1st Pan-American Congress on Computational Mechanics - PANACM 2015
XI Argentine Congress on Computational Mechanics - MECOM 2015
S. Idelsohn, V. Sonzogni, A. Coutinho, M. Cruchaga, A. Lew & M. Cerrolaza (Eds)

AD-HOC GAUSSIAN DICTIONARIES FOR SPARSE REPRESENTATION OF EVOKED RELATED POTENTIALS

VICTORIA PETERSON[1], HUGO L. RUFINER[2] AND RUBEN D. SPIES[3]

[1]Instituto de Investigación en Señales, Sistemas e Inteligencia Computacional, Facultad de Ingeniería y Ciencias Hídricas, Universidad Nacional del Litoral, Consejo Nacional de Investigaciones Científicas y Técnicas (sinc(i)-FICH-UNL-CONICET).
Ruta Nac. No 168, km 472.4 (3000), Santa Fe - Argentina
vpeterson@santafe-conicet.gov.ar

[2]Instituto de Investigación en Señales, Sistemas e Inteligencia Computacional, Facultad de Ingeniería y Ciencias Hídricas, Universidad Nacional del Litoral, Consejo Nacional de Investigaciones Científicas y Técnicas (sinc(i)-FICH-UNL-CONICET)
Ruta Nac. No 168, km 472.4 (3000), Santa Fe - Argentina and Facultad de Ingeniería, Universidad Nacional de Entre Ríos (FI-UNER) Ruta Prov. No 11, km 10 (3100), Oro Verde - Entre Ríos - Argentina
lrufiner@fich.unl.edu.ar

[3]Instituto de Matemática Aplicada del Litoral (IMAL-CONICET-UNL)
Ruta Nac. No 168, Paraje El Pozo (3000), Santa Fe - Argentina and Facultad de Ingeniería Química, Universidad Nacional del Litoral, Consejo Nacional de Investigaciones Científicas y Técnicas (FIQ-UNL)
Santiago del Estero 2829 (3000), Santa Fe - Argentina.
rspies@santafe-conicet.gov.ar

Key words: Brain-Computer Interface, Evoked Related Potentials, Sparse Representation, ℓ_1-minimization, Robust Classification

Abstract. A Brain-Computer Interface (BCI) is a system which provides direct communication between the mind of a person and the outside world by using only brain activity (EEG). A common EEG BCI paradigm is based on the so called Event-Related Potentials (ERP) which are responses of the brain to some external stimuli. For the present work at hand, the innermost part of a BCI is the pattern recognition stage whose aim is to detect the presence of ERPs with high accuracy. In recent years there has been a growing interest in the study of sparse representation of signals. Using a dictionary composed of prototype atoms, signals are written as linear combinations of just a few of those atoms. This sparse representation is found to be appropriate for posterior classification purposes. In this work we propose a sparse representation and posterior classification of ERPs signals by means of an ad-hoc spatio-temporal dictionary composed of bidimensional Gaussian atoms. Based on ℓ_1-minimization we find the sparsest possible solution which allow us to design a robust classification based on nearest representation

1 INTRODUCTION

Sparse representations have received great interest in the recent years due to their success in many applications in signal and image processing [1], [2].

Given a $m \times n$ matrix A, called dictionary, an unknown signal $\mathbf{x} \in \mathbb{R}^n$ and a measurement vector $\mathbf{y} \in \mathbb{R}^m$, we seek to find the sparsest coefficient vector $\mathbf{x}$ such that $\mathbf{y} = A\mathbf{x}$. Recovering $\mathbf{x}$ given A and $\mathbf{y}$ is a non trivial inversion problem since in general the size of $\mathbf{x}$ is greater than the size of $\mathbf{y}$, the measurement $\mathbf{y}$ is contaminated by noise and the problem is usually severally ill-posed. Hence, regularization is required.

