

# Seizure Detection and Advanced Monitoring Techniques

## Epilepsy: A Dynamical Process

The EEG is a complex set of signals with statistical properties variable in terms of both time and space.<sup>1</sup> The individual characteristics of the EEG, such as bursting events (during stage II sleep), limit cycles (alpha activity, mu activity, and ictal activity), amplitude-dependent frequency behavior (the smaller the amplitude, the higher the EEG frequency), and frequency harmonics (particularly in association with photic driving conditions), are but a few of the vast array of concerns related to the properties typical of nonlinear systems. The EEG of the epileptic brain is a nonlinear signal with numerous confusing deterministic and possibly even chaotic properties.<sup>2,3,4</sup>

### QUESTION 14.1: The EEG of the epileptic patient is

- a. Linear
- b. Nonlinear

**ANSWER:** b.

The voltage of the EEG is represented by a series of numerical values over time and space (multi-electrodes), known as a multivariate time series. The standard methods for time series analysis (e.g., power analysis, linear orthogonal transforms, and parametric linear modeling) fail to detect the critical features of a time series generated by a nonlinear system. In some cases this can even result in the false indication that most of the series is random noise.<sup>5</sup> While we are unable to measure all of the relevant variables in the case of a multidimensional, nonlinear system such as the generators of EEG signals, this problem can be addressed mathematically. Since all dynamical systems have variables related over time, one may obtain information about the important dynamical features of the whole system by analyzing a single variable (e.g., voltage) over time. By analyzing more than one variable over time, we can follow the dynamics of the interactions of different parts of the system. Neuronal networks can generate a variety of activities, some of which are characterized by rhythmic or semirhythmic signals which are reflected in the corresponding local EEG field potential. An essential feature of these networks is that variables of the network have both a strong nonlinear range and complex interactions.

Characteristics of the dynamics can depend strongly on small changes in the control parameters and/or the initial conditions. Real neuronal networks behave with complex nonlinear characteristics and can display changes between states such as small-amplitude, quasi-random fluctuations and large-amplitude, rhythmic oscillations. These dynamical state transitions are observed during the transition between interictal state and epileptic seizure onset. A functional system must stay within a given range in order for the system to maintain stable operation. The most essential difference between a normal and an epileptic network can be conceptualized as a decrease in the distance between operating and bifurcation points.

In considering epilepsy as a dynamical disorder of neuronal networks, Lopes da Silva et al.<sup>1</sup> proposed two scenarios of how a seizure could evolve. The first is that a seizure could be caused by a sudden and abrupt state transition, in which case it would not be preceded by detectable dynamical changes in the EEG. Such a scenario would be conceivable for the initiation of seizures in primary generalized epilepsy. Alternatively, this transition could occur as a gradual change or a cascade of changes in dynamics, which could, in theory, be detected and possibly even anticipated.

## Seizure Detection

Most of the current techniques used to detect or “predict” an epileptic seizure involve linear or nonlinear transformation of the signal using one of several mathematical models and subsequently trying to predict or detect the seizure based on the results. These models include some purely mathematical transformations, such as the Fourier transform, and machine learning techniques, like artificial neural networks, or some combination of the two. In this section, we review some of these modeling techniques for detection and prediction of seizures.

Many techniques have been used in an attempt to detect the EEG signature of epileptic seizures. The typical EEG is read visually and the seizure is detected as described in some of the preceding chapters. In recent years, there has been a great deal of research on trying to predict or detect a seizure based off the EEG. The majority of these techniques use some kind of time series analysis method to detect seizures offline. Time series analysis of an EEG falls under one of the following two groups:

1. *Univariate time series analysis* that refers to time series that consist of a single observation recorded sequentially over equal time increments. Time is an implicit variable in the time series. Information on the start time and the sampling rate of the data collection can allow one to visualize the univariate time series graphically as a function of time over the entire duration of data recording. The information contained in the amplitude value of the recorded EEG signal sampled in the form of a discrete time series  $x(t) = x(t_i) = x(i\Delta t)$ , ( $i = 1, 2, \dots, N$  and  $\Delta t$  is the sampling interval) can also be encoded through the amplitude and the phase of the sub-set of harmonic oscillations over a range of different frequencies.

### QUESTION 14.2: Univariate time series analysis consists of

- a. Multiple observations recorded sequentially over equal time increments
- b. A single observation recorded sequentially over variable time increments
- c. Multiple observations recorded sequentially over different time increments
- d. A single observation recorded sequentially over equal time increments

ANSWER: d.

2. *Multivariate time series analysis* that refers to time series that consist of more than one observation recorded sequentially in time. Multivariate time series analysis is used when one wants to understand the interaction between the different components of the system under consideration. As in univariate time series, time is also an implicit variable in the multivariate time series.

