Regression on low-dimension spanned input space applied to neurophysiological signal processing

A. C. Santana, A. V. Barbosa, H. C. Yehia and R. Laboissière

A common problem in neurophysiological signal processing is to extract meaningful information from high-dimension low sample size data (HDLSS). We present RoLDSIS (regression on low-dimension spanned input space), a technique that restrains the regression solution to the subspace spanned by the data points and avoids the need for regularization. We demonstrate its usefulness for finding neurophysiological correlates of phonemic categorization in an electroencephalography (EEG) experiment. Comparison with commonly used regularized regression techniques (LASSO, Ridge and SPLS) are provided.

Introduction: In brain imaging studies, measurements of neuron activity in the central nervous system are collected and related to some hypothesized state of the brain inferred from psychophysical observations. The neuronal activity can be measured using different techniques, like electric potentials (EEG), magnetic fields (MEG) or haemodynamic response (fMRI). Those measurements are usually discretized in time and frequency (through spectro-temporal analysis, like Fourier or wavelet transforms), as well as in the physical space (EEG or MEG sensors, or fMRI voxels), and can be represented as points in a high dimensional space $\mathbb{R}^N$.

Experiments are controlled in such way that the brain is supposed to be in a finite number (let us say, M) of conditions, for example by specifying different perceptual stimuli presented to the subject. Each situation yields a measurement point $\mathbf{x}_i, i = 1, \ldots, M$. Let us suppose that the underlying brain state, in each of the M conditions, can be associated to a continuous, scalar value y_i. This value can be inferred from the physical properties of the stimulus #i or be obtained from some known behavioral response of the subject in the situation #i. We are usually interested in general relationships $y = f(\mathbf{x})$, but we will consider here only the affine relationship $y = a + \mathbf{b}^\intercal\mathbf{x}$. The vector $\mathbf{b} \in \mathbb{R}^N$ represents the y-related neurophysiological axis or, in other words, how the features in the measurement space of $\mathbf{x}$ must be combined in order to yield the value of the brain state y. The vector $\mathbf{b}$, as well as the scalar constant a must be inferred from the M pairs of observations $\{\mathbf{x}_i, y_i\}$.

Since the affine relationship is only an approximation to the real world, the M observations are related through the equation $y_i = a + \mathbf{b}^\intercal\mathbf{x}_i + \epsilon_i$, where the error ϵ_i is assumed to be i.i.d. and normal. The technique for solving this problem is called regression and its goal is to minimize the error function:

$$E(a, \mathbf{b}|\{\mathbf{x}_i, y_i\}) = \sum_{i=1}^{M} (y_i - a - \mathbf{b}^\intercal\mathbf{x}_i)^2 .$$

When $M < N$, the problem is underdetermined, meaning that there is an infinite number of possible a and $\mathbf{b}$ that can yield an optimal solution. This is known as the high dimension, low sample size (HDLSS) problem [1]. As dimensionality increases, the volume of the space also increases so that data observations become sparse (the curse of sparsity). Techniques of regularization or variable selection (as LASSO, Ridge regression [2], and SPLS [3]) can be used to obtain a well-posed problem, formulated as:

$$\min_{\{a,\mathbf{b}\}} \left[E(a, \mathbf{b}|\{\mathbf{x}_i, y_i\}) + \lambda P(\mathbf{b})\right],$$

where λ is a regularization parameter and P is the penalty function for the regression coefficients $\mathbf{b}$. In general, the parameter λ cannot be determined a priori and must be inferred from the data, using some kind of cross-validation procedure. This is possible when there is an abundant number of pairs of observations $\{\mathbf{x}_i, y_i\}$ in order to feed the cross-validation procedure. However, there are situations where the number of observations is scarce, like in EEG experiments, for which each stimulus #i results in a single grand-averaged response and the number of observations M is very small.

