SPARSE ANOMALY REPRESENTATIONS IN VERY HIGH-DIMENSIONAL BRAIN SIGNALS

Catherine Stamoulis, Member, IEEE

Harvard Medical School, Boston, MA 02115, USA

ABSTRACT

Across scales of organization, brain activity is inherently sparse. This is also the case for transiently occurring signal abnormalities associated with neurological disorders, such as epilepsy. Consequently, for the purpose of characterizing these abnormalities, very high-dimensional brain signals may be represented as sparse combinations of the elements of a comprehensive (overcomplete) dictionary. Such a dictionary may be estimated (learned) from the dataset(s) of interest. However, given the statistical, spectral and signature heterogeneity of brain signals recorded over long periods of times, the size of the dictionary may be suboptimal, particularly in terms of its size. In this paper, signal-specific, dataset-specific and individual-specific anomaly dictionaries, estimated via the K-SVD algorithm from noninvasive high-frequency brain signals collected continuously over several days are explored. It is shown that signal-specific dictionaries may yield substantially more accurate representations than those estimated by combining training signals from multiple electrodes.

***Index Terms*—** Sparse dictionaries, K-SVD, brain signals, waveform anomalies

1. INTRODUCTION

There is an ongoing shift in Neuroscience from experimental paradigms in tightly controlled laboratory conditions to more 'unsupervised' paradigms in naturalistic and uncertain settings (e.g, in natural habitats), which aim to study complex multi-domain behaviors and multimodal sensory processing. Novel experiments are thus designed to measure brain activity from freely behaving animals and/or humans over extended periods of time. These hold great promise to provide a wealth of new knowledge on how the brains neural circuitry supports complex behaviors across domains and how it processes multimodal inputs from the outside world. They are also bound to generate very high-dimensional datasets. Such data are also routinely collected during clinical neurophysiological studies, including sleep studies spanning days or weeks and long-term monitoring studies for diagnostic purposes, e.g., in epilepsy patients. They contain a wealth of information on brain dynamics across temporal scales but also electrophysiological markers of underlying neuropathologies.

Comprehensive analyses of very high-dimensional brain signals collected continuously over long periods of time (days or weeks) from multiple electrodes (sometimes over 200 electrodes, particularly in invasive studies) are computationally prohibitive. A typical dataset of invasively recorded brain signals spanning several days may contain $O(10^{10})$ time points with the temporal dimension being the highest (in contrast to high-resolution imaging data where the spatial dimension is highest). However, there is increasing evidence that across scales of organization, from the microscale of individual neurons to that of large ensembles of thousands of cells measured with macroscale electrodes, brain activity is sparse. Furthermore, signal abnormalities, including the recently discovered pathological high-frequency oscillations in invasive and noninvasive EEG recordings are transient and sparse [1, 2, 3, 7, 4, 5, 6]. This property could, therefore, be exploited to substantially reduce the dimension of these signals and derive parsimonious representations of their structure. This would also facilitate the classification of these abnormalities or signal patterns and efficient analysis of their characteristics.

Interpretable sparse representations of high-dimensional brain signals may be difficult to estimate. First, the statistical properties of these signals may vary substantially with time, potentially requiring dynamic estimation or updating of these representations. There is growing evidence that functional activations of neuronal ensembles may have stereotypical, modular and sparse patterns ([9, 10] and references therein). However, it is currently unclear whether these patterns repeat over long periods of time and how they vary between neuron types, brain regions and individuals (or animals). Furthermore, the potential waveform variability of many electrophysiological abnormalities are also unknown. To design robust and computationally efficient detectors of physiological and/or pathological neural activity, whether for characterizing functional neural activations in response to cognitive demands, diagnostic purposes, as part of brain-computer interfaces (BCI) or for next-generation targeted therapies, it is desirable to estimate sparse neural signal representations from a set of fundamental elements (neural primitives, modules or atoms).

