7

EEG Feature Extraction

7.1. Introduction

One of the major steps in the design of a BCI that uses EEG signals is the processing and classification of these signals to identify the user's mental state. As previously seen, EEG signal analysis is divided into three stages: preprocessing (discussed in Chapter 6), feature extraction and classification. This chapter focuses on feature extraction, which consists of describing EEG signals by (an ideally small) core set of values describing the relevant information they contain in order to later classify them. In particular, we will see what type of information to extract from EEG signals for different types of BCIs, and how to extract this information in order for it to best discriminate between different mental states.

7.2. Feature extraction

A feature is a value that describes a property of EEG signals, for example the power of the EEG signal in the μ rhythm for the C3 electrode. The characteristics are usually grouped together in a vector called a "feature vector". As an example, let us look at a BCI using motor imagery (MI) – that is to say that it can recognize imagined movements, for example of the hands. In this case, the two mental states to be identified are imagined movements of the left hand, and imagined movements of the right hand. The feature that is

Chapter written by Fabien LOTTE and Marco CONGEDO.

Brain–Computer Interfaces 1: Foundations and Methods, First Edition.
Edited by Maureen Clerc, Laurent Bougrain and Fabien Lotte.

usually extracted in order to identify these mental states in the EEG signals is band power, i.e. the EEG signal's strength in a specific frequency band. For MI, band powers are usually calculated in μ (about 8–12 Hz) and β (about 16–24 Hz) frequency bands for electrodes located over the sensorimotor cortex (e.g. C3 and C4 electrodes for imagined hand movements) [PFU 01]. Such features can typically be classified using linear discriminant analysis (LDA, see Chapter 8).

In BCI design, EEG signal processing often relies on machine learning techniques. This means that the classifier and/or features are usually adjusted and optimized for each user, using examples of EEG signals produced by each user. These examples of EEG signals are called a training set and are labeled with their membership class – that is to say the user's mental state when the EEG signals were recorded. With this training set, it becomes possible to calibrate a classifier capable of recognizing the class of different EEG signals, as will be described in Chapter 8. Features can also be optimized through examples of EEG signals, for example by selecting the most relevant electrodes to recognize different mental states. Thus, designing a BCI that employs machine learning (which is the case of most BCIs) requires (1) a calibration phase (also called a training phase) that consists of acquiring training EEG signals (i.e. examples) and optimizing the EEG signal processing chain by adjusting the feature settings and/or by adjusting a classifier and (2) a usage phase (also called a test phase), which consists of using the model (features and classifier) obtained during calibration in order to recognize the user's mental state based on new EEG signals (i.e. EEG signals other than those in the training set) to operate the BCI.

As briefly mentioned here, and as we shall see in detail in Chapter 8, a classifier is able to learn what class corresponds to which input features by examining examples. So why not use the values of EEG signals directly as an input to a classifier? The answer is that would most likely not work because of a phenomenon called the "curse of dimensionality": it has been observed in practice that the amount of examples needed to properly describe different classes increases exponentially with the dimension of the feature vector (i.e. the number of features used) [RAU 91]. Some researchers even recommend using 5–10 times more training examples than the size of the feature vector[1] [RAU 91]. What would this recommendation mean if we used the value of EEG signals as features directly? For example, suppose we were using a common EEG system with 32 electrodes sampled at 250 Hz, and an EEG signal example for a given mental task lasts 1 s. We would therefore have a feature vector with a size of $32 \times 250 \times 1 = 8,000$, which would require at least $8,000 \times 5 = 40,000$ training examples by class. Obviously, we cannot ask a BCI user to perform each mental task $40,000$ times in order to calibrate the BCI before he/she can use it. So we need a more compact representation, and thus extract features of EEG signals.

In BCI design, there are three main sources of information that can be used to extract EEG signal features:

– *spatial information*: this describes where (spatially) the relevant signal comes from. In practice, this means selecting specific EEG electrodes or focusing more on some than on other sensors. In other words, this is equivalent to using the signal from one or several specific brain regions, but not (or very little) from any others;

– *spectral (or frequency) information*: this describes how the power of the EEG signal varies in some specific frequency bands. In practice, this is equivalent to using signal band power as features.

– *temporal information*: this describes how EEG signals vary over time. In practice, this means using the values of EEG signals for different specific time intervals (though not all) or different time windows.

