# Spectral Graph Wavelet Transform as Feature Extractor for Machine Learning in Neuroimaging

Yusuf Yigit Pilavci, Nicolas Farrugia

# SPECTRAL GRAPH WAVELET TRANSFORM AS FEATURE EXTRACTOR FOR MACHINE LEARNING IN NEUROIMAGING

*Yusuf Yigit Pilavci [a,b]*

*Nicolas Farrugia [b]*

[a]Politecnico di Milano
Computer Science and Engineering
Milano, Italy

[b]Lab-STICC - IMT Atlantique
Electronics Department
Brest, France

## ABSTRACT

Graph Signal Processing has become a very useful framework for signal operations and representations defined on irregular domains. Exploiting transformations that are defined on graph models can be highly beneficial when the graph encodes relationships between signals. In this work, we present the benefits of using Spectral Graph Wavelet Transform (SGWT) as a feature extractor for machine learning on brain graphs. First, we consider a synthetic regression problem in which the smooth graph signals are generated as input with additive noise, and the target is derived from the input without noise. This enables us to optimize the spectrum coverage using different wavelet shapes. Finally, we present the benefits obtained by SGWT on a functional Magnetic Resonance Imaging (fMRI) open dataset on human subjects, with several graphs and wavelet shapes, by demonstrating significant performance improvements compared to the state of the art.

***Index Terms***— graph signal processing, wavelets, neuroimaging, regression

## 1. INTRODUCTION

The emergence of Graph Signal Processing (GSP) is mostly due to the elegant and powerful analogy between graph laplacian eigenvectors and classical Fourier analysis [1]. As such, application domains involving graphs to model mutltivariate dependencies are naturally adapted to the GSP framework. In particular, recent studies in neuroimaging have leveraged the use of graph theory to study brain networks, giving rise to the field of network neuroscience [2]. However, the majority of network neuroscience studies have focused on analyzing the properties of the graph itself (e.g. graph theoretical descriptors of white matter connectivity [2]), rather than signals of brain activity on a brain graph. In turn, most studies analyzing brain signals relies on the use of massively univariate statistics, analyzing each brain region independently [3]. Recent methodological efforts push towards the use of machine learning as multivariate methods able to capture whole brain dependencies of neural activity [4].

Our aim with the present paper is to demonstrate the potential of simultaneously using brain connectivity and signals from neuroimaging data to boost performance of machine learning tasks, by leveraging the GSP framework. In particular, we aim at testing the benefits of using SGWT as a feature extractor for machine learning on neuroimaging data.

The remainder of the paper is organized as follows. In section 2 we will present previous studies related to SGWT and GSP applied to machine learning in neuroimaging. In section 3 we will present our experimental setup, the formulation of SGWT that will be used in this paper and the procedures used to build our graph models. In section 4 we present experiments and results on a synthetic regression problem and fMRI signals. Finally, we conclude in section 5.

## 2. RELATED WORK

With the development of GSP, frameworks based on wavelets on graphs have become promising, due to their over-complete analysis in both localization on graph nodes and different scales with respect to the spectrum of the graph. To obtain such analysis on irregular domains such as graphs, several formulations of graph wavelets have been proposed [5, 6, 7]. For example, Narang reports a compact comparison of several formulations in terms of significant properties such as perfect reconstruction or orthogonality [5]. In this paper, we use Spectral Graph Wavelet Transform (SGWT) introduced in [7]. Basically, SGWT exploits the Fourier transform analogy that is defined on graphs [1] and defines wavelet kernel functions in the spectral domain.

While there is a growing literature applying GSP to neuroscientific questions, (for a review, see [8]), most studies use GSP to derive descriptors (such as alignement of functional signals to the underlying graph [9]) that are further analyzed using inference based statistics (e.g. [10]). For example, Leonardi [11] shows the statistical relevance of SGWT coefficients in different scales with varying experimental conditions in an fMRI paradigm.