The RoLDSIS technique: For the specific case described above, we propose a technique, called Regression on Low-Dimension Spanned Input Space, or RoLDSIS, that avoids the problem of specifying regularization parameters when the set of observations is very small. The main idea behind this technique is to assume that the neurophysiological axis $\mathbf{b}$ is restricted to the $(M-1)$-dimensional subspace spanned by the M points $\mathbf{x}_i$. This subspace is described by an origin point $\mathbf{x}_0 \in \mathbb{R}^N$ and a $N \times (M-1)$ matrix $\mathbf{V}$, which represents the orthonormal basis of the subspace spanned by the points $\mathbf{x}_i$. The points $\mathbf{x}_i$ are projected onto the spanned subspace through the transformation $\mathbf{z}_i = \mathbf{V}^\intercal(\mathbf{x}_i - \mathbf{x}_0), i = 1, \ldots, M$. The columns of the matrix $\mathbf{V}$ are the vectors that compose the orthonormal basis of the subspace spanned by the data points $\mathbf{x}_i$. Supposing that there is no coplanarity among the M points $\mathbf{x}_i$, the system of M equations $y_i = c + \mathbf{d}^\intercal\mathbf{z}_i$ can be solved exactly for the M free unknowns $c \in \mathbb{R}$ and $\mathbf{d} \in \mathbb{R}^{M-1}$. The neurophysiological axis $\mathbf{b}$, in the original $\mathbb{R}^N$ space can then be simply obtained by doing the inverse transformation $\mathbf{b} = \mathbf{V}\mathbf{d}$. The original observations $\mathbf{x}_i$ can be projected onto the neurophysiological axis $\mathbf{b}$, yielding the representations $\tilde{\mathbf{x}}_i$, whose properties can be further analyzed in order to infer the underlying brain states related to the physical or behavioral measurements y_i.

Application example: neurophysiological correlates of phonemic identification: We apply the RoLDSIS technique to an EEG study aimed at identifying the neural correlates of phonemic categorization, an essential component of speech understanding. Phonemes, the basic elements of speech (like consonants or vowels), have physical properties that largely vary according to context and across different speakers. Nevertheless, listeners seem to factor out the continuous physical variation of signals and always recognize the discrete, idealized phonemes, a phenomenon called categorization [4].

Identification task: In our experiment, we manipulated the voice onset time (VOT) and the consonant release burst creating a continuum of stimuli between the Brazilian Portuguese syllables /da/ and /ta/ [5]. VOT varied in 200 steps from -52 to 17 ms in this continuum. The stimuli created were presented in random order through earphones to eleven subjects, whose task was to identify the heard syllable in a binary forced task. The results of this experiment, for a representative subject, are shown in Fig. 1, where a psychometric, logistic curve was fitted to the subject's response. We selected the stimuli corresponding to 0%, 5%, 50%, 95%, and 100% of /ta/ responses, hereafter called stimuli #1, #2, #3, #4, and #5, respectively. Stimuli #2 and #4 are closer to stimulus #3 from the acoustic (physical) point of view, and closer to the extreme stimuli #1 and #5 from the perceptual (psychophysical) point of view.

EEG experiment: Each subject was subsequently tested in an EEG experiment, where the five selected stimuli were presented in random order, 200 times each. The subject was asked to perform the same phonemic identification task binary choosing between /da/ and /ta/. We recorded the activity of the electrodes placed at the vertex of the head (Cz) and on the mastoid bone, behind the left ear (TP9). The electric potential difference between these two electrodes is our signal of interest and should capture the underlying neuronal information used for phoneme processing and categorization [6, Chapter 12].

Event-related potentials (ERP) Cz$-$TP9 were sampled at 5 kHz, epoched by stimulus response, and baseline corrected. After removing

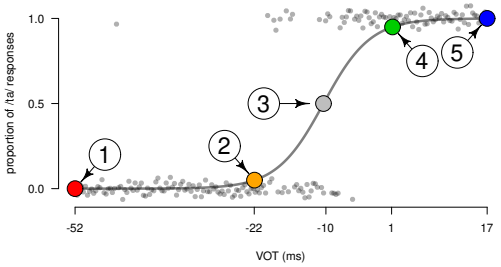

Fig. 1 *Results of the phonemic identification task for a representative subject. Responses to the 200 stimuli, each one for a specific value of voice onset time (along the horizontal axis) are shown as gray dots around 0.0 (for /da/ responses) and around 1.0 (for /ta/ responses). Vertical jitter has been added for the sake of clarity. The gray curve is the theoretical psychometric response fitted to the data. Choices of stimuli #1, #2, #3, #4, and #5, corresponding to 0%, 5%, 50%, 95% and 100% of /ta/ responses, respectively, are shown by colored dots on the psychometric curve.*

