Neural Networks 158 (2023) 83–88

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Neural Networks Letter

A class of doubly stochastic shift operators for random graph signals and their boundedness

Bruno Scalzo [a], Ljubiša Stanković [b], Miloš Daković [b], Anthony G. Constantinides [a], Danilo P. Mandic [a,*]

[a] Imperial College London, London SW7 2AZ, UK
[b] University of Montenegro, Džordža Vašingtona bb, 81000 Podgorica, Montenegro

ARTICLE INFO

Article history:
Received 31 March 2022
Received in revised form 5 September 2022
Accepted 31 October 2022
Available online 13 November 2022

Keywords:
Graph signal processing
Doubly stochastic matrix
Shift operator
Statistical consistency
Boundedness analysis
Graph neural networks

ABSTRACT

A class of doubly stochastic graph shift operators (GSO) is proposed, which is shown to exhibit: (i) lower and upper L_2-boundedness for locally stationary random graph signals, (ii) L_2-isometry for *i.i.d.* random graph signals with the asymptotic increase in the incoming neighbourhood size of vertices, and (iii) preservation of the mean of any graph signal – all prerequisites for reliable graph neural networks. These properties are obtained through a statistical consistency analysis of the proposed graph shift operator, and by exploiting the dual role of the doubly stochastic GSO as a Markov (diffusion) matrix and as an unbiased expectation operator. For generality, we consider directed graphs which exhibit asymmetric connectivity matrices. The proposed approach is validated through an example on the estimation of a vector field.

1. Introduction

Given the rapidly increasing availability of data recorded on irregular domains, it would be extremely advantageous to analyse such unstructured data as signals on graphs and thus benefit from the ability of graphs to incorporate domain-specific knowledge (Chen et al., 2015; Liu, Bao, Zheng, Hayat, et al., 2021; Sandryhaila & Moura, 2013; Shuman, Narang, Frossard, Ortega, & Vandergheynst, 2013). This has motivated the developments in the emerging field of Graph Signal Processing (Ortega, Frossard, Kovacevic, Moura, & Vandergheynst, 2018; Stanković, Mandic, Daković, Brajović, Scalzo, Li, Constantinides, et al., 2020a, 2020b, 2020c; Stankovic et al., 2019), and has spurred the introduction of the graph counterparts of many classical signal processing algorithms.

One such direction is that of the system (or filter) on a graph, which was recently considered in Heimowitz and Eldar (2017), Marques, Segarra, Leus, and Ribeiro (2017), Segarra, Marques, and Ribeiro (2017). In classical signal processing, a system is typically a linear operator that maps an input signal to another (output) signal. However, while the signal *shift operator* (unit time delay) is the lynchpin in discrete-time linear systems, its definition on graphs is not obvious due to the rich underlying connectivity structure. Indeed, the shift of a random graph signal can be viewed as the diffusion of a signal sample from the considered vertex along all edges connected to this vertex. Therefore, to effectively employ a system which operates on random signals acquired on graphs, it is necessary to rigorously consider the statistical properties of such *graph shift operators* (GSOs).

Existing GSOs typically take the form of the graph adjacency or graph Laplacian matrices, which are in general neither bounded nor isometric (distance-preserving) operators Liu, Zhao, and Cai (2020). Without this property, the repeated application of a shift to a graph signal can significantly alter or distort the spectral content of the graph signal, thus making it difficult to design and understand the filter frequency response – a stark contrast to classical signal processing (Gavili & Zhang, 2017). Therefore, for rigour it may not only be desirable but also necessary to preserve the signal energy (L_2-norm) over shifts. For instance, strict-sense stationary graph signals are defined to be statistically invariant to graph shifts, therefore, the isometry property is necessary in order to perform statistical operations on stationary graph signals in a mathematically rigorous and tractable manner (Girault, Goncalves, & Fleury, 2015a, 2015b).

Several isometric shift operators have been recently proposed which satisfy the desirable isometry (distance or energy preserving) property (Gavili & Zhang, 2017; Girault, 2015). These graph shift operators are constructed as a diagonal matrix, with the entries defined as the eigenvalues of the adjacency or Laplacian matrix cast onto a unit circle, thus preserving in this way the isometry property. However, there remain issues that need to be addressed prior to a more widespread application of existing isometric shift operators:

* Corresponding author.
E-mail address: d.mandic@imperial.ac.uk (D.P. Mandic).

https://doi.org/10.1016/j.neunet.2022.10.035
0893-6080/

(i) The isometry property is not guaranteed when the graph is directed as in this case the eigenvalues of the adjacency or weight matrix can be complex-valued (Girault, Goncalves, Narayanan, & Ortega, 2016);
(ii) The eigenvalues used to construct the isometric GSO, which relate to the L_2-norm of graph edge weights, are sensitive to low-probability (outlier) edge weights;
(iii) Important localisation properties of the graph are lost by defining the GSO as a diagonal matrix (Perraudin & Vandergheynst, 2017).