However, studies applying GSP for machine learning in neuroimaging are relatively scarce. In [12], the authors de-

fined a low rank, dimensionality reduction approximation and recovers the underlying graph, showing performance improvements when using the learnt approximation as features for supervised classification. In [13], the authors used GSP as dimensionality reduction and feature extraction in a supervised learning setting with fMRI. The authors show performance improvements in both simulated cases and real data. A recent contributions uses GSP to extract features for brain computer interfaces based of near-infrared spectroscopy brain signals [14], showing significant performance improvements compared to previous work on the same dataset. Nevertheless, to the best of our knowledge, there is no published study exploiting SGWT as feature extractor for machine learning in neuroimaging.

## 3. METHODS

### 3.1. Graph Wavelet Transform

Using graph wavelet transform enables multi-scale signal representations adapted to the underlying graphs, which in turn facilitate the detection of abnormal changes or discontinuities in the original domain, and ease the interpretation of signals in both localization and frequency. With these motivations, we adopt the formulation of SGWT [7].

As in classical signal processing, wavelet functions are defined as $\psi_{s,a}(x)$ for different scales $s$, and translations $a$. Graph Wavelets can be also interpreted in the Fourier domain as:

$$\psi_{s,a}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{\psi}(s\omega)e^{-j\omega a}e^{j\omega x}d\omega \qquad (1)$$

where $\hat{\psi}(s\omega)$ is the Fourier transform of the scaled wavelet, $e^{-j\omega a}$ is the Fourier transform of the spatial translation, (which can be seen as the Fourier Transform of a delta localized at $a$) and $e^{j\omega x}$ is the Fourier basis function. These terms will be addressed to corresponding terms in graph wavelet transform definition.

Let an undirected, weighted graph $G$, with a vertex set $V$ such as $|V| = N$, and weight matrix $\mathbf{W} \in \mathbb{R}^{\mathbb{N}}$. $\mathbf{W}$ is a symmetric matrix of weights $w_{ij}$. The Laplacian operator of $G$ is defined as follows:

$$L = D - W \qquad (2)$$

where $D$ is a diagonal matrix such as $D_{ii} = \sum_j |w_{ij}|$. Here, weights $w_{ij}$ are allowed to be negative, and we deal with issues related to semi-positiveness of the $L$ by using the absolute diagonal degree matrix [11]. Also, the normalized Laplacian is defined as $L_{norm} = D^{-1/2}LD^{-1/2}$. Analogous to classical signal processing, eigenvectors $\mathbf{U}$ and eigenvalues of the Laplacian matrix $\mathbf{L}$ correspond to the Fourier basis and frequency values, respectively. A graph signal $f$ is a vector in $\mathbb{R}^{\mathbb{N}}$ defined in the vertex domain [1], and spectral filtering operations $f$ are done through Graph Fourier Transform, defined by $\hat{f} = \mathbf{U}^T f$.

By interpreting scaling and translation operations, as in equation 1, it is possible to define graph wavelets as follows:

$$\psi_{s,a} = \sum_{n=1}^{N} g(s\lambda_n)\hat{\delta}(n)u_n \qquad (3)$$

where $g$ is a band-pass kernel defined in the spectral domain, corresponding to the Fourier transform of the wavelet at scale $s$ in equation 1, and $\hat{\delta}(n)$ is the Graph Fourier transform of $\delta$ localized at $a$, corresponding to the spatial translation component in equation 1. Finally, $u_n$ are the columns of $\mathbf{U}$, eigenvectors of $\mathbf{L}$. The complete frame for the wavelet transform is computed by adding a low-pass filter, $h(\lambda)$, also called the scaling function. As a result, the obtained transform covers all parts of the graph spectrum. The final transformation for a graph signal $f$ is the following inner product:

$$W_f(s,a) = <\psi_{s,a}, f> \qquad (4)$$

### 3.2. Graph Wavelet Kernel Design

The informative property of SGWT as signal representation is highly related to the selection of the scaling and wavelet functions, $h(\lambda)$ and $g(\lambda)$. Such a choice impacts the stability of reconstruction of the original graph signal. As reported in [7] and [15], a wavelet frame would be tight if the sum of squares of all kernels remains constant through the spectrum (Parseval identity), a necessary condition for perfect reconstruction of a signal. However, it is possible to relax this constraint by accepting some variation, resulting in more freedom of kernel selection (e.g. using cubic splines). Another aspect of kernel design is spectrum coverage. For a graph model, the spectrum is defined by the eigenvalues of $\mathbf{L}$, and any continuous function that is defined in the spectral domain (as a function of $\lambda$) is only evaluated at those eigenvalues, $\lambda_n$, [15]. Therefore, one should examine the eigenvalue locations through the spectrum when designing kernel functions.