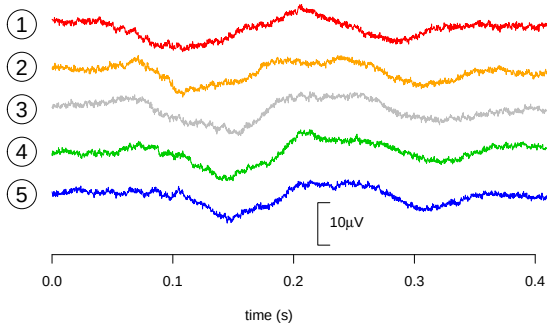

Fig. 2 *Event-related potentials for a representative subject (the same as in Fig. 1). Averaged ERPs for stimuli #1, #2, #3, #4, and #5 are shown from top to bottom. Due to the baseline correction, all signals are close to zero at $t = 0$ s. The signals are displaced vertically for the sake of visualization.*

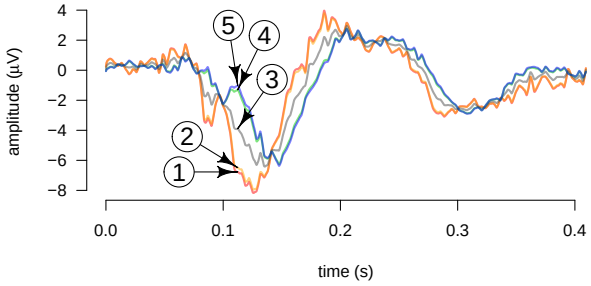

Fig. 4 *Projections of ERPs for stimuli #1, #2, #3, #4, and #5 onto the axis found by the RoLDSIS procedure, for a representative subject (the same as in Figs. 1, 2, and 3). Each projection, represented in the time domain, is drawn with a different color and indicated by the corresponding stimulus number. Note that the signals for stimuli #1 and #2 are almost identical. This also happens for stimuli #4 and #5.*

noisy trials, the remaining trials were used for subsequent processing. Fig. 2 illustrates the averaged ERPs obtained for a representative subject. The experiment was approved by the local Ethics Committee (COEP-UFMG Brazil).

Data processing: Scalograms based on the discrete wavelet transform (DWT) were used to represent electrophysiological responses in time and frequency. Only the DWT bands corresponding to low frequencies (0-156 Hz) were retained. The resulting scalograms were organized as vectors of 128 wavelet coefficients.

Application of RoLDSIS: We could use each one of these 128-dimensional DWT vectors as an observation $\mathbf{x}_i$, in order to find the linear regression that fit the values y_i, associated with the corresponding stimuli in those trails. However, as it is usual in EEG studies, the individual trials are highly noisy and it only makes sense to work with averaged signals. In the extreme case, we would average the ERPs for all the trials corresponding to each one of the five stimuli, resulting in five vectors $\mathbf{x}_i, i = 1, \ldots, 5$, each one associated with stimulus #i and the psychophysical response $y = 0.0$, 0.05, 0.5, 0.95, and 1.0, respectively.

Of course, regularization techniques, as described above (at the end of the Introduction section), cannot be applied in this case. However, it is possible to apply the RoLDSIS technique to this data set. The direction vector $\mathbf{b}$ found by RoLDSIS represents the underlying neurophysiological axis associated with the behavioral task in question. This vector $\mathbf{b}$, which lies in the same 128-dimensional space as the vector $\mathbf{x}$, can be seen as variations of the wavelet coefficients that reveals the time-frequency regions of relevance for phonemic categorization. The vector $\mathbf{b}$ can thus be visualized either as time varying function or by its scalogram in the time-frequency domain. Both representations are depicted in Fig. 3. We can see that phonemic categorization arises as earlier as 80 ms after stimulus onset time, but lasts very shortly in high-frequencies (W6) and lasts progressively longer as frequency decreases (bands W7 to V8). We can also observe some "rebounds" of categorization in bands W8 and V8, long after the stimulus has finished. These results are compatible with the literature [7].