For a wide range of random graph signals, it is desirable to employ instead graph shift operators which exhibit tight boundedness, or even the isometry property with respect to metrics other than the L_2-norm.

To this end, we introduce a class of doubly stochastic GSOs and investigate their statistical and boundedness properties, which are shown to exhibit the following desirable properties:

(i) The L_2-norm of locally stationary random graph signals is upper and lower bounded over graph shifts;
(ii) The L_2-norm isometry is attained for *i.i.d.* random graph signals with an asymptotic increase in the incoming vertex neighbourhood size;
(iii) The mean of the graph signal is preserved over shifts.

These boundedness properties are derived by employing both the *left-stochastic property* of the GSO (each column sums up to unity) which allows for a Markovian (diffusion) interpretation of the graph shift, and the *right-stochastic property* of the GSO (each row sums up to unity). This allows for the GSO to be viewed as an **unbiased expectation operator**. In this way, the examination of the boundedness properties reduces to performing a statistical consistency analysis of the graph shift. This, in turn, provides an analytical procedure for investigating the reliability and efficiency of graph neural networks. Practical utility of the proposed class of GSOs is demonstrated through a physically meaningful and intuitive real-world example of geographically distributed estimation of multi-sensor temperature measurements.

2. Doubly stochastic graph shift operators

2.1. Preliminaries

The signal domain considered in this work are *graphs*, whereby a graph, denoted by $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, is defined as a set of N vertices, $\mathcal{V} = \{1, 2, \ldots, N\}$, which are connected by a set of edges, $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$. The existence of an edge going from vertex m to vertex n is designated by $(m, n) \in \mathcal{E}$.

The *incoming* neighbourhood of a vertex m, denoted by $\mathcal{V}_m \subset \mathcal{V}$, is the subset of vertices which form its 1-neighbourhood for which $(n, m) \in \mathcal{E}$, whereby N_m denotes the size of this neighbourhood, $\mathcal{V}_m$. In other words, N_m represents the number of non-zero weights for the edges connected with a vertex m.

The strength of connectivity of an N-vertex graph can be represented by the *weighted adjacency matrix*, $\mathbf{W} \in \mathbb{R}^{N\times N}$, with its entries defined as

$$W_{mn} \begin{cases} > 0, & (n,m) \in \mathcal{E}, \\ = 0, & (n,m) \notin \mathcal{E}, \end{cases} \tag{1}$$

whereby the amplitude of the entries conveys the *relative* importance of the vertex connections. Regarding the directionality of vertex connections, a graph is undirected if each edge, $(m, n) \in \mathcal{E}$, has its counterpart, $(n, m) \in \mathcal{E}$, such that $\mathbf{W} = \mathbf{W}^{\mathsf{T}}$. For generality, in this work we consider directed graphs, for which this symmetry property does not hold.

2.2. Random signals on a graph

With each vertex, $n \in \mathcal{V}$, we can associate a real-valued random variable, $x_n \in \mathbb{R}$. Upon considering all vertices in $\mathcal{V}$, a random signal on a graph is denoted by $\mathbf{x} = \{x_n | \, n \in \mathcal{V}\} \in \mathbb{R}^N$ and is said to be wide-sense stationary (WSS) if and only if its first- and second-order moments are invariant under the application of a graph shift (Girault, 2015; Girault et al., 2015a, 2015b). By defining the GSO as a matrix $\mathbf{S} \in \mathbb{R}^{N\times N}$, the conditions for graph wide-sense stationarity are given by $E\{\mathbf{x}\} = E\{\mathbf{Sx}\} = E\{\mathbf{S}^n\mathbf{x}\}$, $n = 2, 3, \ldots$, and $E\left\{\mathbf{x}\mathbf{x}^{\mathsf{T}}\right\} = E\left\{\mathbf{S}\mathbf{x}\mathbf{x}^{\mathsf{T}}\mathbf{S}^{\mathsf{T}}\right\} = E\left\{\mathbf{S}^n\mathbf{x}\mathbf{x}^{\mathsf{T}}(\mathbf{S}^{\mathsf{T}})^n\right\}$. Notice that in this way the graph ensemble mean, $E\{\mathbf{x}\}$, is not a constant but a vertex-varying signal. Various Graph Signal Processing applications have been investigated based on this definition of the WSS graph signal (Gama & Ribeiro, 2019; Marques et al., 2017; Perraudin & Vandergheynst, 2017).

