IEEE Transactions on Signal Processing

Orthogonal Fourier Analysis on Directed Acyclic Graphs via Möbius Total Variation

Journal:	*Transactions on Signal Processing*
Manuscript ID	T-SP-32873-2024
Manuscript Type:	Regular Paper (S1)
Date Submitted by the Author:	13-Nov-2024
Complete List of Authors:	Mihal, Vedran; ETH Zürich, Computer Science Püschel, Markus ; ETH Zürich, Computer Science
Subject Category Please select at least one subject category that best reflects the scope of your manuscript:	SIGNAL PROCESSING FOR NETWORKS AND GRAPHS
EDICS:	56. SIPG-REP Transforms, filtering, representation < NEG SIGNAL PROCESSING FOR NETWORKS AND GRAPHS, 57. SIPG-SAMP Sampling, interpolation, denoising and reconstruction < NEG SIGNAL PROCESSING FOR NETWORKS AND GRAPHS

Author's Responses to Custom Submission Questions

Is this manuscript a resubmission of, or related to, a previously rejected manuscript? **Please note that this question refers to all publication venues, not just to this journal or other IEEE journals, and that IEEE Signal Processing Society policies state that authors of a manuscript that has been rejected, from any journal, except for reasons of scope, are allowed to resubmit their manuscript only once.** **Please refer to the IEEE Signal Processing Society Policies and Procedures Manual Section 6.16 for complete information.** **If "Yes", specify the publication venue and manuscript ID of the previous submission and upload a supporting document providing verbatim quotations of all relevant parts of all previous review reports, and outlining how they have been addressed in the revised version. If the manuscript has previously been rejected for reasons of scope, then the authors are required to provide motivation for why the resubmission should be considered to be in scope. If relevant information from previous reviews is withheld, then the manuscript may be immediately rejected.**	No
Is this manuscript an extended version of a conference publication (or a conference article accepted for publication)? **Please note that it is acceptable for a short, up to six (double-column) pages, conference publication to be used as the basis for a more fully developed journal publication. However, authors are required to cite such prior work in the manuscript, either in the introduction or in a footnote, and include an explanation of additional contribution.** **Please refer to the IEEE Signal Processing Society "Information for Authors" for complete information.**	Yes. V. Mihal and M. PÃ¼schel, "MÃ¶bius Total Variation for Directed Acyclic Graphs," ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 2023, pp. 1-5, doi: 10.1109/ICASSP49357.2023.10095435.
Is this manuscript related to any other papers of the authors that are either published, accepted for publication, or currently under review, and that are not included among the	No

references cited in the manuscript? **This also includes papers with other co-authors and covers all publication venues, including other journals, conferences and book chapters.**	
Explain in detail why the contribution of this manuscript is within the scope of Transactions on Audio, Speech and Language Processing? **Please consult the journal's "Aims & Scope" for more detail on topics that are considered to fall within the scope of the journal, noting that some topics may be context-dependent; in particular, contributions related to technique-related topics (e.g. "machine learning") must be applied to a problem within the scope of the journal in order for the contribution to be in scope.** **The absence of a suitable response may be taken into account when making a decision in cases of possible immediate rejection for reasons of scope.**	The paper aims to fill a blind spot in the body of work on graph signal processing: a suitable form of orthogonal Fourier analysis for directed acyclic graphs, which is not well covered by the previous theory.
Why is the contribution significant (What impact will it have)?	The contribution is significant since it provides a workable notion of orthogonal Fourier analysis on arbitrary directed acyclic graphs. This subclass of graphs is of high significance since it captures data with causal relationships. Our work makes it possible to bring Fourier-based signal processing methods to processing such data.
hat are the three papers in the published literature most closely related to this paper? Please provide full citation details, including DOI references where possible.	A. Sandryhaila and J. M. F. Moura, "Discrete Signal Processing on Graphs: Frequency Analysis," in IEEE Transactions on Signal Processing, vol. 62, no. 12, pp. 3042-3054, June15, 2014, doi: 10.1109/TSP.2014.2321121. B. Seifert, C. Wendler and M. PÃ¼schel, "Causal Fourier Analysis on Directed Acyclic Graphs and Posets," in IEEE Transactions on Signal Processing, vol. 71, pp. 3805-3820, 2023, doi: 10.1109/TSP.2023.3324988. R. Shafipour, A. Khodabakhsh, G. Mateos and E. Nikolova, "A Directed Graph Fourier Transform With Spread Frequency Components," in IEEE Transactions on Signal Processing, vol. 67, no. 4, pp. 946-960, 15 Feb.15, 2019, doi: 10.1109/TSP.2018.2886151.
What is distinctive/new about the current	This paper introduces the first orthogonal, scalable

paper relative to these previously published works?	framework specifically designed for signal processing on directed acyclic graphs. Our proposed model is entirely novel, distinguished by the following features: it is exclusively designed for directed acyclic graphs, adopts a Fourier basis that is both orthogonal and scalable, and is built upon a generalization of the classical total variation concept for discrete-time signals.
Are there any preprints of the manuscript (i.e. preprints that are identical to the submitted manuscript, except for minor differences) that have been posted on the authors' personal website, employer's website or institutional repository, arXiv.org, TechRxiv.org, or on any not-for-profit preprint server approved by the IEEE? **Please refer to IEEE Post-Publications Policies and IEEE Post-Publications Policy Infographic for complete information on the IEEE preprint posting policy.** **Please enter "Yes" or "No". If "Yes", please also provide a detailed explanation.** **All other posted preprints (i.e., preprints by the authors that are not identical to the submitted manuscript, and all other preprints by the authors and/or by other authors) are considered to be prior art. Therefore, for the purpose of evaluating novelty as well as potential plagiarism and self-plagiarism, these other preprints should be included in the submitted manuscript's reference list whenever relevant, following common practice. However, if the submitted manuscript is based on one or more other posted preprints, which are not identical to the submitted manuscript but which the authors nevertheless wish to exclude as prior art for the purpose of evaluating novelty, potential plagiarism, and self-plagiarism, authors should explain in detail how these preprints relate to the manuscript, and why these preprints should not be considered prior art. The explanation of how these preprints relate to the manuscript should also be included in a footnote in the manuscript, with the preprints included in the submitted manuscript's reference list.**	No
Are there any other posted preprints that should not be considered to be prior art?	No

Please enter "Yes" or "No". If "Yes", please also provide a detailed explanation.	

Orthogonal Fourier Analysis on Directed Acyclic Graphs via Möbius Total Variation

Vedran Mihal and Markus Püschel

***Abstract*—Signal processing on directed acyclic graphs (DAGs) presents unique challenges. Unlike for undirected graphs, the Laplacian matrix of a DAG lacks a complete eigenbasis in general, and the corresponding adjacency matrix never has one, which prevents the use of conventional graph Fourier analysis. To address these challenges, in this work, we propose the first orthogonal Fourier basis that is uniquely designed for DAGs, referred to as the Möbius basis. This basis is derived from our novel concept of Möbius total variation (Möbius TV) for DAGs, which is a direct generalization of the classical definition of TV in discrete time, which also constitutes a DAG. We then develop two variants of the Möbius basis, which have desirable properties with respect to constant signals. We validate the effectiveness of our model through three types of experiments, applying it to both synthetic and real-world DAGs and DAG signals.**

***Index Terms*—Graph signal processing, orthogonal Fourier basis, Fourier transform, DAG, Möbius inversion**

I. Introduction

Graph signal processing (GSP) generalizes classical signal processing to data (signals) defined on the nodes of a graph. Most of the work in the field is focused on undirected graphs, where there are two commonly used SP frameworks or models: one based on the Laplacian matrix [1] and one based on the adjacency matrix [2]. Both generalize classical SP concepts like Fourier basis, frequency, spectrum and total variation.

Directed graphs. For directed graphs, GSP becomes problematic since both the adjacency matrix and the Laplacian are not necessarily diagonalizable. And if they are, the resulting Fourier basis is usually not orthogonal. Furthermore, the definition of a directed Laplacian is ambiguous.

To overcome these issues, several meaningful signal processing frameworks for directed graphs have been proposed. [3] develops a model based on the directed Laplacian, but the Fourier basis does not exist in general. [4], [5] and [6] obtain an orthogonal basis by solving an optimization problem but this is costly and does not scale well. [7] defines a Hermitian Laplacian for directed graphs, but the Fourier basis becomes complex and is not unique, since it depends on the choice of the required rotation parameter q. The work in [8] adds edges to the graph to make the adjacency matrix diagonalizable, but the obtained Fourier basis is not orthogonal.

Directed acyclic graphs. Directed acyclic graphs (DAGs) are in a sense the hardest case for defining a GSP model since, in particular, their adjacency matrices are strictly upper triangular, i.e., all eigenvalues are zero and there is no eigenbasis. At the same time, DAGs constitute a very important subclass of directed graphs: they model data with causal dependencies (which precludes cycles) and Bayesian networks, as well as multivariate time-series data.

DAGs have a crucial role in two important fields: causal inference [9]–[12] and multivariate time series analysis [13]–[15]. Causal inference is a fundamental aspect of many scientific disciplines [16], where DAGs are used to represent causal relationships and structure of the data. It has applications within many areas, primarily machine learning [17], [18], but also medicine [19], social science [20], etc. Similarly, DAGs are important in multivariate time-series analysis, where one of the primary objectives is to discover causal relationships between temporal components [21], represented by a DAG. Applications include stock price data [22], weather networks [23], and medicine [24].