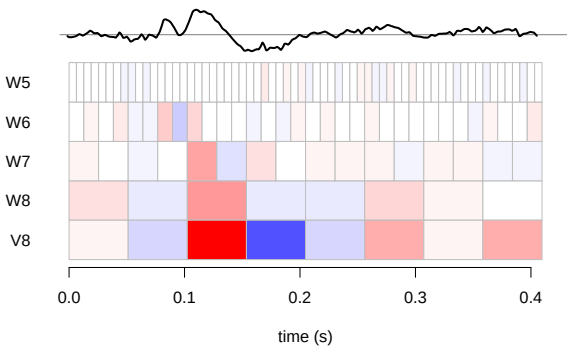

Fig. 3 *Direction obtained for the RoLDSIS procedure for a representative subject (the same as in Figs. 1 and 2). Top panel: temporal representation of the optimal direction vector, obtained by applying the inverse DWT on the RoLDSIS result. The vertical axis are in normalized units (the direction vector has unit norm). Bottom panel: scalogram (time/frequency representation) of the optimal direction vector. The intensity of the DWT coefficients are represented in a color scale, negative values in blue and positive values in red. The more saturated the color in a cell, the higher is the magnitude of the DWT coefficient represented by that cell. Frequency bands of the DWT are shown in increasing order from bottom to top (V8: 0–9.76 Hz, W8: 9.76–19.5 Hz, W7: 19.5–39.1 Hz, W6: 39.1–78.1 Hz, W5: 78.1–156 Hz).*

The original points $\mathbf{x}_i$ can be projected onto the direction defined by the regression vector $\mathbf{b}$ and, through inverse DWT, their representation in the time domain can be investigated. This is illustrated in Fig. 4. The resulting projections of stimuli #1 and #2 are almost indistinguishable, what also happens with stimuli #4 and #5. Stimulus #3 lies in the middle of the others. This is the expected pattern, since the distance between the projected responses should follow the psychometric responses $y_i = 0.0$, 0.05, 0.5, 0.95, and 1.0.

Comparison with regularized linear regression procedures: A legitimate question that may be asked at this point is how the RoLDSIS procedure compares with other regularized regression techniques. To make this comparison, we considered three popular regression techniques, namely LASSO, Ridge Regression and SPLS [2, 3] and performed k-folds cross-validation (CV) procedures, for values of k varying from 3 to 6. For a CV with k folds, we generated $5 \times k$ points by randomly partitioning the set of DWT vectors for each of the five stimuli into k sets with similar amount of trials. The DWT vectors are then averaged inside each set. Each fold contained five DWT vectors corresponding to the five stimuli. At each pass of the CV procedure, one of the folds is put apart as the test set, while the regression model is fitted to the remaining folds (the training set). In our specific implementation of the CV procedure, the $k - 1$ points, in the training set, corresponding to each stimulus were averaged, resulting in a set of five points, to which the model was fitted.

The goal of the CV procedure is to select optimal values for the regularization parameters (λ for LASSO and Ridge Regression, ζ and K for SPLS). Given a set of parameters, the model is fitted to training set and the global CV error is computed as the sum of the mean prediction squared errors (MSE) for the k test sets. Using an optimization procedure, we found the optimal values of the regularization parameters that yield the minimum value of the CV error. Note that RoLDSIS has no regularization parameter, so that the optimization procedure described above does not apply to it. This procedure was applied to each one of the eleven subjects. Fig. 5 shows the population mean CV errors for each number of folds, as well as the 95% confidence intervals of the mean estimations.

In order to assess how differently the regression techniques perform on our data, we fitted a linear mixed model to the results, considering the number of folds as a continuous fixed factor, the regression technique as a fixed discrete factor, and the subject as a random factor. The MSE values, which follow a χ^2 distribution, were transformed to normal [8] and the resulting values were used as the dependent variable of the linear model. The results show a significant increase in MSE with the number of folds ($F[1, 43] = 32.2, p < 0.001$), but there was no effect related to the regression techniques, even though multiple comparison analysis show reliable differences between SPLS and the others ($p < 0.001$). This means that RoLDSIS has an equivalent performance as the traditional regularized regression techniques, with the big advantage that it does not have regularization parameters.