Shuman *et al.* provides tight and spectrum adapted wavelet kernels, called Warped Translates. In [15], a tight frame is generated with a special function:

$$q(\lambda) := \sum_{k=0}^{K} a_k \cos\left(2\pi k(\omega(\lambda) - \frac{1}{2})\right)\mathbf{1}_{0 \leq \lambda < 1} \qquad (5)$$

and its translated versions in the graph spectrum, $q\left(\omega(\lambda) - \frac{m}{R}\right)$ where $\sum_{k=1}^{K}(-1)^k a_k = 0$, and $\omega$ is a non-decreasing warping function that modifies the kernels' behaviour on the spectrum. For example, choosing a warping function as the approximated cumulative distribution function of eigenvalues concentrates the kernels in ranges in which the eigenvalues are densely placed (see Fig. 1). In [15], the superior discriminatory power of warped translates is clearly demonstrated, because different parts of the evaluated spectrum are perfectly covered and segmented by different kernels.
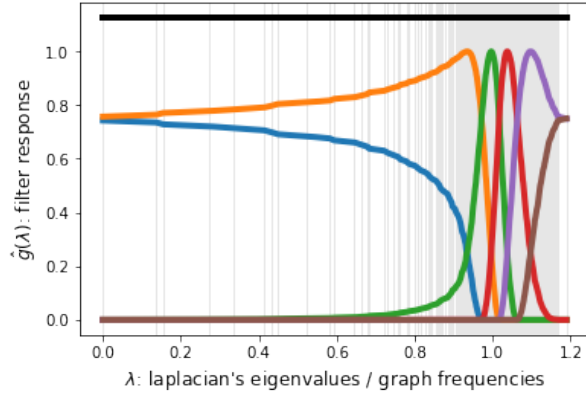
**Fig. 1**. Warped Translate Wavelet Kernels on Spectral Domain, in the case of the KNN-Correlation Brain Graph. Vertical lines depict placement of eigenvalues.

### 3.3. Synthetic Graph Signals and Regression Problem

In this work, synthetic signals are generated for getting more understanding on the behavior of SGWT in a regression problem, in the case where the signals are smooth on the graph. The workflow that generates the input and output starts with creating a Erdos Renyi graph (with a probability of edge presence of $p = 0.1$), $G$ with $N$ nodes and associated weight matrix $\mathbf{W}$. A diffusion operator is computed using a lazy random walk, $\mathbf{A} = (I + D^{-1}W)$. Secondly, we generate a random signal matrix $\mathbf{R}$ of size $M \times N$, where $M$ is number of samples and $N$ is number of features, and the generated set of signals are diffused by multiplying it by the diffusion operator, $\mathbf{A}$, such as the diffused signals are $\hat{\mathbf{R}} = \mathbf{RA}$. The third step is to generate a random vector of regression weights $\beta$, and compute the nonlinear function output as $y = \log(\beta^T \hat{R})$. Finally, zero-mean Gaussian noise is added on the graph signals, so the final observed signals are $\mathbf{X} = \hat{\mathbf{R}} + \mathcal{N}(\mathbf{0}, \sigma^{\mathbf{2}})$, with $\sigma = 0.1$. With this procedure, we are able to obtain graph signals, and also generate a problem that cannot be easily solved by linear regressors. Those signals are smooth on the graph likewise the FMRI signals on the brain nodes.