Given the importance and extensive use of DAGs, there is a need for a robust graph signal processing framework tailored specifically for DAGs. However, the shortcomings of the previously mentioned GSP models for directed graphs become rather amplified for DAGs. For example, the work in [8] necessarily adds cycles, which destroys the DAG structure and is undesirable on causal data. The work in [25] defines a Fourier basis specifically for DAGs, but the corresponding Fourier basis is triangular and thus never orthogonal. It is also based on imposing a causal model on the data which may not be desirable in all applications.

This motivates our proposal of a novel orthogonal Fourier analysis model for directed acyclic graphs.

Contributions. In this paper we propose a GSP model specifically designed for DAGs that is based on a novel definition of orthogonal Fourier transform. We derive this transform via a new form of a difference operator and the associated total variation, by leveraging the theory of Möbius inversion from combinatorics. Thus, we refer to our contribution as the Möbius model. We then experimentally validate the proposed model, demonstrating its practical effectiveness. This paper expands the preliminary work presented in [26].

In particular, we present the following contributions:

- We define a novel difference operator for DAGs, called the Möbius difference operator, together with a corresponding notion of Möbius total variation. We obtain it by generalizing the classical TV for infinite discrete-time signals. The Möbius difference operator leverages the partial order structure provided by DAGs, which requires the absence of cycles.
- Based on the Möbius difference operator we derive a novel Fourier basis for DAGs, called the Möbius basis. This basis has the significant property of being orthogonal.
- We define two variants of the Möbius model, called the sourceless Möbius and the balanced Möbius model, which provide desirable properties related to total variation of constant signals.

- We perform three types of experiments, namely spectral analysis, denoising, and reconstruction from samples, with both synthetic and real-world signals. The results improve over previously established GSP models and demonstrate the utility of our newly proposed Möbius models for DAGs.

II. Signal Processing on Graphs

We recall the basic concepts from graph theory and graph signal processing [1], [27] needed in this paper.

Graph. We consider undirected and directed graphs jointly as $\mathcal{G} = (\mathcal{V}, A)$, where $\mathcal{V} = (v_1, \dots, v_n)$ is the set of n vertices and A is the $n \times n$ adjacency matrix. $A_{ij} = 1$ signifies an edge from v_j to v_i, and is $= 0$ otherwise. If A is symmetric, G is undirected, and directed otherwise. The number of edges is denoted with m. If $\mathcal{G}$ is directed, we denote with $\overline{A}$ the adjacency matrix of the underlying undirected graph (where each directed edge is replaced by an undirected one).

DAG. A directed acyclic graph (DAG) is a directed graph without cycles. In this case we assume the vertex set to be topologically sorted, which means that an edge from v_j to v_i implies $j < i$. This way, A becomes lower triangular with 0s on the diagonal.

Graph signal. A graph signal $\mathbf{x}$ on $\mathcal{G}$ is a vector $\mathbf{x} = (x_1, \dots, x_n)$ where value x_i is associated with vertex v_i.

Degree. For undirected graphs, the degree of a vertex v_i is defined by $\deg(v_i) = \sum_{j=1}^{n} A_{ij}$. The diagonal degree matrix is then $D = \operatorname{diag}(\deg(v_1), \dots, \deg(v_n))$.

For directed graphs, we distinguish between the in-degree and the out-degree of a vertex v_i, defined as $\deg_{\text{in}}(v_i) = \sum_{j=1}^{n} A_{ij}$ and $\deg_{\text{out}}(v_i) = \sum_{j=1}^{n} A_{ji}$, respectively. The in-degree matrix is defined as $D_{\text{in}} = \operatorname{diag}(\deg_{\text{in}}(v_1), \dots, \deg_{\text{in}}(v_n))$ and the out-degree matrix D_{out} analogously.

Laplacian. The graph Laplacian L for undirected graphs is $L = D - A$, where D is the degree matrix and A is the adjacency matrix. For directed graphs, there are several definitions [3]; we adopt $L = D_{\text{in}} - A$.

Incidence matrix. The oriented incidence matrix B of an undirected graph is of size $m \times n$. If an edge e_i connects vertices v_j and v_k where $j < k$, then $B_{ij} = 1$, $B_{ik} = -1$, and all other elements of row i are $= 0$. This definition yields the equality $L = B^\top B$.

Fourier basis. One general approach to obtaining a Fourier basis for any signal domain is through the eigendecomposition of a suitable chosen shift operator [27]. For undirected graphs the two common choices are L in [1] or A in [2]. Both L and A are symmetric and thus have an orthogonal eigenbasis and inverting the eigenmatrix yields the associated Fourier transform. For directed graphs, the eigendecompositions do not exist in general and if they exist, the bases are usually not orthogonal. DAGs are particularly troublesome, since, for example, A has only one eigenvalue zero in this case.

III. Fourier Basis via Total Variation

In this paper we adopt, and then build on the perspective that a graph Fourier basis is derived from a chosen total variation (TV) operator instead of a shift operator, a viewpoint that also underlied the choice of L in [1].

Total variation operators. The general form of TV of a graph signal $\mathbf{x}$ is

$$\operatorname{TV}(\mathbf{x}) = \|\Delta\mathbf{x}\|_p^p, \text{ with } p = 1, 2,$$

where Δ is a chosen linear difference operator. Table I shows in the first three rows the most commonly used TVs. Each has an intuitive interpretation: it accumulates, in different ways, the L1-norm or L2-norm distances between values of nodes connected by a directed or undirected edge. Note that for $p = 2$, $\operatorname{TV}(\mathbf{x}) = \mathbf{x}^\top \Delta^\top \Delta \mathbf{x}$.

The Fourier basis associated with a TV is now again obtained via eigendecomposition where possible, and the eigenvectors sample the range of possible TVs. For the L1-norm based TVs, it is the eigenbasis of Δ; for the L2-norm case, it is the eigenbasis of $\Delta^\top \Delta$. Thus, we obtain the same choices as with the shift view above and with the same restrictions in the case of directed graphs.

Comments.

- The adjacency model is completely undefined for DAGs because all eigenvalues of the adjacency matrix of a DAG are zero, i.e., $\lambda_{\max} = 0$ (where $\lambda_{\max}$ is the eigenvalue of A with the largest magnitude).
- The Laplacian Fourier bases (second and third row) coincide for undirected graphs.
- It is possible to use $\left\| \left(I - \frac{A}{|\lambda_{\max}|} \right) \mathbf{x} \right\|_2^2$ or $\|B\mathbf{x}\|_1$ as TV, but to our knowledge it has not been studied.
- The decomposition of the undirected Laplacian into $L = \Delta^\top \Delta$ is not unique. Thus, B is not the only Δ that would yield the TV in the third row of Table I.
- Fourier bases associated with $\|\Delta x\|_2^2$ enjoy a nice property. They iteratively minimize (or maximize) this TV [28]. Namely, the first basis vector is obtained by finding the $\arg\min_{\mathbf{x}} \|\Delta\mathbf{x}\|_2^2$, and subsequent vectors are obtained by iteratively solving the same minimization problem in the orthogonal complement of the space spanned by previously obtained basis vectors. Thus, in a sense, the basis vectors orthogonally slice the space of possible TVs.
- Other TVs have been proposed, e.g., Table I from [29] shows variants of L and A, and [5] proposes a directed TV, which is then used to obtain a Fourier basis.

Our work (fourth row in Table I) defines a TV operator that is specifically designed for DAGs and uses it to obtain an orthogonal Fourier basis.

IV. Möbius model for DAGs

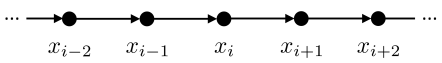
We define the Möbius TV, first proposed in our preliminary work [26] via an intuitive linear difference operator specifically designed for DAGs. In the second step, the Möbius TV then yields an orthogonal Fourier basis as explained before in Section III.

A. Möbius total variation

Our proposed Möbius TV for DAGs is a generalization of the classical TV for infinite discrete-time signals, which can be viewed as associated with the infinite directed path shown in Fig. 1a, which is a DAG.

Model	TV	Undirected graphs	Directed graphs	DAGs
Adjacency	$\left\| \left(I - \frac{A}{\lvert\lambda_{\max}\rvert} \right) \mathbf{x} \right\|_1$	yes	iff A has an eigenbasis	no
Laplacian (1-norm)	$\|(D-A)\,\mathbf{x}\|_1 = \|L\mathbf{x}\|_1$	yes	iff L has an eigenbasis	iff L has an eigenbasis
Laplacian (2-norm)	$\|B\mathbf{x}\|_2^2 = \mathbf{x}^\top L\mathbf{x}$	yes	B is not defined	B not defined
Möbius (this work)	$\|M\mathbf{x}\|_2^2$	M is not defined	M is not defined in general	**yes**

TABLE I: Existence of a Fourier basis via eigendecomposition for different TVs and classes of graphs.