## Univariate Time Series Analysis

### Short-Term Fourier Transform

Power spectral analysis of the EEG is one of the more widely used techniques for detecting or predicting an epileptic seizure. The basic hypothesis is that the EEG signal, when partitioned into its component periodic (sine/cosine waves) elements, has a signature that varies between

the ictal and the interictal states. In order to detect this signature, the Fourier transform of the signal is calculated, and then the frequencies that are most prominent (in amplitude) are identified. It has been shown that there is a relationship between the power spectrum of the EEG signal and the ictal activity.<sup>6</sup> Although there is some correlation between the power spectrum and the ictal activity, the power spectrum is not used as a stand-alone detector of a seizure. It has been coupled with some other time series prediction technique or machine learning modality to detect a seizure.

The Fourier transform breaks up any time varying signal into its frequency components of varying magnitude and is defined in Eq. (14-1).

$$F(k) = \int_{-\infty}^{\infty} f(t)e^{-2\pi ikx} dx \quad (14-1)$$

Euler's formula allows this to be written as shown in Eq. (14-2) for any complex function  $f(t)$ , where  $k$  is the  $k$ th harmonic frequency.

$$F(k) = \int_{-\infty}^{\infty} f(t) \cos(-2\pi kx) dx + \int_{-\infty}^{\infty} f(t) i \sin(-2\pi kx) dx \quad (14-2)$$

Utilizing this system, any time varying signal can be represented as a summation of sine and cosine waves of varying magnitude and frequencies.<sup>7</sup> The Fourier transform is represented with the power spectrum. The power spectrum has a value for each harmonic frequency, which indicates how strong that frequency is in the given signal. The magnitude of this value is calculated by taking the modulus of the complex number that is calculated from the Fourier transform for a given frequency ( $|F(k)|$ ).

Stationarity must be considered when using the Fourier transform. A stationary signal is one that is constant in its statistical parameters over time, and is assumed by the Fourier transform to be present. A signal which is made up of different frequencies at different times will yield the same transform as a signal which is made up of those same frequencies for the entire time period considered. As an example, consider two functions  $f_1$  and  $f_2$  over the domain  $0 \leq t \leq T$  for any two frequencies  $\omega_1$  and  $\omega_2$  shown in Eqs. (14-3) and (14-4).

$$f_1(t) = \sin(2\pi\omega_1 t) + \cos(2\pi\omega_2 t) \quad \text{if } 0 \leq t < T \quad (14-3)$$

and

$$f_2(t) = \begin{cases} \sin(2\pi\omega_1 t) & \text{if } 0 \leq t < T/2 \\ \cos(2\pi\omega_2 t) & \text{if } T/2 \leq t < T \end{cases} \quad (14-4)$$

When using the short-term Fourier transform, the assumption is made that the signal is stationary for some small period of time,  $T_s$ . The Fourier transform is then calculated for segments of the signal of length  $T_s$ . The short-term Fourier transform at time  $t$  gives the Fourier transform calculated over the segment of the signal lasting from  $(t - T_s)$  to  $t$ . The length of  $T_s$  determines the resolution of the analysis. There is a trade-off between time and frequency resolution. A short  $T_s$  yields better time resolution; however, it limits the frequency resolution. The opposite of this is also true; a long  $T_s$  increases frequency resolution while decreasing the time resolution of the output. Other modalities, such as wavelet analysis, can alleviate this limitation. Wavelet analysis provides a model that maintains both time and frequency resolution.

### Discrete Wavelet Transforms

Wavelet transforms follow the principle of superposition, just like Fourier transforms, and assume that EEG signals are composed of various elements from a set of parameterized

basis functions. Wavelets must meet certain mathematical criteria, which allow for the basis functions to be far more general than simple sine/cosine waves as in the Fourier transform. Wavelets make it substantially easier to approximate sharply contoured waveforms such as spikes, as compared to the Fourier transform. Fourier transforms have a limited ability to approximate a spike because of the sine (and cosine) waves' infinite support (i.e., stretch out to infinity in time). In the case of wavelets, there is the possibility of finite support, allowing the estimation of the spike by changing the magnitude of the component basis functions.

The discrete wavelet transform is similar to the Fourier transform in that it will break up any time varying signal into smaller uniform functions, known as the basis functions. The basis functions are created by scaling and translating a single function of a certain form. This function is known as the Mother wavelet. In the case of the Fourier transform, the basis functions used are sine and cosine waves of varying frequency and magnitude. The only requirement for a family of functions to be a basis is that the functions should be both complete and orthonormal under the inner product. Consider the family of functions  $\psi = \{\psi_{ij} \mid -\infty < i, j < \infty\}$ , where each  $i$  value specifies a different scale and each  $j$  value specifies a different translation based off of some mother wavelet function.  $\psi$  is considered to be complete if any continuous function  $f$ , defined over the real line,  $x$ , can be defined by some combination of the functions in  $\psi$  as shown in Eq. (14-5).<sup>7</sup>

