Comparing Spatial and Spectral Graph Filtering for Preprocessing Neurophysiological Signals

Stephan Goerttler[1,2], Min Wu[2], Fei He[1], Daniel Blackburn[3], Ptolemaios Sarrigiannis[4]

[1] Centre for Comp. Science and Math. Modelling, Coventry University, UK

[2] Inst. for Infocomm Research, A*STAR, Singapore

[3] Department of Neuroscience, University of Sheffield, UK

[4] Department of Neurophysiology,
Royal Devon and Exeter NHS Foundation Trust, UK

IEEE SPMB, December 2, 2023

Setting

- Electroencephalography (EEG) recording:
  - Multivariate signal (1)
    - Multiple sensors
    - Each sensor acquires time series
  - Data matrix shape:
    # sensors $\times$ # time samples
- Graph filtering (2; 3)
  - Analogous to filtering in signal processing
  - *Graph*:
    - Encodes connectivity in the data
    - E.g. functional connectivity:
      Pairwise Pearson correlation
  - Filters are defined in terms of graph
  - Applications of graph filtering:
    - Graph denoising (4; 5)
    - Remove correlations
    - Graph filter layer in Graph Neural
      Network (6; 7)

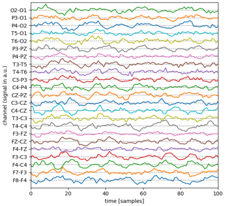


Figure: Multivariate EEG signal

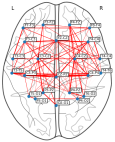
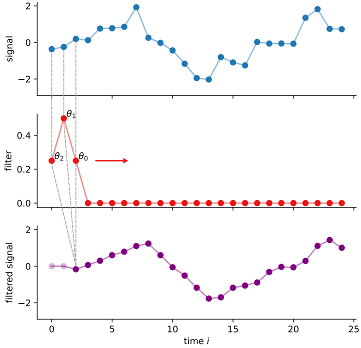


Figure: EEG spatial structure: correlations between channels

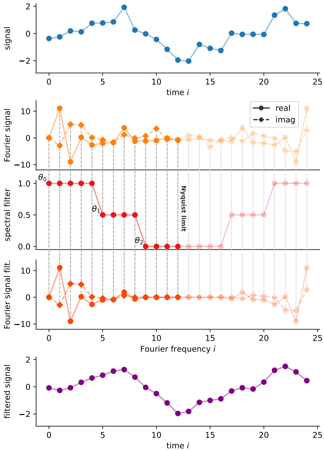
Finite Impulse Response (FIR) filter

- Signal is filtered by convolving signal with (localised) filter $F = [\theta_0, \theta_1, ..., \theta_{N_t-1}]$:
  $\mathbf{x}_{filt} = F * \mathbf{x}$
- Filter with number of parameters $k = 3$:
  $F = [\theta_0, \theta_1, \theta_2]$
- How to deal with boundaries?
  - no padding, padding, **cyclic**, ...
- Filter as matrix: using shift matrix $\mathbf{S}_L$:
  $$\mathbf{F} = \theta_0 \mathbf{1} + \theta_1 \mathbf{S}_L + \theta_2 \mathbf{S}_L^2,$$
  $$\mathbf{x}_{filt} = \mathbf{F}\mathbf{x}$$
- Frequency formulation of FIR filter: Fourier filter

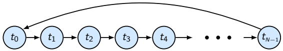
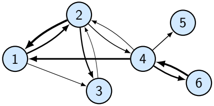


Fourier filter

- Time signal $\mathbf{x}$ is firstly transformed to Fourier domain: $\mathbf{x} \rightarrow \widetilde{\mathbf{x}}$
  - Note: Fourier signal $\widetilde{\mathbf{x}}$ is complex (real and imaginary part)
  - Frequencies past the Nyquist limit mirror lower frequencies
  - Highest frequency is at Nyquist limit
- Fourier signal multiplied with spectral filter $F = [\theta_0, \theta_1, ..., \theta_{N_t-1}]$
  - Alternative: filter with $k = 3 < N_t$ parameters: $F = [\theta_0, \theta_0, ..., \theta_1, ..., \theta_2, \theta_2, ..., \theta_1, ..., \theta_0]$
  - Filtered Fourier signal transformed to time domain: $F \odot \widetilde{\mathbf{x}} \rightarrow \mathbf{x}_{filt}$
- Matrix notation (with discrete Fourier transform matrix $\mathbf{W}$): $\mathbf{x}_{filt} = \mathbf{W}^{-1} diag(F)\,\mathbf{W}\,\mathbf{x}$