It is usually necessary to use different sources of information for different types of BCIs. In particular, BCIs using oscillatory EEG activity (e.g. BCIs based on MI) primarily use spatial and spectral information, while BCIs employing evoked potentials (EP) mainly use spatial and temporal information. The following sections describe feature extraction techniques used for these two types of BCIs.

1 This recommendation was made before the invention of SVM (described in Chapter 8), which are less sensitive to the curse of dimensionality.

7.3. Feature extraction for BCIs employing oscillatory activity

BCIs employing oscillatory activity use mental states that produce amplitude changes in EEG oscillations, that is to say changes in the power of EEG signals in certain frequency bands (ERD/ERS, see Chapter 4). Such BCIs include most notably BCIs employing MI [PFU 01], SSVEP [VIA 10], various mental imagery tasks such as mental arithmetic or mental generation of words [FRI 12] and even different levels of mental workload [MÜH 14].

This section first presents a basic (and overly simplistic) description of that kind of BCI, and then focuses on some more advanced tools including the major algorithm called "common spatial patterns" (CSP).

7.3.1. *Basic design for BCI using oscillatory activity*

BCIs that employ oscillatory activity use the power of the EEG signals in certain frequency bands (spectral information) and certain brain regions (spatial information).

For example, the basic design of an MI-based BCI would use spatial information by extracting features only from EEG sensors located above the brain motor areas, typically C3 sensors for imagined movement of the right hand, Cz for imagined movement of the feet and C4 for imagined movement of the left hand. It would use the spectral information focusing on frequency bands μ (8–12 Hz) and β (16–24 Hz). Specifically, for a BCI that can recognize imagined movements of the left and right hands, basic features would be band power in 8–12 Hz and 16–24 Hz for each of electrodes C3 and C4, which thus constitutes four features. There are many ways to calculate EEG signal band power [BRO 11]. However, a simple but effective technique is to first filter the EEG signal of a given sensor in the target frequency band, and then to calculate the square of the filtered signal to obtain its power, finally averaging the signal over time (e.g. during the last second, using sliding windows). This principle is shown in Figure 7.1.

Unfortunately, this basic design is far from optimal. Indeed, it uses only two electrodes: relevant information measured by other sensors may be missing. In addition, C3 and C4 are perhaps not the best sensors for the subject being studied. Similarly, the fixed frequency bands at 8–12 Hz and 16–24 Hz are perhaps not the best bands for the subject at hand. In general, better performance can be obtained using a specific design for each subject by optimizing the best electrodes and frequency bands. Using more than two sensors also makes it possible to obtain better performance.

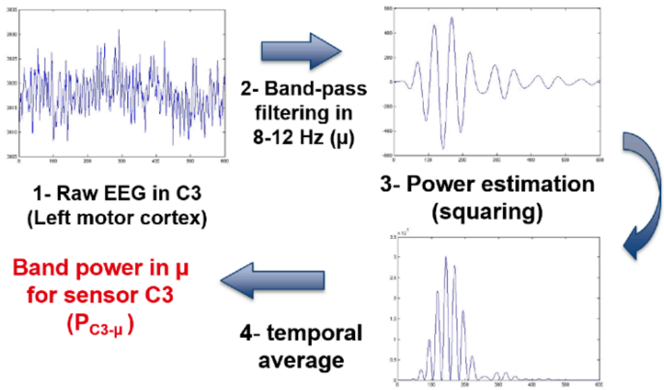

Figure 7.1. *Band power feature extraction from a raw EEG signal. The EEG signal illustrated here was recorded during an instance of imagined right hand movement (beginning at $t = 0$ s). The contralateral ERD during the imagined movement is clearly visible: the signal power of the C3 electrode (left motor cortex) in 8–12 Hz band clearly decreases during imagined movement*

7.3.2. *Toward more advanced, multiple electrode BCIs*

The need to use more than two sensors as well as specific sensors in each subject has lead to the design of BCIs that use multiple electrodes. This need has been confirmed by various studies that suggest that, for MI, maximum performance is achieved with a large number of electrodes, for example with 48 sensors according to [SAN 10]. However, simply adding sensors will not solve performance issues. In fact, using more sensors means extracting more features, which makes the curse of dimensionality even more likely. Therefore, just adding sensors can even reduce performance if the number of available training examples is too low. In order to efficiently employ multiple sensors, there are three main approaches, each contributing to reducing the dimension of the feature vector:

– *feature selection algorithms*: these are methods for automatically selecting a relevant subset of features among the features initially extracted [GUY 03];

– *sensor selection algorithms*: these are methods similar to those employed in feature selection and are aimed at automatically selecting a relevant subset of sensors from among all available sensors;

– *spatial filtering algorithms*: these are methods for combining several sensors, generally through linear combination, in order to form a new (virtual) sensor from the extracted features.

