Discrete Signal Processing on Graphs (DSP_G): Big Data Processing

Deepak Goyadi, Jashan Singhal, Utkarsh Sharma and Mustafa Lokhandwala

Department of Electrical Engineering, IIT Bombay

Temperature Dataset

To demonstrate a couple of applications of DSP_G, we take a dataset consisting of average daily temperatures across 18 cities in western USA.

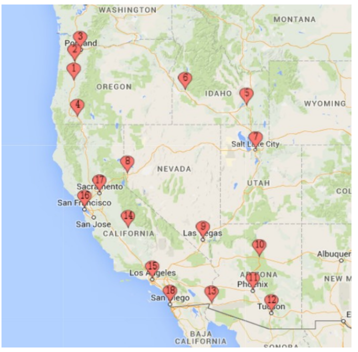


Figure 1: The 18 cities for which the average daily temperature dataset has been processed using DSP_G.

This dataset was chosen because the adjacency matrix of weather sensor networks can be constructed in a straightforward manner from the geographical distances between the cities.

$$A_{n,m} = \frac{e^{-d_{nm}^2}}{\sqrt{\sum_k e^{-d_{nk}^2} \sum_l e^{-d_{ml}^2}}}$$

The matrix $\mathbf{A}$ in our case is symmetric by construction, hence its eigenvectors form an orthogonal basis, and the graph Fourier transform matrix $\mathbf{F}$ is orthogonal.[2]

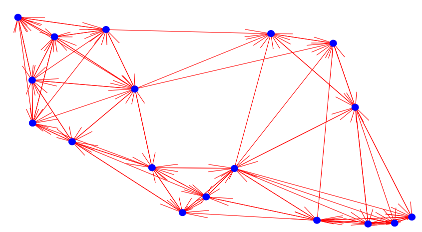


Figure 2: The graph created using the adjacency matrix calculated using the distances between nodes. Note that only 5 nearest neighbours are considered.

As a preliminary proof-of-concept, we apply the DSP_G algorithms for *Linear Prediction* and *Signal Compression* on the daily temperature data of the month January, 2000.

Signal Compression

Efficient signal representation is required in multiple DSP areas, such as storage, compression, and transmission. A widely-used technique is based on expanding signals into orthonormal basis with the expectation that most information is captured with few basis functions.

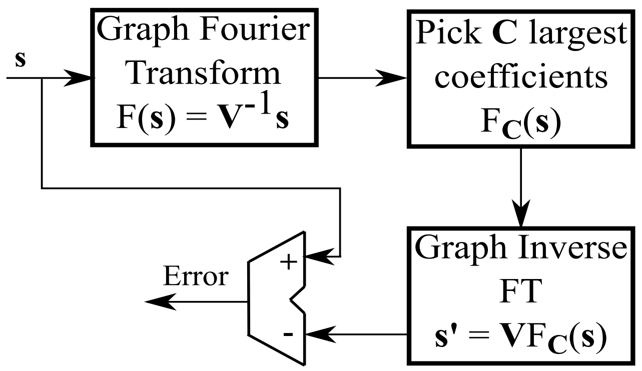


Figure 3: A flow chart depicting the process of signal compression and percentage error calculation.

If the transform represents a Fourier transform in some model, it means that signals are sparse in the frequency domain in this model, i.e., they contain only few frequencies. The original signal $\mathbf{s}$ can be then approximated by $\mathbf{s'}$ as,

$$\mathbf{s'} = \mathbf{F}^{-1}(\hat{\mathbf{s}}_\mathbf{0}, \cdots, \hat{\mathbf{s}}_{\mathbf{C-1}}, 0, \cdots, 0)$$

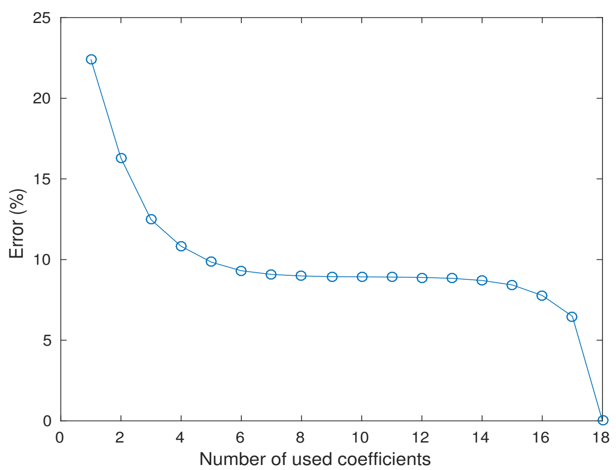
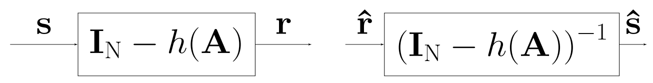