### 3.4. fMRI Datasets

We consider here an open dataset of fMRI data on human subjects who rated pictures with emotional content [16]. We fetched statistical maps of whole brain activity during single trials for each subject, from neurovault [17] (collection number 1964). The supervised learning task consists in predicting the rating given by the subject from brain maps. As this dataset didn't include connectivity data that could be used to compute subject-specific graphs, we estimated average brain graphs from resting state fMRI data (i.e. spontaneous fluctuations of the brain at rest) from 158 subjects of an open resting-state dataset [18]. We used

the preprocessed resting-state data, described here `https://neuroanatomyandconnectivity.github.io/opendata/`. As both datasets were spatially normalized in the standard MNI space, we defined regions of interest (ROI) to enable the correspondance between the graph, defined on the resting-state dataset, and the signals from the emotional rating dataset. Therefore, we parcellated all resting-state time-series and brain maps on 523 non-overlapping ROIs from the finest scale of BASC atlas (444 networks) [19].

### 3.5. Graph Construction

We explain here the different graph inference methods applied on the parcellated resting state time-series. The weight matrix, $\mathbf{W}$ formulations are as follows:

- Functional Connectivity with Covariance or Correlation Graph: We derive graphs from temporal covariance and correlation. As variant of these graphs, thresholded and binary versions of these adjacency matrices are also defined.

- KNN Covariance/Correlation Graph: Another variant to previous graphs is generated by limiting the numbers of neighbors per node. For each node, the K strongest edges are kept and rest are set to a weight of 0. For the symmetry, when $\mathbf{W}_{ij}$ is set with nonzero edge, same value is kept in $\mathbf{W}_{ji}$. Binary versions are also generated by setting nonzero values to 1.

- Semi-Local Graph: Another strategy that exploits both the geometrical structure and functional connectivity is the semi-local graph, as tested in [13]. The semi-local graph uses a threshold (set to keep the graph connected) on the euclidean distance between baricenters of nodes, and sets connection weights corresponding to long distances to 0.

- Kalofolias Graph: We used the method from Kalofolias [20], which relies on defining a smoothness prior to infer the graph from the data.

### 3.6. Dimensionality Reduction and Regression

For the synthetic dataset, generated inputs $X$ are first transformed using SGWT, followed by dimensionality reduction using 100-best feature selection. Performances of linear regression predicting $y$ are compared with and without SGWT for different kernels. For the fMRI dataset, we reproduced the methods from the original paper [16] to be able to compare results. A PCA with 121 components is used for dimensionality reduction, and regression is performed with Lasso. Careful cross-validation (CV) is performed over subjects and regularization parameter for LASSO, using leave-one-subject-out, enabling reliable generalization. We compared the original pipeline that uses no parcellation (original signal), parcellated signals on BASC ROI, and SGWT representations.

| Data | CV Results | | | Test Results | | | Transform Properties | |
|------|------|------|------|------|------|------|------|------|
| Repr. | RMSE | $R^2$-score | Pearson | RMSE | $R^2$-score | Pearson | Graph | Kernel |
| Original | 1.080 ±0.028 | 0.204 ±0.073 | 0.810 ±0.018 | 1.036 | 0.451 | 0.693 | NA | NA |
| ROI | 1.054 ±0.031 | 0.231 ±0.079 | 0.827±0.017 | 1.022 | 0.466 | 0.701 | NA | NA |
| SGWT-1 | 1.047 ±0.029 | 0.234±0.082 | 0.832 ±0.015 | **0.983** | **0.506** | 0.725 | Corr. | Warped |
| SGWT-2 | 1.067±0.030 | 0.205 ±0.085 | 0.840 ±0.014 | 0.987 | 0.502 | **0.737** | KNN Corr. | Warped |
| SGWT-3 | 1.056±0.034 | 0.205±0.085 | 0.839±0.014 | 0.988 | 0.500 | 0.722 | KNN Cov | Meyer |
| SGWT-4 | **1.033 ±0.030** | **0.245 ±0.082** | 0.828±0.016 | 0.990 | 0.498 | 0.720 | Kalofolias | Cubic Spline |
| SGWT-5 | 1.062 ±0.031 | 0.196 ±0.107 | **0.843 ±0.014** | 0.991 | 0.487 | 0.730 | KNN Corr. | Iter. Sinus |