Among a large number of dimensionality reduction methods for compressed sensing and sparse signal/image representation, various dictionary-based approaches have been proposed [11, 12, 13, 14, 15, 16]. All highlight two important is-

This work is supported by the National Science Foundation, Grant ACI #1649865

sues: computational cost and the dimension of the dictionary. In the case of pre-defined dictionaries ([17, 18, 19, 20] among many others), the set of atoms based on which sparse signal representations are estimated is a priori known or assumed. These dictionaries are typically overcomplete and thus sufficiently large to approximate a set of signals fairly well but are not data-derived and may be thus be non-optimal. Dictionaries that can be learned from the data have received considerable attention (e.g., [21, 14, 15] and [22, 23] for method reviews), with more recent work focusing on the estimation of optimally sparse dictionaries [24]. However, similarly to all learning-based approaches, the size and heterogeneity of the data used to train the algorithms and derive the dictionary may vary substantially between applications. In the case of high-dimensional brain signals, there are very limited studies on dictionary-based sparse representations [25, 26] (mostly for BCI applications) and, to the best of our knowledge, no studies on dictionaries for sparse brain signal anomalies.

This paper investigated dictionaries of high-frequency (>100 Hz) anomalies in very high-dimensional electrophysiological signals for sparse representations of the latter. These anomalies are thought to be intrinsically sparse in time and possibly in space (they occur transiently and in a subset of recording electrodes) and may be sensitive and specific biomarkers of abnormal neurodynamic processes, e.g., seizures. The K-SVD method [15], and for comparison the Method of Optimal Direction (MOD) [12], were used to learn these dictionaries. The paper addresses primarily the heterogeneity of estimated dictionaries, when the training data are selected from individual electrodes, all electrodes, a single patient or multiple patients, and associated accuracy of the resulting sparse representations.

2. MATERIALS AND METHODS

2.1. The K-SVD and MOD methods

Both methods are only briefly summarized here. Details on their development and characteristics may be found in [15] and [12]. The K-SVD method aims to derive (learn) a dictionary of atoms from a representative (training) dataset. For a set of N vectors $Y = \{y_i(t)\}$, $i = 1, ...N$, each of length n, a dictionary of K atoms, with $K \ll N$, is estimated so as to minimize the error:

$$min\{\| \mathbf{Y} - \mathbf{DX} \|_F^2\} \tag{1}$$

$\forall i, \| \mathbf{x_i} \|_0 \leq T_0$, the latter being a typically small set of pre-defined non-zero entries. $\| \cdot \|_F$ refers to the Frobenius norm and $\| \cdot \|_0$ the l^0 norm, the count of non-zero entries in a vector. $\mathbf{D} \in R^{nXn}$ is the dictionary and $\mathbf{X} = \{\mathbf{x_i}\}$ is the coefficient matrix of sparse representations, which are both updated to ultimately converge to a local minimum. Similarly, the MOD method involves sparsification (individually for each signal) and dictionary update, using standard methods, e.g. Orthogonal Matching Pursuit (OMP) [27]. It is a computationally efficient method that converges after a small number of iterations. The primary challenge in the context of high-dimensional brain signals is how to optimally estimate the dictionary. In the case of spatially localized high-frequency pathological waveform anomalies with little a priori information of their origin of onset (at least non in noninvasive signals) and co-occurrence of transient physiological, and thus normal, activity, it is unclear if electrode-specific, region-specific and/or patient specific dictionaries versus overcomplete global dictionaries are most appropriate.

2.2. Electrophysiological data

The study was approved by the BIDMC institutional review board. Noninvasive (scalp) EEGs were analyzed from three adult epilepsy patients with continuous data collected at the Comprehensive Epilepsy Center, Beth Israel Deaconess Medical Center (BIDMC), as part of clinically indicated studies, using a 10-20 EEG system (22 electrodes). One patient had additional sub-temporal electrodes (a total of 28 electrodes). Signals were sampled at a rate of 500 samples/s and were re-referenced to an average reference montage prior to analysis. Given that the study focused on the high-frequency (> 80 Hz) part of the EEG spectrum, signals were high-pass filtered with a 3rd order elliptical filter (cutoff at 80 Hz, 0.5 dB ripple in the passband and 20 dB attenuation in the stopband). Examples of pathological, transient and spatially localized high-frequency waveforms are shown in Figure 1.

