Adaptive Message Passing Sign Algorithm

Changran Peng, Yi Yan, and Ercan Engin Kuruoglu *
Tsinghua Shenzhen International Graduate School, Tsinghua University

Abstract

A new algorithm named the Adaptive Message Passing Sign (AMPS) algorithm is introduced for online prediction, missing data imputation, and impulsive noise removal in time-varying graph signals. This work investigates the potential of message passing on spectral adaptive graph filters to define online localized node aggregations. AMPS updates a sign error derived from l_1-norm optimization between observation and estimation, leading to fast and robust predictions in the presence of impulsive noise. The combination of adaptive spectral graph filters with message passing reveals a different perspective on viewing message passing and vice versa. Testing on a real-world network formed by a map of nationwide weather stations, the AMPS algorithm accurately forecasts time-varying temperatures.

1 Introduction

Recently, network and graph-structured data have become increasingly popular across various research fields including computer science, social science, biology, engineering, and finance, owing to their unique ability to represent multivariate irregularities [1–4]. Time-varying data on the nodes are recorded on graphs constructed based on geographical locations, for example, a map of 5G signal reception strength [5] or temperature recorded at multiple locations [6]. In GSP, a combination of classical adaptive filters with graph shift operations performs efficiently on the online processing of time-varying graph signals owing to its simplicity of implementation. Adaptive graph filtering in the spectral domain utilizes a predefined bandlimited filter on the global level derived from Graph Fourier Transform (GFT), which could avoid the time-consuming training process, and then update in the direction opposite to the error at each time step based on convex optimization. The graph least mean squares (GLMS) algorithm, first proposed among all the adaptive graph filters, estimates graph signals using l_2-norm optimization with the presumption of Gaussian noise [5]; several extensions of the GLMS algorithm have emerged, including Normalized GLMS (GNLMS) [6], and Graph-Sign algorithm [7]. Spatial graph algorithms can be approximated from the before-mentioned spectral algorithms using Chebyshev polynomials to transform spectral filtering into spatial graph diffusion, notable examples are the Graph Diffusion LMS [8] and the Graph-Sign-Diffusion (GSD) [9]. Merging time series analysis techniques with GSP points the direction to another problem-solving solution, leading to the introduction of the graph Vector Autoregressive model [10], the graph Vector Autoregressive–Moving-Average model [11], and the graph GARCH model [12].

Gaussian noise assumption is seen in most noise models, and l_2-norm optimization is the go-to option for Gaussian noise because minimizing the squared error corresponds to the maximum likelihood estimate solution [13]. However, the underlying noise in a variety of realistic applications, including meteorological recordings [14] and powerline communication [15], is verified to possess impulsive behaviors that could be represented by heavy-tailed, non-Gaussian distributions, such as generalized Gaussian, Student's t, and α-stable distributions [13]. The impulsiveness presented in heavy-tailed distributions can be characterized by large or infinite variance, causing l_2-norm optimization-based algorithms like GLMS and GNLMS to be unstable [16]. For the purpose of eliminating the downsides

*Changran Peng and Yi Yan contributed equally. Corresponding Author: Ercan E. Kuruoglu. This work was supported by the Tsinghua Shenzhen International Graduate School Startup Fund under Grant QD2022024C and the Shenzhen Science and Technology Innovation Commission under Grant JCYJ20220530143002005.

caused by performing l_2-norm optimization under non-Gaussian noise, the graph Least Mean p^{th} power (GLMP) [17], the Normalized GLMP [18], and Graph-Sign algorithm [7] were proposed to use l_p-norm optimization instead.