$$f(x) = \sum_{i,j=-\infty}^{\infty} c_{ij} \psi_{ij}(x) \quad (14-5)$$

A family of functions must meet two criteria to be orthonormal under the inner product. It must be the case that for any  $i, j, l$ , and  $m$  where  $i \neq l$  and  $j \neq m$  that  $\langle \psi_{ij}, \psi_{lm} \rangle \geq 0$  and  $\langle \psi_{ij}, \psi_{ij} \rangle \geq 1$ , where  $\langle f, g \rangle$  is the inner product and is defined as shown in Eq. (14-6) and  $f(x)^*$  is the complex conjugate of  $f(x)$ .

$$\langle f, g \rangle = \int_{-\infty}^{\infty} f(x)^* g(x) dx \quad (14-6)$$

The wavelet basis is very similar to the Fourier basis, with the exception that the wavelet basis does not have to be infinite. In a wavelet transform, the basis functions can be defined over a certain window and then be zero everywhere else. As long as the family of functions defined by scaling and translating the mother wavelet is orthonormally complete, that family of functions can be used as the basis. With the Fourier transform, the basis is made up of sine and cosine waves that are defined over all values of  $x$  where  $-\infty < x < \infty$ .

One of the simplest wavelets is the *Haar Wavelet* (Daubechies 2 wavelet). In a manner similar to the Fourier series, any continuous function  $f(x)$  defined on  $[0,1]$  can be represented using the expansion shown in Eq. (14-7).  $h_{j,k}(x)$  is known as the Haar wavelet function and is defined as shown in Eq. (14-8) and  $p_{j,k}(x)$  is known as the Haar scaling function and is defined as shown in Eq. (14-9).<sup>7</sup>

$$f(x) = \sum_{j=J}^{\infty} \sum_{k=0}^{2^j-1} \langle f, h_{j,k} \rangle h_{j,k}(x) + \sum_{k=0}^{2^J-1} \langle f, p_{J,k} \rangle p_{J,k}(x) \quad (14-7)$$

$$h_{j,k}(x) = \begin{cases} 2^{j/2} & \text{if } 0 \leq 2^j x - k < 1/2 \\ -2^{j/2} & \text{if } 1/2 \leq 2^j x - k < 1 \\ 0 & \text{otherwise} \end{cases} \quad (14-8)$$

$$p_{J,k}(x) = \begin{cases} 2^{J/2} & \text{if } 0 \leq 2^J x - k < 1 \\ 0 & \text{otherwise} \end{cases} \quad (14-9)$$

The combination of the Haar scaling function at the largest scale along with the Haar wavelet functions creates a set of functions which is an orthonormal basis for functions in  $R^2$ .

Spectral entropy calculates some features based on the power spectrum. Entropy was first used in physics as a thermodynamic quantity describing the amount of disorder in a system but can also be used to calculate the entropy for a given probability distribution.<sup>8</sup> The entropy measure that Shannon developed can be expressed as shown in Eq. (14-10).

$$H = -\sum p_k \log p_k \quad (14-10)$$

Entropy is a measure of how much information there is to learn from a random event occurring. Events that are unlikely to occur yield more information than events that are very probable. For spectral entropy, the power spectrum is considered to be a probability distribution. The spectral entropy is an indicator of the number of frequencies that make up a signal. A signal made up of many different frequencies (e.g., white noise), would have a relatively uniform distribution and therefore yield high spectral entropy. Conversely, a signal made up of a single frequency would yield low spectral entropy.

A wavelet filter can be used to partition the EEG between seizure and nonseizure states. It triggered any increase in power or shift in frequency regardless of cause, whether this change in the signal was caused by an artifact, normal EEG activity, interictal epileptiform discharges, or ictal activity. The signals were then passed through a second filter that tried to isolate the seizures from the other activity. By decomposing the signal into components and passing it through the second step of isolating the seizures, the authors were able to detect all seizures with an average of 2.8 false positives per hour.<sup>9</sup> This system was unable to reliably predict seizure onset.

### Statistical Moments

It is possible to describe an approximation to the distribution of a random variable using moments and functions of moments, even when a cumulative distribution function for such a variable cannot be determined.<sup>10</sup> Statistical moments relate information about the distribution of the amplitude of a given signal. In probability theory, the  $k$ th moment is defined as shown in Eq. (14-11), where  $E[x]$  is the expected value of  $x$ .

$$\mu_k = E[x^k] = \int x^k p(x) \quad (14-11)$$

The first statistical moment is the mean of the distribution being considered.

