre-processing steps. Conic decoding performance for subject 8 is at least as good as for spectral decoding whenever artifactual trials are removed. (B) Group-level performance was very similar across pre-processing choices. (C) The pairwise decoding accuracy is generally unaffected by pre-processing choices (as opposed to the number of significant pairs). For subject 8, however, accuracy generally increased with artifact rejection.

but significantly above chance (t(7) = 3.54, p = .0095, 2-tailed) for the six-way classification. When we censored high motion trials using Samaha and colleague's (Samaha et al., 2016) criteria the anterior classifier did not reach significance ($\mu = .1911 \pm .0296; t(7) = 2.33$, p = .0525, 2-tailed). Performance for the deconstructed conic classifiers (variance and correlation) was even weaker. These results confirm Samaha and colleague's (Samaha et al., 2016) conclusion that activity in anterior channels does not reflect the attended location. Moreover, combining the very anterior and very posterior channels together did not improve classification accuracy above the posterior channels alone (t(7) = -1.85, p = .106, 2-tailed). In fact, the accuracy tended to decrease ($\Delta \mu = -.019 \pm .029$) indicating that the anterior channels were simply adding noise. Thus, results using the conic method agree with the previous findings that anterior-posterior coupling does contribute additional information regarding the attended location from what is found using the posterior channels alone.

5.6.3. Testing conic time series

Conic time series were formed by taking the classification accuracy over moving windows of 6 data points (30 ms) to form predictions. Finer resolutions (i.e. single data points) are also possible, but we chose a (very small) sliding window instead to aid visualization. Accuracy is for leave-one-out cross-validation. Classifiers were trained on the interval 750–1900 ms post-cue and tested on the full trial interval. Thus the first 750 ms was not present in the training data for any trial. The resultant accuracies were averaged across class for each subject and smoothed with length 7 (35 ms), unit variance Gaussian kernel. The kernel was scaled to have an area of exactly one so it did not bias overall performance.

In order to test the significance of the conic time series we performed permutations based upon scrambled class labels (spatial location) with all else identical. The number of observations per scrambled class was the same as in the original data. We performed 10,000 permutations for each combination of subject × classifier. Using the individual subject distributions we then constructed the group-level distribution from the means of subject-wise permuted data repeated 350,000 times for each classifier. None of the resultant significance thresholds displayed substantial variation with respect to classifier choice or time post-cue (all $\sigma^2 < .05\%$) so we set them constant at the mean value.

5.6.4. Ruling out visually-evoked and central/parietal influences to posterior decoding

To control for the possibility that our earlier onset of significant posterior decoding was due to visually-evoked activity we considered the similarity between activity during the window of primary visual responses (i.e. the first 250 ms after stimulus onset = 50 ms post-cue; Makeig et al., 2002) and the delay period. This activity does not resemble the learned patterns during delay as reflected in insignificant decoding accuracy (Fig. 6A) so it is unlikely that the onset of significant posterior decoding is related to visually-evoked responses. Another possibility is that the earlier onset of significant decoding reflects later processing of the cue such as its semantic significance as opposed to reflecting spatial attention. We do not believe this explanation is likely either due to the spatial profile of decoding - namely a mismatch between central/parietal and posterior decoding time courses. Event-related activity associated with semantic processing has been previously described (in other experiments) over the relevant interval (300-500 ms post-cue) but is centered over central and parietal electrodes (Kutas and Hillyard, 1980; Kutas and Federmeier, 2011) while we only considered occipital and parieto-occipital (the 13 most posterior). To test whether our early decoding onset reflected processes linked to the parietal lobes, we considered the parietal electrodes (an additional 11) both alone and in addition to posterior electrodes. In both cases classification accuracy was above chance. However, the addition of parietal electrodes did not augment the early decoding accuracy and parietal electrodes alone did not exhibit a sustained response starting at that interval. Moreover, combining posterior and parietal electrodes did not improve mean classification accuracy relative the posterior electrodes alone ($\Delta \mu = -.0298 \pm .0614;t$ (7) = -1.375, p = .212). As parietal electrodes did not add additional information for decoding (as reflected in accuracy), posterior conic decoding is unlikely to be driven by parietal processes. However, parietal electrodes did reveal a different pattern of results with an even earlier, but weak and transient, component which reached significance (p < .05) beginning at 170 ms and lasted approximately 100 ms. This component was driven by the correlation of derivatives between parietal electrodes (as opposed to variance) and demonstrated a different time course than the decoding performance at posterior electrodes (namely a lack of sustained response). Thus, our earlier onset of significant posterior decoding cannot be explained by processes linked to parietal sites that tend to occur near the time interval of interest (i.e. semantic event-related activity Kutas and Hillyard, 1980; Kutas and Federmeier, 2011).