Data from 2 adult patients [one male and one female, ages 29 and 47, respectively] with diagnosed focal epilepsy. One patient had seizures originating in the left temporal lobe and one patient had seizures original in bilateral frontal lobes. It is currently unclear whether stereotypical high-frequency signal abnormalities occur across electrodes, brain regions or even patients. Thus, patients with electrophysiological abnormalities in different parts of the brain were included. Scalp EEG recordings spanned ~50 - ~70 h.

Scalp EEGs recorded over long periods of time are typically contaminated by various artifacts, including eye-blinking (typically high-amplitude but low-frequency), muscle activity (high-amplitude and broadband twitching, chewing, etc) and movement (typically lower frequency). The algorithm presented in [8] was used to suppress these artifacts, though it is possible that residual, muscle-related high-frequency activity is still detectable in high-pass filtered signals. Training datasets were carefully selected from periods not containing high-frequency waveforms that are likely to be associated with muscle artifacts. This was done by examining the broadband data at the same time segments as those containing measurable high-frequency activity, since artifacts have high amplitudes and are usually easily detectable by visual inspection. For each electrode, a training dataset spanning an interictal period of 2 h was used to estimate and

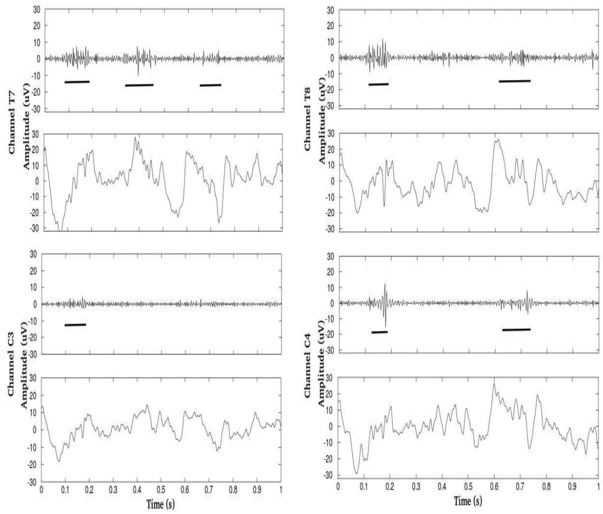


Fig. 1. One-second interictal (non-seizure) high-pass and corresponding low-pass filtered segments from electrodes T7 (top left), T8 (top right), C3 (bottom left) and C4 (bottom right). Low-pass filtered segments are included to show that identified high-frequency waveforms are unlikely to be related to muscle artifacts (which would be detectable at low frequencies as well). Note that high-frequency activity in electrode C3 has negligible amplitude.

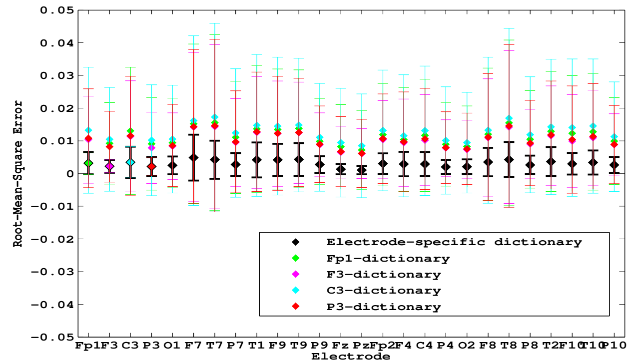
update the corresponding dictionary.