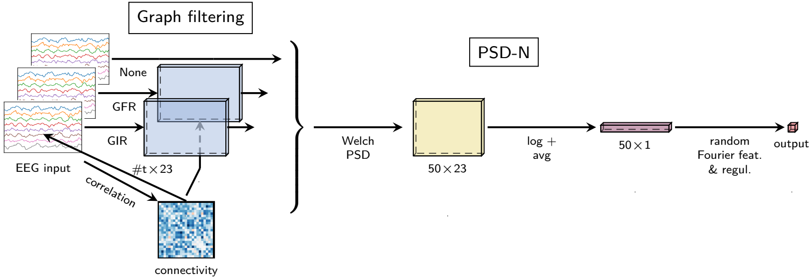


Analogy classical filtering - graph filtering

	time filtering	graph filtering (2; 3)
graph		
connect. matrix	$\mathbf{A}_{cyc} = \begin{bmatrix} 0 & 0 & 0 & \dots & 1 \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & & \dots & 1 & 0 \end{bmatrix}$	$\mathbf{A} = \begin{bmatrix} 0 & 2 & 0 & \dots & 0 \\ 1.5 & 0 & 0.5 & \dots & 0 \\ 1 & 0.5 & 0 & \dots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & & \dots & 0 & 0 \end{bmatrix}$
FIR filter (k=3)	$\theta_0 \mathbf{1} + \theta_1 \mathbf{A}_{cyc} + \theta_2 \mathbf{A}_{cyc}^2$	$\theta_0 \mathbf{1} + \theta_1 \mathbf{A} + \theta_2 \mathbf{A}^2$
eigendecompos.	$\mathbf{A}_{cyc} = \left(\mathbf{W}^{-1}\right) \Lambda_{cyc} \mathbf{W}$	$\mathbf{A} = \left(\mathbf{W}_{GFT}^{-1}\right) \Lambda \, \mathbf{W}_{GFT}$
spectral filter	$\mathbf{W}^{-1} \begin{bmatrix} \theta_0 & 0 & \dots & 0 \\ 0 & \theta_1 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & & \dots & \theta_{N-1} \end{bmatrix} \mathbf{W}$	$\mathbf{W}_{GFT}^{-1} \begin{bmatrix} \theta_0 & 0 & \dots & 0 \\ 0 & \theta_1 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & & \dots & \theta_{N-1} \end{bmatrix} \mathbf{W}_{GFT}$

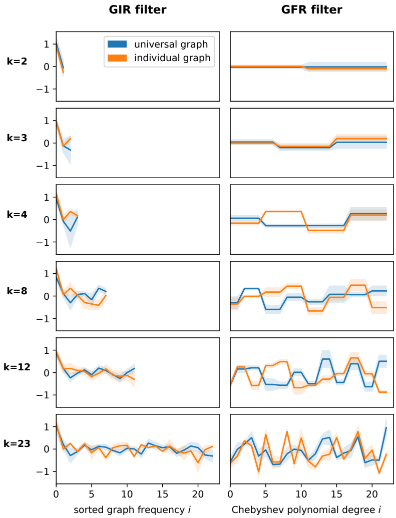
Graph filter preprocessing

- Task: EEG Alzheimer's disease classification
- Use neural network to train filter coefficients!
- Base graph:
  - Pairwise Pearson correlation
  - Universal or individual
- (Trainable) graph filtering
  - GFR filter ("graph frequency response", Fourier filter)
  - GIR filter
- Extract features (power spectral densities)
- Classifier network (random fourier features layer (8), SVM-like)



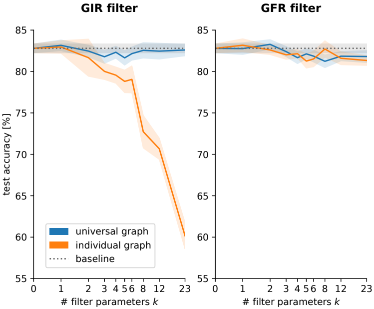
Results - filter shape

- Test two filters: GIR and GFR
- Use universal graph (blue) and individual graph (orange)
- Vary # filter coefficients
- Run each configuration 30 times (3 sample sizes × 10 repeats)
- Results:
  - same filters learned across repeats
  - same filters learned even if # coeffs. varied
- Do trained filters generalise to unseen data?