5.6.5. Interpreting local vs. distributed contributions to decoding in EEG data

In general, dissociating the source of EEG signals is a nontrivial endeavor, although several methods have been developed to localize source signals (i.e. Grech et al., 2008). The main confound is volume conductance, by which changes in focal cortical activity can lead to changes in correlations between regions. For instance, scalp regions which have high conductance paths to the site of local activation will appear more correlated with each other. Therefore, simply identifying differences in decoding accuracy between variance and correlation is insufficient to suggest that the neural generators are more focal vs. distributed. However, in the present case we can separate these possibilities as the variance and correlation decoding time-courses are anticorrelated during this early interval. In contrast, a focal neural generator will always have a positively correlated impact on variance and correlation as they both depend monotonically upon the generative signal's amplitude. We consider the early period of significant decoding to be from 300 to 500 ms corresponding to the first 200 ms after the start of significant decoding performance. Over this interval the group average time series for correlation-based and variance-based decoding accuracy are negatively correlated (r(39) = -.324, p = .0108, 2-tailed) and the same holds for individual subjects after detrending: (mean $r = -.22 \pm .28$ after detrending). Thus, the time courses of variancebased and correlation-based decoding accuracies during this early interval are inconsistent with a common focal neural source.

5.6.6. Generating spatial maps

All spatial maps displayed were generated using the group-averaged covariances of the derivative for each class. Data in Fig. 4 is displayed from a posterior coronal view (as indicated in center). For each location we considered the cone generated by the contrast location "x" vs. the hemifield opposite "x" (the average covariance across 3 locations). The method in which we visualized the conic matrix is by conversion to a vector which forms a static head-map. As with functional connectivity, the more natural way to visualize the matrix is as a weighted graph between nodes, however, we chose to use a headmap to better visualize the evident retinotopic organization. To do this, we defined each channels contribution to the conic matrix as the sum of its squared contributions to each eigenvector v_j weighted by the corresponding eigenvalue λ_j . Thus the *i*th channel's weight (w_i) was:

$$w_i = \sum_{j=1}^n \lambda_j [v_j]_i^2 \tag{36}$$

Smooth maps were created by interpolating the channel-wise measurements after coronal projection. Linear triangulation-based interpolation was performed with the built-in MATLAB function 'griddata'.

5.6.7. Eigenlengths and eigencomponents

In the two class case (labels = ± 1), the conic boundary corresponds to $\partial C(P) := {\dot{x} | \dot{x}^T P \dot{x} = 0}$ with *P* the weighted inverse of covariance matrices:

$$P \coloneqq \frac{\sum_{(\dot{x}_{l-1})}^{-1}}{\det |\Sigma_{(\dot{x}_{l-1})}^{-1}|^{1/n}} - \frac{\sum_{(\dot{x}_{l+1})}^{-1}}{\det |\Sigma_{(\dot{x}_{l+1})}^{-1}|^{1/n}}$$
(37)

The class assignment function is $i(t) = \operatorname{sgn}(\dot{x}(t)^T P \dot{x}(t))$. As *P* is real symmetric, it has a spectral decomposition UDU^{H} . Splitting the diagonal matrix of eigenvalues (*D*) into positive and negative components: $D = D^+ - D^-$ produces:

$$\dot{x}^{T}P\dot{x} = \|\sqrt{D^{+}}(U^{H}\dot{x})\|_{2}^{2} - \|\sqrt{D^{-}}(U^{H}\dot{x})\|_{2}^{2}$$
(38)

As the right side of Eq. (38) involves the difference of two norms, we refer to them as the positive (with D^+) and negative (with D^-) eigenlengths. These eigenlengths are useful for generating a 2-dimensional projection of the data by which to compare separability as in Fig. 5C. Similarly we generate "eigencomponents" of the system by simply rotating it into the cone's coordinate-axes ($U^H \dot{x}(t)$). We call the eigencomponents positive or negative based upon the sign of the corresponding eigenvalue in *D* and these may be used to linearly decompose the signal into class-sensitive components as in Fig. 5A and B. In general, the magnitude of each component's eigenvalue is related to how tuned that component is for a specific class.