Given a functional $J(\mathbf{x})$, which penalizes certain undesired properties of the solution, a regularized solution can then be obtained by solving the following constrained optimization problem:

$$(P_J): \quad \min_{\mathbf{x}\in\mathbb{R}^n} J(\mathbf{x}) \quad s.t. \quad \mathbf{y} = A\mathbf{x}. \tag{1}$$

Most of the existing works on sparse learning are based on variants of the ℓ_1-norm regularization ($J(\mathbf{x}) = ||\mathbf{x}||_1$) due to its sparsity-inducing property, convenient convexity, wide and strong theoretical support, and great success in several applications [3]. Thus, the regularized sparse solution are obtained by solving the following minimization problem:

$$(P_1): \quad \min_{\mathbf{x}\in\mathbb{R}^n} ||\mathbf{x}||_1 \quad s.t. \quad \mathbf{y} = A\mathbf{x}. \tag{2}$$

It has been shown, under certain conditions, that the minimum ℓ_1-norm solution of an undetermined linear system is also the sparsest possible solution [4]. In practice, however, $\mathbf{y}$ is contaminated by noise and therefore the equality constraint in (2) must be relaxed:

$$(P_{1,2}): \quad \min_{\mathbf{x}\in\mathbb{R}^n} ||\mathbf{x}||_1 \quad s.t. \quad ||\mathbf{y} - A\mathbf{x}||_2 \leq \epsilon, \tag{3}$$

where ϵ is noise level.

Problem $(P_{1,2})$, known as the *basis pursuit denoising* problem (BPDN), is equivalent to the following unconstrained minimization problem:

$$(P_\lambda): \quad \min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}||\mathbf{y} - A\mathbf{x}||_2^2 + \lambda||\mathbf{x}||_1, \tag{4}$$

which can be viewed as a generalized Tikhonov-Phillips regularization functional [5] or in a statistical context as a Least Absolute Shrinkage and Selection Operator (LASSO) [6].

In this article we pretend to use the virtues of sparse representation framework together with the ideas proposed in [7] in order to develop a robust classification method in the context of Event Related Potential (ERP) detection.

The problem at hand comes from Brain-Computer Interfaces (BCI) systems. BCI can significantly improve the quality of life of a person who cannot control his/her own body or even is not able to communicate. By using only brain activity, BCI provides a person a new way of communication and control without needing any peripheral nerves or muscles

[8]. The most common and non-invasive method used to decode the intention of a BCI user consists of detecting the presence of ERP signals in electroencephalogram (EEG) records.

In the “oddball” paradigm it is well known that when a person is stimulated with some external and “rare” item (which can be auditory, visual or somatosensorial) an ERP is elicited. One of the main components of ERP signals is an enhanced positive-going component with a latency of about 300 ms (called P300 wave) [9], [10]. Unfortunately detecting a P300 wave (which means detecting the ERP signal) is not an easy task. Mainly due to the fact that SNR between ERP and EEG signals is very low (about -50 db) and also due to the large variation present in P300 wave records among different trails.

In order to use all the information that can be found in EEG records we shall construct an ad-hoc dictionary whose elements are composed by two bidimensional Gaussian atoms representing the main spatio-temporal variation of EEG records with and without P300, respectively. As explained in the next section, the dictionary was built based on a certain neurologically well-supported assumptions on the P300 wave.

2 MATERIALS AND METHODS

2.1 Database

An Open-Access P300 speller database was used [11]. EEG records from 18 subjects were acquired by 10 electrodes in the positions shown in Fig.1. The records were digitalized at a rate of 256 Hz.

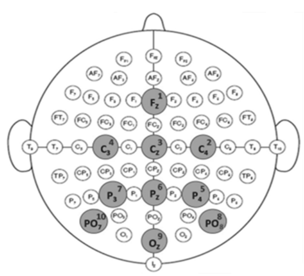


Figure 1: Positions of the electrodes used in EEG acquisition. Modified from [11].

The Farwell and Donchin oddball paradigm was used [10], where a 6-by-6 matrix containing letters and numbers were displayed on a computer screen. During the experiment, a subject is asked to spell different words. The person must focus on one character at the time. As stimulus, a row or a column of the matrix is randomly flashed. In each stimulating block, every row and column of the matrix is intensified only once. If the person is well concentrated, when the chosen character is illuminated, a relevant event occurs, i.e. an ERP signal is elicited [12]. Since there are two classes involved, the 6-by-6 matrix result in twelve possible events (one per each row and one per each column), of which only two are relevant, i.e. with ERP, in every intensification block.

Each subject participated in 4 sessions, where the first two ones were copy-spelling runs. For this reason, in this work we used the first and second sessions as training and testing sets, respectively. Each session consisted of 15 trials, resulting in 2880 EEG epochs per channel (480 target records and 2400 non-target records) for the training set and 900 EEG epochs per channel (150 of them being target) for the testing set. An epoch is a EEG record of one second duration extracted at the beginning of an intensification.