3. RESULTS

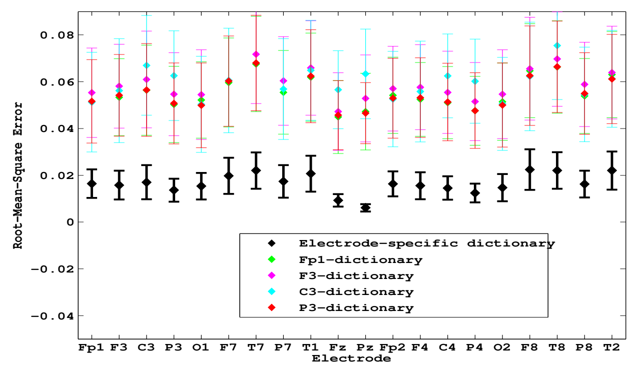
Electrode-specific dictionaries and corresponding sparse reconstructions were first estimated. The signal reconstruction root-mean square errors (RMSE) for each electrode, averaged over all recordings for each patient, are shown in Figure 2. RMSEs when a 'mismatched' signal dictionary was used in the reconstructions (i.e, a dictionary estimated based on signals from one electrode was used to estimate sparse representations of signals from a different electrodes) are also superimposed. Although electrode-specific dictionaries may share common waveform anomaly atoms, given that a relatively large area of cortex must be simultaneously active for low-amplitude, high-frequency signals to be measurable at the scalp, signal reconstruction errors increased substantially when a particular electrode-specific dictionary was used to represent signals are other electrodes. Note that across estimations, differences between results based on the K-SVD and the MOD were negligible, in the sense of the accuracy of the sparse representations.

Training signals from all electrodes were then combined to estimate a common dictionary for the entire dataset. Despite an average 20-fold increase in the size of the dictionary, reconstruction RMSEs were substantially increased ($\sim 25\%$ higher) when a common dictionary was estimated and used for sparse representations of the entire dataset compared to the smaller electrode-specific dictionaries. Thus, increasing the size and heterogeneity of the dictionary did not improve the accuracy of the sparse representation. The results for patient # are shown in Figure 3.

Finally, sparse representations based on dictionaries from multiple patients were estimated. For each electrode, high-pass filtered training signals the 2 patients were combined. Increasing the heterogeneity of the dictionary at the electrode level, i.e., combining training waveforms from multiple patients but maintaining electrode-specific dictionaries. The size of these dictionaries increased modestly (on average slightly more than 50%), indicating both a potentially non-linear relationship between the number of datasets (patients) and the increase in dictionary size as well as the existence of common anomaly atoms between patients. Interestingly,



(a) Patient 1.



(b) Patient 2.

Fig. 2. RMSE of sparse signal representations based on individual (electrode-specific dictionaries) are shown for each electrode (black) and corresponding RMSEs using a mismatched dictionary, i.e., the dictionaries for electrode Fp1 (green), F3 (magenta), C3 (central), P3 (red), as a common dictionary to estimate representations of all electrodes. Error bars indicate the standard deviation of the RMSE over all recordings.

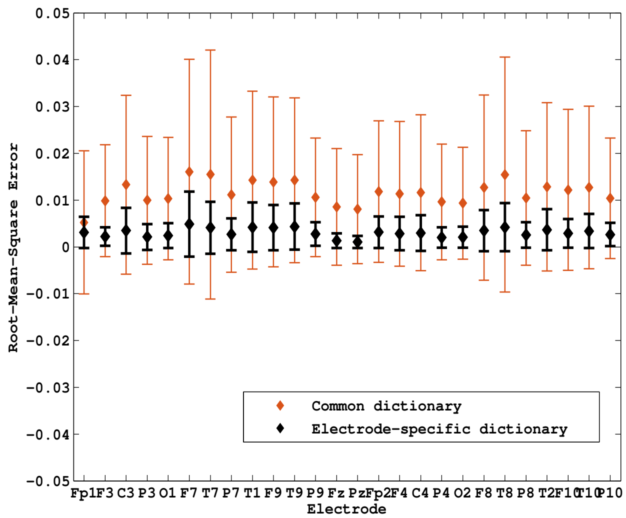


Fig. 3. RMSE of sparse signal representations based on individual (electrode-specific dictionaries) are shown for each electrode (black) and corresponding RMSEs using a common dictionary are superimposed (red).

based on this dictionary the accuracy of the sparse representations increased for patient #2 and decreased for patient #1, relative to those based on patient- and electrode-specific dictionaries. This may be due to the fact that the training dataset was adequate for one of the patients' recordings but not for the other. These results are summarized in Figure 4.