5.6.8. Testing the conic sensitivity to noise

To illustrate the invariance properties of our conic classifier we considered trials number 21, 24, 27, and 30 for the lower right and upper left locations each for subject 3. Using this data we calculated the discriminating cone and projected all subsequent data onto the cone's main eigenvectors for visualization (the first two corresponding to lower right and the first for upper left). Thus, all transformed data are plotted on the same axis generated from the original cone. We then considered three noise-transformations on the data: multiplicative noise, dynamic temporal warping (DTW), and multivariate box noise. To simulate multiplicative noise we simply multiplied the pre-derivative data with a random spline before computing derivatives. We also used a random spline for the dynamic temporal warping example in which the spline values served as the time values at which we evaluated the data through interpolation. Thus, the transformation under multiplicative noise between a univariate spline s(t) and multivariate data X (t) was calculated as s(t)X(t), while the transformation under DTW was calculated as X(s(t)) with non-integer values assessed through interpolation. In the latter case splines were constrained to be positive-valued and monotone. Random splines was calculated by first generating a set of random time spacing between interpolation control points (distributed ||N(0, 10 ms)|| and ||N(0, 50 ms)|| for multiplication and DTW, respectively. The x-values for interpolation control points were likewise randomly generated from $\mathcal{N}(0, .2)$ and $\|\mathcal{N}(0, .05)\|$, respectively. To generate a temporal scaling function, the *x*-values generated from DTW were converted to their cumulative sum to generate a monotone time sequence and then rescaled to match the data's temporal bounds. Box noise was applied additively using a randomly generated box function. To create these functions we first specified a fixed number of jump points (20) and a vector-valued variance for the jump amplitudes of each channel. We generated these variances from the distribution $2||\mathcal{N}(0, 1)||$. The spacing between jumps was generated from the distribution $5 \text{ ms} + N(0, 5 \text{ ms})^2$ separately for each channel and values were rescaled and floored to produce integer values within the data range. The amplitudes of each jump were generated from a normal distribution with the standard deviation mentioned previously (randomly drawn from 2||N(0, 1)|| for each channel). After noise was added (independently for each class) data was projected into the original conic eigenspace to determine whether the original geometry still held.

Acknowledgements

M.S. was funded by NSF-DGE-1143954 from the US National Science Foundation. T.B. acknowledges R37 MH066078 from the US National Institute of Health. S.C. holds a Career Award at the Scientific Interface from the Burroughs-Wellcome Fund. We thank J. Samaha for sending us his data and set of trials excluded from classification due to artifact. Portions of this work were supported by AFOSR 15RT0189, NSF ECCS 1509342 NSF CMMI 1537015, and NSF CMMI 1653589, from the US Air Force Office of Scientific Research and the US National Science Foundation, respectively.