(a) Infinite directed path graph.

$$\Delta = \begin{pmatrix} \ddots & \vdots & \vdots & \vdots & \vdots & \\ \cdots & 1 & 0 & 0 & 0 & \cdots \\ \cdots & -1 & 1 & 0 & 0 & \cdots \\ \cdots & 0 & -1 & 1 & 0 & \cdots \\ \cdots & 0 & 0 & -1 & 1 & \cdots \\ & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

(b) Difference operator

$$\Sigma = \begin{pmatrix} \ddots & \vdots & \vdots & \vdots & \vdots & \\ \cdots & 1 & 0 & 0 & 0 & \cdots \\ \cdots & 1 & 1 & 0 & 0 & \cdots \\ \cdots & 1 & 1 & 1 & 0 & \cdots \\ \cdots & 1 & 1 & 1 & 1 & \cdots \\ & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

(c) Integration operator

Fig. 1: Infinite directed path and the matrix representations of the corresponding difference and integration operators.

Infinite discrete-time. The classical TV [30] of a discrete-time signal $\mathbf{x} = (x_i)_{i\in\mathbb{Z}}$ is

$$\mathrm{TV}(\mathbf{x}) = \sum_i |x_i - x_{i-1}| = \|\Delta\mathbf{x}\|_1\,, \tag{1}$$

where Δ is the corresponding difference operator:

$$(\Delta\mathbf{x})_i = x_i - x_{i-1}.$$

The (infinite) matrix Δ is shown in Fig. 1b.

Crucial for our derivation later, Δ is the inverse of the integration operator

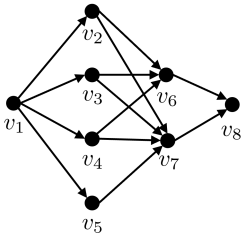
$$(\Sigma\mathbf{x})_i = \sum_{j\le i} x_j \tag{2}$$

shown as a matrix in Fig. 1c. Due to the infinite domain, $\Sigma\mathbf{x}$ only exists if $\mathbf{x}$ is summable.

TV for arbitrary (finite) DAGs. The infinite discrete-time integration operator Σ in (2) sums, for a given node, the values of all predecessor nodes. We generalize it by doing the same on arbitrary DAGs with n nodes. We write $v_j \preceq v_i$ if v_j is a predecessor of v_i, which means there is a directed path from v_j to v_i. Relation $\preceq$ defines a partial order on the vertex set $\mathcal{V} = (v_1, \ldots, v_n)$ since DAGs have no cycles. We define the integration operator as

$$(Z\mathbf{x})_i = \sum_{v_j\preceq v_i} x_j.$$

We name it Z instead of Σ since it is known as the zeta function in combinatorics [31].



(a) DAG on 8 vertices

$$M = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 2 & -1 & -1 & -1 & 0 & 1 & 0 & 0 \\ 3 & -1 & -1 & -1 & -1 & 0 & 1 & 0 \\ -2 & 1 & 1 & 1 & 0 & -1 & -1 & 1 \end{pmatrix}$$

(b) Möbius difference operator

$$Z = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

(c) Zeta integration operator

Fig. 2: A DAG on 8 vertices and the corresponding difference and integration operators. The elements of the Möbius matrix can be arbitrary integers.

The inverse of Z is the corresponding difference operator M, which is known as the Möbius transform in combinatorics [31]. It can be computed using the Möbius inversion formula [31]:

$$M_{ij} = \begin{cases} 1 & \text{if } i = j, \\ -\sum\limits_{v_j \preceq v_k \prec v_i} M_{kj} & \text{otherwise.} \end{cases}$$

Both, Z and M are lower triangular with 1s on the diagonal.

As a result, the Möbius difference operator M defines a notion of TV for a DAG signal $\mathbf{x}$ that we call the Möbius TV:

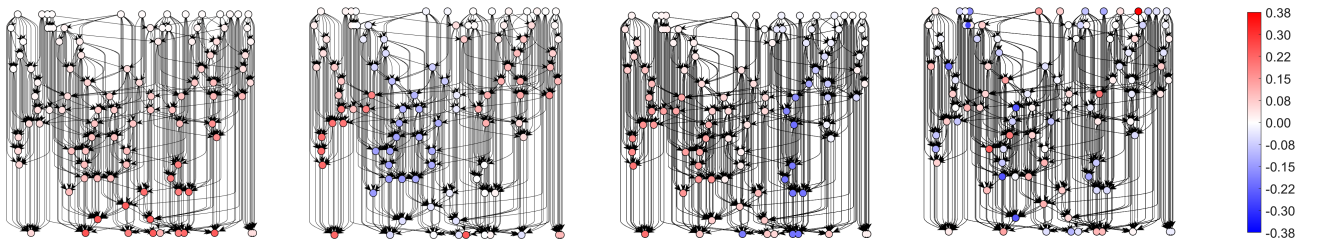
$$\mathrm{TV}_M(\mathbf{x}) = \|M\mathbf{x}\|_2^2\,.$$

We choose the L2-norm to obtain an orthogonal basis, as explained next.

Example. Fig. 2 shows an example, which also demonstrates that M is different from prior difference operators on graphs.

B. Möbius basis

Based on the Möbius total variation, we define the associated orthogonal Fourier basis on a DAG $\mathcal{G}$ with n vertices:



(a) First basis vector (b) Second basis vector (c) Third basis vector (d) 50th basis vector (e) Scale

Fig. 3: Example Möbius basis vectors on a random geometric DAG with 100 vertices. The first three vectors illustrate the notion of smoothness introduced by the Möbius basis.

Definition IV.1. *A DAG $\mathcal{G}$ with n vertices and Möbius difference operator M has an associated orthogonal Fourier basis V, which is the eigenbasis of $M^\top M$, i.e., $V^\top M^\top M V = \Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. We call V the Möbius basis, so $\mathcal{F} = V^{-1} = V^\top$ is the associated Fourier transform.*

Example. As an example, we generate in Fig. 3 a random geometric DAG on 100 vertices (nodes ordered by edge direction from top to bottom). Namely, we generate $n = 100$ random points $p_1, \ldots, p_{100}$ in the square $[0,1] \times [0,1]$, which represent vertices $v_1, \ldots, v_{100}$. We add an edge from v_i to v_j iff $\|p_i - p_j\|_2^2 < 0.035$ and $i < j$, to obtain a DAG. We order the associated Möbius Fourier basis by TV from low to high. The first three basis vectors in Fig. 3 are thus considered smooth and the 50th basis vector is an example of a non-smooth signal.

C. Properties

We discuss some properties of the proposed Möbius TV and its associated Fourier basis.

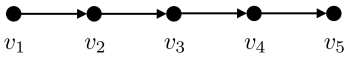
Uniqueness. The Möbius basis is obtained via an eigendecomposition of $M^\top M$ and there is only one Möbius difference operator M that yields this matrix:

Theorem IV.2. *The Möbius difference operator M is uniquely retrievable from $M^\top M$.*

Proof. All diagonal entries of M are $= 1$ and $M^\top M$ is positive definite so it has a unique Cholesky decomposition (upper triangular version) into $M^\top$ and M when fixing the diagonal entries of M to be positive. □

Invertibility and zero TV. Prior difference operators for undirected or directed graphs (Table I) are typically not invertible, which is equivalent to the existence of signals with $\mathrm{TV} = 0$. For example, for the Laplacian-based TVs, constant signals have $\mathrm{TV} = 0$, a property that is often considered desirable. As a consequence, in these cases the associated integration operator is not uniquely defined, just as in standard continuous calculus.

In contrast, M is invertible since it has determinant 1 and thus only the zero signal has zero Möbius TV, which may be undesirable. To tackle this issue, we present in the next section several, arguably intuitive variants of the Möbius difference operator (which differ from the ones we originally proposed in [26]). As a starting point to motivate these model variants, we conclude this section by instantiating the Möbius TV on a (finite) directed path graph.



(a) Directed path on 5 vertices

$$M = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix} \qquad Z = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

(b) Möbius difference operator (c) Zeta integration operator

Fig. 4: Directed path graph and the corresponding difference and integration operators.

D. Möbius model on a finite directed path

We analyze the Möbius TV on the simplest example of a DAG: the finite directed path graph shown in Fig. 4 together with M and Z. The property that stands out is that the value of the (unique in this case) source vertex remains unaltered after applying M, i.e., no notion of difference is taken. For example, when applied to a constant signal $c\mathbf{1} = (c, \ldots, c)^\top$, we get

$$M \cdot c\mathbf{1} = (c, 0, \ldots, 0)^\top \quad \text{and thus} \quad \mathrm{TV}_M(c\mathbf{1}) = c^2.$$

This is counterintuitive and prevents constant signals from having zero TV. It serves as the motivation for the Möbius model variants presented next.

V. Möbius Model Variants

In this section we present two variants of the Möbius difference operator and thus the Möbius TV. The first one, called the *sourceless Möbius TV*, solves the aforementioned problem that occurs for directed paths, i.e., the constant signals on directed paths (and on any single-source DAG) have the sourceless Möbius TV equal to zero. The second variant, called the *balanced Möbius TV*, has the property that constant signals have zero total variation for any directed acyclic graph. We point out that these variants are different from the ones defined in [26].