Despite the desirable mathematical tractability of WSS graph signals, they do not appropriately model the smoothness of the nonstationarity in the vertex-domain which is inherent to real-world graph signals. Consequently, a class of *locally* stationary graph signals was introduced in Girault, Narayanan, and Ortega (2017a, 2017b), whereby the statistical properties of the vertex signals within a neighbourhood are assumed to be identical. While the statistical conditions for local graph stationarity in Girault et al. (2017a, 2017b) are defined in terms of the local graph power spectral density, we consider a milder definition of local stationarity in the vertex domain based on the work in Dahlhaus (1996), whereby the vertex signals in a neighbourhood $\mathcal{V}_m$ exhibit the same first- and second-order moments, that is

$$\mu_m = E\{x_n\}, \quad \sigma_m^2 = \mathrm{var}\{x_n\}, \quad \forall n \in \mathcal{V}_m. \tag{2}$$

In addition, we allow for a non-zero correlation between the vertex signals in this neighbourhood, that is, the correlation coefficient has the form

$$\mathrm{corr}\{x_n, x_k\} = \begin{cases} \rho_m, & n \neq k, \\ 1, & n = k, \end{cases} \quad \forall n, k \in \mathcal{V}_m. \tag{3}$$

2.3. Doubly stochastic graph shift operators

We next consider a class of doubly stochastic GSOs, denoted by $\mathbf{S} \in \mathbb{R}^{N\times N}$, which exhibit the following properties

$$S_{mn} \geq 0, \quad \mathbf{S1} = \mathbf{1}, \quad \mathbf{S}^{\mathsf{T}}\mathbf{1} = \mathbf{1} \tag{4}$$

that is, $\mathbf{S}$ is a square, non-negative, matrix with columns and rows which sum up to unity.

Remark 1. The left-stochasticity property ($\mathbf{S}^{\mathsf{T}}\mathbf{1} = \mathbf{1}$, columns sum up to unity) allows for a Markov (diffusion) matrix interpretation of the shift operator, whereby the (m, n)-th entry, S_{mn}, can be thought of as the transition probability of a random walker going from a vertex n to a vertex m. Intuitively, the probability of going from a vertex n to any vertex $m \in \mathcal{V}$ is equal to unity, i.e. $\sum_{m=1}^{N} S_{mn} = 1$ (the nth column of $\mathbf{S}$ sums up to unity).

Remark 2. The L_p-norm of a matrix, $\mathbf{A}$, is defined as

$$\|\mathbf{A}\|_p = \sup_{\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{Ax}\|_p}{\|\mathbf{x}\|_p} = \begin{cases} \max_j \sum_i |A_{ij}|, & p = 1, \\ \sqrt{\lambda_{max}(\mathbf{A}^{\mathsf{T}}\mathbf{A})}, & p = 2, \\ \max_i \sum_j |A_{ij}|, & p = \infty. \end{cases} \tag{5}$$

Since the largest eigenvalue of a doubly stochastic matrix is equal to unity (Bapat & Raghavan, 1997), and the rows and columns also sum up to unity, then $\|\mathbf{S}\|_p = 1$ for all $p = 1, 2, \infty$. In this way, following from the above definition of the matrix norm, we obtain the following result

$$\|\mathbf{S}\mathbf{x}\|_p \leq \|\mathbf{S}\|_p \|\mathbf{x}\|_p = \|\mathbf{x}\|_p, \quad \forall p = 1, 2, \infty. \tag{6}$$

Remark 3. The doubly stochastic shift operator preserves the mean of the graph signal values, a very desirable property in graph neural networks. This can be seen by considering a graph signal, $\mathbf{x} \in \mathbb{R}^N$, whose mean is equal to $\mu\mathbf{1}$, that is, $\mathbf{x} = \mu\mathbf{1} + \mathbf{v}$, where $\mathbf{v} \in \mathbb{R}^N$ is a signal such that $\mathbf{1}^T\mathbf{v} = 0$. For this signal $\mathbf{1}^\mathsf{T}\mathbf{x} = \mathbf{1}^\mathsf{T}\mu\mathbf{1} + \mathbf{1}^\mathsf{T}\mathbf{v} = N\mu$ holds. Upon applying the doubly stochastic shift to this graph signal, we have

$$\mathbf{y} = \mathbf{S}\mathbf{x} = \mathbf{S}(\mu\mathbf{1} + \mathbf{v}) = \mu\mathbf{S}\mathbf{1} + \mathbf{S}\mathbf{v} = \mu\mathbf{1} + \mathbf{S}\mathbf{v} \tag{7}$$

Observe that the mean of $\mathbf{y}$ is also $\mu\mathbf{1}$, since $\mathbf{1}^\mathsf{T}\mathbf{y} = \mathbf{1}^\mathsf{T}(\mu\mathbf{1}+\mathbf{S}\mathbf{v}) = \mathbf{1}^\mathsf{T}\mu\mathbf{1} + \mathbf{1}^\mathsf{T}\mathbf{S}\mathbf{v}$. Based on (4), $\mathbf{1}^\mathsf{T}\mathbf{S} = \mathbf{1}^\mathsf{T}$ holds, producing $\mathbf{1}^\mathsf{T}\mathbf{S}\mathbf{v} = \mathbf{1}^\mathsf{T}\mathbf{v} = \mathbf{0}$ and $\mathbf{1}^\mathsf{T}\mathbf{y} = N\mu$.