Later, we will focus on spatial filtering, for which algorithms specific to EEG and BCIs have been developed. Feature selection is indeed a set of general tools in machine learning, which is not specific to EEG or BCI (see [GUY 03] for details). As for sensor selection, the algorithms used are usually derived from the feature selection algorithms. Readers interested in the topic can refer, for example, to [SCH 05] or [ARV 11] and their references to learn more about the subject.

7.3.2.1. *Spatial filtering*

Spatial filtering consists of using a small number of new channels (virtual channels) that are defined as a linear combination of the original sensors. Formally, spatial filtering is described by $\tilde{x} = \sum_i w_i x_i = wX$, where $\tilde{x}$ is the spatially filtered signal, x_i is the EEG signal from sensor i, w_i is the weight given to this sensor in the spatial filtering and X is a matrix whose $i\mathrm{th}$ row is x_i, i.e. X is the EEG signal matrix for each sensor. Spatial filtering is useful not only because it reduces the size of the problem (passing from many initial EEG sensors to a small number of spatially filtered signals – far fewer spatial filters than the number of original sensors are typically used) but also because it has a neurophysiological meaning. Indeed, as discussed above (Chapters 2 and 3), EEG signals measured on the surface of the scalp are the result of noisy mixtures of EEG signals from different brain regions. In other words, since the EEG signal from the cortex is diffused as it passes through the skull or scalp, when it arrives at the EEG sensors, said signal is diffused and dispersed on several EEG sensors. Therefore, spatial filtering makes it possible to help recover the original signal (from the cortex) by gathering relevant information that was scattered over different sensors.

There are different ways to define the spatial filters. In particular, the weight of a filter w_i can be fixed beforehand, according to neurophysiological knowledge, or it can be optimized using training examples. Among the fixed spatial filters, we can mention bipolar and Laplacian filters in particular, which have already been described in the previous chapter. It has been shown that extracting features from bipolar or laplacian channels rather than from the original EEG sensors significantly increases classification performance [MCF 97]. Methods for reconstruction of distributed sources (also presented in the previous chapter) can also be used to define fixed spatial filters in order to analyze the EEG signal from very specific brain regions. Extracting features from spatial filter obtained by source reconstruction also allows for better classification performance than when extracting features from the original EEG sensors [CON 06].

The second category of spatial filters, which are based on data, contains filters that are optimized for each subject on the training data. This category contains in particular the spatial filters constructed through independent component analysis (ICA) [KAC 08], as described in the previous chapter. These techniques make it possible to obtain the weight w_i of spatial filters in an unsupervised manner, that is to say without knowing the labels (classes) of the training data. Alternatively, the weight of the spatial filters can be defined in a supervised manner (i.e. knowing the label of each training example) in order to optimize a measure of separation between classes. One such algorithm has been developed for BCI based on EEG oscillatory activity: the CSP algorithm [RAM 00], which is described below.

7.3.3. *The CSP algorithm*

Informally, the CSP algorithm optimizes spatial filters w such that the variance of the filtered EEG signal is maximum for one class and minimal for another class. Since the variance of a filtered signal in the frequency band b is equal to the power of the signal in the band b, this means that CSP optimizes the spatial filters to obtain the band power features that are optimally discriminant because their value is maximally different between the two classes. CSP is therefore especially useful for BCIs based on oscillatory activity, since the most useful type of features for their design is precisely band power. For example, for BCIs employing MI, EEG signals are typically filtered in the 8–30 Hz band (μ and β rhythms) before being spatially filtered by CSP [RAM 00]. Formally, CSP optimizes the spatial filters w by extremizing (i.e. minimizing and maximizing) the following function:

$$J_{CSP1}(w) = \frac{wX_1X_1^Tw^T}{wX_2X_2^Tw^T} \qquad [7.1]$$

which is equivalent to extremizing

$$J_{CSP2}(w) = \frac{wC_1w^T}{wC_2w^T} \qquad [7.2]$$

where T is the matrix transpose, X_i are EEG training signals for class i, which were previously bandpass filtered (matrix with EEG samples as columns and sensors as rows) and C_i the spatial covariance matrix for class i. In practice, the covariance matrix C_i is defined as the average of the covariance matrices of each example of the class i [RAM 00]. In this equation, wX_i is the spatially filtered EEG signal for class i, and $wX_iX_i^Tw^T$ is therefore the variance of the spatially filtered signal, that is to say its band power. So maximizing and minimizing $J_{CSP}(w)$ makes it possible to obtain spatially filtered signals whose band power is maximally different between classes. $J_{CSP}(w)$ is in the form of what is mathematically called a generalized Rayleigh quotient. Therefore, maximizing and minimizing this function can be solved by Generalized Eigen Value Decomposition (GEVD). The spatial filters w that maximize or minimize $J_{CSP}(w)$ are therefore the eigenvectors corresponding to the largest and smallest values of the GEVD of matrices C_1 and C_2. Typically, six CSP filters are used (i.e. three pairs of filters), which correspond to the three largest and three smallest eigenvalues. Once the filters have been obtained, a CSP feature f is calculated as follows:

$$f = \log(wCw^T) \approx \log(var(wX)) \qquad [7.3]$$

i.e. a CSP feature is simply the band power of the signal spatially filtered with one of the CSP filters w. The use of CSP is shown in Figure 7.2. In this figure, a marked difference in variance (and hence band power) can be observed between signals spatially filtered with CSP for each of the two classes, which provides good classification performance.

The CSP algorithm has many advantages: first, it provides relatively high classification performance for BCIs. This is a fairly flexible algorithm, since it can be used for any BCI using the ERD/ERS. Finally, it is numerically efficient and a simple algorithm to implement. All of this means that CSP has become one of the most popular and efficient algorithms for designing BCIs based on oscillatory activity [LOT 11]. Their performance is illustrated in the following section.

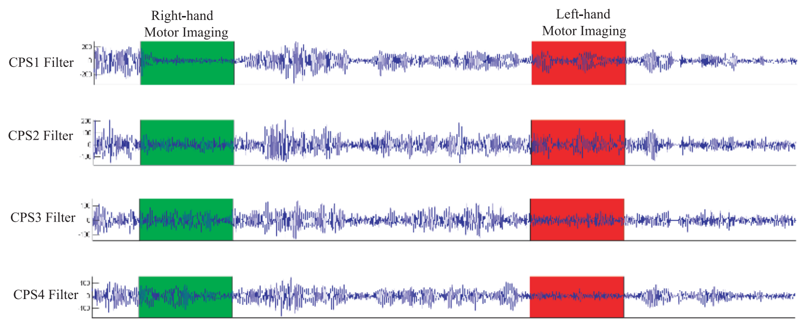

Figure 7.2. *EEG signals spatially filtered with the common spatial patterns (CSPs) algorithm. The first two spatial filters (both signals from the top) are those that maximize the variance of the signals in the "imagined movement of the left hand" class (in red) while minimizing those of the "imagined movement of the right hand" class (in green). The last two filters (the two signals on the bottom) do the opposite, i.e. they maximize the variance of the "imagined movement of the right hand" class, while minimizing the variance of the "imagined movement of the left hand" class. For a color version of this figure, see www.iste.co.uk/clerc/interfaces1.zip*

7.3.4. *Illustration on real data*

In order to illustrate the impact of spatial filters on BCI classification performance, we compared the performances obtained by different filters on the dataset IIA from "BCI competition IV" [TAN 12]. During this competition, a training set and a test set were made available to competitors. Each set contains 72 examples of EEG signals for each class (imagining a movement of the left hand, right hand, tongue and feet) for nine different subjects. The training set examples were labeled with their class, while the examples for the test set were not labeled: competitors had to calibrate their algorithms (e.g. the classifier) on the training set, and use it to guess the class of test examples. The aim of the competition was to identify the best algorithms for recognizing the different mental tasks. Data from this competition is now available for testing, evaluating and comparing different methods.

We compared four different BCI designs offline, each based on one of four different spatial filters and aimed at distinguishing imagined movements of the right and left hands. The four filters used were (1) no filter, just EEG signals from electrodes C3 and C4; (2) bipolar filter around C3 and C4; (3) Laplacian filter around C3 and C4 and (4) CSP filters (three pairs of filters). In order to classify the signals, we first performed a spectral filtering in the 8–30 Hz band, and then filtered spatially with one of the four filters. We then calculated the signal's average band power over a time window of 2 s starting 0.5 s after the beginning of the imagined task. These band powers for each filter are then given to a classifier (LDA) to identify the task performed. CSP and LDA are calibrated on the training set, and each method is then tested on the test set. Figure 7.3 shows the classification performance obtained on the test set by each method.