Notation. Before we derive the Möbius operator variants, we need additional notation. Let $\mathbf{e}_i$ denote the ith canonical base vector ($= 1$ at position i and $= 0$ otherwise). Let $\mathcal{S} = \{i_1, \ldots, i_k\} \subseteq \mathcal{V}$ be the sources of $\mathcal{G}$, so $|\mathcal{S}| = k$. Let $\mathbf{1}_\mathcal{S} \in \mathbb{R}^n$ be the indicator vector of the set of sources $\mathcal{S}$, i.e., $(\mathbf{1}_\mathcal{S})_{i_j} = \sum_{j=1}^k \mathbf{e}_{i_j}$, and let $\mathbf{1}_{\mathcal{V}\setminus\mathcal{S}} = \mathbf{1} - \mathbf{1}_\mathcal{S}$ be the indicator vector of the set $\mathcal{V} \setminus \mathcal{S}$.

A. Sourceless Möbius difference operator

The directed path example (Fig. 4) illustrates that the source values are carried over unaltered after the Möbius differentiation. To address this issue, the idea is to remove the source values when applying the Möbius difference operator. The rows of M corresponding to the source nodes have 1 on the diagonal and are $= 0$ elsewhere, so we set these diagonal elements to 0 to obtain the sourceless Möbius difference operator:

Definition V.1. *If M is the Möbius difference operator, the sourceless Möbius difference operator $\widehat{M}$ is defined as*

$$\begin{aligned}\widehat{M} &= \operatorname{diag}\left(\mathbf{1}_{\mathcal{V}\setminus\mathcal{S}}\right)\cdot M\\ &= \operatorname{diag}\left(\mathbf{1}-\mathbf{1}_{\mathcal{S}}\right)\cdot M.\end{aligned}$$

The corresponding notion of the sourceless Möbius TV is thus

$$TV_{\widehat{M}}(\mathbf{x}) = \left\|\widehat{M}\mathbf{x}\right\|_2^2,$$

and the corresponging Fourier basis is the eigenbasis of $\widehat{M}^\top\widehat{M}$.

We consider a few examples.

Example V.2. *For a directed path on n vertices, the sourceless Möbius operator has the following form:*

$$\widehat{M} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 & 0\\ -1 & 1 & 0 & \cdots & 0 & 0 & 0\\ 0 & -1 & 1 & \cdots & 0 & 0 & 0\\ & & & \ddots\ \ddots\ \ddots & & & \\ 0 & 0 & 0 & \cdots & -1 & 1 & 0\\ 0 & 0 & 0 & \cdots & 0 & -1 & 1 \end{pmatrix}$$

This ensures that constant signals, and in this case only those, have TV $= 0$.

Example V.3. *For a graph with n vertices and no edges (which is a DAG), the Möbius operator is the identity matrix. On the other hand, the sourceless Möbius operator of the same graph is the all-zero matrix, so the corresponding total variation is always zero (as expected for a graph with no edges).*

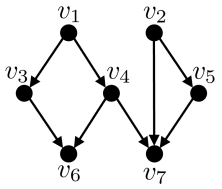
Example V.4. *Figure 5 shows the Möbius and the sourceless Möbius operator for a DAG with two sources. Figure 5d shows two signals which have sourceless Möbius TV equal to zero.*

In general, the rank of $\widehat{M}$ is $n-k$ (where k is the number of sources), so the space of signals with zero TV is k-dimensional. A basis for that space is associated with the source nodes:

Theorem V.5. *The source nodes introduce a basis for the space of signals with the sourceless Möbius TV $= 0$. In particular, every source i_j introduces one basis vector which is $= 1$ at i_j and all its successors, and $= 0$ elsewhere.*

Proof. First we show that these vectors have sourceless Möbius TV equal to zero.

The vector introduced by the source i_j corresponds to the column i_j of the zeta matrix Z, which we denote with $\mathbf{v}_{i_j}$. Since Z is the inverse of M, it follows that $M\mathbf{v}_{i_j} = \mathbf{e}_{i_j}$, so $\widehat{M}\mathbf{v}_{i_j} = \operatorname{diag}\left(\mathbf{1}_{\mathcal{V}\setminus\mathcal{S}}\right)\mathbf{e}_{i_j} = \mathbf{0}$.



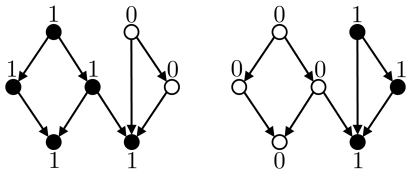
(a) DAG with two sources

$$M = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 1 & 0 & 0 & 0 & 0 & 0\\ -1 & 0 & 1 & 0 & 0 & 0 & 0\\ -1 & 0 & 0 & 1 & 0 & 0 & 0\\ 0 & -1 & 0 & 0 & 1 & 0 & 0\\ 1 & 0 & -1 & -1 & 0 & 1 & 0\\ 0 & 0 & 0 & -1 & -1 & 0 & 1 \end{pmatrix}$$

(b) Möbius difference operator

$$\widehat{M} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 0\\ -1 & 0 & 1 & 0 & 0 & 0 & 0\\ -1 & 0 & 0 & 1 & 0 & 0 & 0\\ 0 & -1 & 0 & 0 & 1 & 0 & 0\\ 1 & 0 & -1 & -1 & 0 & 1 & 0\\ 0 & 0 & 0 & -1 & -1 & 0 & 1 \end{pmatrix}$$

(c) Sourceless Möbius difference operator



(d) Two signals with zero sourceless Möbius TV. Black vertices are successors of sources v_1 and v_2 respectively.

Fig. 5: Directed acyclic graph with two sources, the Möbius and the sourceless Möbius operator, signals with zero TV.

As a submatrix of the invertible Z, these k vectors (one for every source) are linearly independent, and thus form a basis. □

As an example, the signals in Fig. 5d form a basis for the space of signals with the sourceless Möbius TV $= 0$ on the DAG in Fig. 5a.

As a corollary, we conclude that for graphs with only one source, the only signals with zero sourceless Möbius TV are the constant signals. However, the existence of several sources changes the notion of being constant for DAGs.

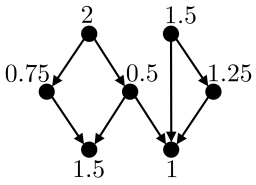
B. Balanced Möbius difference operator

We still want to define a variant of the Möbius TV for which only constant signals have TV $= 0$. For the sourceless Möbius TV only graphs with one source have that property, which motivates our construction: we consolidate all sources into one source.

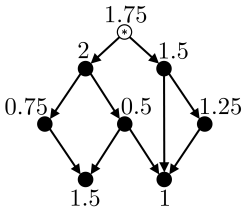
Namely, we add a virtual supersource vertex v^* to a DAG $\mathcal{G}$, connect it to all sources, and set its value to be the average of all source values, i.e., $(1/k)\sum_{j=1}^{k} x_{i_j} = \mathbf{1}_{\mathcal{S}}^\top\mathbf{x}/\left|\mathcal{S}\right|$. We denote the extended DAG with $\mathcal{G}^*$ and the extended signal with $\mathbf{x}^* = \left(\mathbf{1}_{\mathcal{S}}^\top\cdot\mathbf{x}/\left|\mathcal{S}\right|,\ \mathbf{x}\right)$. The construction is illustrated with an example in Fig. 6.

The zeta and the Möbius operator of the extended DAG $\mathcal{G}^*$ are

$$Z^* = \begin{bmatrix} 1 & \mathbf{0}^\top\\ \mathbf{1} & Z \end{bmatrix},\quad M^* = \begin{bmatrix} 1 & \mathbf{0}^\top\\ -M\cdot\mathbf{1} & M \end{bmatrix}.$$

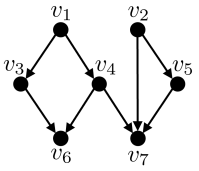


(a) DAG $\mathcal{G}$ with two sources (top nodes) and signal $\mathbf{x}$



(b) Extended DAG $\mathcal{G}^*$ with extended signal $\mathbf{x}^*$

Fig. 6: Construction of the extended DAG to obtain the balanced Möbius difference operator.



(a) DAG $\mathcal{G}$

$$\overline{M} = \begin{pmatrix} 0.5 & -0.5 & 0 & 0 & 0 & 0 & 0 \\ -0.5 & 0.5 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & -1 & -1 & 0 & 1 & 0 \\ 0.5 & 0.5 & 0 & -1 & -1 & 0 & 1 \end{pmatrix}$$

(b) Balanced Möbius diff. operator

Fig. 7: Directed acyclic graph with two sources and its balanced Möbius difference operator.

Now we apply the sourceless Möbius operator to the extended DAG (and the extended signal) to obtain the proposed balanced Möbius operator $\widetilde{M}$:

$$\begin{aligned} \widetilde{M}\mathbf{x} = \widehat{M^*}\mathbf{x}^* &= \begin{bmatrix} 0 & \mathbf{0}^\top \\ \mathbf{0} & I \end{bmatrix} M^*\mathbf{x}^* \\ &= \begin{bmatrix} -M \cdot \mathbf{1} & M \end{bmatrix} \cdot \begin{bmatrix} \frac{\mathbf{1}_{\mathcal{S}}^\top}{|\mathcal{S}|} \cdot \mathbf{x} \\ \mathbf{x} \end{bmatrix} \\ &= M \cdot \left(I - \frac{\mathbf{1}\mathbf{1}_{\mathcal{S}}^\top}{|\mathcal{S}|} \right) \cdot \mathbf{x}. \end{aligned}$$

Definition V.6. *The balanced Möbius operator of a DAG $\mathcal{G}$ with the set of sources $\mathcal{S}$ is defined as*

$$\widetilde{M} = M \cdot \left(I - \frac{\mathbf{1}\mathbf{1}_{\mathcal{S}}^\top}{|\mathcal{S}|} \right).$$

The corresponding notion of total variation is defined as

$$\mathrm{TV}_{\widetilde{M}}(\mathbf{x}) = \left\| \widetilde{M}\mathbf{x} \right\|_2^2,$$

and the corresponding Fourier basis is the eigenbasis of $\widetilde{M}^\top \widetilde{M}$.