2.2 Gaussian Dictionary Generation

2.2.1 Image Pre-processing

In order to use all the information available in an epoch, we used not only the information time-to-time but also the spatial information given by the electrode's positions. In this way, we constructed one image per trial in the time-channel plane. Since a trial is an epoch of one second duration extracted at the beginning of an intensification, and the number of channels was ten, our images consisted of 256×10 pixels.

Analyzing the images of the great averages for both classes (target records and non-target records), we noticed that the shape of the P300's peaks could be significantly improve by re-ordering the channels. In order to do this we proceeded as follows. Since the registered amplitude of the P300 wave differs according to the sensors's position, the channels were re-ordered based on decreasing signal energy, i.e, we chose as the first channel the one possessing greatest ℓ_2-norm, and the rest were decreasingly ordered by their Euclidean distance at this "first" channel. Next we filtered the image with a median filter. An example of a resulting re-ordered and filtered image are shown in column 2 of both figures Fig.2a and Fig.2b.

After analyzing each re-ordered and filtered target image, we observed that for all of them there were one or two notable peaks between 0.2s and 0.6s. On the other hand, the peaks for non-target images always came from the oscillatory background. For those reasons and because the 0.2-0.6 range is in agreement with the latency window of the P300 wave, we cropped all images between 0.2s and 0.6s, resulting in images of 104×10 pixels.

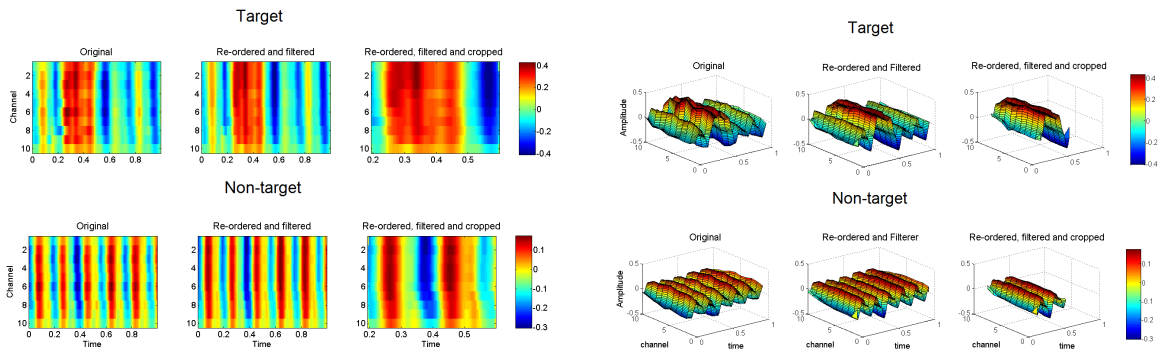
Images belonging to the three different pre-processing stages are shown in Fig.2 for the first subject in the database. There are the original image, the re-ordered and filtered image and the re-ordered, filtered and cropped image for both classes in two different views.

In the sequel, a template will referred to the re-ordered, filtered and cropped images.

2.2.2 Levenberg-Marquardt Estimation of Mother Elements

Inspired by the ideas of pre-defined dictionaries (like wavelet dictionaries), we want to generate a dictionary by means of variations of an appropriate "mother" element.

In our case we shall consider one mother element per class and per subject, given by a



(a) 2D view (b) 3D view

Figure 2: Target and non-target images for subject N° 1. From left to right: original image, re-ordered and filtered image and re-ordered, filtered and cropped image. Fig. 2a show a 2D view while Fig.2b show a 3D view

linear affine combination of two Gaussian functions (atoms). More precisely, let $\mathcal{P} = \{(p_1, p_2, ..., p_{11}) \in \mathbb{R}^{11} : p_5, p_6, p_{10}, p_{11} > 0\}$ and for $\mathbf{p} \in \mathcal{P}$ define:

$$z(t, c; \mathbf{p}) = p_1 + p_2 \exp\left(\frac{(t-p_3)^2}{p_5^2} + \frac{(c-p_4)^2}{p_6^2}\right) + p_7 \exp\left(\frac{(t-p_8)^2}{p_{10}^2} + \frac{(c-p_9)^2}{p_{11}^2}\right). \quad (5)$$

In the sequel we shall always identify the scalar field $z(t, c; \mathbf{p})$ with the vector $z(\mathbf{p}) \in \mathbb{R}^{1040}$ obtained after stacking in a column vector the matrix resulting of the evaluation of $z(t, c; \mathbf{p})$ over 104×10 image grid.