References

- Adeli, H., Ghosh-Dastidar, S., Dadmehr, N., 2007. A wavelet-chaos methodology for analysis of EEGs and EEG subbands to detect seizure and epilepsy. IEEE Trans. Biomed. Eng. 54 (2), 205–211. https://doi.org/10.1109/TBME.2006.886855.
- Adeli, H., Ghost-Dastidar, S., Dadmehr, N., 2008. A spatio-temporal wavelet-chaos methodology for EEG-based diagnosis of Alzheimer's disease. Neurosci. Lett. 444 (2), 190–194. https://doi.org/10.1016/j.neulet.2008.08.008.
- Ahnert, K., Abel, M., 2007. Numerical differentiation of experimental data: local versus global methods. Comput. Phys. Commun. 177, 764–774. https://doi.org/10.1016/j. cpc.2007.03.009.
- Andrzejak, R.G., Lehnertz, K., Mormann, F., Rieke, C., David, P., Elger, C.E., 2001. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state. Phys. Rev. E 64 (061907), 1–8. https://doi.org/10.1103/PhysRevE.64.061907.
- Boser, B.E., Guyon, I.M., Vapnik, V.N., 1992. A training algorithm for optimal margin classifiers. Proc. of the Fifth Annual Workshop on Computational Learning Theory 144–152. https://doi.org/10.1145/130385.130401.
- Canolty, R.T., Edwards, E., Dalal, S.S., Soltani, M., Nagarajan, S.S., Kirsch, H.E., Berger, M.S., Barbaro, N.M., Knight, R.T., 2006. High gamma power is phase-locked to theta oscillations in human neocortex. Science 313, 1626–1628. https://doi.org/10.1126/ science.1128115.
- Chang, C., Glover, G.H., 2010. Time-frequency dynamics of resting-state brain connectivity measured with fMRI. NeuroImage 50 (1), 81–98. https://doi.org/10.1016/j. neuroimage.2009.12.011.
- Cohen, J.D., Perlstein, W.M., Braver, T.S., Nystrom, L.E., Noll, D.C., Jonides, J., Smith, E.E., 1997. Temporal dynamics of brain activation during a working memory task. Nature 386, 604–608. https://doi.org/10.1038/386604a0.
- Etzel, J., Zachs, J., Braver, T., 2013. Searchlight analysis: promise, pitfalls, and potential. NeuroImage 78, 261–269. https://doi.org/10.1016/j.neuroimage.2013.03.041.
- Feng, L., Wang, Y., Wu, J., 2017. Semiflows "monotone with respect to high-rank cones" on a Banach space. SIAM J. Math. Anal. 49 (1), 142–161. https://doi.org/10.1137/ 16M1064295.
- Foster, J.J., Sutterer, D.W., Serences, J.T., Vogel, E.K., Awh, E., 2016. The topography of alpha-band activity tracks the content of spatial working memory. J. Neurophysiol. 115 (1), 168–177. https://doi.org/10.1152/jn.00860.2015.
- Friedman, J., Hastie, T., Tibshirani, R., 2008. Sparse inverse covariance estimation with the graphical lasso. Biostatistics 9 (3), 432–441. https://doi.org/10.1093/ biostatistics/kxm045.
- Fuentemilla, L., Penny, W.D., Cashdollar, N., Bunzeck, N., Düzel, E., 2010. Theta-coupled periodic replay in working memory. Curr. Biol. 20 (7), 606–612. https://doi.org/10. 1016/j.cub.2010.01.057.
- Güler, N.F., Übeyli, E.D., Güler, İ., 2005. Recurrent neural networks employing Lyapunov exponents for EEG signals classification. Expert Syst. Appl. 29 (3), 506–514. https:// doi.org/10.1016/j.eswa.2005.04.011.
- Górecki, T., Łuczak, M., 2013. Using derivatives in time series classification. Data Min. Knowl. Discov. 26 (2), 310–331. https://doi.org/10.1007/s10618-012-0251-4.
- Górecki, T., Łuczak, M., 2015. Multivariate time series classification with parametric derivative dynamic time warping. Expert Syst. Appl. 42 (5), 2305–2312. https://doi. org/10.1016/j.eswa.2014.11.007.
- Garcia, J.O., Srinivasan, R., Serences, J.T., 2013. Near-real-time feature-selective modulations in human cortex. Curr. Biol. 23 (6), 515–522. https://doi.org/10.1016/j.cub. 2013.02.013.
- Grech, R., Cassar, T., Muscat, J., Camilleri, K.P., Fabri, S.G., Zervakis, M., Xanthopoulos, P., Sakkalis, V., Vanrumste, B., 2008. Review on solving the inverse problem in EEG source analysis. J. NeuroEng. Rehabil. 5 (1), 25. https://doi.org/10.1186/1743-0003-5-25.
- 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. https://doi.org/10.1016/j.neuroimage.2013.10. 067.
- Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P., 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293 (5539), 2425–2430. https://doi.org/10.1126/science.1063736.
- Haxby, J.V., Connolly, A.C., Gutupalli, J.S., 2014. Decoding neural representational spaces using multivariate pattern analysis. Annu. Rev. Neurosci. 37 (1), 435–456. https://doi.org/10.1146/annurev-neuro-062012-170325.
- Hindmarsh, J.L., Rose, R.M., 1984. A model of neuronal bursting using three coupled first order differential equations. Proc. R. Soc. Lond: B 221, 87–102. https://doi.org/10.

M.F. Singh et al.

1098/rspb.1984.0024.