Figure 4: Average reconstruction error $||s - s'||^2/||s||^2$ for the compression of 31 daily temperature snapshots based on the graph Fourier transform using $1 \leqslant C \leqslant 18$ coefficients

Linear Prediction

Linear prediction (LP) is an efficient technique for representation, transmission, and generation of time series. Using DSP_G, we can extend Linear Prediction to graph signals.[2]



For each snapshot $\mathbf{s}$ of 31 measurements, we construct a prediction filter $h(\mathbf{A})$ with L taps by minimizing the energy of the residual

$$\mathbf{r} = s - h(\mathbf{A})s = (\mathbf{I}_N - h(\mathbf{A}))s$$

The residual $\mathbf{r}$ is then quantized using B bits, and the quantized residual $\hat{\mathbf{r}}$ is processed with the inverse filter to synthesize an approximated signal

$$\hat{\mathbf{s}} = (\mathbf{I}_N - h(\mathbf{A}))^{-1}\hat{\mathbf{r}}$$

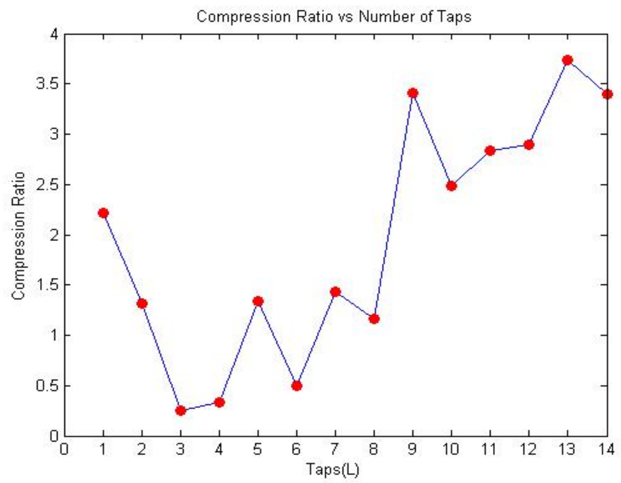


Figure 5: The plot of compression ratio vs L for the daily average temperature dataset by varying L from 1 to 14

With the goal of minimising the compression ratio, we plot the latter with respect to the number of tap coefficients, L. The plot shows a minima, which is the optimum value of L. At this value of L, we obtain a compression ratio which is less than 0.4 in all days considered and less than 0.3 for almost all days.

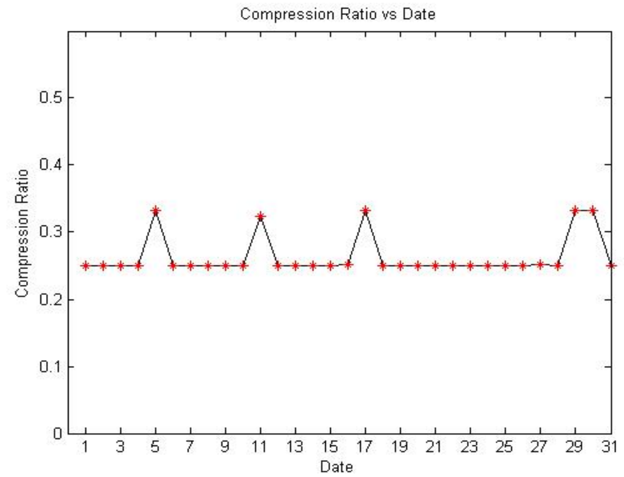


Figure 6: The plot shown is that of compression ratio plotted against the 31 days. The compression ratio is found to lie between 0.25 and 0.33 . Therefore the worst case compression ratio is 0.33.

Acknowledgement

We acknowledge our friend and classmate Khushhall Chandra Mahajan for finding the temperature dataset which we have used in this assignment.

References

1. EPFL-Inria *Graph Signal Processing* Available from World Wide Web: (https://project.inria.fr/epfl-Inria/projects/fast-transformation-graph/)
2. Aliaksei Sandryhaila and José M. F. Moura *Discrete Signal Processing on Graphs*. Available from World Wide Web: (http://arxiv.org/abs/1210.4752)
3. Aliaksei Sandryhaila and José M. F. Moura *Big Data Analysis with Signal Processing on Graphs*. Available from World Wide Web: (https://users.ece.cmu.edu/ asandryh/papers/spm14.pdf)
4. The University of Dayton *Average Daily Temperature Archive*.