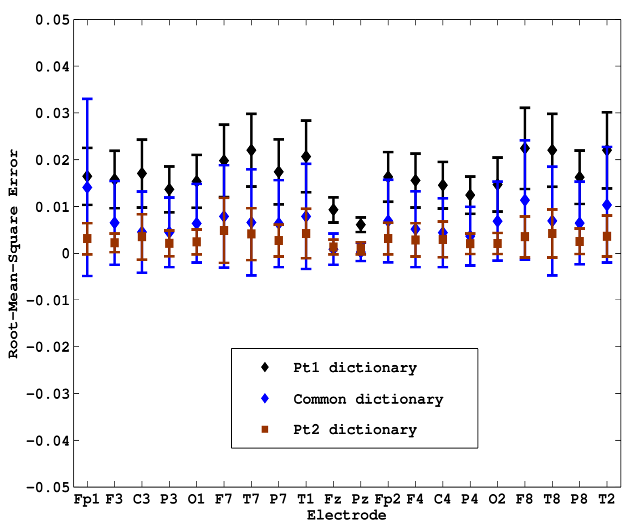


Fig. 4. RMSE of sparse signal representations based on individual (electrode-specific dictionaries) are shown for each electrode, for patients #1 (brown) and #2 (black) respectively, and corresponding RMSEs using a common dictionary for both patients superimposed (blue).

4. CONCLUSION

In a preliminary investigation of sparse representations of high-frequency waveform anomalies in very high-dimensional electrophysiological signals based on data-derived dictionaries, we have assessed the impact of the training data used to learn these dictionaries and the latter's heterogeneity on the accuracy of the sparse representations. Electrode-specific dictionaries yielded substantially more accurate representations than those estimated by combining training signals from multiple electrodes to derive a larger and theoretically more heterogeneous and more redundant dictionary. This suggests that waveform anomalies may vary substantially between electrodes with potentially little overlap. In contrast, when training data from both patients were combined to estimate electrode-specific dictionaries, the accuracy of the estimated sparse representations only decreased slightly for one patient and increased substantially for the other. It is possible that this is in part due to the choice of the training dataset, which may not have adequately captured the variability of waveform anomalies across the entire recordings for one of the patients. There findings are based on a very small patient sample (although the data spanned $\leq$50 h). A substantially larger sample is necessary to assess the relationship between the size of the dictionary size and the number of datasets used to estimate it, common atoms in electrode-specific dictionaries and the impact of different training data. Ideally, a single (overcomplete) dictionary is desirable that can be used across datasets and electrodes. This preliminary investigation suggests that augmenting the training dataset to include data from multiple electrodes does not result in improved sparse representations, highlighting the potential heterogeneity of high-frequency waveforms across the brain. Finally, systematic simulations are necessary to assess issues in dictionary learning and the accuracy of sparse representations in a controlled way.

5. REFERENCES

[1] A Bragin., JJ Engel, CL Wilson, I Fried, G Buzsaki, High-frequency oscillations in human brain, *Hippocampus*, 9:137142, 1999.

[2] J Jacobs, P LeVan, R Chander, et al, Interictal high-frequency oscillations (80-500 Hz) are an indicator of seizure onset areas independent of spikes in the human epileptic brain, *Epilepsia*, 49(11): 1893-1907, 2008.

[3] J Jacobs, R Zelmann, J Jirsch, et al, High frequency oscillati ons (80-500 Hz) in the preictal period in patients with focal seizures, *Epilepsia*, 50(7): 1780-1892, 2009.

[4] LP Andrade-Valenca, F Dubeau, F Mari, et al, Interictal scalp fast oscillations as a marker of the seizure onset zone, *Neurology*, 77(6): 524-531, 2011.

[5] C Stamoulis, L Gruber, D Schomer, B Chang, High-frequency neuronal network modulations encoded in scalp EEG precede the onset of focal seizures, *Epilepsy Behav*, 23:471-80, 2012.

[6] C Stamoulis, D Schomer, B Chang, Information theoretic measures of network coordination in high-frequency scalp EEG reveal dynamic patterns associated with seizure termination, *Epilepsy Res*, 105:299-315, 2013.