Remark 4. From Remark 3, the doubly stochastic shift exhibits the L_1-isometry for non-negative graph signals, since $\|\mathbf{x}\|_1 = N\mu$. From the Birkhoff–von Neumann theorem, a doubly stochastic matrix can be factorised into a convex combination of k permutation matrices, i.e. $\mathbf{S} = \sum_{i=1}^{k} a_i\mathbf{P}_i$ with the coefficients $0 \leq a_i \leq 1$, and $k \leq (N-1)^2 + 1$. Since permutation matrices exhibit the L_1-isometry, then so too does the doubly stochastic GSO, i.e. $\|\mathbf{S}\mathbf{x}\|_1 = \|\mathbf{x}\|_1$.

Remark 5. Intuitively, Remarks 2–3 describe the behaviour of a doubly stochastic graph shift as a diffusion toward a uniform graph signal, since the noise component of a signal $\mathbf{x} = \mu\mathbf{1} + \mathbf{v}$ diffuses over a graph shift, i.e. $\|\mathbf{S}\mathbf{v}\|_p \leq \|\mathbf{v}\|_p$, while the mean of the graph signal, μ, is preserved over shifts. To see this, recall that the largest eigenvalue of $\mathbf{S}$ is equal to unity, while the remaining eigenvalues lie on the closed unit disk. Furthermore, the eigenvector associated with the unit eigenvalue is $\frac{1}{\sqrt{N}}\mathbf{1}$. As a result, we obtain the following convergence bound

$$\lim_{k\to\infty} \mathbf{S}^k\mathbf{x} = \tfrac{1}{N}\mathbf{1}\mathbf{1}^\mathsf{T}\mathbf{x} = \mu\mathbf{1} \tag{8}$$

Observe also that this describes the characteristic behaviour of a diffusion process asymptotically approaching a uniform (constant) signal.

In practice, the actual probabilities of vertex transition are often unknown, however, these can be inferred using the available information of the graph domain topology, implied by the weight matrix, $\mathbf{W}$. The graph edges weights, W_{mn}, can be defined through domain knowledge, based on the geometry of vertex positions, or based on data similarity methods for learning the underlying graph topology (Coifman & Lafon, 2006; Heimowitz & Eldar, 2017; Stankovic et al., 2019). Once a weight matrix, $\mathbf{W}$, is defined, there exist several techniques for obtaining a doubly stochastic GSO, $\mathbf{S}$, with the properties in (4). The most well-known procedure is the *Sinkhorn–Knopp* algorithm which can retrieve $\mathbf{S}$ through an alternating normalisation of $\mathbf{W}$ by its row and column sums (Sinkhorn, 1964). This iterative procedure is summarised in Alg. 1, with the operator $\mathcal{D} : \mathbb{R}^N \mapsto \mathbb{R}^{N\times N}$ defined as $\mathcal{D}(\mathbf{x}) = \mathrm{diag}(\mathbf{x})$.

The proposed method, summarised in Algorithm 1, effectively performs a sequence of contraction mappings in order to reach a fixed point — the desired doubly stochastic GSO. For more detail about fixed points in this context we refer to Mandic and Yamada (2007). It is important to note that such an operation not change the structure of the graph. The non-zero weights in $\mathbf{W}$ remain non-zero in $\mathbf{S}$, while the vertices which are not connected (with $W_{mn} = 0$), remain zero valued in $\mathbf{S}$. The summation in (11) is

Algorithm 1. Sinkhorn–Knopp algorithm

procedure SINKHORN–KNOPP($\mathbf{W}$)
  $\mathbf{r} \leftarrow \mathbf{1}$
  while not converged **do**
    $\mathbf{c} \leftarrow \mathcal{D}(\mathbf{W}^\mathsf{T}\mathbf{r})^{-1}\mathbf{1}$
    $\mathbf{r} \leftarrow \mathcal{D}(\mathbf{W}\mathbf{c})^{-1}\mathbf{1}$
  $\mathbf{S} \leftarrow \mathcal{D}(\mathbf{r})\mathbf{W}\mathcal{D}(\mathbf{c})$
  return $\mathbf{S}$

over the non-zero values of S_{mn} only. Therefore, relation (11) follows directly from (4). Notice that in the summation (11), the zero weights (between vertices which are not connected) are not included.