Theorem V.7. *For any directed acyclic graph, only constant signals have the balanced Möbius TV $= 0$.*

Proof.

$$\begin{aligned} M \cdot \left(I - \frac{\mathbf{1}\mathbf{1}_{\mathcal{S}}^\top}{|\mathcal{S}|} \right) \mathbf{x} = 0 &\Leftrightarrow \left(I - \frac{\mathbf{1}\mathbf{1}_{\mathcal{S}}^\top}{|\mathcal{S}|} \right) \mathbf{x} = \mathbf{0} \\ &\Leftrightarrow \mathbf{x} = \frac{\mathbf{1}_{\mathcal{S}}^\top \mathbf{x}}{|\mathcal{S}|} \cdot \mathbf{1} \\ &\Leftrightarrow \mathbf{x} \text{ is constant.} \end{aligned}$$

□

Fig. 7 shows an example of the balanced Möbius difference operator for a DAG with two sources. The elements are not necessarily integers.

For directed paths, the sourceless and the balanced Möbius operator are identical. More generally we have the following result:

Theorem V.8. *The sourceless and the balanced Möbius operators coincide for DAGs with only one source.*

Proof. For a DAG with one source, the first column of matrix Z is $\mathbf{1}$ and the first row of matrix M is $\mathbf{e}_1^\top$. The following holds:

$$\begin{aligned} \widehat{M} &= \mathrm{diag}(\mathbf{1} - \mathbf{e}_1) \cdot M \\ &= M - \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} \end{aligned}$$

$$\begin{aligned} \widetilde{M} &= M \cdot (I - \mathbf{1} \cdot \mathbf{e}_1^\top) \\ &= M - M \cdot \begin{bmatrix} 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \dots & 0 \end{bmatrix} \\ &= M - \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}. \end{aligned}$$

The last equality follows from the fact that the first column of matrix Z (which is the inverse of M) is $\mathbf{1}$. □

VI. Experiments and Applications

We evaluate our proposed orthogonal Fourier bases for DAGs in several experiments. First, we perform spectral analysis of synthetic and real-world DAG signals, comparing the obtained spectra to those of prior GSP models. Second, we apply our Möbius bases in a signal denoising experiment, showing superior performance compared to prior bases. Finally, we perform a sampling and reconstruction experiment, where the choice of the sampled vertices and the reconstruction depend on the chosen Fourier basis.

To validate the viability of the Möbius framework for DAGs, we will perform all experiments using both the Möbius models and several other previously proposed GSP models for directed and undirected graphs. The chosen models and associated Fourier bases are listed in Table II. Here $\overline{A}$ denotes the adjacency matrix of the underlying undirected graph (obtained by forgetting the direction of edges) and B is the oriented incidence matrix associated with $\overline{A}$. Note that when using the directed Laplacian, we use the eigenbasis of $(D_{\text{in}} - A)^\top (D_{\text{in}} - A)$ as the associated Fourier basis, to ensure existence. Note that all considered Fourier bases are orthogonal, i.e., the Parseval's theorem holds.

A. Graphs and signals

We introduce the DAGs and signals that we use in our experiments.

Synthetic signals. For synthetic signals we consider a linear structural equation model (SEM), which has been the most common one used for generating DAG signals, e.g. [32], [33], including in work on causal inference on DAGs [34]. It is defined by the following equation:

$$\mathbf{x} = W\mathbf{x} + \mathbf{n}, \tag{3}$$

Model	Difference operator Δ	Fourier basis $V = \mathcal{F}^{-1}$
Möbius	M	eigenbasis of $M^\top$
Sourceless Möbius	$\widehat{M}$	eigenbasis of $\widehat{M}^\top \widehat{M}$
Balanced Möbius	$\widetilde{M}$	eigenbasis of $\widetilde{M}^\top \widetilde{M}$
Directed Laplacian	$D_{\text{in}} - A$	eigenbasis of $(D_{\text{in}} - A)^\top (D_{\text{in}} - A)$
Undirected adjacency	$I - \frac{\overline{A}}{\lvert\lambda_{\max}\rvert}$	eigenbasis of $\overline{A}$
Undirected Laplacian	B	eigenbasis of $B^\top B = L$

TABLE II: GSP models used in the experiments. The difference operators Δ are used with $\text{TV}(\mathbf{x}) = \|\Delta\mathbf{x}\|_2$ for the denoising experiments, and the associated Fourier bases V are used for spectral analysis and the sampling and reconstruction experiments. All Fourier bases are orthogonal, i.e., the Parseval's theorem holds.

where $\mathbf{x} \in \mathbb{R}^n$ is the signal, W is the weighted adjacency matrix of the DAG, and $\mathbf{n} \sim \mathcal{N}(0, I)$.

As a synthetic graph we consider a random geometric DAG obtained by generating $n = 100$ random points $p_1, \ldots, p_{100}$ in square $[0, 1] \times [0, 1]$, which represent vertices $v_1, \ldots, v_{100}$. We add an edge from vertex v_i to vertex v_j if and only if $\|p_i - p_j\|_2 < \delta = 0.28$ and $i < j$ to obtain a DAG. To obtain the weighted adjacency matrix W, we assign to each edge a weight from $[0, 1]$, uniformly at random.

In addition, we consider the following three types of real-world DAGs and associated signals used in prior work.

Thames river network. We consider the network of the river Thames and its major tributaries [35], where measurements are taken at 13 different sites once a week. Nodes represent pairs (site, time point), edges connect adjacent sites at the same time point directed as the water flows, and a site with itself at consecutive time points. The resulting graph is a DAG since water flows only downstream, and the signals we consider represent mean daily water flow measured once a week for the years 2013–2016 (four signals).

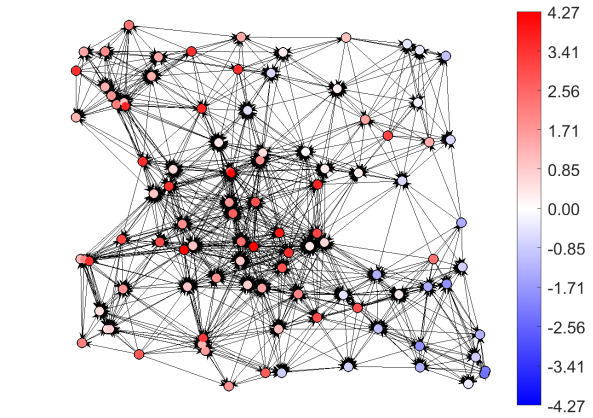
Arth150 network. We consider the gene causal network for *Arabidopsis thaliana* plant (Arth150) [36] with 107 nodes. Nodes represent genes, and directed edges are determined by a complex statistical method described in [37], which creates a DAG. The signals assign to each gene the degree of expression in a cell.

Magic-irri network. Lastly, we consider the dependency Bayesian network for the indica rice population (Magic-irri) [36] with 64 nodes. Nodes represent single nucleotide polymorphisms and phenotypic traits, and edges represent directed stochastic dependencies between them. The network is a DAG, and the signals are allele frequencies of single nucleotide polymorphisms and phenotypic trait measurements [38].

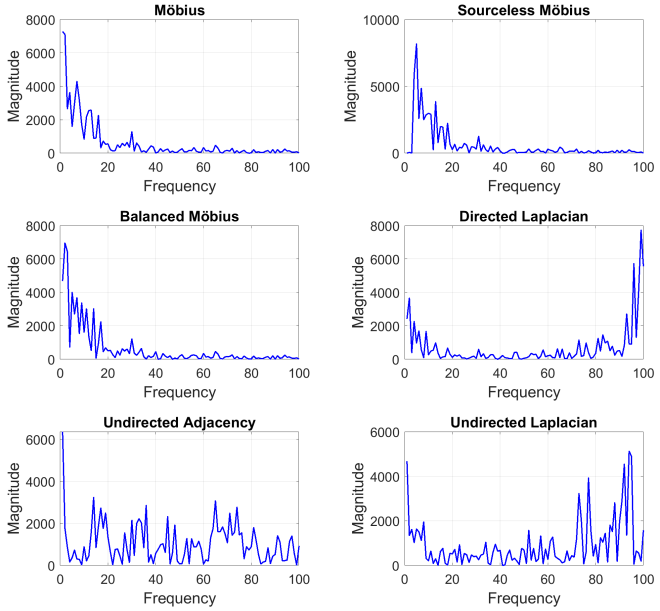
B. Spectral analysis

We analyze the spectra of the considered synthetic and real-world signals, where the spectrum of a signal $\mathbf{x}$ is obtained by performing the Fourier transform ($\mathcal{F}\mathbf{x} = V^{-1}\mathbf{x}$). We show spectra based on the models listed in Table II, which include our Möbius models. The goal is to investigate how well the spectra can capture structure, and in particular identify signals as low-frequency, a property one would expect from many real-world signals.

Synthetic signals. We generate a random geometric DAG and a synthetic signal according to the linear SEM in (3).