**Table 1**. Results for the fMRI Dataset

| Wavelet Kernels | MSE | $R^2$-score |
|------|------|------|
| Cubic Spline | 1514.45e-05 | 0.392438 |
| Meyer | 1518.24e-05 | 0.390937 |
| Iterated sinusoidal | 1515.33e-05 | 0.392079 |
| **Warped Translate** | **1507.40e-05** | **0.395246** |
| No Wavelet | 1533.48e-05 | 0.384859 |

**Table 2**. Results for the Synthetic dataset

## 4. RESULTS

For the evaluation of results on the synthetic dataset, $R^2$-score and Mean Square Error (MSE) are used, whereas for the fMRI dataset, Root mean square error (RMSE) and Pearson correlation are employed , to be consistent with [16]. Variability in generalization is reported using the average and the standard error of the mean (SE) over CV folds for each score. In addition, we report results on a generalization test set using a predefined, separate hold-out test dataset as explained in [16].

Table 1 presents results obtained for the fMRI dataset. Result for SGWT are picked from the set of best results for different graphs and wavelet function. In Table 1, the best five results for SGWT are reported and these scores stands for the results of different graph and kernel selections. SGWT provides significant performance gains when compared to using only parcellated ROI signals, or original signals with no parcellation. In particular, warped translate have a better potential for generalization to the test dataset. We suggest that SGWT enables an efficient exploitation of underlying multivariate dependencies, using spectrum-adapted wavelet kernels on a brain graph. While the magnitude of performance gain is strongly dependent to underlying information inferred by the graph, warped translates are able to extract more informative features in terms of regression problems.

Comparisons for the synthetic dataset are reported in Table 2. The experiment for this analysis is repeated 500 times with randomly generated graphs with 500 nodes, and average scores are interpreted. As indicated by Table 2, SGWT, and in particular warped translate kernels are able to generate features that are significantly more informative to solve the regression problem. This indicates that the regression problem
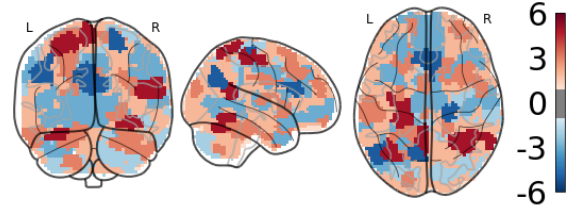


**Fig. 2**. Significant Scale-Localization map on brain. Positive and negative weights are denoted with ± sign.

considerably benefits from using discriminative, spectrum-adaptive coverage of kernels w.r.t. others.

We depict in Fig. 2 the different scale and spatial localization of the largest LASSO coefficients that solve the regression problem. Conceptually, this representation shows the most important graph scales for each localization in brain. Negative and positive weights are separately examined and the largest contribution for each node is denoted with its sign and scale in Fig.2.

## 5. CONCLUSION

In this contribution, we have demonstrated the potential of combining SGWT and machine learning using synthetic data and fMRI data from open datasets. A key point of the proposed approach is to rely on Warped Translated kernels for wavelet definitions, which optimizes spectral coverage of SGWT. We showed that using features from SGWT can boost performance in a challenging regression task on neuroimaging data. In future work, we plan to better examine how the estimated features can be used to enhance interpretability of the trained models. Another perspective is to define new models based on dynamic graphs in order to fully exploit the temporal dimension of brain activity, instead of relying solely on spatial maps as graph signals.

## 6. REFERENCES

[1] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 83–98, 2013.

[2] Danielle S Bassett and Olaf Sporns, "Network neuroscience," *Nature neuroscience*, vol. 20, no. 3, pp. 353, 2017.

[3] Karl J Friston, "Statistical parametric mapping.," 1994.

[4] Gael Varoquaux and Bertrand Thirion, "How machine learning is shaping cognitive neuroimaging," *GigaScience*, vol. 3, no. 1, pp. 28, 2014.

[5] Sunil K Narang and Antonio Ortega, "Perfect reconstruction two-channel wavelet filter banks for graph structured data," *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 2786–2799, 2012.