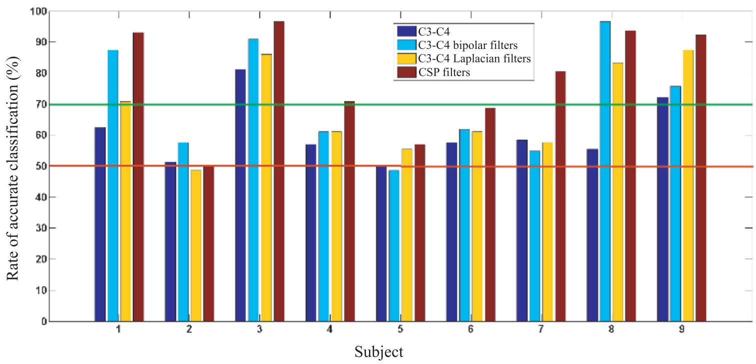

Figure 7.3. *Classification performance (rate of accurate classification) obtained on data set IIa for "BCI competition IV" [TAN 12] in classifying imagined movement of the right hand using different spatial filters. For a color version of this figure, see www.iste.co.uk/clerc/interfaces1.zip*

As can be seen in this figure, using a fixed spatial filter, such as bipolar or Laplacian filters, makes it possible to increase performance (average performance for bipolar filters: 70.5%, for Laplacian filters 68%) as compared to using only C3 and C4 (average performance 60.7%). Using a data-based filter such as CSP will increase classification performance even more (average performance of 78.1%). It should, nevertheless, be noted that CSP is not a perfect algorithm, far from it. It is especially sensitive to the presence of noise and artifacts, and does not work well with little training data. Many variations of CSP have been proposed to remedy this, and research on the subject is still very active [LOT 11, SAM 14].

7.4. Feature extraction for the BCIs employing EPs

A typical example of an EP used in BCIs is P300 [FAZ 12], as described in Chapter 4. EPs are characterized by specific temporal variations appearing in response to a stimulus. Thus, unlike BCIs using brain activity, BCIs using the EP mainly employ temporal information rather than spectral information. Nevertheless, such as BCI employing oscillatory activity, those exploiting EP can also exploit spatial information.

For example, in the case of a BCI using P300, spatial information is used that focuses mainly on parietal and occipital electrodes (i.e. by extracting features from these electrodes only), where P300 comes from. Krusienski *et al.* recommend, for example, using a set of eight sensors, located in Fz, Cz, P3, Pz, P4, PO7, Oz and PO8 [KRU 06]. Once the relevant spatial information has been identified – i.e. the electrodes (as in the example above) – features can be extracted from each of their signals. For EPs in general and therefore also for P300, the features used usually reflect the signals' temporal information, that is to say how the signal amplitude varies with time. This is done by using the value of the different points of EEG signals preprocessed as features. More specifically, features are usually extracted from an EP (1) by performing a low-pass or bandpass filtering of the signals (e.g.in 1–12 Hz for P300, given that EPs are usually slow waves), (2) by subsampling the filtered signal to reduce the number of EEG signal points and therefore the dimension of the problem, and (3) by collecting the values of the remaining EEG points for all selected sensors into a single feature vector which will be used as the input of a classifier. This process is illustrated in Figure 7.4 in the extraction of features from the Pz electrode for a BCI using P300.

Extracting several time points from several sensors as features makes BCIs employing EP usually have a higher dimension than those using oscillatory activity. Therefore, it is important to use classifiers that can handle high dimension (see the following section), or to use, as mentioned above, feature or sensor selection algorithms, which are the same for the two types of BCIs. Spatial filters devoted to EP have also been proposed.

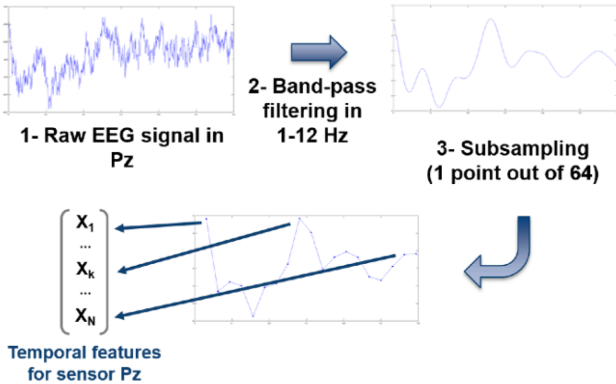

Figure 7.4. *Traditional process for extracting features from an EEG sensor for a BCI using EP (here P300). The stimulus that can evoke the EP appears at t = 0 s*