Results - performance

- GIR filter:
  - Universal graph
    - Test accuracy constant, below baseline
  - Individual graph
    - sharply decreases with # parameters
  - Explanation:
    - E.g. GIR filter coefficient θ_{17} corresponds to $\mathbf{A}^{17}$
    - Universal θ_{17} different for individual graphs $\mathbf{A}_{p1}^{17}$ and $\mathbf{A}_{p2}^{17}$



- GFR filter (Fourier filter):
  - Test accuracy constant
  - No difference between universal and individual graph
  - Below baseline:
    - null result

Interpretation

- Null result:
  - Function of filtering not needed for classification network
- # filter parameters for GIR filter based on individual graph:
  - More coefficients: more detailed filter
  - Less coefficients: better generalisation

  $\rightarrow$ Trade-off between detail and generalisation
  - Interpretation in the literature:
    - Less coefficients: less parameters (7)
    - Only partly true!
- Optimal GIR filter likely not higher than k=3 or k=2
  - Similar findings in the literature ($k <= 3$) (9)
  - Holds even for large networks

References I

[1] E. Pereda, R. Q. Quiroga, and J. Bhattacharya, "Nonlinear multivariate analysis of neurophysiological signals," *Progress in neurobiology*, vol. 77, no. 1-2, pp. 1–37, 2005.

[2] A. Sandryhaila and J. M. Moura, "Discrete signal processing on graphs," *IEEE transactions on signal processing*, vol. 61, no. 7, pp. 1644–1656, 2013.

[3] A. Ortega, P. Frossard, J. Kovačević, J. M. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 808–828, 2018.

[4] S. Chen, A. Sandryhaila, J. M. Moura, and J. Kovacevic, "Signal denoising on graphs via graph filtering," *2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2014, pp. 872–876.

[5] A. Pentari, G. Tzagkarakis, K. Marias, and P. Tsakalides, "Graph-based denoising of eeg signals in impulsive environments," *2020 28th European Signal Processing Conference (EUSIPCO)*. IEEE, 2021, pp. 1095–1099.

References II

[6] M. Defferrard, X. Bresson, and P. Vandergheynst, “Convolutional neural networks on graphs with fast localized spectral filtering,” *Advances in neural information processing systems*, vol. 29, 2016.

[7] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” *arXiv preprint arXiv:1609.02907*, 2016.

[8] A. Rahimi and B. Recht, “Random features for large-scale kernel machines,” *Advances in neural information processing systems*, vol. 20, 2007.

[9] F. Wu, A. Souza, T. Zhang, C. Fifty, T. Yu, and K. Weinberger, “Simplifying graph convolutional networks,” *International conference on machine learning*. PMLR, 2019, pp. 6861–6871.

Limitations of analogy

	Time graph	**Arbitrary graph**

FIR filter	• Highly localised: FIR filter with $k = 3$ only covers 3 nodes of time graph	• Generally not localised:
		  ◦ Example: each node connected to 10 nodes
		  ◦ impulse response of filter with $k = 3$ can cover up to $10 \times 10 = 100$ nodes!

Fourier filter	• Graph is directed $\rightarrow$ eigendecomposition is complex
	• eigenvalues have same magnitude:
	  ◦ ordering of frequency not by their eigenvalue magnitude
	  ◦ “High” frequencies past Nyquist limit are actually low frequencies
		• Graph is typically undirected $\rightarrow$ eigendecomposition is real
		• eigenvalues with different magnitude:
		  ◦ Clear ordering of frequency
		  ◦ But higher frequencies carry less meaning