In general, the statistical moments are taken about the mean. This is also known as the  $k$ th central moment and is defined by Eq. (14-12), where  $\mu$  is the mean of the data set considered.<sup>10</sup>

$$\mu_k = E[(x - \mu)^k] = \int (x - \mu)^k p(x) \quad (14-12)$$

The second moment about the mean would give the variance. The third and fourth moments about the mean would produce the skew and kurtosis, respectively. The skew of a distribution indicates the amount of asymmetry in that distribution, while the kurtosis shows the degree of “peakedness” of that distribution. The absolute value of the skewness  $|\mu_3|$  was used for seizure prediction in a review by Mormann et al.<sup>4</sup> Skewness was not able to significantly predict a seizure by detecting the state change from interictal to preictal.

### Recurrence Time Statistics

The recurrence time statistic (RTS), T1, is a characteristic of trajectories in an abstract dynamical system. T1 has been calculated for some ECoG data in an effort to detect seizures, with significant success. With two different patients and a total of 79 h of data, researchers were able to detect 97% of the seizures with only an average of 0.29 false negatives per hour.<sup>11</sup> Results from our preliminary studies on human EEG signals showed that the RTS exhibited significant change during the ictal period that distinguishes from the background interictal period. It may be possible to use RTS in the development of an automated seizure-warning algorithm.

### Lyapunov Exponent

Patients go through a preictal transition approximately 1/2 to 1 h before a seizure occurs. This preictal state can be characterized using the *Lyapunov exponent*.<sup>12-20</sup> The Lyapunov exponent measures the speed of divergence of nearby trajectories in a dynamical system. The noted approach therefore treats the epileptic brain as a dynamical system.<sup>21-23</sup> It considers a seizure as a transition from a chaotic state (where trajectories are sensitive to initial conditions) to an ordered state (where trajectories are insensitive to initial conditions) in the dynamical system. The Lyapunov exponent is a nonlinear measure of the average rate of divergence/convergence of two neighboring trajectories in a dynamical system dependent on the sensitivity of initial conditions. It has been successfully used to identify preictal changes in the EEG data.<sup>12-14</sup> Lyapunov exponents can be estimated from the equation of motion describing the time evolution of a given dynamical system. However, in the absence of equation of motion describing the trajectory of the dynamical system, Lyapunov exponents are determined from observed scalar time series data,  $x(t_n) = x(n \delta t)$ , where  $\delta t$  is the sampling rate for the data acquisition. In this situation, the goal is to generate a higher dimensional vector embedding of the scalar data  $x(t)$  that defines the state space of the multivariate brain dynamics from which the scalar EEG data are derived. This is achieved by constructing a higher dimensional vector  $\mathbf{x}_i$  from the data segment  $x(t)$  of given duration  $T$ , as shown in Eq. (14-13), with  $\tau$  defining the embedding delay used to construct a higher dimensional vector  $x$  from  $x(t)$  and  $d$  is the selected dimension of the embedding space and  $t_i$  is the time instance within the period  $[T - (d - 1)\tau]$ .

$$\mathbf{x}_i = [x(t_i), x(t_i - \tau), \dots, x(t_i - (d - 1)\tau)] \quad (14-13)$$

The geometrical theorem<sup>24</sup> states that for an appropriate choice of  $d > d_{\min}$ ,  $\mathbf{x}_i$  provides a faithful representation of the phase space for the dynamical systems from which the scalar time series was derived. A suitable practical choice for  $d$ , the embedding dimension, can be derived from the “false nearest neighbor” algorithm. In addition, a suitable prescription for selecting the embedding delay,  $\tau$ , is also given by Abarbanel.<sup>25</sup> From  $\mathbf{x}_i$ , the most stable short-term estimation of the largest Lyapunov exponent can be performed, which is referred to as “short-term largest Lyapunov exponent” (STL<sub>max</sub>).<sup>14</sup> The estimation  $L$  of STL<sub>max</sub> is obtained using Eq. (4-14) where  $\delta x_{ij}(0) = x(t_i) - x(t_j)$  is the displacement vector, defined at time points  $t_i$  and  $t_j$ , and  $\delta x_{ij}(\Delta t) = x(t_i + \Delta t) - x(t_j + \Delta t)$  is the same vector after time  $\Delta t$ .  $N$  is the total number of local STL<sub>max</sub> that will be estimated within the time period  $T$  of the data segment, where  $T = N\Delta t + (d - 1)\tau$ .

$$L = \frac{1}{N\Delta t} \sum_{i=1}^N \log_2 \left| \frac{\delta x_{ij}(\Delta t)}{\delta x_{ij}(0)} \right| \quad (14-14)$$

A decrease in the Lyapunov exponent indicates this transition to a more ordered state. The assumptions underlying this methodology have been experimentally observed in the

Q1

STL<sub>max</sub> time series data from both human patients and rodents. This characterization by the Lyapunov exponent has, however, been successful only for EEG data recorded from particular areas in the neocortex and hippocampus and has been unsuccessful for other areas. Unfortunately, these areas can vary from seizure to seizure even in the same patient. The method is therefore very sensitive to the electrode sites chosen. When the correct sites were chosen, the preictal transition was seen in more than 91% of the seizures. On average, this led to a prediction rate of 80.77% and an average warning time of 63 min.<sup>19</sup> Unfortunately for the reasons stated above, this method has been plagued by problems limiting its predictive capacity.