7.4.1. *Spatial filtering for BCIs employing EPs*

As with BCIs employing oscillatory activity, BCIs using the EP can also benefit from spatial filtering to identify brain sources whose features are discriminatory. So why not also use the CSP for EPs? This is due to the fact that one crucial bit of information for classifying EPs is information on EEG signals' time course. Unfortunately, CSP completely ignores this information because it only considers average signal power (and therefore not its time course) in optimizing filters. Thus, CSP is not suitable for EP classification. Fortunately, there are other spatial filtering algorithms specifically for EPs. We can most prominently mention Fisher spatial filtering, which is devoted to EPs, and was proposed by Hoffman *et al.* [HOF 06], or xDAWN, proposed by Rivet *et al.* [RIV 09]. The objective of these two spatial filters is to obtain spatially filtered signals such that the EPs are more visible and more discriminable than the original EEG signals. These two methods use different objective functions to achieve this goal. We will describe xDAWN filtering to illustrate spatial filtering of EPs for BCI. The xDAWN spatial filtering algorithm, which has proved very effective for EP classification, seeks to maximize the signal to noise ratio. Informally, this means that xDAWN seeks to highlight the EP in order to make it more visible amidst the noise.

Formally, xDAWN optimizes spatial filters by maximizing the following function:

$$J_{xDAWN} = \frac{wSS^T w^T}{wXX^T w^T} \qquad [7.4]$$

where S is the estimated average time course of the EP (averaged over the number of repetitions). The average EP can be estimated more accurately using least squares, where the EPs temporally overlap (which is the case with "P300-Speller") [RIV 09]. In this equation, the numerator represents the signal, that is to say the relevant information that it seeks to highlight. Indeed, $wSS^T w^T$ is the power of the time course of the EP after spatial filtering. In contrast, in the denominator $wXX^T w^T$ is the power of all EEG signals after spatial filtering. The denominator thus contains both the signal (the target EP) and noise. Therefore, maximizing J_{xDAWN} requires simultaneously maximizing the signal – that is to say highlighting the EP – and minimizing the signal plus the noise – that is to say making noise as small as possible [RIV 09]. This approach makes it possible to significantly improve EP classification performance, especially when few training examples are available.

7.5. Alternative methods and the Riemannian geometry approach

This chapter has presented the main tools for recognizing a user's mental state based on the EEG oscillations (spectral information) and EPs (temporal information). These tools are relatively simple, widely used and effective. These are not, however, the only tools available. We can mention, for example, other methods for extracting information, such as Hjorth temporal parameters [OBE 01] or "time domain parameters" [VID 09], methods for measuring the complexity of signals [BRO 12] or methods measuring connectivity information – that is to say measuring how the various sensor signals are connected (e.g. synchronized or correlated) [CAR 14].

Recent work has explored a new approach to feature extraction for BCIs [BAR 12, CON 13]. The idea is to summarize the relevant information, whether spatial, frequential and/or temporal, for a particular point in a multidimensional space. This space is the Riemannian manifold of symmetric positive definite (SPD) matrices.

An SPD matrix is a square and symmetric matrix for which all eigenvalues are positive. For example, the spatial covariance matrices in EEG are SPD (Figure 7.5, on the left). The Riemannian manifold of SPD matrices is their native and curved space (see Figure 7.5, on the right). On this manifold, a suitable metric gives us a way to calculate the distance between any two points (Figure 7.5, on the right).