[6] Mark Crovella and Eric Kolaczyk, "Graph wavelets for spatial traffic analysis," in *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*. IEEE, 2003, vol. 3, pp. 1848–1857.

[7] David K. Hammond, Pierre Vandergheynst, and Rmi Gribonval, "Wavelets on graphs via spectral graph theory," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 129 – 150, 2011.

[8] Weiyu Huang, Thomas AW Bolton, Alejandro Ribeiro, J.D. Medaglia, D.S. Bassett, and Dimitri Van De Ville, "A graph signal processing view on functional brain imaging," *arXiv preprint arXiv:1710.01135*, 2017.

[9] John D Medaglia, Weiyu Huang, Elisabeth A Karuza, Apoorva Kelkar, Sharon L Thompson-Schill, Alejandro Ribeiro, and Danielle S Bassett, "Functional alignment with anatomical networks is associated with cognitive flexibility," *Nature Human Behaviour*, vol. 2, no. 2, pp. 156, 2018.

[10] Keith Smith, Benjamin Ricaud, Nauman Shahid, Stephen Rhodes, John M. Starr, Agustn Ibanez, Mario A. Parra, Javier Escudero, and Pierre Vandergheynst, "Locating temporal functional dynamics of visual short-term memory binding using graph modular dirichlet energy," *Scientific Reports*, vol. 7, 2017.

[11] Nora Leonardi and Dimitri Van De Ville, "Wavelet frames on graphs defined by fmri functional connectivity," in *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*. IEEE, 2011, pp. 2136–2139.

[12] Liu Rui, Hossein Nejati, Seyed Hamid Safavi, and Ngai-Man Cheung, "Simultaneous low-rank component and graph estimation for high-dimensional graph signals: Application to brain imaging," in *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*. IEEE, 2017, pp. 4134–4138.

[13] Mathilde Ménoret, Nicolas Farrugia, Bastien Pasdeloup, and Vincent Gripon, "Evaluating graph signal processing for neuroimaging through classification and dimensionality reduction," *arXiv preprint arXiv:1703.01842*, 2017.

[14] Panagiotis C Petrantonakis and Ioannis Kompatsiaris, "Single-trial nirs data classification for brain–computer interfaces using graph signal processing," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 9, pp. 1700–1709, 2018.

[15] David I. Shuman, Christoph Wiesmeyr, Nicki Holighaus, and Pierre Vandergheynst, "Spectrum-adapted tight graph wavelet and vertex-frequency frames," *IEEE Transactions on Signal Processing*, vol. 63, pp. 4223–4235, 2015.

[16] Luke J Chang, Peter J Gianaros, Stephen B Manuck, Anjali Krishnan, and Tor D Wager, "A sensitive and specific neural signature for picture-induced negative affect," *PLoS biology*, vol. 13, no. 6, pp. e1002180, 2015.

[17] Krzysztof J Gorgolewski, Gael Varoquaux, Gabriel Rivera, Yannick Schwarz, Satrajit S Ghosh, Camille Maumet, Vanessa V Sochat, Thomas E Nichols, Russell A Poldrack, Jean-Baptiste Poline, et al., "Neurovault. org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain," *Frontiers in neuroinformatics*, vol. 9, pp. 8, 2015.

[18] Natacha Mendes, Sabine Oligschlaeger, Mark E Lauckner, Johannes Golchert, Julia M Huntenburg, Marcel Falkiewicz, Melissa Ellamil, Sarah Krause, Blazej M Baczkowski, Roberto Cozatl, et al., "A functional connectome phenotyping dataset including cognitive state and personality measures," *bioRxiv*, p. 164764, 2017.

[19] Pierre Bellec, Pedro Rosa-Neto, Oliver C. Lyttelton, Habib Benali, and Alan C. Evans, "Multi-level bootstrap analysis of stable clusters in resting-state fMRI," *NeuroImage*, vol. 51, no. 3, pp. 1126–1139, 2010.

[20] Vassilis Kalofolias, "How to learn a graph from smooth signals," in *The 19th International Conference on Artificial Intelligence and Statistics (AISTATS 2016)*. Journal of Machine Learning Research (JMLR), 2016.