### Multivariate Measures

Multivariate analysis measures multiple channels of EEG simultaneously. This system considers the interactions between the channels and the correlation between them rather than analyzing individual channels. This is useful if there is some interaction (e.g., synchronization) between different regions of the brain preceding a seizure. Of the techniques discussed below, the simple synchronization measure and the lag synchronization measure are bivariate measures. Bivariate measures only consider two channels at a time and define how those two channels correlate. The other measures account for all of the EEG channels simultaneously. This is accomplished via a dimensionality reduction technique known as principal component analysis (PCA).

### Simple Synchronization Measure

Since there are abnormally large amounts of highly synchronous activity during seizures, possibly beginning hours before ictal onset, analysis of synchronicity may be of benefit. Quiroga et al.<sup>26</sup> suggested a multivariate method to calculate the synchronization between two EEG channels. First, it defines certain “events” for a pair of signals. Then, once the “events” have been defined in the signals, this method then counts the number of times the “events” in the two signals occur within a specified amount of time ( $\tau$ ) of each other. It then divides this count by a normalizing term equivalent to the maximum number of events that could be synchronized in the signals.

For two discrete EEG channels  $x_i$  and  $y_j$ ,  $i = 1, \dots, N$ , where  $N$  is the number of points making up the EEG signal for the segment considered, event times are defined to be  $t_i^x$  and  $t_j^y$  ( $i = 1, \dots, m_x; j = 1, \dots, m_y$ ). An event can be defined to be anything; however, events should be chosen so that the events appear simultaneously across the signals when they are synchronized. Quiroga et al.<sup>27</sup> define an event to be a local maximum over a range of  $K$  values. In other words, the  $i$ th point in signal  $x$  would be an event if  $x_i > x_{i \pm k}$ ,  $k = 1, \dots, K$ .  $\tau$ , which is the time within which events from  $x$  and  $y$  must occur in order to be considered synchronized, needs to be less than half of the minimum inter-event distance; otherwise, a single event in one signal could be considered to be synchronized with two different events in the other signal.

Finally, the number of events in  $x$  that appear “shortly” (within  $\tau$ ) after an event in  $y$  is counted as shown in Eq. (14-15), and  $J_{ij}^\tau$  is defined as shown in Eq. (14-16).

$$c^\tau(x|y) = \sum_{i=1}^{m_x} \sum_{j=1}^{m_y} J_{ij}^\tau \quad (14-15)$$

$$J_{ij}^\tau = \begin{cases} 1 & \text{if } 0 < t_i^x - t_j^y \\ 1/2 & \text{if } t_i^x = t_j^y \\ 0 & \text{else} \end{cases} \quad (14-16)$$

Similarly, the number of events in  $y$  that appear shortly after an event in  $x$  can also be defined in an analogous way. This would be denoted as  $c^\tau(y | x)$ . With these two values, the synchronization measure  $Q_\tau$  can be calculated. This measure is shown in Eq. (14.17).

$$Q_\tau = \frac{c^\tau(x | y) + c^\tau(y | x)}{\sqrt{m_x m_y}} \quad (14-17)$$

The metric is normalized so that  $0 \leq Q_\tau \leq 1$  and  $Q_\tau$  is 1 if and only if  $x$  and  $y$  are fully synchronized (always have corresponding events within  $\tau$ ).

### Lag Synchronization

When two different systems are identical with the exception of a shift by some time lag,  $\tau$ , they are said to be lag synchronized.<sup>27</sup> To calculate the similarity of two signals, they used a normalized cross correlation function Eq. (14.18).

$$C(s_a, s_b)(\tau) = \left| \frac{\text{corr}(s_a, s_b)(\tau)}{\sqrt{\text{corr}(s_a, s_a)(0) \cdot \text{corr}(s_b, s_b)(\tau)}} \right| \quad (14-18)$$

where  $\text{corr}(s_a, s_b)(\tau)$  represents the linear cross correlation function between the two time series  $s_a(t)$  and  $s_b(t)$  computed at lag time  $\tau$  as defined in Eq. (14.19).

$$\text{corr}(s_a, s_b)(\tau) = \int_{-\infty}^{\infty} s_a(t + \tau) s_b(t) dt \quad (14-19)$$