[7] B Frauscher, F Bartolomei, K Kobayashi et al, High-frequency oscillations: The state of clinical research, *Epilepsia* 58(8): 1316-1329, 2017.

[8] C Stamoulis, B Chang, Application of matched-filtering to extract EEG features and decouple signal contributions from multiple seizure foci in brain malformations, *IEEE Proc 4th Int Conf Neural Eng*, 514-517, 2009.

[9] S Ganguli, H Sompolinsky, Compressed sensing, sparsity and dimensionality in neuronal information processing and data analysis, *Ann Rev of Neuroscience*, 35:485508, 2012.

[10] H Lee, DS Lee, H Kang, BN Kim, MK Chung, Sparse Brain Network Recovery under Compressed Sensing, *IEEE Trans Med Imag*, 30(5):1154-1165, 2011.

[11] DL Donoho, M Elad, Optimally sparse representation in general (nonorthogonal) dictionaries via l-1 minimization, *Proc Natl Acad Sci*, 100(5): 2197-2202, 2003.

[12] K Engan, SO Aase, JH Husoy, BMethod of optimal directions for frame design, *Proc IEEE Int Conf Acoust, Speech, Signal Process*, 5:24432446, 1999.

[13] K Engan, SO Aase, JH Husoy, Multi-frame compression: Theory and design, *EURASIP Signal Process*, 80(10): 21212140, 2000.

[14] K Kreutz-Delgado, JF Murray, BD Rao, K Engan, T Lee, TJ Sejnowski, Dictionary learning algorithms for sparse representation, *Neural Comp*, 15(2):349396, 2003.

[15] M Aharon, M Elad, A Bruckstein, K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation, *IEEE Trans Sig Proc*, 54(11): 4311-4322, 2006.

[16] H Lee, A Battle, R Raina, AY Ng, Efficient sparse coding algorithms, *Advances in neural information processing systems*, 2006

[17] S Mallat, Z Zhang, Matching Pursuit with Time-Frequency Dictionaries, *IEEE Trans Sig Proc*, 41(12): 3397-3415, 1993.

[18] N Cho, CJ Kuo, Sparse Music Representation With Source-Specific Dictionaries and Its Application to Signal Separation, *IEEE Trans Audio Speech Proc*, 19(2): 337-348, 2011.

[19] M Yang, L Zhang, Gabor Feature based Sparse Representation for Face Recognition with Gabor Occlusion Dictionary, *Eur Conf Comp Vis*, 448-461, 2010.

[20] D Barchiesi, MD Plumbley, Learning Incoherent Dictionaries for Sparse Approximation using Iterative Projections and Rotations, *IEEE Trans Sig Proc*, 61(8): 2055-2065, 2013.

[21] BA Olshausen, DJ Field, Emergence of simple-cell receptive field properties by learning a sparse code for natural images,*Nature*, 381(6583): 607609, 1996.

[22] R Rubinstein, AM Bruckstein, M Elad, Dictionaries for Sparse Representation Modeling, *Proc IEEE*, 98(6):1045-1057, 2010.

[23] Z Zhang, Y Xu, J Yang, X Li, D Zhang, A Survey of Sparse Representation: Algorithms and Applications, *IEEE Access*, 3:490-530, 2015.

[24] R Rubinstein, M Zibulevsky, M Elad,Double Sparsity: Learning Sparse Dictionaries for Sparse Signal Approximation, *IEEE Trans Sig Proc*, 58(3):1553-1564, 2010.

[25] B Hamner, R Cavarriaga, J Milan, Learning dictionaries of spatial and temporal EEG primitives for brain-computer interfaces *Workshop on Structured Sparsity: Learning and Inference*, 1-4, 2011.

[26] W Zhou, Y Yang, Z Yu, Discriminative dictionary learning for EEG signal classification in Brain-computer interface, *ICARCV*, 1-5, 2013.

[27] TT Cai, L Wang, Orthogonal Matching Pursuit for Sparse Signal Recovery With Noise, *IEEE Trans Sig Proc*, 57(7):4680-4688, 2011.