Given the template $f(t, c)$ (or simply $f \in \mathbb{R}^{1040}$) we formulate the corresponding nonlinear fitting problem associated to f and the mother element as:

$$(F_{\mathbf{p}}) : \quad \min_{\mathbf{p} \in \mathcal{P}} ||f - z(\mathbf{p})||_2^2. \quad (6)$$

The Levenberg-Marquardt (LM) method [13], [14] is a standard and efficient technique to solve nonlinear least squares problems. We used it to find $\mathbf{p}$ in (6).

The estimated mother elements for the same subject used in Fig.2 are shown in Fig.3. The templates are shown on the left while the two estimated Gaussian elements fitted by LM algorithm are on the right.

It is timely to observe the good fitting of the mother elements to the corresponding templates, especially for the target class.

2.2.3 Dictionary Build Up

As the dictionary must capture the main variation of the P300 wave for each subject, in order to construct an appropriate dictionary it is very important to analyze the sensitivity

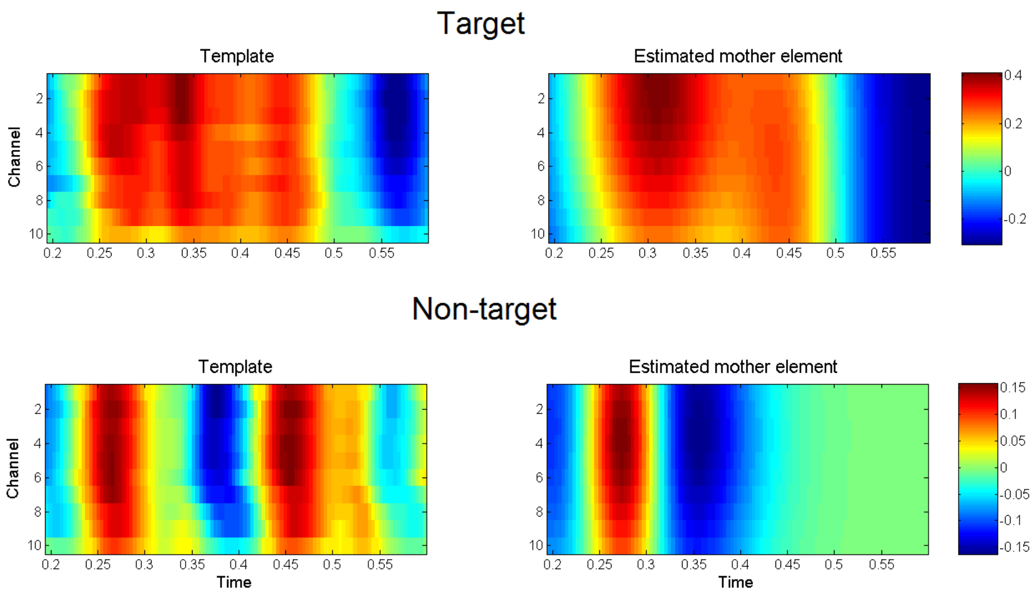


Figure 3: Templates for subject N° 1 (left) and mother elements estimated by LM method (right).

of each one of the components of the parameters vector **p** with respect to the epochs in each one both classes. With this in mind, we performed a rough analysis in this direction following the next steps:

1. **Averaging and balancing:** take averages of 5 trials for target records and averages of 25 trials for non-target records (see Remark (1) below).

2. **Template generation:** for each one of the averaged trials generated in step (1) construct the corresponding template image, as described in Section 2.2.1.

3. **Parameters generation:** for each one of the templates described in step (2) estimate the corresponding parameter **p** by solving problem (6) with the LM method.

4. **Parameters range estimation:** step (3) generates a family of vectors $\mathbf{p} \in \mathbb{R}^{11}$ for each one of both classes. The range variation for each one of the 11 components of **p** is estimated within each one of both families in the usual way.