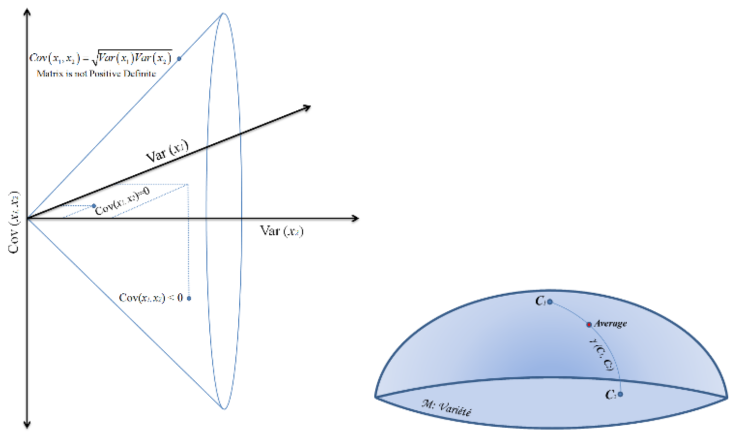

Figure 7.5. *Left: Covariance matrices with dimension $N \times N$ are constrained by their symmetry and the positive sign of their diagonal elements (variance), among others. We can easily visualize the topology in the case of 2 $\times$ 2 matrices; the cone on the left represents each of these matrices as a point in Euclidean space in three dimensions, where two coordinates correspond to the two variances (the two diagonal elements) and the third corresponds to the covariance (off-diagonal elements, which are equal since the matrix is symmetrical). By construction, the point must remain within the cone. As soon as that point touches the edge of the cone, the matrix is no longer SPD. Right: The geodesic going through two points C1 and C2 on the Riemannian manifold of SPD matrices is the minimum path length between them. The geometric mean of these two points is the halfway point on the geodesic, which is usually far from the arithmetic mean $\frac{1}{2}(C1 + C2)$*

With this notion of distance, we can estimate the geometric mean of a cloud of points (each point being an SPD matrix representing the EEG signal) corresponding to the different classes of a BCI obtained in the calibration phase. We can thus classify new signals simply by evaluating their distance from the geometric mean of each class (Figure 7.6). This approach proved to be as simple as it is effective, providing classification results as good as those obtained by the state-of-the art for MI and P300, while allowing a better generalization between sessions and subjects [BAR 12, CON 13]. Moreover, this strength has also helped to design BCIs that do not require calibration. For this, a database is used to initialize the BCI and the BCI then adapts to the subject during use. Moreover, with this approach, the processing chain is identical for all types of BCIs: only the definition of the point that summarizes EEG signals on the manifold changes depending on the type of BCI. For example, for BCIs based on MI spatial information is used, and for those based on the EP, temporal information is used [CON 13].

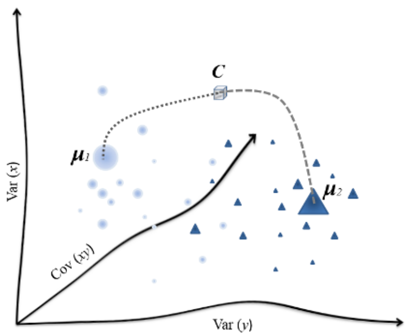

Figure 7.6. *Consider a two-class BCI with 2 × 2 matrices. With the calibration data, we estimate the geometric mean for Class 1 (μ_1) and Class 2 (μ_2). In the testing phase, a test (C) represented by an SPD matrix is classified by evaluating its distance to each geometric mean and labeled as belonging to the nearest class. The same process is used for arrays of any size and for any number of classes*

7.6. Conclusions

In this chapter, we tried to show the reader the kinds of relevant information that can be extracted from EEG signals in order to later be classified, as well as how to extract this information. In particular, we saw that the three main sources of information are (1) spectral information that is used with band power features, (2) temporal information, typically represented as the amplitude of preprocessed EEG signals in short time and (3) spatial information, which can be exploited by focusing on certain sensors or using spatial filters (CSP for BCI using oscillatory activity, or xDAWN for those operating EPs). It is important to note that there is still much work on BCI feature extraction, and the ideal method, which is both robust, invariant (over time) and universal (works well for all users), still remains to be found.

7.7. Bibliography

[ARV 11] ARVANEH M., GUAN C., ANG K. *et al.*, "Optimizing the channel selection and classification accuracy in EEG-based BCI", *IEEE Transactions on Biomedical Engineering*, vol. 58, p. 1865–1873, 2011.

[BAR 12] BARACHANT A., BONNET S., CONGEDO M. *et al.*, "Multiclass brain–computer interface classification by Riemannian geometry", *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 4, pp. 920–928, 2012.

[BRO 11] BRODU N., LOTTE F., LÉCUYER A., "Comparative study of band-power extraction techniques for motor imagery classification", *Proceedings of IEEE CCMB*, pp. 1–6, 2011.

[BRO 12] BRODU N., LOTTE F., LÉCUYER A., "Exploring two novel features for EEG-based brain-computer interfaces: multifractal cumulants and predictive complexity", *Neurocomputing*, vol. 79, no. 1, pp. 87–94, 2012.

[CAR 14] CARAMIA N., LOTTE F., RAMAT S., "Optimizing spatial filter pairs for EEG classification based on phase synchronization", *Proceedings of ICASSP*, 2014.