By introducing nonlinear activation functions into GSP, architectures such as Graph Convolutional Neural Networks (GCN) and Graph Attention Networks have extended the spatial and spectral GSP methods to machine learning tasks such as node classification, link prediction, and image classification [19–22]. The Spatio-Temporal Graph Convolutional Networks extend GCN by introducing the ability to process time-varying data [23]. These graph neural network (GNN) architectures can be generalized by the Message Passing Neural Networks (MPNN) where the graph representations and operations are defined locally on the nodes by a message passing scheme instead of globally on the graph topology [24, 25]. Compared to specific architectures such as the GCN, MPNN provides more degrees of freedom by having several choices of aggregation functions such as sum, mean, or max and allowing us to tune the magnitude of the localized aggregation flexibly [25]. Compared to graph representation done by the Adjacency matrix or the Laplacian matrix that gives the global view of the graph, MPNNs provide localized representations based on node neighborhood relationships. In addition, edge weights can also be incorporated into the aggregations through message passing.

With insights from the simplicity of implementation of adaptive GSP algorithms and the expressiveness power of message passing of GNNs, we would like to take a step further by breaking the convention of using only global information to define adaptive graph filters. In this paper, we propose the Adaptive Message Passing Sign (AMPS) algorithm, which is a novel adaptive graph filter defined using the localized node message passing scheme with high robustness under impulsive noise.

2 Preliminaries

A graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is defined with node set $\mathcal{V}$, where the N nodes are $v_1...v_N$, and an edge set $\mathcal{E}$, with edge weights $e_1...e_E$. The function value or the data features $\boldsymbol{x}[t]$ defined on the nodes of $\mathcal{G}$ is a time-varying graph signal. The neighborhood relationship of how nodes are connected by edges can form an adjacency matrix $\mathbf{A}$: the ik^{th} entry of $\mathbf{A}$ is the edge weight of the edge between nodes v_i and v_k. The degree matrix $\mathbf{D}$ can be formed by summing all the rows of $\mathbf{A}$ and then forming a diagonal matrix of size N by N. The well known graph Laplacian matrix $\mathbf{L}$ is $\mathbf{L} = \mathbf{D} - \mathbf{A}$. If we perform eigendecomposition with $\mathbf{L}$, we can define the graph Fourier transform (GFT) $\mathbf{L} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$ where $\mathbf{U}$ is the orthonormal eigenvectors of $\mathbf{L}$ and $\boldsymbol{\Lambda}$ is the eigenvalue matrix. The eigenvector eigenvalue pairs are sorted in increasing order of eigenvalues to create a notion of graph frequency. A filtering operation defines a function $h(\boldsymbol{\Lambda})$ on the frequency components to manipulate the frequency content through graph convolution $h(\mathbf{L})\boldsymbol{x}[t] = \mathbf{U}h(\boldsymbol{\Lambda})\mathbf{U}^T\boldsymbol{x}[t]$. Missing values can be defined using a diagonal masking matrix $\mathbf{D}_\mathcal{S}$ of 0s and 1s, with the 0s indicating missing [5].

The symmetric α-stable (SαS) distribution is used in this paper as the impulsive noise model. SαS is governed by the characteristic exponent α, the location parameter μ, and the scale parameter γ. The SαS has no analytic PDF but has the characteristic function $\boldsymbol{\phi}(t) = \exp\{j\mu t - \gamma|t|^\alpha\}$. The mean and variance of SαS are undefined unless only when $1 < \alpha \leqslant 2$ so the mean can be defined or when α = 2 so the variance can be defined. Setting $\alpha = 2$ will make SαS into Gaussian distribution, and $\alpha = 1$ makes SαS the Cauchy distribution [13].

3 Methodology

In this section, we focus on how to use message passing to derive a robust adaptive graph algorithm. Let $\boldsymbol{x}[t]$ denote the ground truth signal, and $\boldsymbol{\eta}[t]$ be the SαS noise $i.i.d.$ among the nodes. To achieve robust estimation, AMPS uses an adaptive update message passing strategy that minimizes the error between a noisy partial observation $\boldsymbol{y}[t] = \mathbf{D}_\mathcal{S}(\boldsymbol{x}[t] + \boldsymbol{\eta}[t])$ and the estimation $\hat{\boldsymbol{x}}[t]$. Following adaptive filter conventions, instead of treating the node signals as the message, the error of each node observation from its neighborhood will be aggregated in AMPS, meaning that $\boldsymbol{z} = \mathbf{D}_\mathcal{S}(\boldsymbol{y}[t] - \boldsymbol{x}[t])$ in (3). To let AMPS output robust estimation unaffected by the SαS noise $\boldsymbol{\eta}[t]$, the minimum dispersion criterion is used to define an optimization problem that leads to l_1-norm [26]:

$$J(\hat{\boldsymbol{x}}[t]) = \mathbb{E}\left\|\boldsymbol{y}[t] - \mathbf{D}_\mathcal{S}\mathbf{U}h(\boldsymbol{\Lambda})\mathbf{U}^T\hat{\boldsymbol{x}}[\mathbf{t}]\right\|_1^1. \tag{1}$$

In GSP, adaptive graph filters such as the Graph Diffusion LMS [8] or GSD [9] use a global update strategy along the graph convolution can be approximated by the Chebyshev polynomial approximation derived in [20] to conduct online filtering:

$$\mathbf{U}h(\mathbf{\Lambda})\mathbf{U}^T \approx \sum_{p=0}^{P} \theta_p T_p(\mathbf{L}) = \sum_{p=0}^{P} \hat{\theta}_p \mathbf{L}^p; T_p(\mathbf{L}) = \begin{cases} 1, & \text{if } p = 0 \\ \frac{2\mathbf{L}-\lambda_N \mathbf{I}}{\lambda_N}, & \text{if } p = 1 \\ \frac{4\mathbf{L}-2\lambda_N \mathbf{I}}{\lambda_N} T_{p-1}(\mathbf{L}) - T_{p-2}(\mathbf{L}), & \text{if } p \geq 2. \end{cases} \tag{2}$$

In (2), θ_0 and $\hat{\theta_0}$ are the coefficients, P is the number of polynomials used in the approximation, and $T_p(\mathbf{L})$ is the shifted Chebyshev polynomial.

It is easy to check that (2) is a global method: all the node signals are being processed at the same time by $\mathbf{U}h(\mathbf{\Lambda})\mathbf{U}^T \boldsymbol{x}[t] \approx \sum_{p=0}^{P} \hat{\theta}_p \mathbf{L}^p \boldsymbol{x}[t]$. In AMPS, we would like to replace this global filtering with localized message passing to achieve online data imputation and denoising. The message passing on node v_i is an aggregation of the graph signal or node feature in the local neighborhood [25]:

$$\text{agg}(v_i) = \Omega_{k=1\ldots K}(z_i, m_{i,k}, z_k), \tag{3}$$

where z_k is the signal on the k^{th} neighboring node that is directly connected by an edge with a distance 1-hop away or the self-loop, and $m_{i,k}$ is the weight of the message. The function Ω is a differentiable and permutation invariant operation suitable for aggregation; common choices seen in previous literature include sum, mean, or max [25]. This message passing scheme is naturally an data imputation algorithm because the missing data can be estimates by properly setting the weights and then aggregating the neighborhood graph signal. Fig. 1 shows an example of message passing.

To replace global filtering in (2) with localized message passing in (3), the choice of Ω is sum aggregation. Setting weights $m_{i,k} = -e_{i,k}$ and $m_{i,i} = L_{i,i}$ leads to

$$\mathbf{L}\boldsymbol{z} = \text{vec}\left(\text{agg}(v_i)\big|_{i=0\ldots N}\right) = \text{vec}\left(z_i - \sum_{j=0}^{N_j} e_{i,k} z_k \Bigg|_{i=0\ldots N} \right) = F(\boldsymbol{z}), \tag{4}$$

where vec() is the vectorize operation. After some algebraic manipulations, we can confirm that the p^{th} power multiplication $\mathbf{L}^p \boldsymbol{z}$ can be achieved by recursively applying (4) p times; we will denote this as $F^p(\boldsymbol{z})$. The update strategy of AMPS follows the graph adaptive filters seen in [7] by calculating the gradient of (1) and replacing the approximation (2) with message passing shown in (3):