Remark (1)**.** *The different number of elements, 5 and 25, for taking averages in each one of the classes has the objective of balancing the different number of epochs among classes.*

For the dictionary generation itself we varied one parameter at the time while the others were kept constant. The variations were made in order to cover the whole range in each one of the components of $\mathbf{p}$ in one hundred equal increments[1].

From the sparse representation point of view, it is highly desirable not to have dictionary's elements which are "too similar". The Mutual Coherence (MC) of a dictionary, denoted by $\mu(A)$, measures the similarity between dictionary's elements. It is defined as the maximal absolute scalar product between two different ℓ_2-normalized elements of the dictionary A [1], that is:

$$\mu(A) \doteq \max_{i \neq j} | < \mathbf{a}_i, \mathbf{a}_j > |. \tag{7}$$

Notice that if two dictionary's elements are parallel then $\mu(A) = 1$, otherwise $\mu(A) \geq 0$.

In the dictionary generation algorithm we discarded an element if its MC was greater than some predefined number κ. To avoid classification bias, the dictionary sizes should be the same for both classes. With this objective in mind, we randomly eliminated the necessary number of elements from the larger dictionary whose MCs were greater than another predefined value (e.g. $\kappa - 0.5$).

At this point we have for any subject one dictionary per class, both dictionaries having the same number of elements and with some predefined MC=κ. It is appropriate to mention here that the size of the dictionary growths with κ.

2.3 Classification Based on Sparse Representation

Let us define a new matrix A as the concatenation of the n elements from both target and non-target dictionaries, A_1 and A_2, respectively, that is:

$$A \doteq [A_1 \ A_2] = [\mathbf{a}_{1,1}, \mathbf{a}_{1,2}, ..., \mathbf{a}_{1,n}, \mathbf{a}_{2,1}, ..., \mathbf{a}_{2,n}]. \tag{8}$$

For given ϵ, A and $\mathbf{y}$, let $\hat{\mathbf{x}}$ be the solution of the problem $(P_{1,2})$ in (3). Ideally one would expect that the nonzero entries of $\hat{\mathbf{x}}$ will correspond to columns of A belonging all to the same class. In that case the association of $\mathbf{y}$ to one of both classes is clearly trivial. Noise and modeling errors, nonetheless, may lead to nonzero entries associated to the wrong class. It is therefore clear that any classification criterion will have to take into account additional information such as the goodness of fit, that is, a measure of how well the coefficients associated with each one of the classes reproduced $\mathbf{y}$.

For each class i, let $\delta_i : \mathbb{R}^n \to \mathbb{R}^n$ be the lifting function that selects the coefficients associated with the i^{th} class, that for $\mathbf{x} \in \mathbb{R}^n$, $\delta_i(\mathbf{x}) \in \mathbb{R}^n$ is the vector whose only nonzero entries are the elements in $\mathbf{x}$ that are associated to class i. The representation of a given test sample $\mathbf{y}$ in class i is then $\hat{\mathbf{y}}_i = A\delta_i(\hat{\mathbf{x}})$. The classification of $\mathbf{y}$ proceeds by assigning it to the class that minimizes the residual, i.e., we associate to $\mathbf{y}$ the class given by:

$$\underset{i=1,2}{\text{argmim}} \ r_i(\mathbf{y}) \doteq ||\mathbf{y} - A\delta_i(\hat{\mathbf{x}})||_2. \tag{9}$$

The Sparse Representation-based Classification (SRC) algorithm (as proposed in [7]) can then be written as follows:

[1] The variation in p_1, corresponding to the offset parameter in (5) was negligible and therefore it was kept constant for the dictionary generation.

1. **Input**: The train dictionary $A = [A_1, A_2] \in \mathbb{R}^{m\times n}$, a test sample $\mathbf{y} \in \mathbb{R}^m$ and ϵ tolerance or λ regularized parameter.

2. Normalize the columns of A to have unit $\ell_2 - norm$.

3. Solve the ℓ_1-minimization problem:

$$\min_{\mathbf{x}\in\mathbb{R}^n} ||\mathbf{x}||_1 \quad s.t. \quad ||\mathbf{y} - A\mathbf{x}||_2 \leq \epsilon.$$

   Or equivalently solve:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}||\mathbf{y} - A\mathbf{x}||_2^2 + \lambda||\mathbf{x}||_1.$$

4. Compute the residuals $r_i(\mathbf{y}) = ||\mathbf{y} - A\delta_i(\hat{\mathbf{x}})||_2$, for i=1,2.

5. **Output**: Identity($\mathbf{y}$)=$\underset{i=1,2}{\text{argmim}}\ r_i(\mathbf{y})$.