Notice that, in general, $\mathbf{W}$ can be directed. In this case this procedure is also known as the *iterative proportional scaling algorithm* (Stephan, 1942).

3. Boundedness of doubly stochastic GSOs

Next, we investigate the application of the doubly stochastic matrix as a shift operator for random graph signals, and demonstrate its boundedness properties – a prerequisite to its use in real world applications. We begin by considering a doubly stochastic shift applied to the mth vertex signal, x_m, to give

$$\mathcal{S}(x_m) = \sum_{n\in\mathcal{V}_m} S_{mn}x_n \tag{9}$$

The variance of the so shifted signal is then given by

$$\mathrm{var}\{\mathcal{S}(x_m)\} = E\left\{\mathcal{S}(x_m)^2\right\} - E\{\mathcal{S}(x_m)\}^2 \tag{10}$$

Upon rearranging the above equation, we obtain the expression for the *expected* power of the shifted random graph signal

$$E\left\{\mathcal{S}(x_m)^2\right\} = E\{\mathcal{S}(x_m)\}^2 + \mathrm{var}\{\mathcal{S}(x_m)\} \tag{11}$$

In this way, the examination of the boundedness properties over graph shifts boils down to a *statistical consistency* analysis of the graph shift operator.

Given the difficulty of evaluation of the statistical consistency of the graph shift for an arbitrary random graph signal, we consider a locally stationary graph signal as described in Section 2.2, whereby the signal at a vertex $n \in \mathcal{V}_m$ is assumed to be marginally distributed according to $x_n \sim \mathcal{N}(\mu, \sigma^2)$, and the correlation between vertex signals in this neighbourhood is assumed to be $\rho = \mathrm{corr}\{x_n, x_k\}$ for all $n, k \in \mathcal{V}_m$, $n \neq k$.

3.1. Bias

The doubly stochastic shift in (9) is an *unbiased* estimator of the mean, since each of its rows sums up to unity so that for $E\{x_n\} = \mu$ we have

$$E\{\mathcal{S}(x_m)\} = \sum_{n\in\mathcal{V}_m} S_{mn}E\{x_n\} = \mu\sum_{n\in\mathcal{V}_m} S_{mn} = \mu \tag{12}$$

3.2. Asymptotic consistency

Consider the variance of the graph shift in (9), given by

$$\begin{aligned}\mathrm{var}\{\mathcal{S}(x_m)\} &= \sum_{n\in\mathcal{V}_m}\sum_{k\in\mathcal{V}_m} S_{mn}S_{mk}\,\mathrm{cov}\{x_n, x_k\} \\ &= \sigma^2\Bigg(\sum_{n\in\mathcal{V}_m} S_{mn}^2 + \rho\sum_{\substack{n\in\mathcal{V}_m\\ n\neq k}}\sum_{\substack{k\in\mathcal{V}_m\\ k\neq n}} S_{mn}S_{mk}\Bigg),\end{aligned} \tag{13}$$

where N_m is the neighbourhood size of $\mathcal{V}_m$.

The Cauchy–Schwarz inequality can be employed on $\mathbf{1}^T\mathbf{S}$, to assert the following bounds on the last term in the above expression

$$\sum_{\substack{n\in\mathcal{V}_m\\n\neq k}}\sum_{\substack{k\in\mathcal{V}_m\\k\neq n}} S_{mn}S_{mk} \leq \left(\sum_{n\in\mathcal{V}_m} S_{mn}\right)^2 \leq N_m \sum_{n\in\mathcal{V}_m} S_{mn}^2 \tag{14}$$

This inequality can also be directly obtained from $S_{mn} \geq 0$. Therefore, the variance of the doubly stochastic GSO is bounded from above according to

$$\text{var}\{\mathcal{S}(x_m)\} \leq \sigma^2(1+N_m\rho)\sum_{n\in\mathcal{V}_m} S_{mn}^2 \tag{15}$$

Furthermore, an upper bound to the term $\sum_{n\in\mathcal{V}_m} S_{mn}^2$ can be obtained from the Kantorovich inequality (Kantorovich, 1948), which yields

$$\left(\sum_{n\in\mathcal{V}_m} S_{mn}^2\right)\left(\sum_{n\in\mathcal{V}_m} 1^2\right) \leq \frac{(L+U)^2}{4LU}\left(\sum_{n\in\mathcal{V}_m} S_{mn}\right)^2 \tag{16}$$

where L and U denote respectively the lower and upper bounds on the possible values of the elements $S_{mn} \in \mathbf{S}$, which in our case are

$$0 < L \leq S_{mn} \leq U < 1, \quad \forall m, n \tag{17}$$

We finally obtain the following result

$$\sum_{n\in\mathcal{V}_m} S_{mn}^2 \leq \frac{1}{N_m}\frac{(L+U)^2}{4LU} < 1 \tag{18}$$

Remark 6. The lower and upper bounds of S_{mn}, L and U, are invariant to the neighbourhood size, N_m, as they are determined solely from the underlying physics of the problem.