[CON 06] CONGEDO M., LOTTE F., LÉCUYER A., "Classification of movement intention by spatially filtered electromagnetic inverse solutions", *Physics in Medicine and Biology*, vol. 51, no. 8, pp. 1971–1989, 2006.

[CON 13] CONGEDO M., EEG source analysis, PhD thesis, HDR presented at Doctoral School EDISCE, University of Grenoble, 2013.

[FAZ 12] FAZEL-REZAI R., ALLISON B., GUGER C. *et al.*, "P300 brain computer interface: current challenges and emerging trends", *Frontiers in Neuroengineering*, vol. 5, no. 14, 2012.

[FRI 12] FRIEDRICH E., SCHERER R., NEUPER C., "The effect of distinct mental strategies on classification performance for brain-computer interfaces", *International Journal of Psychophys*iology, vol. 84, no. 1, pp. 86–94, 2012.

[GUY 03] GUYON I., ELISSEEFF A., "An introduction to variable and feature selection", *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.

[HOF 06] HOFFMANN U., VESIN J., EBRAHIMI T., "Spatial filters for the classification of event-related potentials", *Proceedings of ESANN*, 2006.

[KAC 08] KACHENOURA A., ALBERA L., SENHADJI L. *et al.*, "ICA: a potential tool for BCI systems", *IEEE Signal Processing Magazine*, vol. 25, no. 1, pp. 57–68, 2008.

[KRU 06] KRUSIENSKI D., SELLERS E., CABESTAING F. *et al.*, "A comparison of classification techniques for the P300 Speller", *Journal of Neural Engineering*, vol. 3, pp. 299–305, 2006.

[LOT 11] LOTTE F., GUAN C., "Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms", *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 2, pp. 355–362, 2011.

[MCF 97] McFarland D.J., McCane L.M., David S.V. *et al.*, "Spatial filter selection for EEG-based communication", *Electroencephalographic Clinical Neurophysiology*, vol. 103, no. 3, pp. 386–394, 1997.

[MÜH 14] Mühl C., Jeunet C., Lotte F., "EEG-based workload estimation across affective contexts", *Frontiers in Neuroscience*, vol. 8, p. 114, 2014.

[OBE 01] Obermeier B., Guger C., Neuper C. *et al.*, "Hidden Markov models for online classification of single trial EEG", *Patt Rec Letters*, vol. 22, no. 12, pp. 1299–1309, 2001.

[PFU 01] Pfurtscheller G., Neuper C., "Motor imagery and direct brain–computer communication", *Proceedings of the IEEE*, vol. 89, no. 7, pp. 1123–1134, 2001.

[RAM 00] Ramoser H., Muller-Gerking J., Pfurtscheller G., "Optimal spatial filtering of single trial EEG during imagined hand movement", *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 4, pp. 441–446, 2000.

[RAU 91] Raudys S.J., Jain A.K., "Small sample size effects in statistical pattern recognition: recommendations for practitioners", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 3, pp. 252–264, 1991.

[RIV 09] Rivet B., Souloumiac A., Attina V. *et al.*, "xDAWN algorithm to enhance evoked potentials: application to brain computer interface", *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 8, pp. 2035–2043, 2009.

[SAM 14] Samek W., Kawanabe M., Muller K., "Divergence-based framework for common spatial patterns algorithms", *IEEE Review in Biomedical Eng*ineering, vol 7, pp. 50–72, 2014.

[SAN 10] Sannelli C., Dickhaus T., Halder S. *et al.*, "On optimal channel configurations for SMR-based brain–computer interfaces", *Brain Topography*, 2010.

[SCH 05] Schröder M., Lal T., Hinterberger T. *et al.*, "Robust EEG channel selection across subjects for brain–computer interfaces", *EURASIP Journal of Applied Signal Process*ing, pp. 3103–3112, 2005.

[TAN 12] Tangermann M., Müller K., Aertsen A. *et al.*, "Review of the BCI competition IV", *Frontiers in Neuros*cience, 2012.

[VIA 10] Vialatte F., Maurice M., Dauwels J. *et al.*, "Steady-state visually evoked potentials: focus on essential paradigms and future perspectives", *Progress in Neurobiology*, vol. 90, pp. 418–438, 2010.

[VID 09] Vidaurre C., Krämer N., Blankertz B. *et al.*, "Time domain parameters as a feature for EEG-based brain computer interfaces", *Neural Networks*, vol. 22, pp. 1313–1319, 2009.