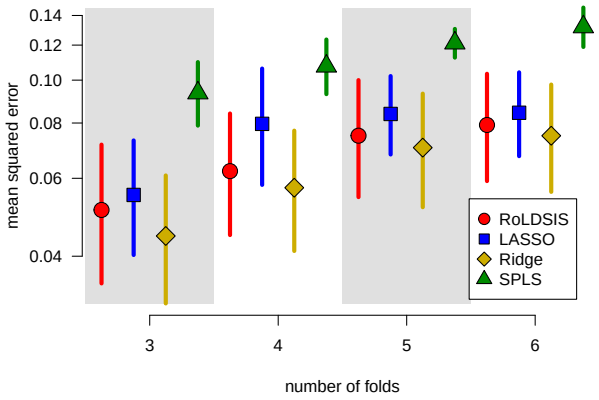

Fig. 5 *Cross-validation errors for the proposed regression method (RoLDSIS) and the methods of regularized linear regression. Results for 3, 4, 5 and 6 folds are shown. The mean squared errors for the test set of the CV are shown with dots. Confidence intervals at 95% are represented by vertical bars.*

Finally, we assess the differences among the regression techniques, in terms of the neurophysiological axis obtained by the regression. To do that, we took the neurophysiological direction, as determined by the vector of regression coefficients $\mathbf{b}$, for each subject and each regression technique, and normalize them (i.e. computed the vectors $\hat{\mathbf{b}} = \mathbf{b}/||\mathbf{b}||$. For the RoLDSIS technique, we used the regression coefficients obtained for the reduced set of five DWT vectors, while for the other techniques, we used the result of the 3-fold CV. For each regression technique, the neurophysiological directions for all subjects were combined by computing the root mean square (RMS) for each DWT component, across the population. These RMS values for the four regression techniques are shown in Fig. 6 in the form of time-frequency scalograms. The darker a DWT component appears in the scalogram, the more frequently it appears in the neurophysiological direction found by the regression across the population. We can notice that the RoLDSIS procedure yields results similar to the Ridge Regression technique. The main difference is that Ridge's solutions are more dispersed in the scalogram, whereas they appear more concentrated in the regions of the time-frequency space that should be related to the phonemic categorization process (see discussion above in section "Application of RoLDSIS"). The scalograms for LASSO looks like a chopped version of the scalogram for RoLDSIS. At least, LASSO seem to be capturing the important aspects of the neurophysiological axis, but since it also does feature selection, besides regularizing the regression, fewer DWT components appear in the scalogram. On the other hand, the SPLS technique produces a very dispersed scalogram, even though it captures the relevant components between 0.1 and 0.3 s in the W8 and V8 bands.

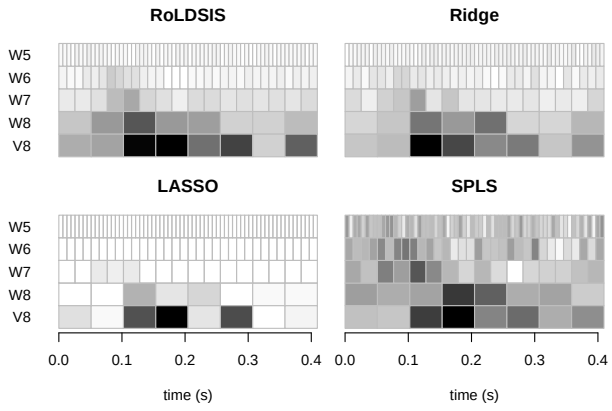

Fig. 6 *Scalograms of the regression results. The scalogram for the root mean squared regression coefficients for each component of the DWT, across the population, are shown for the proposed regression method (RoLDSIS) and the methods of regularized linear regression. Shades of gray represent the cumulative RMS (white for zero and black for the maximum value). Frequency bands of the DWT are the same as those in Fig. 3.*

It must be emphasized that the results shown in Fig 6 for the RoLDSIS technique are consistent with the literature on neurophysiological correlates of speech categorization and processing of theta oscillations (our V8 DWT band), beta oscillations (W8 and W7 bands) and gamma oscillations (W6 and W5 bands) [7, 9, 10].