The normalized cross correlation function yields a value between 0 and 1, which indicates how similar the two signals ( $s_a$  and  $s_b$ ) are. If the normalized cross correlation function produces a value close to 1 for a given  $\tau$ , then the signals are considered to be lag synchronized by a phase of  $\tau$ . Hence the final feature used to calculate the lag synchronization is the largest normalized cross correlation over all values of  $\tau$ , shown in Eq. (14-20). A  $C_{\max}$  value of 1 indicates totally synchronized signals within some time lag  $\tau$  and unsynchronized signals produce a value very close to 0.

$$C_{\max} = \max_{\tau} \{C(s_a, s_b)(\tau)\} \quad (14-20)$$

### Principal Component Analysis

PCA attempts to solve the problem of excessive dimensionality by combining features to reduce the overall dimensionality. PCA takes a data set in a multi-dimensional space, finds the most prominent dimensions in that data set, and linearly transforms the original data set to a lower dimensional space using the most prominent dimensions from the original data set. ~~PCA is used as a seizure detection technique itself.<sup>28</sup> It is also used as a~~ tool to extract the most important dimensions from a data matrix containing paired correlation information for all EEG channels. Only an outline of the derivation of PCA is given here. The reader should refer to Duda et al.<sup>29</sup> for a more detailed mathematical derivation.

Given a  $d$ -dimensional data set of size  $n$  ( $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ ), we first consider the problem of finding a vector  $\mathbf{x}_0$  to represent all of the vectors in the data set. This comes down to the problem of finding the vector  $\mathbf{x}_0$ , which is closest to every point in the data set. The vector is identified by minimizing the sum of the squared distances between  $\mathbf{x}_0$  and all of the points in the data set. The goal is to identify the value of  $\mathbf{x}_0$  that minimizes the criterion function  $J_0$  shown in Eq. (14-21).

$$J_0(\mathbf{x}_0) = \sum_{k=1}^n \|\mathbf{x}_0 - \mathbf{x}_k\|^2 \quad (14-21)$$

It can be shown that the value of  $\mathbf{x}_0$  which minimizes  $J_0$  is the sample mean ( $1/N \sum \mathbf{x}_k$ ) of the data set.<sup>29</sup> The sample mean has zero dimensionality and therefore provides no information regarding the spread of the data. In order to represent this information, the data set would need to be projected onto a space with some dimensionality. In order to project the original data set onto a one-dimensional space, it must be projected onto a line in the original space that runs through the sample mean. The data points in the new space can then be defined by  $\mathbf{x} = \mathbf{m} + a\mathbf{e}$ . Here,  $\mathbf{e}$  is the unit vector in the direction of the line and  $a$  is a scalar, which represents the distance from  $\mathbf{m}$  to  $\mathbf{x}$ . A second criterion function  $J_1$  can now be defined which calculates the sum of the squared distances between the points in the original data set and the projected points on the line Eq. (14-22).

$$J_1(a_1, \dots, a_n, \mathbf{e}) = \sum_{k=1}^n \|(\mathbf{m} + a_k \mathbf{e}) - \mathbf{x}_k\|^2 \quad (14-22)$$

Taking into consideration that  $\|\mathbf{e}\| = 1$ , the value of  $a_k$  which minimizes  $J_1$  is found to be  $a_k = \mathbf{e}^t(\mathbf{x}_k - \mathbf{m})$ . In order to find the best direction,  $\mathbf{e}$ , for the line, this value of  $a_k$  is substituted back into Eq. (14-22) to get Eq. (14-23).  $J_1$  from Eq. (14-23) can now be minimized with respect to  $\mathbf{e}$  to find the direction of the line. It turns out that the vector which minimizes  $J_1$  is the one that satisfies the equation  $\mathbf{S}\mathbf{e} = \lambda\mathbf{e}$ , for some scalar value  $\lambda$ .  $\mathbf{S}$  is the scatter matrix of the original data set as defined in Eq. (14-24).

$$J_1(\mathbf{e}) = \sum_{k=1}^n a_k^2 - 2 \sum_{k=1}^n a_k^2 + \sum_{k=1}^n \|\mathbf{x}_k - \mathbf{m}\|^2 \quad (14-23)$$

$$\mathbf{S} = \sum_{k=1}^n (\mathbf{x}_k - \mathbf{m})(\mathbf{x}_k - \mathbf{m})^t \quad (14-24)$$

Since  $\mathbf{e}$  must satisfy  $\mathbf{S}\mathbf{e} = \lambda\mathbf{e}$ , it is easy to realize that  $\mathbf{e}$  must be an eigenvector of the scatter matrix  $\mathbf{S}$ . In addition to  $\mathbf{e}$  being an eigenvector of  $\mathbf{S}$ , Duda et al.<sup>29</sup> also showed that the eigenvector which will yield the best representation of the original data set is the one that corresponds to the largest eigenvalue. By projecting the data onto the eigenvectors of the scatter matrix that correspond to the  $d'$  highest eigenvalues, the original data set can be projected down to a space of size  $d'$  dimensionality.