- Honey, C.J., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J.P., Meuli, R., Hagmann, P., 2009. Predicting human resting-state functional connectivity from structural connectivity. PNAS 106 (6), 2035–2040. https://doi.org/10.1073/pnas.0811168106.
- Hopfield, J., 1984. Neurons with graded response have collective computational properties like those of two-state neurons. PNAS 81, 3088–3092. https://doi.org/10. 1073/pnas.81.10.3088.
- Iasemidis, L.D., Sackellares, J.C., Zaveri, H.P., Williams, W.J., 1990. Phase space topography and the Lyapunov exponent of electrocorticograms in partial seizures. Brain Topogr. 2 (3), 187–201. https://doi.org/10.1007/BF01140588.
- Jafarpour, A., Horner, A.J., Fuentemilla, L., Penny, W.D., Duzel, E., 2013. Decoding oscillatory representations and mechanisms in memory. Neuropsychologia 51 (4), 772–780. https://doi.org/10.1016/j.neuropsychologia.2012.04.002.
- Jerbi, K., Ossandón, T., Hamamé, C.M., Senova, S., Dalal, S.S., Jung, J., Minotti, L., Bertrand, O., Berthoz, A., Kahane, P., Lachaux, J.-P., 2009. Task-related gamma-band dynamics from an intracerebral perspective: review and implications for surface EEG and MEG. Hum. Brain Mapp. 30 (6), 1758–1771. https://doi.org/10.1002/hbm. 20750.
- Kafashan, M., Ching, S., Palanca, B.J.A., 2016. Sevoflurane alters spatiotemporal functional connectivity motifs that link resting-state networks during wakefulness. Front. Neural Circuits 10, 107. https://doi.org/10.3389/fncir.2016.00107.
- King, J.-R., Dehaene, S., 2014. Characterizing the dynamics of mental representations: the temporal generalization method. Trends Cogn. Sci. (4), 203–210. https://doi.org/10. 1016/j.tics.2014.01.002.
- Kutas, M., Federmeier, K.D., 2011. Thirty years and counting: finding meaning in the n400 component of the event-related brain potential (ERP). Annu. Rev. Psychol. 62 (1), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123.
- Kutas, M., Hillyard, S., 1980. Reading senseless sentences: brain potentials reflect semantic incongruity. Science 207 (4427), 203–205. https://doi.org/10.1126/science. 7350657.
- Lloyd, S.P., 1982. Least squares quantization in PCM. IEEE T IT (2), 129–137. https://doi. org/10.1109/TIT.1982.1056489.
- Lorenz, E.N., 1963. Deterministic nonperiodic flow. J. Atmos. Sci. 130–141. https://doi. org/10.1175/1520-0469(1963)020 < 0130:DNF > 2.0.CO;2.
- Łuczak, M., 2016. Hierarchical clustering of time series data with parametric derivative dynamic time warping. Expert Syst. Appl. 62, 116–130. https://doi.org/10.1016/j. eswa.2016.06.012.
- Makeig, S., Westerfield, M., Jung, T.-P., Enghoff, S., Townsend, J., Courchesne, E., Sejnowski, T.J., 2002. Dynamic brain sources of visually evoked responses. Science 295 (5555), 690–694. https://doi.org/10.1126/science.1066168.
- Masterton, R.A., Abbott, D.F., Fleming, S.W., Jackson, G.D., 2007. Measurement and reduction of motion and ballistocardiogram artefacts from simultaneous EEG and fMRI recordings. NeuroImage 37, 202–211. https://doi.org/10.1016/j.neuroimage. 2007.02.060.
- Monto, S., Palva, S., Voipio, J., Palva, J.M., 2008. Very slow EEG fluctuations predict the dynamics of stimulus detection and oscillation amplitudes in humans. J. Neurosci. 28 (33), 8268–8272. https://doi.org/10.1523/JNEUROSCI.1910-08.2008.
- Norman, K.A., Polyn, S.M., Detre, G.J., Haxby, J.V., 2006. Beyond mind-reading: multivoxel pattern analysis of fMRI data. Trends Cogn. Sci. 10 (9), 424–430. https://doi. org/10.1016/j.tics.2006.07.005.
- Riehl, J.R., Palanca, B.J., Ching, S., 2017. High energy brain dynamics during anesthesiainduced unconsciousness. Netw. Neurosci. https://doi.org/10.1162/NETN a 00023.