$$\hat{\boldsymbol{x}}[t+1] = \hat{\boldsymbol{x}}[t] - \mu \frac{\partial J(\hat{\boldsymbol{x}}[t])}{\partial \hat{\boldsymbol{x}}[t]} = \hat{\boldsymbol{x}}[t] + \mu \left(\sum_{p=1}^{P} \hat{\theta}_p F^p(\hat{\boldsymbol{z}}[t]) + \hat{\theta}_0 \hat{\boldsymbol{z}[t]} \right), \tag{5}$$

where $\hat{\boldsymbol{z}}[t] = \mathbf{D}_{\mathcal{S}} \text{sign}(\boldsymbol{y}[t] - \hat{\boldsymbol{x}}[t])$ and μ is the parameter that controls the magnitude of the update following classical adaptive filtering convention.

Each aggregation in (5) is a message passing based on the estimation error controlled by μ in the direction opposite to the difference between $\hat{\boldsymbol{x}}[t]$ and $\boldsymbol{y}[t]$. We should point out that the N node aggregations at $p = 1$ are conducted on the sign(), $F^1(\hat{\boldsymbol{z}}[t]) = \mathbf{L}\mathbf{D}_{\mathcal{S}} \text{sign}(\boldsymbol{y}[t] - \hat{\boldsymbol{x}}[t])$, which means that the messages passed by $F^1(\hat{\boldsymbol{z}}[t])$ are sign-errors fixed in magnitude therefore unaffected by the impulsive noise $\boldsymbol{\eta}[t]$. The proceeding $p = 2 \ldots P$ aggregations will recursively aggregate based on results $F^1(\hat{\boldsymbol{z}}[t])$, which means that AMPS will update based on a p-hop message passing, with

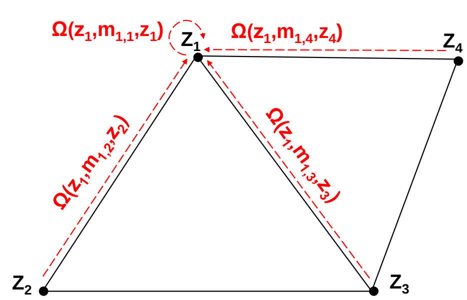

Figure 1: An illustration of message passing.

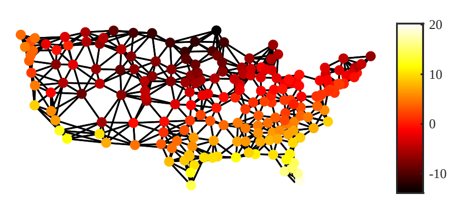

Figure 2: A time instance of the time-varing temperature graph signal of the U.S.

the very first messages being the sign-error unaffected by $\boldsymbol{\eta}[t]$. In this way, missing data can be interpolated simply by aggregating the neighborhood signals. AMPS assumes that the graph signals have frequency components mainly in the low-frequency bands, in other words, the graph signals are smooth [27]. Ideally, we can use a low pass filter with passband $[0, \lambda_l]$ to define the filter $h(\boldsymbol{\Lambda})$ =diag($\boldsymbol{h_l}$). The constant l is the l^{th} eigenvalue index in $\boldsymbol{\Lambda}$ and $\boldsymbol{h}_l$ is a N by 1 vector with the first l elements being 1 and last $N - l$ elements being 0. The coefficients θ of (2) can be efficiently solved by first using an iterative approach in a distributed manner shown in [28]. Combining low pass filters with the l_1-optimization results in (5) will allow AMPS to effectively remove the noise $\boldsymbol{\eta}[t]$.