Conclusion: In this paper, we propose a novel regression technique, called RoLDSIS, that addresses the HDLSS problem. Other popular regularized regression techniques need the specification of regularization parameters and, consequently, enough data points for running reliable cross-validation procedures. In contrast, RoLDSIS assumes that the regression solution will be embedded in the subspace spanned by the data points. This allows an exact solution for problem, even when the number of data points is extremely small. In particular, this new technique may be useful for EEG experiments, where ERPs must be averaged for many repetitions of a small number of presented stimulus. We applied RoLDSIS to the analysis of an EEG experiment data, that aimed at finding the neurophysiological correlates of phonemic categorization. The obtained results, by regressing the wavelet-transformed ERPs against the psychophysical responses of eleven subjects, showed relevant characteristics of speech perception in the time-frequency domain. Comparison with LASSO, SPLS and Ridge Regressions were also presented, showing that RoLDSIS is a suitable alternative for the processing of neurophysiological signals.

Acknowledgments: The authors thank Ludovic Bellier for insightful comments on a previous version of this work. This study was funded by grants from IXXI (Institut Rhônalpin des Systèmes Complexes), PEPS Grenoble-CNRS, CNPq, and FAPEMIG (APQ-03701-16). Support from CAPES and Universidade Federal de Ouro Preto is also acknowledged. There are no conflicts of interest.

Code availability: The data used in the present paper, as well as the code in R used for the analysis and the production of the figures in this paper is available at `https://github.com/RoLDSIS`.

A. C. Santana (Dep. Control and Automation Engineering, Universidade Federal de Ouro Preto, Ouro Preto-MG, Brazil)
✉ E-mail: adrielle@ufop.edu.br
A. V. Barbosa, H. C. Yehia and A. C. Santana (Grad. Prog. Elec. Eng., Universidade Federal de Minas Gerais, Belo Horizonte-MG, Brazil)
✉ E-mail: hani@cpdee.ufmg.br ✉ E-mail: vilela@cpdee.ufmg.br
R. Laboissière and A. C. Santana (Laboratory of Psychology and NeuroCognition, CNRS & Université Grenoble Alpes, Saint-Martin-d'Hères, France)
✉ E-mail: rafael.laboissiere@univ-grenoble-alpes.fr

References

1 Hall, P., Marron, J.S. and Neeman, A.: 'Geometric representation of high dimension, low sample size data', *J Roy Stat Soc B*, 2005, **67**, (3), pp. 427–444
2 Friedman, J., Hastie, T. and Tibshirani, R.: 'Regularization paths for generalized linear models via coordinate descent', *J Stat Softw*, 2010, **33**, (1)
3 Chun, H. and Keleş, S.: 'Sparse partial least squares regression for simultaneous dimension reduction and variable selection', *J R Stat Soc Series B Stat Methodol*, 2010, **72**, (1), pp. 3–25
4 Simon, C. and Fourcin, A.J.: 'Crosslanguage study of speechpattern learning', *J Acoust Soc Am*, 1978, **63**, (3), pp. 925–935
5 Wood, C.C.: 'Discriminability, response bias, and phoneme categories in discrimination of voice onset time', *J Acoust Soc Am*, 1976, **60**, (6), pp. 1381–1389
6 Hall, J.W.: 'New handbook of auditory evoked responses'. (Boston, Massachusetts, USA: Pearson, 2007)
7 Bouton, S., Chambon, V., Tyrand, R., Guggisberg, A.G., Seeck, M., Karkar, S., et al.: 'Focal versus distributed temporal cortex activity for speech sound category assignment', *Proc Natl Acad Sci USA*, 2018, **115**, (6), pp. E1299–E1308
8 Hawkins, D.M. and Wixley, R.A.J.: 'A note on the transformation of chi-squared variables to normality', *Am Stat*, 1986, **40**, (4), pp. 296
9 Giraud, A.L. and Poeppel, D.: 'Cortical oscillations and speech processing: emerging computational principles and operations', *Nat Neurosci*, 2012, **15**, (4), pp. 511
10 Bidelman, G.M.: 'Induced neural beta oscillations predict categorical speech perception abilities', *Brain Lang*, 2015, **141**, pp. 62–69