3 RESULTS AND DISCUSSIONS

The sparse vectors $\hat{\mathbf{x}}$ for each test observation $\mathbf{y}$ were estimated using different functions by the application of SLEP 4.1 toolbox [15]. More precisely we estimated the sparse vectors $\hat{\mathbf{x}}$ as follows:

1. ℓ_1-ball constrained least squares problem (LeastC):

$$\min_{\mathbf{x}} \frac{1}{2}||\mathbf{y} - A\mathbf{x}||_2^2 \quad s.t. \quad ||\mathbf{x}||_1 < \epsilon, \tag{10}$$

   where ϵ is the radius of the ℓ_1-ball.

2. ℓ_1-norm regularized least squares problem (LeastR):

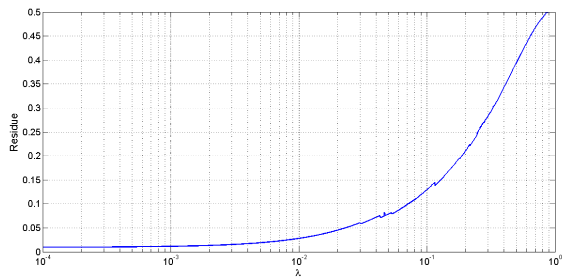
$$\min_{\mathbf{x}} \frac{1}{2}||\mathbf{y} - A\mathbf{x}||_2^2 + \lambda||\mathbf{x}||_1. \tag{11}$$

3. Non-negative ℓ_1-ball constrained least squares problem (NNLeastC):

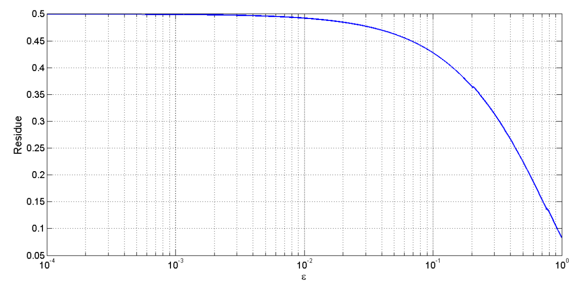
$$\min_{\mathbf{x}} \frac{1}{2}||\mathbf{y} - A\mathbf{x}||_2^2 \quad s.t. \quad ||\mathbf{x}||_1 < \epsilon, \mathbf{x} \geq 0. \tag{12}$$

4. Non-negative ℓ_1-norm regularized least squares problem (NNLeastR):

$$\min_{\mathbf{x}\geq 0} \frac{1}{2}||\mathbf{y} - A\mathbf{x}||_2^2 + \lambda||\mathbf{x}||_1. \tag{13}$$



(a)



(b)

Figure 4: Residue as a function of λ for the LeastR algorithm (4a) and residue as a function of ϵ for the LeastC algorithm (4b)

.

The optimal values of λ and ϵ were fixed after analyzing the residue as a function of λ or ϵ, respectively. We varied the parameters between 10^{-4} and 1 in increment of 10^{-4}. We finally chose $\lambda = 0.05$ and $\epsilon = 0.8$ as optimal values. The behavior of the residue is shown in Fig.4.

In order to analyze the impact of the choice of MC value on generation of the dictionary and, consequently, on the classification rate obtained by SRC algorithm, we varied the MC values from 0.500 to 0.995 in increments of 0.005.

The unweighted accuracy rate (UAR) was used as the performance classification measure. Fig.5 shows the classification results reached by each optimization problem (LeastC, LeastR, NNLeastC and NNLeastR) for four different database's subjects.

There does not seem to be a clear tendency between classification results and MC values. Note that while for subject N° 10 the UAR shows a clearly increasing tendency for increasing MC values, the opposite happens for subject N° 1. This observation constitutes a solid reason for developing a subject depended classification tools.

Table 1 summarizes the best UAR results per each subject with the different optimization problems. The corresponding MC value is shown between parenthesis. An analysis of Table 1 seems to suggest that better classification result are obtained by imposing $\mathbf{x} \geq 0$. Although further analysis is required, we strongly believe that this is due to the fact that the elements in the target dictionary present mainly positive peaks.