4 Experimental results and discussion

The dataset in the study represents a dynamic, time-evolving graph signal consisting of 95 hourly-recorded temperatures collected from 197 weather stations across the United States [29]; 130 out of the 197 stations are observed and the rest are assumed to be missing. For the observed signal, we manually added SαS noise with $\alpha = 1.3$ and $\gamma = 0.1$. The missing temperatures are treated as zero. All tested algorithms follow a zero initialization of $\hat{\boldsymbol{x}}[0] = \bar{0}$ and the experiments are repeated 100 times. We use 8-nearest-neighbor according to the latitude and longitude of all weather stations to form the global graph topology, following [6]. An illustration of the temperature graph at one selected time instance is shown in Figure 2.

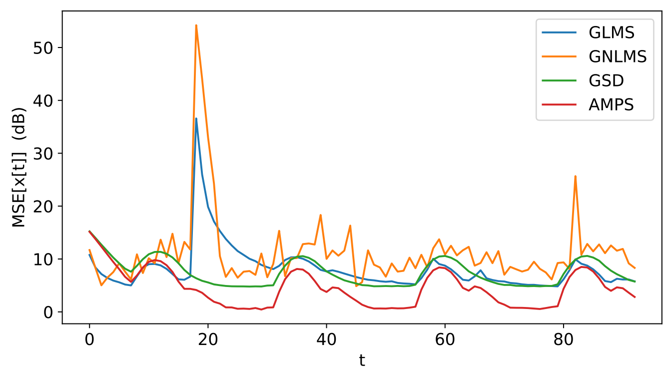

Figure 3: The MSE from $t = 1$ to 95.

The experiment aims to forecast temperature at time $t + 1$ given the missing value and noisy observation of temperature at time t. The parameter choice of step size μ in AMPS is set using grid search; the best-performing value is acquired as $\mu = 1.9$. The performance of all algorithms will be measured in the spatial domain using the mean squared error (MSE) at each time step between the estimation value and the ground truth graph signal: $\text{MSE}[\boldsymbol{x}[t]] = \frac{1}{N}\sum_{i=1}^{N}(x_i[t] - \hat{x}_i[t])^2$. Here, subscript i indicates the i^{th} node in the graph. The MSEs of all the algorithms are calculated at each time point for the forecasted temperature as illustrated in Fig. 3. We can find that AMPS has the lowest MSE compared with GLMS, GNLMS, and GSD at most time points. The reason that GLMS and GNLMS perform worse is that the distribution of noise here is SαS rather than Gaussian, reflecting the fact that l_1-norm optimization is more suitable than l_2-norm optimization on impulsive noise. The aggregation used in AMPS is a message passing of sign-error in (5), with the initial message being fixed magnitude from the optimization results of the minimum dispersion criterion, behaving robustly under SαS noise. Our proposed AMPS algorithm outperforms GSD due to adopting a message passing scheme on top of graph sign aggregation, which is more flexible and expressive at leveraging the localized information at each graph node.

5 Conclusion

In this paper, we proposed the AMPS algorithm for robust time-varying graph signal estimation under the presence of SαS noise by adopting a localized node message passing scheme. This preliminary work examined the potential for leveraging the combination of message passing, l_1-norm optimization, and adaptive filters to form a robust, localized, flexible, and expressive algorithm for processing time-varying graph signals under impulsive noise.

References

[1] A. Sandryhaila and J. M. Moura, "Big data analysis with signal processing on graphs: Representation and processing of massive data sets with irregular structure," *IEEE Signal Processing Magazine*, vol. 31, no. 5, pp. 80–90, 2014.

[2] 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.

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

[4] X. Dong, D. Thanou, L. Toni, M. Bronstein, and P. Frossard, "Graph signal processing for machine learning: A review and new perspectives," *IEEE Signal Processing Magazine*, vol. 37, no. 6, pp. 117–127, 2020.

[5] P. D. Lorenzo, S. Barbarossa, P. Banelli, and S. Sardellitti, "Adaptive least mean squares estimation of graph signals," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 2, no. 4, pp. 555 – 568, 2016.