### Correlation Structure

One method of seizure analysis is to consider the correlation over all of the recorded EEG channels. In order to define the correlation matrix, segments of the EEG signal are considered for each window of specified time. The signal is then normalized for each channel within this window. Given  $z$  channels, the correlation matrix,  $\mathbf{C}$  is defined in Eq. (14-25), where  $w_i$  specifies the length of the given window ( $w$ ) and  $\text{EEG}_i$  is the  $i$ th channel.  $\text{EEG}_i$  has also been normalized to have zero mean and unit variance.<sup>30</sup>  $C_{ij}$  will yield a value of 0 when  $\text{EEG}_i$  and  $\text{EEG}_j$  are uncorrelated, a value of 1 when they are perfectly correlated, and a value of  $-1$  when they are anticorrelated. It should also be noted that the correlation matrix is symmetrical since  $C_{ij} = C_{ji}$ . In addition,  $C_{ii} = 1$  for all values of  $i$  since any signal will be perfectly correlated with itself. It follows that the trace of the matrix ( $\sum C_{ii}$ ) will always equal the number of channels ( $m$ ).

$$C_{ij} = \frac{1}{w_i} \sum_{t \in w} \text{EEG}_i(t) \cdot \text{EEG}_j(t) \quad (14-25)$$

To simplify the representation of the correlation matrix, the eigenvalues of the matrix are calculated. The eigenvalues tell which dimensions of the original matrix have the highest correlation. When the eigenvalues ( $\lambda_1, \lambda_2, \dots, \lambda_z$ ) are sorted so that  $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_{\max}$ , they can then be used to produce a spectrum of the correlation matrix  $C$ .<sup>34</sup> This spectrum is sorted by the intensity of correlation and used to track how the dynamics of all EEG channels are affected when a seizure occurs.

### Self-Organizing Map

The self-organizing map (SOM) is a machine learning–based technique to detect seizures. The SOM is a particular kind of an artificial neural network that uses unsupervised learning to classify data. It is simply provided with the data and the network learns on its own. One reported result transformed the EEG signal using a Fast Fourier Transform (FFT) and subsequently used the FFT vector as input to an SOM. With the help of some additional stipulations on the amplitudes and frequencies, the SOM was able to detect 90% of the seizures with an average of 0.71 false positives per hour.<sup>35</sup> This was utilized for seizure detection and not as an attempt at seizure prediction.

### Support Vector Machine

The support vector machine (SVM) is an advanced machine learning technique that has been used for seizure detection. SVM is a reinforcement learning technique, that is, it requires data that are labeled with the class information. An SVM is a classifier that partitions the feature space (or the kernel space in the case of a kernel SVM) into two classes using a hyper-plane. Each sample is represented as a point in the feature space and is assigned a class depending on which side of the hyper-plane it lies on. The classifier that is yielded by the SVM-learning algorithm is the optimal hyper-plane that minimizes the expected risk of misclassifying the unseen samples. Once noise and artifacts are removed, kernel SVMs have been applied to EEG with reasonable results: detection of 97% of the seizures. Of the seizures that were detected, the author reported that he was able to predict 40% of the ictal events by an average of 48 s before the onset of the seizure.<sup>36</sup> The need for artifact and noise removal is, however, a notable limitation.

### Conclusion

Epilepsy is a dynamic disease, with a wide variety of seizures and presentations. There is a rich set of electrographical records to analyze, assisted by advanced signal processing techniques. An array of univariate and multivariate methods are employed. While these methods have shown some success in detecting seizures, seizure prediction has been much more problematic. Though advanced techniques are used, the analysis of the EEG requires further intensive research.

### REFERENCES

1. Lopes da Silva FH. EEG analysis: Theory and practice; computer-assisted EEG diagnosis: Pattern recognition techniques. In: Niedermeyer E, Lopes da Silva FH, eds. *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Baltimore, MD: Williams & Wilkins, 1987:871–897.
2. Iasemidis LD. On the Dynamics of the Human Brain in Temporal Lobe Epilepsy. Ann Arbor: University of Michigan, 1991.
3. Le Van Quyen M, Martinerie J, Baulac M, et al. Anticipating epileptic seizure in real time by a nonlinear analysis of similarity between EEG recordings. *Neuroreport* 1999;10:2149–2155.
4. Mormann F, Kreuz T, Rieke C, et al. On the predictability of epileptic seizures. *Clin Neurophysiol* 2005;116(3):569–587.
5. Oppenheim AV, Wornell GW, Isabelle SH, et al. Signal processing in the context of chaotic signals. *IEEE Int Conf ASSP* 1992;4:117–120.