(a) Geometric DAG and a synthetic signal in the symmetric logarithmic scale



(b) Spectra of the synthetic signal

Fig. 8: Synthetic signal obtained by a linear DAG model and its spectra according to different Fourier bases.

Fig. 8a shows the generated graph and the signal in the symmetric logarithmic scale. Fig. 8b shows the magnitude spectrum of the signal according to the Möbius, the sourceless Möbius and the balanced Möbius basis. We additionally show the magnitude of the frequency components according to the directed Laplacian, and the adjacency and the Laplacian model of the underlying undirected graph. The x-axis represents the frequency components from low to high frequencies, and the y-axis represents the magnitude of the corresponding frequency component.

The signal is distinctly low-frequency with respect to all Möbius bases. However, it does not exhibit any particular characteristic according to the directed Laplacian, undirected Laplacian and the undirected adjacency model. Thus, our proposed Fourier bases can be a good choice for processing signals based on linear SEMs of the form in (3).

Thames river network signals. Fig. 9 shows the spectra of the mean daily water flow signal from 2013 on the Thames river network. Once again, the signal is low-frequency according to the Möbius models, but not for the others.

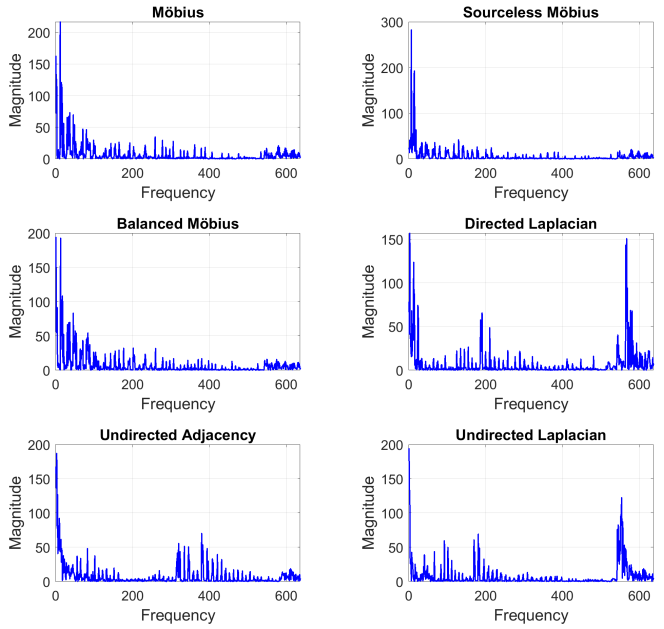


Fig. 9: Spectra of the Thames river network signal according to different bases.

Arth150 network. In Fig. 10 we see the spectra of a signal on the Arth150 network. The signal is low-frequency according to all presented models, except from the undirected adjacency model. This example illustrates that there is an overlap in the notion of low frequency among different models.

Magic-irri network. We show the spectra of an example signal on the Magic-irri network in Fig. 11. Although not distinctly low-frequency in any of the bases, we see that more of the signal is contained in low frequencies with respect to the Möbius and balanced Möbius basis in comparison with other bases. However, it shows that, not surprisingly, not all real-world signals are low-frequency.

The spectra shown in this subsection will help with understanding the denoising experiment that follows.

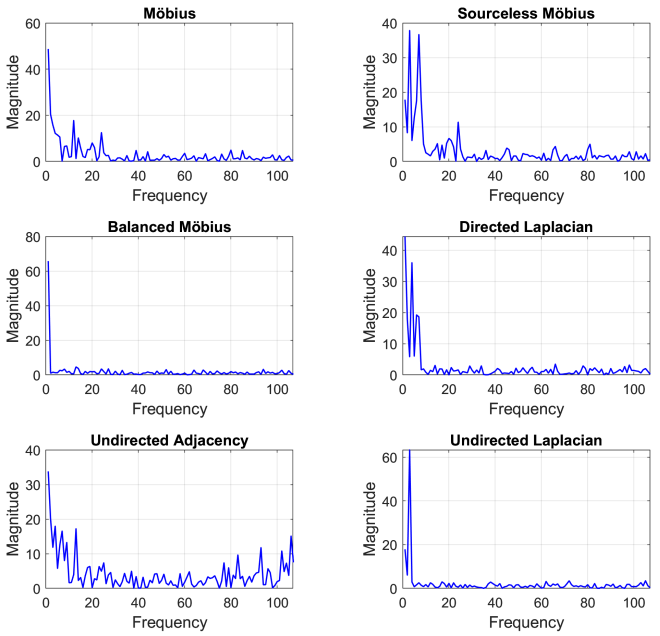


Fig. 10: Spectra of the Arth150 network signal according to different bases.

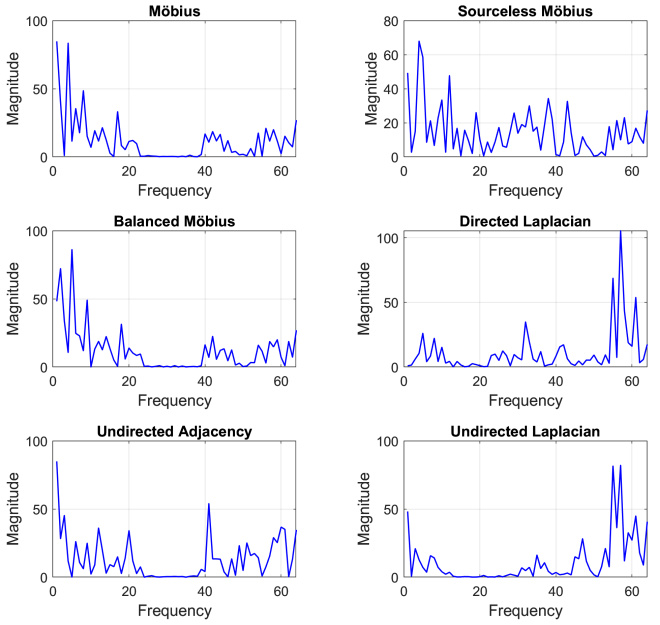
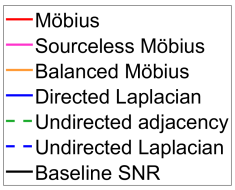


Fig. 11: Spectra of the Magic-irri network signal according to different bases.

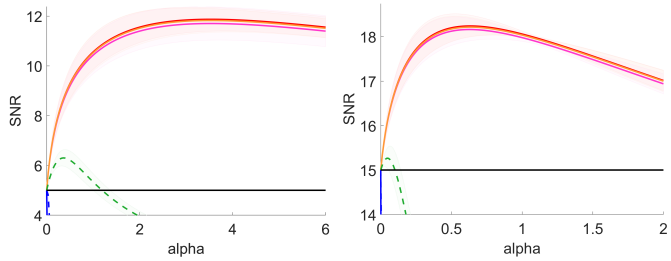
C. Signal denoising

We perform a total-variation-based DAG signal denoising experiment. It depends on the chosen graph difference operator, so we perform it using the newly proposed Möbius difference operators and three prior difference operators listed in Table II.

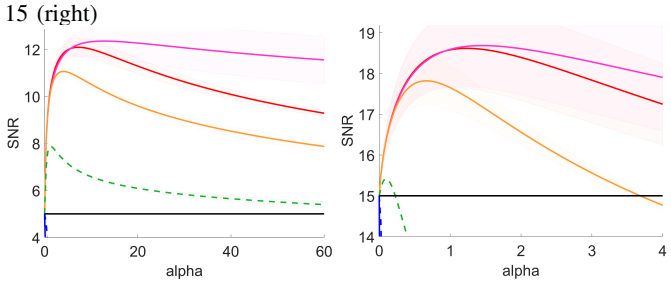
Experimental setup. Let $\mathcal{G}$ be an n-vertex DAG with adjacency matrix A. Let $\mathbf{x}$ be a signal on $\mathcal{G}$, and let Δ be a difference operator on $\mathcal{G}$ from Table II.



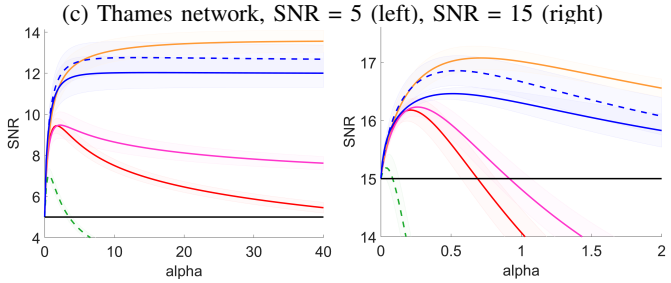
(a) Legend



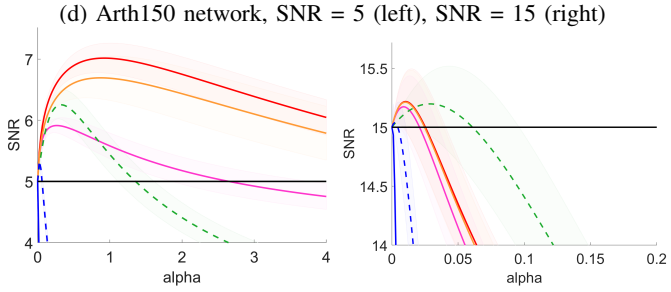
(b) Geometric DAG with synthetic signals, SNR = 5 (left), SNR = 15 (right)



(c) Thames network, SNR = 5 (left), SNR = 15 (right)



(d) Arth150 network, SNR = 5 (left), SNR = 15 (right)



(e) Magic-irri network, SNR = 5 (left), SNR = 15 (right)

Fig. 12: TV denoising on four classes of DAG data with twelve different TV operators. The scales on the axes are different.