4 CONCLUSIONS

In this paper, we constructed an ad-hoc Gaussian dictionary for representing the P300 wave in a channel-time space. We focused our work in constructing a suitable dictionary per each subject in order to represent the variation of the P300 wave of that particular subject. It is not of our interest to find a generalized (all subjects) representation of the P300 wave. Moreover, we have good reasons to believe that better classification result can be obtained by improving the representation of non-target signals.

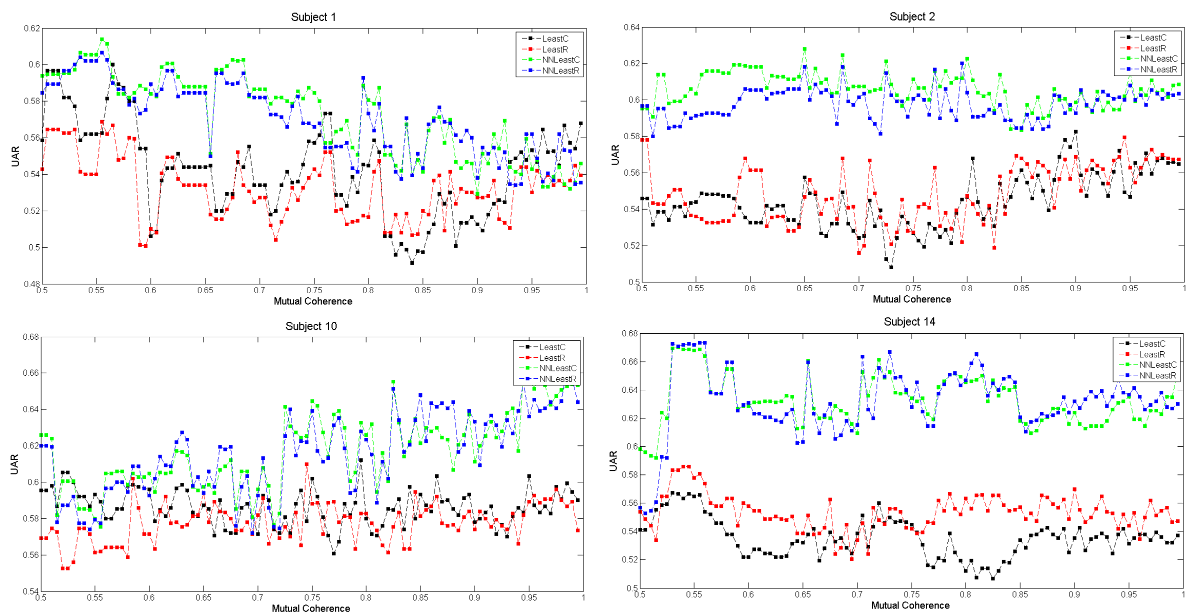


Figure 5: UAR results as a function of MC values for the four different algorithms used.

The resulting ℓ_1-minimization problems were solved using two different optimization approach both with and without a positive constraint, and used the sparse information of the $\hat{\mathbf{x}}$ solution for classification purposes. Although the classification performances are far for being optimal, we point out that the classification method used is very simple and moreover noisy single trail epochs were used. Much further research is clearly needed in this regard. In particular it is of great interest to find alternative and/or complementary ways to the ℓ_1-minimization approach, which could allow classification improvement. We are currently devoting efforts in this direction.

5 ACKNOWLEDGEMENTS

This work was supported in part by Consejo Nacional de Investigaciones Científicas y Técnicas, CONICET, through PIP 2014-2016 Nro. 11220130100216-CO, the Air Force Office of Scientific Research, AFOSR/SOARD, through Grant FA9550-14-1-0130 and by CAID-UNL 2011.

REFERENCES

[1] Elad, M. Sparse and Redundant Representation, from theory to applicationsin signal and image processing. *Springer*. (2010). ISBN 978-1-4419-7010-7.

[2] Bruckstein, J.M, Donoho, D. and Elad, M. From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images. *Society for Industrial and Applied*

Table 1: Best UAR results for the different optimization problems.