From Remark 6, with an increase in the incoming neighbourhood size, N_m, in the limit we obtain the following upper bound on the graph shift variance, since from (15) and (18)

$$\begin{aligned}\lim_{N_m\to\infty}\text{var}\{\mathcal{S}(x_m)\} &\leq \lim_{N_m\to\infty}\sigma^2(1+N_m\rho)\sum_{n\in\mathcal{V}_m} S_{mn}^2 \\ &\leq \lim_{N_m\to\infty}\frac{\sigma^2(1+N_m\rho)}{N_m}\frac{(L+U)^2}{4LU} \\ &= \rho\sigma^2\frac{(L+U)^2}{4LU}\end{aligned} \tag{19}$$

This proves that, as desired, the degree of statistical inconsistency of the doubly stochastic graph shift is upper bounded.

Remark 7. For an *i.i.d.* random graph signal ($\rho = 0$), the upper bound on the shift variance vanishes, since from (19)

$$\lim_{N_n\to\infty}\text{var}\{\mathcal{S}(x_m)\} = 0 \tag{20}$$

Therefore, the doubly stochastic graph shift is *statistically consistent* for *i.i.d.* graph signals.

3.3. L_2-norm upper boundedness

An upper L_2-boundedness of the shift operator can also be proven for an asymptotic increase in the neighbourhood size, N_m. Starting from (11)–(12) and by employing the inequality in (19), we obtain the following asymptotic behaviour

$$\lim_{N_m\to\infty} E\left\{\mathcal{S}(x_m)^2\right\} \leq \mu^2 + \rho\sigma^2\frac{(L+U)^2}{4LU} \tag{21}$$

which proves the asymptotic L_2-norm upper boundedness of the doubly stochastic graph shifted signal.

Remark 8. Notice that the bias term in (21), given by

$$\frac{(L+U)^2}{4LU} = \left(\frac{\frac{1}{2}(L+U)}{\sqrt{LU}}\right)^2 \geq 1 \tag{22}$$

is simply the square of the ratio of the arithmetic mean (AM) to the geometric mean (GM) of the bounds L and U. The AM–GM inequality therefore asserts that this bias term is bounded by unity from below, with the equality attained for $L = U$. In other words, the magnitude of this bias term is minimised by maximising the ratio of the bounds, $\frac{L}{U}$, or equivalently, by promoting the homogeneity of the graph edge weights, W_{mn}, within each neighbourhood, $\mathcal{V}_m$. This suggests that the design of neighbourhoods, or the location of additional vertices, may be chosen so as to maximise the ratio, $\frac{L}{U}$, thus in turn tightening the upper boundedness of the GSO.

3.4. L_2-norm lower boundedness

Since the variance of the graph shifted signal is strictly non-negative, $\text{var}\{\mathcal{S}(x_m)\} \geq 0$, from (11)–(12) we obtain the lower bound of the L_2-norm of the shifted signal in the form

$$E\left\{\mathcal{S}(x_m)^2\right\} \geq \mu^2 \tag{23}$$

which is a direct consequence of Jensen's inequality.

Remark 9. The shift operator is asymptotically L_2-norm isometric for *i.i.d.* graph signals, since for $\rho = 0$ the lower and upper bounds of $\lim_{N_m\to\infty} E\left\{\mathcal{S}(x_m)^2\right\}$ coincide, leading to the following desired result

$$\lim_{N_m\to\infty} E\left\{\mathcal{S}(x_m)^2\right\} = \mu^2 \tag{24}$$

Therefore, the doubly stochastic GSO preserves the power of the mean for *i.i.d.* graph signals.