- Samaha, J., Sprague, T.C., Postle, B.R., 2016. Decoding and reconstructing the focus of spatial attention from the topography of alpha-band oscillations. J. Cogn. Neurosci. 28 (8), 1090–1097. https://doi.org/10.1162/jocn_a_00955.
- Sanchez, L.A., 2009. Cones of rank 2 and the Poincaré–Bendixson property for a new class of monotone systems. J. Differ. Equ. 246, 1978–1990. https://doi.org/10.1016/j.jde. 2008.10.015.
- Sanchez, L.A., 2010. Existence of periodic orbits for high-dimensional autonomous systems. J. Math. Anal. Appl. 363 (2), 409–418. https://doi.org/10.1016/j.jmaa.2009. 08.058.

Sauer, T., Yorke, J., Casdagli, M., 1991. Embedology. J. Stat. Phys. 65, 579–616. https:// doi.org/10.1007/BF01053745.

- Schmidt, N.M., Blankertz, B., Treder, M.S., 2011. α-Modulation induced by covert attention shifts as a new input modality for EEG-based BCIS. 2010 IEEE International Conference on Systems Man and Cybernetics (SMC) 481–487. https://doi.org/10. 1109/ICSMC.2010.5641967.
- Sitges, C., Bornas, X., Llabrés, J., Noguera, M., Montoya, P., 2010. Linear and nonlinear analyses of EEG dynamics during non-painful somatosensory processing in chronic pain patients. Int. J. Psychophysiol. 77 (2), 176–183. https://doi.org/10.1016/j. ijpsycho.2010.05.010.
- Smith, R.A., 1980. Existence of periodic orbits of autonomous ordinary differential equations. Proc. R. Soc. Edinb. Sect. A: Math. 85 (1–2), 153–172. https://doi.org/10. 1017/S030821050001177X.

Smith, H.L., 2008. Monotone Dynamical Systems: An Introduction to the Theory of Competitive and Cooperative Systems, No. 41. American Mathematical Society.

- Spampinato, C., Palazzo, S., Kavasidis, I., Giordano, D., Souly, N., Shah, M., 2017. 2017 IEEE Conference on Computer Vision and Pattern Recognition. https://doi.org/10. 1109/CVPR.2017.479.
- Tajima, S., Yanagawa, T., Fujii, N., Toyoizumi, T., 2015. Untangling brain-wide dynamics in consciousness by cross-embedding. PLoS Comput. Biol. https://doi.org/10.1371/ journal.pcbi.1004537.

Takens, F., 1981. Detecting strange attractors in turbulence. In: Rand, D., Young, L.-S. (Eds.), Lecture Notes in Mathematics 898. Springer-Verlag, Berlin, pp. 366–381.

- Torta, A.B.L., Komorowskie, R.W., Mannsf, J.R., Kopellc, N.J., Eichenbaume, H., 2009. Theta-gamma coupling increases during the learning of item-context associations. PNAS 106 (49), 20942–20947. https://doi.org/10.1073/pnas.0911331106.
- Treder, M.S., Bahramisharif, A., Schmidt, N.M., van Gerven, M.A., Blankertz, B., 2011. Brain-computer interfacing using modulations of alpha activity induced by covert shifts of attention. J. NeuroEng. Rehabil. 8 (24). https://doi.org/10.1186/1743-0003-8-24.
- van de Nieuwenhuijzen, M.E., Backus, A.R., Bahramisharif, A., Doeller, C.F., Jensen, O., van Gerven, M.A.J., 2013. MEG-based decoding of the spatiotemporal dynamics of visual category perception. NeuroImage 1063–1073. https://doi.org/10.1016/j. neuroimage.2013.07.075.
- Wang, F., Gelfand, A.E., 2013. Directional data analysis under the general projected normal distribution. Stat. Methodol. 10 (1), 113–127. https://doi.org/10.1016/j. stamet.2012.07.005.
- Wang, J., 2007. Wavelet approach to numerical differentiation of noisy functions. Commun. Pure Appl. Anal. 6 (3), 873–897. https://doi.org/10.3934/cpaa.2007.6. 873.
- Wen, H., Shi, J., Zhang, Y., Lu, K.-H., Cao, J., Liu, Z., 2017. Neural encoding and decoding with deep learning for dynamic natural vision. Cereb. Cortex 1–25. https://doi.org/ 10.1093/cercor/bhx268.