Subject N^{o}	LeastC	LeastR	NNLeastC	NNLeastR
1	0.600 (0.565)	0.569 (0.555)	0.614 (0.555)	0.607 (0.555)
2	0.583 (0.900)	0.579 (0.945)	0.628 (0.650)	0.620 (0.795)
3	0.531 (0.680)	0.531 (0.730)	0.503 (0.780)	0.509 (0.680)
4	0.558 (0.970)	0.546 (0.645)	0.583 (0.955)	0.590 (0.965)
5	0.528 (0.670)	0.540 (0.700)	0.559 (0.570)	0.567 (0.600)
6	0.547 (0.855)	0.533 (0.750)	0.539 (0.855)	0.538 (0.855)
7	0.554 (0.645)	0.558 (0.900)	0.545 (0.980)	0.548 (0.995)
8	0.576 (0.920)	0.561 (0.790)	0.537 (0.995)	0.539 (0.995)
9	0.531 (0.650)	0.544 (0.595)	0.551 (0.700)	0.604 (0.635)
10	0.612 (0.795)	0.610 (0.745)	0.663 (0.960)	0.664 (0.985)
11	0.557 (0.910)	0.565 (0.940)	0.595 (0.995)	0.624 (0.940)
12	0.519 (0.680)	0.517 (0.600)	0.545 (0.635)	0.547 (0.640)
13	0.580 (0.865)	0.577 (0.840)	0.639 (0.740)	0.647 (0.625)
14	0.567 (0.530)	0.586 (0.540)	0.670 (0.535)	0.673 (0.555)
15	0.573 (0.700)	0.561 (0.645)	0.585 (0.960)	0.559 (0.855)
16	0.523 (0.585)	0.523 (0.590)	0.614 (0.995)	0.609 (0.995)
17	0.577 (0.590)	0.581 (0.855)	0.589 (0.995)	0.595 (0.605)
18	0.546 (0.920)	0.545 (0.760)	0.577 (0.975)	0.577 (0.980)

Mathematics. (2009) **51**:34–81.

[3] Donoho, D. For Most Large Underdetermined Systems of Equations,the Minimal l1-norm Near-Solution Approximates the Sparsest Near-Solution. *Communications on Pure and Applied Mathematics*. (2006) **59**:907–934.

[4] Donoho, D. and Elad, M. Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization. *Proceeding of the National Academy of Sciences of the United States of America*. (2003) **100**:2197–2202.

[5] Engl, H.W., Hanke, M. and Neubauer, A. Regularization of inverse problems. *Kluwer Academic Publishers*. (2010). ISBN 0-7923-4157-0.

[6] Tibshirani, R. Regression, Shrinkage and Selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*. (1996) **58**:267–288.

[7] Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S. and Ma, Y. Robust Face Recognition via Sparse Representation. *IEE Transactions on Pattern Analysis and Machine Intelligence*. (1996) **31**:1–18.

[8] Wolpaw J.R,, Birbaumer N., McFarland D.J., Pfurtscheller G., Vaugham T.M. Brain computer interfaces for communication and control. *Clinical Neurophysiology.* (2002) **113**:767–791.

[9] Hiyard S.A., Kutas M. Electrophysiology of congnitive processing *Annual Reviews Phychol.* (1983) **34**:33–61.

[10] Farwell L.A., Donchin E. Talking off the top of your head: toward a metal prosthesis utilizing event-related brain potentials *Electroencephalography and clinical neurophysiology.* (1988) **70**:510–523.

[11] Ledesma-Ramirez C., Bojorges-Valdez E., Yañez-Suarez O., Saavedra C., Boygrain L., and Gentiletti G. An Open-Access P300 Speller Database. *Fourth international BCI meeting, Monterrey, USA, California* (2010).

[12] Donchin E., Spencer K.M. and Wijesinghe R. The mental prosthesis: Assessing the Speed of a P300-Based Brain-Computer Interface. *Transactions on Rehabilitation Engineering.* (2000)**8**:174–179.

[13] Levenberg, K. A method for the solution of certain problems in least squares. *Quart. Appl. Math.* (1944)**2**:164-168.

[14] Marquardt, D. An algorithm for least-squares estimation of nonlinear parameters *SIAM J. Appl. Math.* (1963)**11**:431-441.

[15] Liu J., Ji S. and Ye J. SLEP: Sparse Learning with Efficient Projections. *Arizona State University* (2009) http://www.public.asu.edu/ jye02/Software/SLEP.