6. Blanco S. Applying time-frequency analysis to seizure EEG activity. A method to help to identify the source of epileptic seizures. *IEEE Eng Med Biol Mag* 1997;16:64–71.
7. Walnut DF. *An Introduction to Wavelet Analysis*. Boston–Basel–Berlin: Birkhauser, 2002.
8. Shannon CE. A mathematical theory of communication. *Bell Syst Tech J* 1948;27:379–423.
9. Osorio I. Real-time automated detection and quantitative analysis of seizures and short-term prediction of clinical onset. *Epilepsia* 1998;39(6):615–627.
10. Wilks SS. *Mathematical Statistics*. New York: Wiley, 1962.
11. Liu H. Epileptic seizure detection from ECoG using recurrence time statistics. *Proceedings of the 26th Annual International Conference of the IEEE EMBS* 2004, 29–32.
12. Iasemidis LD, Shiau DS, Chaivalitwongse W, et al. Adaptive epileptic seizure prediction system. *IEEE Trans Biomed Eng* 2003;50:616–627.
13. Iasemidis LD, Shiau DS, Sackellares JC, et al. Dynamical resetting of the human brain at epileptic seizures: Application of nonlinear dynamics and global optimization techniques. *IEEE Trans Biomed Eng* 2004; 51:493–506.
14. Iasemidis LD, Pardalos PM, Sackellares JC, et al. Quadratic binary programming and dynamical systems approach to determine the predictability of epileptic seizures. *J Comb Optim* 2001;5:9–26.
15. Iasemidis LD, Sackellares JC. The temporal evolution of the largest Lyapunov exponent on the human epileptic cortex. In: Duke DW, Pritchard WS, eds. *Measuring Chaos in the Human Brain*. Singapore: World Scientific, 1991:49–82.
16. Iasemidis LD, Sackellares JC. Long time scale temporo-spatial patterns of entrainment of preictal electrocorticographic data in human temporal lobe epilepsy. *Epilepsia* 1990;31(5):621.
17. Iasemidis LD. Time dependencies in the occurrences of epileptic seizures. *Epilepsy Res* 1994;17(1):81–94.
18. Pardalos PM. Seizure warning algorithm based on optimization and nonlinear dynamics. *Math Program* 2004; 101(2):365–385.
19. Sackellares JC. Epileptic seizures as neural resetting mechanisms. *Epilepsia* 1997;38(Suppl. 3):189.
20. Degan H, Holden A, Olsen LF. *Chaos in Biological Systems*. New York: Plenum, 1987.
21. Marcus M, Aller SM, Nicolis G. *From Chemical to Biological Organization*, Berlin: Springer-Verlag, 1988.
22. Sackellares JC, Iasemidis LD, Shiau DS, et al. Epilepsy—when chaos fails. In: Lehnertz K, Arnold J, Grassberger P, et al., eds. *Chaos in the Brain*. Singapore: World Scientific, 2000.
23. Takens F. *Detecting Strange Attractors in Turbulence of Dynamical Systems and Turbulence*. Berlin: Springer, 1981.
24. Abarbanel HDI. *Analysis of Observed Chaotic Data*. New York: Springer-Verlag, 1996.
25. Milton J, Jung P. *Epilepsy as a Dynamic Disease*. New York: Springer, 2003.
26. Rosenblum MG, Pikovsky AS, Kurths J. From phase to lag synchronization in coupled chaotic oscillators. *Phys Rev Lett* 1997;78(22):4193–4196.
27. Mormann F, Andrzejak RG, Kreuz T, et al. Automated detection of a pre-seizure state based on a decrease in synchronization in intracranial electroencephalogram recordings from epilepsy patients. *Phys Rev E* 2003;67(2):2003.
28. Quiroga RQ, Kreuz T, Grassberger P. Event synchronization: A simple and fast method to measure synchronicity and time delay patterns. *Phys Rev E* 2002;66(041904):2002.
29. Duda RO, Hart PE, Stork DG. *Pattern Classification*. New York: Wiley-Interscience, 1997:114–117.
30. Schindler K, Leung H, Elger CE, et al. Assessing seizure dynamics by analysing the correlation structure of multichannel intracranial EEG. *Brain* 2007;130(1):65.
31. Gabor AJ. Automated seizure detection using a self-organizing neural network. *Electroencephalogr clin Neurophysiol* 1996;(99):257–266.
32. Gardner AB. *A Novelty Detection Approach to Seizure Analysis from Intracranial EEG*. Atlanta: Georgia Institute of Technology, 2004.
33. Tass P. *Phase resetting in medicine and biology: Stochastic modelling and data analysis* Berlin: Springer Verlag, 1999.

## Queires

[Q1] Abarbanel has been listed as no. 24 in the reference list, but it has been cross-referred with no. 25 in the text. Please check the reference numbers from 24 to 33 in the list and in the text. Please check the reference Quiroga et al. also with the text and the list.