3.5. Boundedness of systems on a graph

For a random graph input signal, a linear system of order K is defined as (Sandryhaila & Moura, 2013)

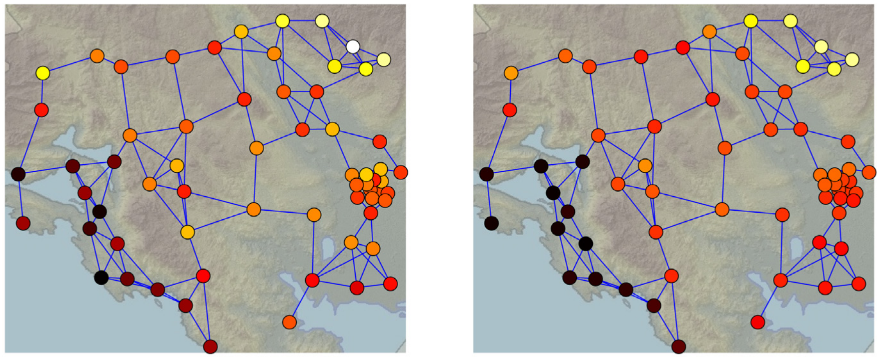
$$\mathbf{y} = \sum_{k=0}^{K} h_k\mathbf{S}^k\mathbf{x} \tag{25}$$

where h_k are the system coefficients. Based on the above derived boundedness properties, the class of systems based on the doubly stochastic shift also exhibits desirable boundedness properties. Starting from (6), we obtain the inequality $\|\mathbf{S}^k\mathbf{x}\|_p \leq \|\mathbf{x}\|_p$ for $p = 1, 2, \infty$ and all $k \geq 0$. Together with Minkowski's inequality, we can show that the graph system output, $\mathbf{y}$, is L_2-norm upper bounded as follows

$$\|\mathbf{y}\|_p \leq \sum_{k=0}^{K}|h_k|\|\mathbf{S}^k\mathbf{x}\|_p \leq \sum_{k=0}^{K}|h_k|\|\mathbf{x}\|_p, \quad \forall p = 1, 2, \infty. \tag{26}$$

4. Numerical example

Consider a multi-sensor setup, described in Stankovic et al. (2019), for measuring a temperature field in a geographic region. This temperature field consists of $N = 64$ sensor measurements in total, as shown in Fig. 1(a). Each measured sensor signal was corrupted with synthetic Gaussian noise to emulate the possible adverse effects of the local environment on sensor readings or faulty sensor activity. The signal at the nth vertex can therefore be mathematically expressed as $x_n \sim \mathcal{N}(\mu_n, \sigma^2)$, where μ_n is the true temperature at the nth sensor (vertex). In our study, the



(a) Observed temperature field (SNR = 14.0 dB)

(b) GSO expectation. (SNR = 19.8 dB)

Fig. 1. Performance of the proposed doubly-stochastic graph shift operator for signal estimation from noisy observed measurements.

standard deviation of the additive noise was set to $\sigma = 2$, to yield the signal-to-noise ratio in x_n of SNR = 14.0 dB.

The doubly stochastic GSO was employed as an expectation operator to estimate the true temperature from the observed temperature field. The weight matrix entries, $W_{mn} = e^{-r_{mn}^2}$, were specified based on the geographical distance between vertices, r_{mn}, thus accounting for the difference in latitude, longitude and altitude. The matrix $\mathbf{S}$ (shift operator) was obtained from $\mathbf{W}$ using the Sinkhorn-Knopp algorithm described in Alg. 1. The denoised temperature field is illustrated in Fig. 1(b), whereby the shifted graph signal attained a 5.8 dB SNR gain.

5. Conclusions

We have introduced a class of isometric doubly stochastic graph shift operators (GSOs), which preserve the mean of random signals along graph shifts. Their boundedness property has been established by performing a statistical consistency analysis of the graph shift. This has been achieved based on the dual role of the doubly stochastic GSO as a Markov (diffusion) matrix and as an unbiased graph expectation operator. The usefulness of doubly stochastic GSOs for operating on random graph signals has been demonstrated through asymptotic statistical consistency analysis, and supported by a practical real-world multi-sensor temperature estimation example.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Bapat, R. B., & Raghavan, T. E. S. (1997). *Non-negative matrices and applications*. Cambridge University Press.

Chen, S., Varma, R., Sandryhaila, A., Kovacevic, J., Moura, J. M. F., & Vandergheynst, P. (2015). Discrete signal processing on graphs: Sampling theory. *IEEE Transactions on Signal Processing*, *63*(24), 6510–6523.

Coifman, R. R., & Lafon, S. (2006). Diffusion maps. *Applied and Computational Harmonic Analysis*, *21*, 5–30.

Dahlhaus, R. (1996). On the Kullback-Leiber information divergence of locally stationary processes. *Stochastic Processes and their Applications*, *62*, 139–168.

Gama, F., & Ribeiro, A. (2019). Ergodicity in stationary graph processes: A weak law of large numbers. *IEEE Transactions on Signal Processing*, *67*(10), 2761–2774.

Gavili, A., & Zhang, X. P. (2017). On the shift operator, graph frequency, and optimal filtering in graph signal processing. *IEEE Transactions on Signal Processing*, *65*(23), 6303–6318.