We add uncorrelated zero-mean Gaussian noise $\mathbf{e}$ to $\mathbf{x}$ to obtain the noisy signal $\mathbf{y} = \mathbf{x} + \mathbf{e}$, considering two different signal-to-noise ratios (SNRs): 5 and 15. SNR is defined as usual: $10 \cdot \log_{10}\left(\|\mathbf{x}\|_2^2 \,/\, \|\mathbf{e}\|_2^2\right)$.

For a chosen difference operator Δ, we perform TV-based denoising by solving the following problem [39], [40]:

$$\begin{aligned}\hat{\mathbf{x}} &= \arg\min_{\mathbf{x}} \left(\|\mathbf{x}-\mathbf{y}\| + \alpha\, \mathrm{TV}(\mathbf{x})^2\right) \\ &= \arg\min_{\mathbf{x}} \left(\|\mathbf{x}-\mathbf{y}\| + \alpha\, \|\Delta\mathbf{x}\|_2^2\right).\end{aligned}$$

Here, $\hat{\mathbf{x}}$ is the denoised signal and α is a parameter which balances between smoothness (according to Δ) and the reconstruction error. The closed-form solution to this problem is:

$$\hat{\mathbf{x}} = (I + \alpha \cdot \Delta^\top \Delta)^{-1}\mathbf{y}$$

The experiment is designed to effectively denoise signals that are low-frequency according to the selected difference operator Δ. We consider difference operators from Table II and vary α, since finding the optimal α is non-trivial [41], [42].

Results. We perform denoising on the same four DAGs explained previously, and for each DAG we consider four different signals.

The results are shown in Fig. 12. For each DAG there are two plots: one for the low SNR of the noisy signal (SNR = 5) on the left, and one for the high SNR of the noisy signal (SNR = 15) on the right. The x-axis represents the chosen parameter α, and the y-axis represents SNR after denoising (the average among four signals). We vary the scale on the x- and y-axis, to focus on the range of interest for α. Undirected TVs correspond to dashed lines and our Möbius TVs are shown in different red/orange tones. The baseline SNR of $\mathbf{y}$ is a horizontal black line. The legend is in Fig. 12a. Shaded areas around lines represent the standard deviation over the four signals.

The Möbius TVs exhibit the best performance in most cases and, generally, the method's effectiveness is roughly correlated with the extent to which the signals are low-frequency within this model (see previous subsection). Thus, the somewhat poor performance of all TVs on the Magic-irri network is due to the fact that the signals are not really low-frequency according to any of the TV operators used for denoising (see Fig. 11).

In particular, the inferior performance of the undirected models indicates that ignoring directions in a DAG is not always a viable approach, highlighting the necessity for a specialized GSP model for DAGs. We also performed denoising using additional variants of the Laplacian and other difference operators, but the results were not competitive. See [26] for details.

D. Signal sampling and reconstruction

In this experiment we sample signals on DAGs and reconstruct them from samples. The experiment is designed to work well on low-frequency and bandlimited signals, and the choice of the samples and the reconstruction depend on the Fourier basis (and the sampling algorithm). Thus, we perform the experiment using the newly proposed Möbius bases and three other bases for comparison. The Fourier bases are listed in Table II.

Experimental setup. Let $\mathbf{x} \in \mathbb{R}^n$ be a signal on an n-vertex DAG $\mathcal{G}$ with the Fourier basis V. Let K be the number of samples and $V_{(K)}$ the first K columns of V. For the Fourier basis V we choose the sampling operator $\Psi : \mathbb{R}^n \to \mathbb{R}^K$ by applying algorithm 2 from [43], which iteratively removes rows from matrix $V_{(K)}$ in a greedy way to minimize the spectral norm of its Moore-Penrose pseudoinverse, while

maintaining its rank. The matrix obtained after removing $n-K$ rows is $\Psi V_{(K)}$ (with Ψ minimizing $\left\|(\Psi V_{(K)})^{-1}\right\|_2$). Thus the algorithm produces a sampling operator Ψ that is effective for low-frequency signals, as explained next.

We fix a cut-off frequency K and assume signals are low-frequency, so we can represent a signal $\mathbf{x}$ as

$$\mathbf{x} = V_{(K)}\mathbf{s}_{(K)} + \mathbf{e},$$

where $\mathbf{s}_{(K)}$ are the first K frequencies of the spectrum of $\mathbf{x}$, and $\|\mathbf{e}\|$ is not too large compared to $\|\mathbf{x}\|$.

The sampled signal $\Psi\mathbf{x}$ is reconstructed using the following expression:

$$\hat{\mathbf{x}} = V_{(K)}(\Psi V_{(K)})^{-1}\Psi\mathbf{x},$$

where $\hat{\mathbf{x}}$ is the reconstructed signal. For strictly bandlimited signals ($\mathbf{x} = V_{(K)}\mathbf{s}_{(K)}$) the reconstruction is perfect, and for low-frequency signals we have the following:

$$\begin{aligned}\|\hat{\mathbf{x}} - \mathbf{x}\|_2 &= \left\|(V_{(K)}(\Psi V_{(K)})^{-1}\Psi - I)\mathbf{e}\right\|_2 \\ &= \left\|V_{(K)}(\Psi V_{(K)})^{-1}\Psi\mathbf{e}\right\|_2 \\ &\leq \left\|V_{(K)}(\Psi V_{(K)})^{-1}\Psi\right\|_2 \|\mathbf{e}\|_2 \\ &\leq \left\|V_{(K)}\right\|_2 \left\|(\Psi V_{(K)})^{-1}\right\|_2 \|\Psi\|_2 \|\mathbf{e}\|_2 ,\end{aligned}$$

where the upper bound is minimized by the choice of the sampling operator Ψ (minimizing $\left\|(\Psi V_{(K)})^{-1}\right\|_2$ with algorithm 2 from [43]).

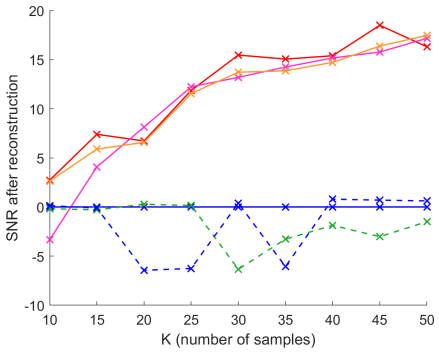
The sampling and reconstruction process depends on the selected Fourier basis V. Therefore, we conduct the experiment using the newly proposed Möbius bases, as well as three other Fourier bases for comparison (see Table II).

Results. We perform the sampling experiment on the same four DAGs and four signals used in the denoising experiment. For each DAG we report the average SNR (among four signals) after sampling and reconstruction in Figure 13. The y-axis represents SNR of the reconstructed signal, and the x-axis represents the number of sampled vertices K. The cut-off number of samples is different for each graph.

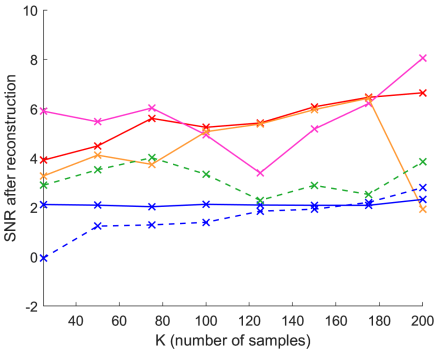
The results show that one of the Möbius bases provides the best reconstruction in most instances, suggesting that the Möbius models are suitable candidates for developing a sampling operator for the DAGs and signals used.

VII. Conclusion

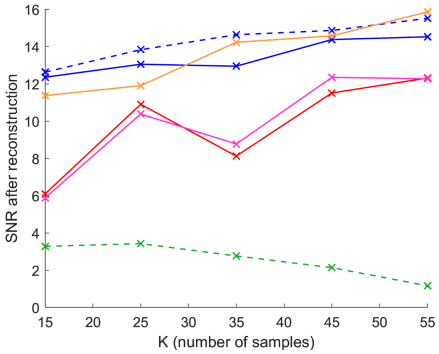
Signal processing on directed acyclic graphs poses significant challenges because the adjacency matrix is non-diagonalizable and all its eigenvalues are zero, leading to a collapse of the spectrum. This work introduces a new orthogonal framework for graph signal processing on DAGs, establishing an orthogonal Fourier basis, named the *Möbius basis*, derived from the Möbius total variation. Two variants of the Möbius basis are proposed, each possessing unique properties related to signals with zero total variation. The practical effectiveness of the proposed framework is demonstrated through experiments on both synthetic and real-world DAG signals, including applications such as signal denoising, sampling, and reconstruction.