[6] M. J. M. Spelta and W. A. Martins, "Normalized lms algorithm and data-selective strategies for adaptive graph signal estimation," *Signal Processing*, vol. 167, p. 107326, 2020.

[7] Y. Yan, E. E. Kuruoglu, and M. A. Altinkaya, "Adaptive sign algorithm for graph signal processing," *Signal Processing*, vol. 200, 2022.

[8] R. Nassif, C. Richard, J. Chen, and A. H. Sayed, "A graph diffusion lms strategy for adaptive graph signal processing," *Asilomar*, 2017.

[9] Y. Yan and E. E. Kuruoglu, "Fast and robust wind speed prediction under impulsive noise via adaptive graph-sign diffusion," in *IEEE CAI*, pp. 302–305, 2023.

[10] J. Mei and J. M. F. Moura, "Signal processing on graphs: Causal modeling of unstructured data," *IEEE Trans. Signal Process.*, vol. 65, no. 8, pp. 2077–2092, 2017.

[11] E. Isufi, A. Loukas, A. Simonetto, and G. Leus, "Autoregressive moving average graph filtering," *IEEE Trans. Signal Process.*, vol. 65, no. 2, pp. 274–288, 2017.

[12] J. Hong, Y. Yan, E. E. Kuruoglu, and W. K. Chan, "Multivariate time series forecasting with garch models on graphs," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 9, pp. 557–568, 2023.

[13] Y. Chen, H. C. So, and E. E. Kuruoglu, "Variance analysis of unbiased least l_p-norm estimator in non-gaussian noise," *Signal Processing*, vol. 122, pp. 190–203, 2016.

[14] F. V. Brock, "A nonlinear filter to remove impulse noise from meteorological data," *Journal of Atmospheric and Oceanic Technology*, vol. 3, no. 1, pp. 51 – 58, 1986.

[15] O. Karakuş, E. E. Kuruoğlu, and M. A. Altınkaya, "Modelling impulsive noise in indoor powerline communication systems," *Signal, image and video processing*, vol. 14, no. 8, pp. 1655–1661, 2020.

[16] E. E. Kuruoğlu, P. J. Rayner, and W. J. Fitzgerald, "Least lp-norm impulsive noise cancellation with polynomial filters," *Signal Processing*, vol. 69, no. 1, pp. 1–14, 1998.

[17] N. Nguyen, K. Doğançay, and W. Wang, "Adaptive estimation and sparse sampling for graph signals in alpha-stable noise," *Digital Signal Processing*, vol. 105, p. 102782, 2020.

[18] Y. Yan, R. Adel, and E. E. Kuruoglu, "Graph normalized-lmp algorithm for signal estimation under impulsive noise," *J. Signal Process. Syst.*, pp. 1–12, 2022.

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

[20] 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.

[21] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, "Spectral networks and locally connected networks on graphs," *ICLR*, 2014.

[22] B. P. Chamberlain, S. Shirobokov, E. Rossi, F. Frasca, T. Markovich, N. Hammerla, M. M. Bronstein, and M. Hansmire, "Graph neural networks for link prediction with subgraph sketching," *ICLR*, 2022.

[23] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," in *IJCAI*, 2018.

[24] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, "Neural message passing for quantum chemistry," in *ICML*, pp. 1263–1272, PMLR, 2017.

[25] M. Fey and J. E. Lenssen, "Fast graph representation learning with pytorch geometric," *ICLR Workshop*, 2019.

[26] M. Shao and C. Nikias, "Signal processing with fractional lower order moments," *Proc. IEEE*, vol. 81, no. 7, pp. 986–1010, 1993.

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

[28] D. I. Shuman, P. Vandergheynst, and P. Frossard, "Chebyshev polynomial approximation for distributed signal processing," *DCOSS*, pp. 1–8, 2011.

[29] Arguez et. al *NOAA's U.S. Climate Normals (1981-2010)*, 2010.