Girault, B. (2015). Stationary graph signals using an isometric graph translation. In *In Proceedings of the European Signal Processing Conference* (pp. 1516–1520).

Girault, B., Goncalves, P., & Fleury, E. (2015a). Translation and stationarity for graph signals. [Research Report] RR-8719, Ecole Normale Superieure de Lyon, INRIA.

Girault, B., Goncalves, P., & Fleury, E. (2015b). Translation on graphs: An isometric shift operator. *IEEE Signal Processing Letters*, *22*(12), 2416–2420.

Girault, B., Goncalves, P., Narayanan, S. S., & Ortega, A. (2016). Localization bounds for the graph translation. In *Proceedings of the IEEE Global Conference on Signal and Information Processing* (pp. 331–335).

Girault, B., Narayanan, S. S., & Ortega, A. (2017a). Local stationarity of graph signals: Insights and experiments. In *Proceedings of the SPIE Conference on Optical Engineering and Applications* (pp. 1–17).

Girault, B., Narayanan, S. S., & Ortega, A. (2017b). Towards a definition of local stationarity for graph signals. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 4139–4143).

Heimowitz, A., & Eldar, Y. C. (2017). A unified view of diffusion maps and signal processing on graphs. In *Proceedings of the International Conference on Sampling Theory and Applications (SampTA)* (pp. 308–312).

Kantorovich, L. V. (1948). Functional analysis and applied mathematics. *Uspekhi Matematicheskikh Nauk*, *3*(6), 89–185.

Liu, J.-B., Bao, Y., Zheng, W.-T., Hayat, S., et al. (2021). Network coherence analysis on a family of nested weighted n-polygon networks. *Fractals*, *29*(08), 1–15.

Liu, J.-B., Zhao, J., & Cai, Z.-Q. (2020). On the generalized adjacency, Laplacian and singless Laplacian spectra of the weighted edge corona networks. *Physica A. Statistical Mechanics and its Applications*, *540*, 1–11.

Mandic, D. P., & Yamada, I. (2007). Machine learning and signal processing applications of fixed point theory. In *Tutorial at the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'07)* (pp. 1–144).

Marques, A. G., Segarra, S., Leus, G., & Ribeiro, A. (2017). Stationary graph processes and spectral estimation. *IEEE Transactions on Signal Processing*, *65*(22), 5911–5926.

Ortega, A., Frossard, P., Kovacevic, J., Moura, J. M. F., & Vandergheynst, P. (2018). Graph signal processing: Overview, challenges, and applications. *In Proceedings of the IEEE*, *106*(5), 808–828.

Perraudin, N., & Vandergheynst, P. (2017). Stationary signal processing on graphs. *IEEE Transactions on Signal Processing*, *65*(13), 3462–3477.

Sandryhaila, A., & Moura, J. M. F. (2013). Discrete signal processing on graphs. *IEEE Transactions on Signal Processing*, *61*(7), 1644–1656.

Segarra, S., Marques, A. G., & Ribeiro, A. (2017). Optimal graph-filter design and applications to distributed linear network operators. *IEEE Transactions on Signal Processing*, *65*(15), 4117–4131.

Shuman, D. I., Narang, S. K., Frossard, P., Ortega, A., & Vandergheynst, P. (2013). The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine*, *30*, 83–98.

Sinkhorn, R. (1964). A relationship between arbitrary positive matrices and doubly stochastic matrices. *The Annals of Mathematical Statistics*, *35*, 876–879.

Stanković, L., Mandic, D. P., Daković, M., Brajović, M., Scalzo, B., Li, S., et al. (2020a). Data analytics on graphs. Part I: Graphs and spectra on graphs. *Foundations and Trends® in Machine Learning*, *13*(1), 1–157.

Stanković, L., Mandic, D. P., Daković, M., Brajović, M., Scalzo, B., Li, S., et al. (2020b). Data analytics on graphs. Part II: Signals on graphs. *Foundations and Trends® in Machine Learning*, *13*(2-3), 158–331.

Stanković, L., Mandic, D. P., Daković, M., Brajović, M., Scalzo, B., Li, S., et al. (2020c). Data analytics on graphs. Part III: Machine learning on graphs, from graph topology to applications. *Foundations and Trends® in Machine Learning*, *13*(4), 332–530.

Stankovic, L., Mandic, D. P., Dakovic, M., Kisil, I., Sejdic, E., & Constantinides, A. G. (2019). Understanding the basis of graph signal processing via an intuitive example-driven approach. *IEEE Signal Processing Magazine*, *36*(6), 133–145.

Stephan, F. F. (1942). An iterative method of adjusting sample frequency tables when expected marginal totals are known. *The Annals of Mathematical Statistics*, *13*(2), 166–178.