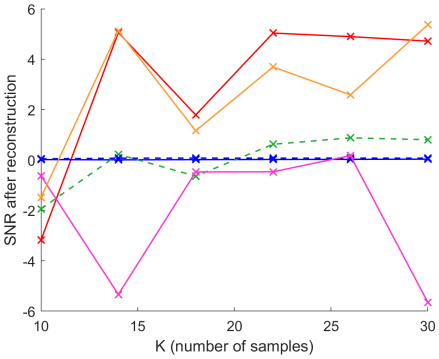
(a) Geometric DAG with synthetic signals



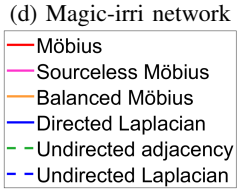
(b) Thames river network



(c) Arth150 network



(d) Magic-irri network



(e) Legend

Fig. 13: SNR after sampling and reconstruction on four classes of DAG data with nine different Fourier bases. The scales on the axes are different.

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 Process. Mag.*, vol. 30, no. 3, pp. 83–98, 2013.

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

[3] R. Singh, A. Chakraborty, and B. S. Manoj, "Graph Fourier transform based on directed Laplacian," in *Int. Conf. Signal Process. Commun.*, 2016, pp. 1–5.

[4] R. Shafipour, A. Khodabakhsh, G. Mateos, and E. Nikolova, "Digraph Fourier transform via spectral dispersion minimization," in *Int. Conf. Acoust., Speech, Signal Process.* IEEE, 2018, pp. 6284–6288.

[5] R. Shafipour, A. Khodabakhsh, G. Mateos, and E. Nikolova, "A directed graph Fourier transform with spread frequency components," *IEEE Trans. Signal Process.*, vol. 67, no. 4, pp. 946–960, 2018.

[6] S. Sardellitti, S. Barbarossa, and P. Di Lorenzo, "On the graph Fourier transform for directed graphs," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 6, pp. 796–811, 2017.

[7] S. Furutani, T. Shibahara, M. Akiyama, K. Hato, and M. Aida, "Graph signal processing for directed graphs based on the Hermitian Laplacian," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discov. Databases*, 2019, pp. 447–463.

[8] B. Seifert and M. Püschel, "Digraph signal processing with generalized boundary conditions," *IEEE Trans. Signal Process.*, vol. 69, pp. 1422–1437, 2021.

[9] P.-L. Loh and P. Bühlmann, "High-dimensional learning of linear causal networks via inverse covariance estimation," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 3065–3105, 2014.

[10] W. Chen, M. Drton, and Y. S. Wang, "On causal discovery with an equal-variance assumption," *Biometrika*, vol. 106, no. 4, pp. 973–980, 2019.

[11] M. Chevalley, Y. Roohani, A. Mehrjou, J. Leskovec, and P. Schwab, "Causalbench: A large-scale benchmark for network inference from single-cell perturbation data," *arXiv preprint arXiv:2210.17283*, 2022.

[12] M. Kaiser and M. Sipos, "Unsuitability of NOTEARS for causal graph discovery when dealing with dimensional quantities," *Neural Process. Letters*, vol. 54, no. 3, pp. 1587–1595, 2022.

[13] C. K. Assaad, E. Devijver, and E. Gaussier, "Survey and evaluation of causal discovery methods for time series," *Journal of Artificial Intelligence Research*, vol. 73, pp. 767–819, 2022.

[14] J. Runge, A. Gerhardus, G. Varando, V. Eyring, and G. Camps-Valls, "Causal inference for time series," *Nature Reviews Earth & Environment*, vol. 4, no. 7, pp. 487–505, 2023.

[15] U. Hasan, E. Hossain, and M. O. Gani, "A survey on causal discovery methods for i.i.d. and time series data," *arXiv preprint arXiv:2303.15027*, 2023.

[16] M. J. Vowels, N. C. Camgoz, and R. Bowden, "D'ya like dags? A survey on structure learning and causal discovery," *ACM Computing Surveys*, vol. 55, no. 4, pp. 1–36, 2022.

[17] M. J. Vowels, N. C. Camgoz, and R. Bowden, "Targeted VAE: Structured inference and targeted learning for causal parameter estimation," 2020.

[18] Y. Wang, V. Menkovski, H. Wang, X. Du, and M. Pechenizkiy, "Causal discovery from incomplete data: a deep learning approach," *arXiv preprint arXiv:2001.05343*, 2020.

[19] D. C. Castro, I. Walker, and B. Glocker, "Causality matters in medical imaging," *Nature Communications*, vol. 11, no. 1, pp. 3673, 2020.

[20] M. J. Vowels, "Limited functional form, misspecification, and unreliable interpretations in psychology and social science," *arXiv preprint arXiv:2009.10025*, vol. 2, 2020.

[21] C. Gong, D. Yao, C. Zhang, W. Li, and J. Bi, "Causal discovery from temporal data: An overview and new perspectives," *arXiv preprint arXiv:2303.10112*, 2023.

[22] Y. Jiang and S. Shimizu, "Linkages among the Foreign Exchange, Stock, and Bond Markets in Japan and the United States," in *Causal Analysis Workshop Series*. PMLR, 2023, pp. 1–19.

[23] X. Yang, Z.-H. Wang, C. Wang, and Y.-C. Lai, "Detecting the causal influence of thermal environments among climate regions in the United States," *Journal of Environmental Management*, vol. 322, pp. 116001, 2022.

[24] S. M. Smith, K. L. Miller, G. Salimi-Khorshidi, M. Webster, C. F. Beckmann, T. E. Nichols, J. D. Ramsey, and M. W. Woolrich, "Network modelling methods for FMRI," *Neuroimage*, vol. 54, no. 2, pp. 875–891, 2011.

[25] B. Seifert, C. Wendler, and M. Püschel, "Causal Fourier analysis on directed acyclic graphs and posets," *IEEE Trans. Signal Process.*, vol. 71, pp. 3805–3820, 2023.

[26] V. Mihal and M. Püschel, "Möbius total variation for directed acyclic graphs," in *Int. Conf. Acoust., Speech, Signal Process.*, 2023, pp. 1–5.

[27] 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.

[28] L. Yang, A. Qi, C. Huang, and J. Huang, "Graph Fourier transform based on L1 norm variation minimization," *Appl. Comput. Harm. Anal.*, vol. 52, pp. 348–365, 2021.

[29] A. Anis, A. Gadde, and A. Ortega, "Efficient sampling set selection for bandlimited graph signals using graph spectral proxies," *IEEE Trans. Signal Process.*, vol. 64, no. 14, pp. 3775–3789, 2016.

[30] S. Mallat, *A wavelet tour of signal processing*, Elsevier, 1999.

[31] G.-C. Rota, "On the Foundations of Combinatorial Theory I. Theory of Möbius functions," *Z. Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 2, no. 4, pp. 340–368, 1964.

[32] I. Ng, A. Ghassami, and K. Zhang, "On the role of sparsity and DAG constraints for learning linear DAGs," *Advances in Neural Information Processing Systems*, vol. 33, pp. 17943–17954, 2020.

[33] K. Harada and H. Fujisawa, "Sparse estimation of linear non-Gaussian acyclic model for causal discovery," *Neurocomputing*, vol. 459, pp. 223–233, 2021.

[34] S. Shimizu, P. O. Hoyer, A. Hyvärinen, A. Kerminen, and M. Jordan, "A linear non-Gaussian acyclic model for causal discovery," *Journal of Machine Learning Research*, vol. 7, no. 10, 2006.

[35] M. J. Bowes, L. K. Armstrong, S. A. Harman, D. J. E. Nicholls, H. D. Wickham, P. M. Scarlett, and M. D. Juergens, "Weekly water quality data from the River Thames and its major tributaries (2009-2017)," 2020, https://doi.org/10.5285/cf10ea9a-a249-4074-ac0c-e0c3079e5e45, (accessed Nov. 6, 2024).

[36] M. Scutari, "bnlearn - an R package for Bayesian network learning and inference, Bayesian Network Repository," https://www.bnlearn.com/bnrepository/, (accessed Nov. 6, 2024).

[37] R. Opgen-Rhein and K. Strimmer, "From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data," *BMC systems biology*, vol. 1, no. 1, pp. 1–10, 2007.

[38] M. Scutari, P. Howell, D. J. Balding, and I. Mackay, "Multiple quantitative trait analysis using Bayesian networks," *Genetics*, vol. 198, no. 1, pp. 129–137, 2014.

[39] S. Chen, A. Sandryhaila, G. Lederman, Z. Wang, J. M. F. Moura, P. Rizzo, J. Bielak, J. H. Garrett, and J. Kovačevic, "Signal inpainting on graphs via total variation minimization," in *Int. Conf. Acoust., Speech, Signal Process.* IEEE, 2014, pp. 8267–8271.

[40] S. Chen, A. Sandryhaila, J. M. F. Moura, and J. Kovačević, "Signal recovery on graphs: Variation minimization," *IEEE Trans. Signal Process.*, vol. 63, no. 17, pp. 4609–4624, 2015.

[41] W. C. Karl, "Regularization in image restoration and reconstruction," in *Handbook of Image and Video Process.*, pp. 183–V. Elsevier, 2005.

[42] S. Ramani, T. Blu, and M. Unser, "Monte-Carlo SURE: A black-box optimization of regularization parameters for general denoising algorithms," *IEEE Trans. Image Process.*, vol. 17, no. 9, pp. 1540–1554, 2008.

[43] H. Avron and C. Boutsidis, "Faster subset selection for matrices and applications," *SIAM Journal on Matrix Analysis and Applications*, vol. 34, no. 4, pp. 1464–1499, 2013.